SURFACE ROUGHENING OF HETEROEPITAXIAL THIN FILMS

Huajian Gao

Division of Mechanics and Computation, Department of Mechanical Engineering, Stanford University, Stanford, California 94305; e-mail: gao@am-sun2.stanford.edu

William D. Nix

Department of Materials Science and Engineering, Stanford University, Stanford, California 94305; e-mail: nix@soe.stanford.edu

ABSTRACT

Heteroepitaxial structures with strained semiconductor thin films are widely used in electronic and optoelectronic devices. One of the more important defect creation processes in these films is related to a stress-induced morphological instability that tends to roughen the film surface by mass diffusion during film growth or annealing. Interestingly, the same mechanism of surface roughening can be utilized for fabrication of quantum dot devices. This article gives an overview of a series of theoretical and experimental studies on surface roughening in heteroepitaxial films. It is shown that the strain caused by lattice mismatch drives the diffusional atomic flux along the film surface in such a way that an initially flat film evolves into an undulating profile with cusp-like surface valleys with singular stress concentration near the cusp tip. The essential features of this evolution process are described by a family of mathematical curves called cycloids. The fundamental length and time scales associated with surface roughening can be obtained from thermodynamic and kinetic considerations. The stress concentration at cycloid-like surface valleys caused by roughening is found to create dislocations of various characters that participate in the overall strain relaxation of a heterostructure.

173

INTRODUCTION

Heteroepitaxial structures with strained semiconductor thin films have found wide applications in electronic and optoelectronic devices (1, 2). The mechanism of how defect nucleation might occur in a defect-free epitaxial film under stress has been a mystery that has attracted much attention in the last few decades. Device fabrication requires that defects be kept to a minimum, because otherwise the electrical properties or the performance of the device will be adversely affected. Defects are most probably introduced during the film deposition process itself or during subsequent thermal processing, such as dopant drive-in. Hence, understanding the processes by which the various defects are formed is imperative for developing methodologies to minimize their density.

In recent years, surface roughening caused by elastic strain from lattice mismatch between film and substrate has been recognized as a dominant mechanism for defect formation in heteroepitaxial films. Theoretical studies (3–8) have shown that an initially flat surface of a stressed solid is unstable with respect to morphological variations in surface shape. As a consequence of this morphological instability, the surface is driven toward a shape with cusp-like surface valleys (9–14). The phenomenon of surface roughening has been observed in a growing number of experiments, most of which have involved observations of film roughening under various growth conditions, with surface morphology forming a range of configurations, including islands (15–18), smoothly undulating surfaces (19–21), and cusped surfaces (22). Another class of experiments has involved controlled annealing of defect-free, flat as-grown films (23, 24), which have been used to test the characteristic length and time scales, as well as other features predicted by the theoretical models.

There are close interactions between the surface-roughening process and dislocation formation in heteroepitaxial films. The stress concentration caused by film roughening can aid the nucleation of dislocations in a dislocation-free film. It has been established (25–27) that the elastic mismatch stress in heteroepitaxial films provides a driving force for threading dislocations that extend from the free surface of the film to the interface between the film and substrate. The propagation of threading dislocations in the film layer leaves behind misfit dislocation lines along the interface. This process relaxes the elastic mismatch strain and increases the total length of the dislocation. Below a critical film thickness h_{cr}, called Matthews' critical thickness (25), the strain energy reduction from relaxation of lattice mismatch is insufficient to compensate for the cost of increasing dislocation length, in which case a coherent film is energetically preferred over an incoherent film. It is therefore possible to conduct surface-roughening experiments on subcritical films with no dislocation formation (23, 24). For supercritical films with thicknesses larger than h_{cr}, the

propagation of threading dislocations causes an overall relaxation of elastic strain in the film. The velocities of the threading dislocations have been measured, and other basic aspects of strain relaxation have been studied (28, 29). Controlled annealing of defect-free, flat as-grown supercritical films (23, 24) has shown that surface-roughening precedes dislocation formation. Misfit dislocations are observed to form during the surface-roughening process and, as the dislocation density increases, the surface undulations are rotated to align with the dislocation network (23). The study of surface roughening has provided an understanding of the mechanisms by which dislocations can be formed in a defect-free film. Without roughening, the calculated activation energy for forming a critical-sized dislocation nucleus is much too large for thermal nucleation to occur (22, 24, 30).

A positive aspect of surface roughening in nanostructures is that it could be used in the fabrication of quantum dots. Controlled annealing of a very thin heteroepitaxial film causes the film to break up into islands. A quantum dot refers to arrays of nanometer-sized islands consisting of 10^3–10^5 atoms, which exhibit unusual electrical and optical properties (31). These structures afford the possibility of developing a range of novel microelectronic devices, such as single-electron transistors, quantum semiconductor lasers, light-emitting diodes, and photovoltaic cells (32). Quantum dots are significantly smaller than the micrometer-sized structures found in current microelectronic devices, which makes it difficult to manufacture them by standard lithographic techniques. The process of surface roughening is a promising approach by which controlled growth and annealing processes can be developed to allow the quantum dots to spontaneously form from a very thin heteroepitaxial layer.

This paper provides an overview of the recent theoretical and experimental studies on surface roughening in heteroepitaxial thin films. We have chosen to focus our discussions on a few representative studies, rather than enumerate the work contained in a large body of literature. On the theoretical side, we focus on the essential features of surface roughening as described by a cycloid surface evolution model. On the experimental side, we focus on a series of controlled annealing experiments with subcritical films for isolating surface roughening from other phenomena such as growth and dislocation formation and on annealing experiments with supercritical films in illustrating the role of surface roughening in defect formation.

THEORY OF SURFACE ROUGHENING

Lattice Mismatch in a Heteroepitaxial Film

Figure 1 shows a schematic representation of a heteroepitaxial thin-film structure in which a uniform film of lattice spacing a_f and thickness h is coherently bonded to a substrate of lattice spacing a_s. In a dislocation-free film, the lattice
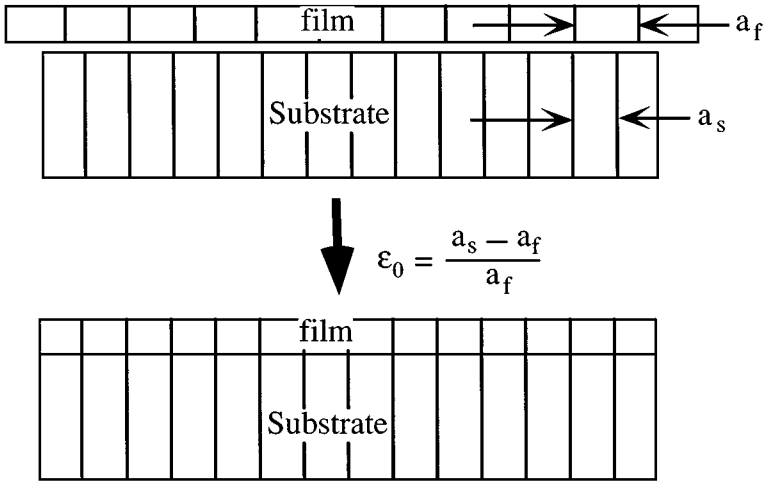
Figure 1 A heteroepitaxial thin-film structure. The mismatch between the lattice parameters of the film and substrate is accommodated by an elastic strain in the film.

mismatch is fully accommodated by a biaxial elastic strain $\varepsilon_0 = (a_s - a_f)/a_f$, which for $\mathrm{Si}_{1-x}\mathrm{Ge}_x/\mathrm{Si}$ and $\mathrm{Ga}_{1-x}\mathrm{In}_x\mathrm{As}/\mathrm{GaAs}$ heterostructures is approximately equal to $-4.2x\%$ and $-6x\%$, respectively, where x denotes the alloy fraction. Correspondingly, a biaxial stress $\sigma_0 = E\varepsilon_0/(1 - v)$ is expected in a dislocation-free $\mathrm{Si}_{1-x}\mathrm{Ge}_x/\mathrm{Si}$ film of uniform thickness, where $E/(1 - v)$ is the biaxial plane modulus, and E and v are Young's modulus and Poisson's ratio of the film material. This stress is very large compared with that of a bulk material. The biaxial-plane modulus of Si(100) is around 180 GPa and that of Ge(100) is around 141 GPa. For $\mathrm{Si}_{0.9}\mathrm{Ge}_{0.1}/\mathrm{Si}(100)$, the lattice mismatch is $\varepsilon_0 = -0.42\%$ and the stress is $\sigma_0 = -0.74$ GPa. For $\mathrm{Si}_{0.65}\mathrm{Ge}_{0.35}/\mathrm{Si}(100)$, the mismatch strain is $\varepsilon_0 = -1.47\%$, and the stress is $\sigma_0 = -2.44$ GPa. In comparison, typical values of stresses in a bulk material are in the mega-Pascal range.

Surface-Roughening Instability

Owing to the small length scales and the high-temperature environment of heteroepitaxial films, diffusive mass transport along the surface of these films is very fast and causes the morphology of the film surface to evolve continuously with time. The dominant thermodynamic forces that control the morphological evolution are variations in elastic strain energy and surface energy of the film. It is often a good approximation to neglect heat terms and take the sum of elastic strain energy and surface energy as the total free energy of the system. Under this assumption, the surface diffusion process is governed by a chemical

potential function $\chi = \Omega(\gamma\kappa + w)$ where Ω is the atomic volume, γ is the surface energy density, κ is the curvature, and w is the strain energy density along the surface (3). A perfectly flat film with a constant potential $\chi = \Omega w_0$ along the surface is at an unstable equilibrium because the system can lower its free energy by rearranging atomic positions via mass transport along the surface. This instability leads to roughening of the film surface.

The onset of surface roughening in heteroepitaxial films can be understood by considering the stability of a flat film (Figure 2a) with respect to an infinitesimal-cosine wavy perturbation (Figure 2b). This problem can be investigated as a two-dimensional cosine surface under plane strain conditions (3–5) or as a three-dimensional wavy surface under biaxial stress (33, 34). These studies indicate that there exists a critical wavelength that depends on the lattice mismatch strain ε_0, as $\lambda_{cr} = \pi(1-v)\gamma/[E(1+v)\varepsilon_0^2]$. If $\lambda < \lambda_{cr}$, the chemical potential is higher at a surface peak than at a surface trough, and the process of surface diffusion tends to smooth out the wavy surface by transporting matter from
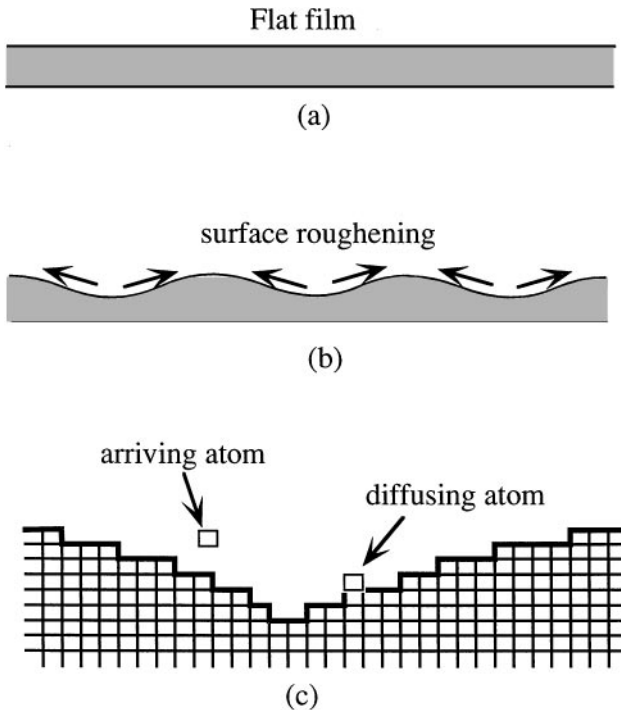

Figure 2 (*a*) A flat heteroepitaxial film is unstable with respect to (*b*) surface roughening by diffusive mass transport. (*c*) An atomistic view of diffusion during film growth and annealing.

peaks to valleys. If $\lambda > \lambda_{cr}$, the chemical potential is lower at a surface peak than at a surface trough, and atoms are transported away from valleys and deposited onto peaks, hence magnifying the initial perturbation and causing the surface to roughen. Within a numerical factor, the critical wavelength λ_{cr} is equal to the ratio between the surface energy density γ, which has the dimension of energy per unit area, and the strain energy density w_0, which has the dimension of energy per unit volume. A dimensional analysis will show that this ratio sets an intrinsic length scale for surface roughening. The strain energy dominates in perturbations with sufficiently long wavelengths and tends to destabilize a flat surface.

The mechanism of surface roughening has general implications in many other material systems that are dominated by competition between strain energy and surface energy (35). However, such generalizations may include other thermodynamic forces and may involve different mass transport mechanisms. For example, in bulk materials, the gravitational force can completely suppress the surface roughening (6, 36). In very thin films consisting of a few monolayers of atoms, the effect of substrate needs to be considered (13). In polycrystalline films, mass transport along grain boundaries could completely alter the mechanism of surface roughening.

Fastest Growing Perturbation Wavemode

The surface profile of a slightly undulating surface can be expanded in a Fourier series consisting of an infinite number of perturbation wavemodes. The surface evolution can be affected by different kinetic processes, such as surface diffusion, bulk diffusion, film deposition, and evaporation. The kinetics of these processes determine a fastest growing perturbation wavemode as the dominant roughening profile observable in experiments. For typical conditions during growth and annealing of heteroepitaxial films, surface diffusion along the film surface is the dominant mechanism of mass transport. Inserting the elasticity solution of a slightly undulating surface into the linearized equation of surface diffusion, it is found (3, 5) that the growth rate of a perturbation wavemode $\cos kx$ is proportional to the exponential factor $\exp[(k_{cr} - k)ck^3 t]$, where k is the wavenumber of perturbation and $k_{cr} = 2\pi/\lambda_{cr}$ is the critical wavenumber of the initial instability, $c = D_s \Omega^2 \delta_s \gamma / k_B T$ is a constant dependent on the temperature T and other material properties, D_s is the surface diffusivity, Ω is the atomic volume, δ_s is the number of atoms (sites) per unit surface area, and k_B is the Boltzmann constant. The fastest growing wavemode corresponds to maximizing the growth rate among all possible choices of the perturbation wavenumber k. This is easily found to be $k_{\text{diffusion}} = 3k_{cr}/4$, with corresponding wavelength $\lambda_{\text{diffusion}} = 4\lambda_{cr}/3$. The linearized kinetic equation also suggests that the characteristic time for surface roughening is proportional

to $k_B T \gamma^3/[D_s \Omega^2 \delta_s E^4 \varepsilon_0^8]$, which is highly sensitive to the lattice mismatch strain ε_0.

Anisotropy Effects in Surface Roughening

The surface roughening process is also affected by anisotropy in elastic properties and surface energy. These effects have been investigated (6) by using a two-dimensional linear perturbation analysis of an anisotropic solid with a cosine wavy surface. An explicit solution for the critical wavelength is found, which can be expressed in terms of a surface energy function $\gamma(\theta)$ and the mismatch stress σ_0 as $\lambda_{cr} = 2\pi[\gamma(0) + \gamma''(0)]/(Y\sigma_0^2)$, where Y is a component of the so-called surface admittance tensor in the direction of surface undulation (6). The surface energy $\gamma(\theta)$ corresponds to the Wulff plot of surface energy as a function of the orientation θ of the surface, and we have assumed $\theta = 0$ represents the unperturbed flat surface. When surface energy anisotropy is strong enough to cause crystal facets to form, as in sufficiently low temperatures, the function $\gamma(\theta)$ has a cusp structure at $\theta = 0$ so that $\gamma''(0) = \infty$. In this case, the surface roughening is completely suppressed by the anisotropy effects. This can also be understood from the perspective of the creation of sharply defined atomic steps along a rough surface, which tends to stabilize a flat surface (37). Under most experimental conditions, it appears that the temperature is sufficiently high that a smooth Wulff function $\gamma(\theta)$, rather than one with sharp cusps, is more realistic. If this assumption holds, the anisotropy effects will not completely suppress the surface roughening, but rather will modify the critical wavelength.

The surface admittance Y is a measure of elastic compliance of an elastically anisotropic solid. This quantity also appears in the self-energy of a dislocation and the energy release rate of a crack in an anisotropic solid (6). For cubic crystals, an explicit expression for Y exists. Calculations (24, 38) have shown that the energy release rate associated with surface roughening of a $Si_{1-x}Ge_x/Si(100)$ film is larger for surface undulations in $\langle 100 \rangle$ type crystallographic directions than those in $\langle 110 \rangle$ type directions. This result has been confirmed by controlled annealing experiments (24).

Singular Stress Concentration at Cusp-Like Surface Valleys

Linear perturbation analysis (33) of the stress concentrations at slightly undulating surfaces indicated that a cosine wavy surface with an amplitude-to-wavelength ratio of <0.1 is sufficient to double the nominal stress level at a valley. Further insight into the dramatic effect of a rough surface on stress concentration was provided by the study of cycloid-shaped surfaces (9). It was discovered that, under uniform lateral stress, the cusped cycloid surface in Figure 3a induces a crack-like singular stress field with identical singularity strength (i.e. stress intensity factor) to that of the periodic array of semiinfinite
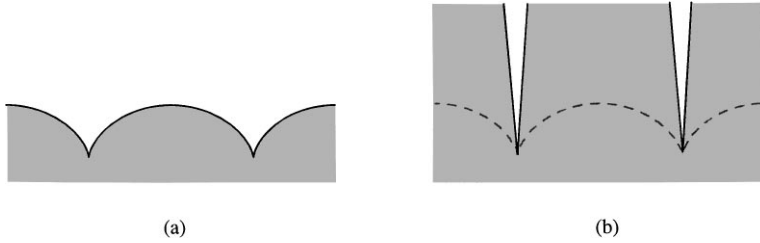
(a)   (b)

Figure 3 Geometry of (*a*) a cusped cycloid surface and (*b*) a periodical array of semi-infinite cracks.

cracks shown in Figure 3*b*. Application of fracture mechanics concepts (39) predicts that the cusped cycloid surface should fail (i.e. induce nucleation of other defects) at the same stress level as the periodic cracks, even though there are no apparent "cracks" associated with the cycloid surface!

While investigating the mechanics of surface roughening, several general conservation properties have been discovered (34, 40a,b). It was found that, during the surface roughening, the projected average of strain energy density along the plane of the film remains constant in time. This property can be written as $\langle w \rangle = w_0$, where w_0 is the strain energy density of a perfectly flat film and the operator $\langle w \rangle$ means averaging w along the plane of the film. Since the strain energy is quadratic in stress, this conservation property means that the square average of stress is independent of the geometry of the rough surface, $\langle \sigma^2 \rangle = \sigma_0^2$. Hence, reduction in stress along any part of a rough surface must be accompanied by an increase in stress at some other part of the surface. Any nonuniformity in stress distribution caused by surface roughening inevitably causes stress concentration. The same conservation property also indicates that, among all periodic rough surfaces, the worst one with the strongest possible stress concentration has zero stress everywhere along the surface except at singularity points. The cycloid surface is such a configuration; it causes singular stress concentration at cusps by completely relaxing the stress along the rest of the surface. Stress concentration by any periodic rough surfaces is necessarily bound by the cycloid surface shown in Figure 3*a*.

Figure 4 compares the contours of the maximum principal stress and the maximum shear stress within one half period of the cycloid surface with the corresponding stress contours near a periodical array of semi-infinite cracks (9). Very similar stress patterns are observed for both cases, with identical stress intensity factor $K = \sigma_0 \sqrt{\lambda/2}$, where λ is the spacing between singularity points. This similarity could be explained as follows. In the solid with semi-infinite cracks, the material outside the cycloid contour (shown as a dotted line in Figure 3*b*) does not carry load and can be eliminated without affecting the
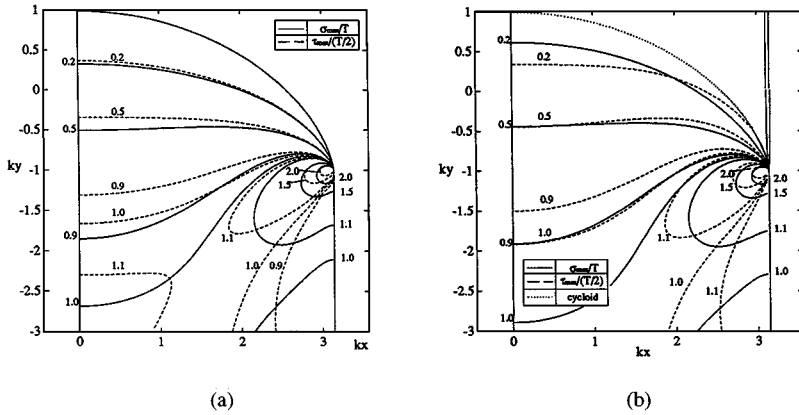
Figure 4   Contour plots of the maximum principal stress σ_{max} and the maximum shear stress τ_{max} for (*a*) a cusped cycloid surface and (*b*) a periodic array of semi-infinite cracks (9) *Solid line*, σ_{max}/T; *dashed line*, $\tau_{max}/(T/2)$; *dotted line*, cycloid.

stress distribution in the rest of the solid. It was calculated (9) that stresses along the cycloid line in the periodically cracked solid are indeed <1%–2% of the nominal stress.

Cycloid Model of Surface Roughening

Surface roughening of lattice-mismatched heteroepitaxial films induces significant stress concentration at surface valleys. The stress concentration at a surface valley forces the atoms to drift farther away from the valley, which causes deepening and sharpening of the valley into cusps. A schematic lattice view of stress concentration near a cusped surface valley under compressive stress is shown in Figure 5*a*. The lattice near the cusp is under large stress, whereas the lattice near the peak is relaxed. Atoms are driven away from the valley region by stress concentration there. At a critical point, the cusp formation process triggers the nucleation of a dislocation loop, which could expand to create misfit dislocations along the film/substrate interface, as shown in Figure 5*b*.

An analytical model that captures the essential features of surface roughening and cusp formation is the cycloid model (9) shown in Figure 6*a*. In this model, the surface evolution is described by a family of cycloid curves, $kx = k\xi + \varepsilon \sin k\xi$, $ky = \varepsilon \cos k\xi - \varepsilon^2/2$, which is characterized by a single roughening parameter $\varepsilon = Ak$, where A and k are the amplitude and wavenumber of the cycloid surface. As ε evolves from 0 to 1, this curve describes a perfectly flat surface evolving, with mass conserved, into a cusped surface with period equal to λ, as shown in Figure 6*a*.
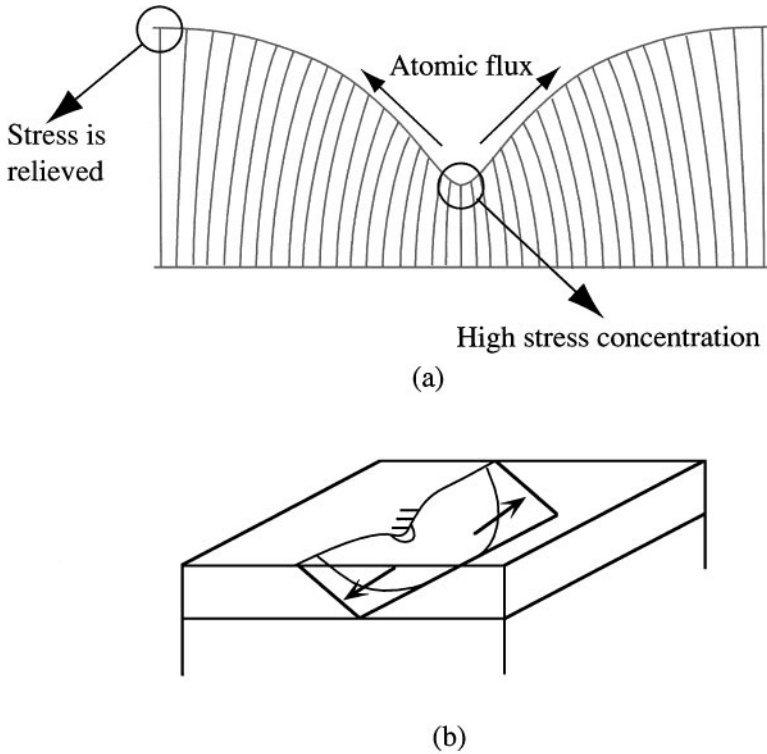
Figure 5 (*a*) A schematic lattice view of stress concentration near a cusp valley that leads to (*b*) formation of misfit dislocations in a heteroepitaxial film.

In the cycloid model (9), the variation of free energy per wavelength as the surface evolves from the perfectly flat state $\varepsilon = 0$ to a wavy state ε is found to be $k\Delta E_{tot}/\gamma = -\pi k_{cr}\varepsilon^2/2k + \int_0^{2\pi}(\sqrt{1 + 2\varepsilon\cos\theta + \varepsilon^2} - 1)\,d\theta$. This expression correctly predicts all the critical length scales. At the initial instability ($\varepsilon \ll 1$), one can show that $k\Delta E_{tot}/\gamma \cong -\pi(k_{cr} - k)\varepsilon^2/2k$. At the cusp formation limit ($\varepsilon \cong 1$), the free energy behaves as $k\Delta E_{tot}/\gamma \cong -2(k_{lcr} - k)\varepsilon^2/k$, where k_{lcr} is the critical wavelength for cusp formation, $k_{lcr} = \pi k_{cr}/4$. The corresponding critical wavelength is $\lambda_{lcr} = 4\lambda_{cr}/\pi$.

The condition for cusp formation is intimately connected to the Griffith condition for crack growth in fracture mechanics. The Griffith condition requires a balance between the strain energy release rate at a crack tip and the surface energy going into creating two free surfaces during a fracture process.
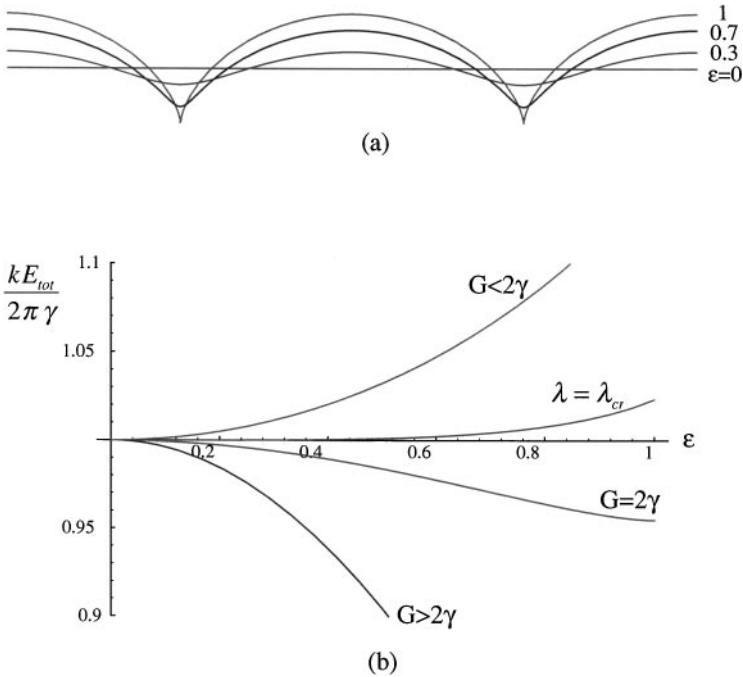
Figure 6   (*a*) The cycloid model of surface roughening. (*b*) The variation of free energy in the cycloid model. Note that there is no energy barrier as soon as the Griffith condition is satisfied.

This condition can be written as $G = 2\gamma$. We can check this condition for the cycloid model by using the Irwin relation $G = (1 - v^2)K^2/E$, where $K = E\varepsilon_0\sqrt{\lambda/2}/(1 - v)$ is the stress intensity factor. One finds that the condition $\lambda = \lambda_{lcr}$ is equivalent to $G = 2\gamma$.

Figure 6*b* plots the energy variation associated with surface roughening for different values of the perturbation wavelength. Several ranges of behavior are observed. For very short wavelengths, $\lambda < \lambda_{cr}$, the free energy increases with the surface roughness. In the range $\lambda_{cr} < \lambda < \lambda_{lcr}$, the free energy first decreases with the surface roughness and then increases after a minimum is reached. In the range $\lambda > \lambda_{lcr}$, there is no longer an energy barrier for an initially perfectly flat surface to develop into a cusped configuration via mass transport. The fastest growing wavemode under surface diffusion falls in the range $\lambda > \lambda_{lcr}$, and we expect to observe cusp formation in experiments.

In the cycloid model (9), the local curvature κ_v of a surface valley changes as $\kappa_v = \kappa_v^0(1 - t/t^*)^{-1/4}$, where κ_v^0 is the local curvature at the onset of diffusion

and t^* is a characteristic time for the surface to reach the cusp formation. Note that a finite amount of time is needed for cusp formation.

Some General Properties of Cusp Formation

In the cycloid model, the condition for cusp formation is equivalent to the Griffith condition of fracture mechanics. To understand this connection, it is insightful to consider the variation of free energy during a surface diffusion process. By definition, the time-rate of total energy change is related to the chemical potential by $\dot{E}_{tot} = \int \chi(s) v_n ds / \Omega$, where $\chi(s)$ is the chemical potential at a position s along the surface and v_n denotes the velocity of the surface motion in the outward normal direction caused by accumulation or removal of atoms. The integration is to be carried out over the entire surface. If there is a cusp along the boundary, χ behaves as a Dirac delta function (40). Without loss of generality, one may write χ near the cusp as $\chi = \Omega(G - 2\gamma)\delta_{\mathrm{Dir}}(x) + \tilde{\chi}$, where $\tilde{\chi}$ is a well-behaved function, x denotes a local axis perpendicular to the local tangent of the cusp tip, and $x = 0$ is the cusp tip. The diffusive mass flux causes the surface to move with velocity $v_n \propto \partial^2 \chi / \delta s^2$ (41, 42). From these relations, we see that the Griffith condition $G = 2\gamma$ is equivalent to assuming that the energy release rate $\dot{E}_{tot}$ remain finite in surface diffusion. This assumption is not unreasonable since the first law of thermodynamics requires that $\dot{E}_{tot}$ be balanced by the heat flux, which cannot be infinite in a physical process. The Griffith condition can also be understood from the expression of chemical potential, which suggests that a cusp with $G < 2\gamma$ acts like a very strong "sink" point that attracts diffusive flux, which, if it exists, will be instantly eliminated by the diffusion process. In other words, the very large curvature at a cusp can be maintained only if the Griffith condition is satisfied.

Several additional properties can be deduced from the assumption that $\dot{E}_{tot}$ remains finite during surface diffusion. First, we can show that the local stress field must be symmetric with respect to the cusp tip; i.e. a diffusively formed cusp must be of mode I symmetry. If a cusp has a mode II stress intensity factor K_{II}, the tangential stress along the surface near the cusp behaves as $\sigma_{tt} \propto K_{II}/\sqrt{s} \propto K_{II}/\sqrt{x}$, where s is the arc length distance from the cusp tip. The resulting strain energy density contains a singular term of order x^{-1} which cannot be balanced by the surface energy term. The case of a mode III stress intensity factor K_{III} leads to similar behavior. Thus, the chemical potential near a cusp generally behaves as $\chi = \Omega(G - 2\gamma)\delta_{\mathrm{Dir}}(x) + c_1 K_{II}^2/x + c_2 K_{III}^2/x + \cdots$, where c_1 and c_2 are numerical factors. To avoid a divergent $\dot{E}_{tot}$, the Griffith condition $G = 2\gamma$ must be satisfied, and both K_{II} and K_{III} must vanish.

Similarly, we can show that it is impossible to form a surface notch with a finite opening angle $2\beta > 0$ ($\beta = 0$ corresponds to a cusp) during stress-controlled surface diffusion. This is because a notch has a weaker stress singularity than a crack has (43). The curvature at the notch contains a Dirac singularity $\kappa = -2\cos\beta\delta_{\mathrm{Dir}}(x)$, whereas the strain energy contains a power singularity $w = c/x^{\alpha}$, with singularity strength $0 < \alpha < 1$. The singular terms in chemical potential behave as $\chi = -2\Omega\cos\beta\delta_{\mathrm{Dir}}(x) + c\Omega/x^{\alpha}$. In this case, it is impossible to balance the Dirac singularity in the surface energy term with the power singularity in the strain energy term.

Critical Wavelengths for Two-Dimensional and Three-Dimensional Cusp Formation

The cycloid model predicts a critical wavelength λ_{lcr} for cusp formation under two-dimensional conditions. Is this prediction exact? It may be helpful to examine the significance of λ_{lcr} without referring to a specific model. Consider the evolution of a wavy surface with wavelength λ. If λ is only slightly larger than the critical wavelength λ_{cr}, a smoothly undulating and nearly straight equilibrium surface exists. As λ increases, the equilibrium surface becomes more deeply (and sharply) grooved. This scenario continues until the wavelength is large enough for us to find a surface profile along which the Griffith condition $G = 2\gamma$ can be satisfied. In that case, cusp formation will occur because the stress concentration at a surface valley always tends to deepen and sharpen the valley by driving the atoms out of the valley region. This is the essence of the critical wavelength λ_{lcr}. The conservation law of surface roughening $\langle w \rangle = w_0$ indicates that it is possible to satisfy $G = 2\gamma$ only if $\lambda \geq \lambda_{lcr}$. At the critical wavelength $\lambda = \lambda_{lcr}$, the surface profiles that satisfy $G = 2\gamma$ have a Dirac singular distribution of strain energy at the cusps and zero stress along the rest of the surface. Such surfaces are not unique. The cycloid surface is one of such configurations that captures the essential features of cusp formation. The critical wavelength λ_{lcr} can be obtained from the conservation law $\langle w \rangle = w_0$ without resorting to the details of a specific model or a numerical simulation (34).

To understand the process of cusp formation along the surface of a biaxially stressed film, we also need to consider cusping in three-dimensional configurations. A wavy surface undulating in one direction causes the strain in that direction to be relaxed. However, the strain in the transverse direction will remain. More complete strain relaxation can be achieved by forming a three-dimensional cusping network. The critical wavelength for forming a three-dimensional cusping network can be similarly determined from the conservation law $\langle w \rangle = w_0$ without resorting to a three-dimensional numerical

simulation (34). The essence of this calculation is briefly outlined below. Consider a three-dimensional network of surface islands which are spaced by wavelength λ and aligned in some periodic pattern. A sufficiently large λ will induce cusping along the junction contours between adjacent surface islands. The cusp contour surrounding each island must have the unit cell configuration of a polygon to satisfy the symmetry requirements. The assembly of these unit cells forms a three-dimensional cusp network. One example is a hexagonal cusp array along the surface. The three-dimensional cusping wavelength is denoted as Λ_{lcr}. Based on the conservation law $\langle w \rangle = w_0$, the surface configuration that satisfies the Griffith condition at $\lambda = \Lambda_{lcr}$ must have a Dirac singular distribution of strain energy along the cusp lines and zero stress along the rest of the surface. The energy release rate G for such a surface is equal to the strain energy density multiplied by the area of the unit cell and divided by the cusp length per unit cell. This energy release rate is found to be $G = w_0\lambda/2$, regardless of the shape of the unit cell (34). The Griffith condition $G = 2\gamma$ then determines that the critical wavelength for the formation of a three-dimensional cusp network along the surface of a biaxially stressed film is $\Lambda_{lcr} = (1 + v)\lambda_{lcr}$.

For most materials, the Poisson ratio v lies in the range between 0.2 and 0.4. Under this condition, the dominant wavelength during surface diffusion falls in the range $\lambda_{lcr} < \lambda_{\text{diffusion}} < \Lambda_{lcr}$, which is sufficient for two-dimensional cusp formation but too small for the formation of a three-dimensional cusp network. This result has been confirmed by experiments (23, 44).

Numerical Simulation of Surface Roughening

Numerical simulations of stress-driven surface evolution and cusp formation have been conducted by several research groups (10–14) by solving the coupled equations of elasticity and surface diffusion. These simulations are consistent with the essential features of the cycloid model. In one numerical approach (12–14), a Fourier series is used to describe the surface morphology as $h(x, t) = \sum_n a_n \cos(2n\pi x/\lambda)$; a_0 remains constant owing to mass conservation, and a_n is computed as the surface evolves.

Figure 7 plots the results of a representative numerical analysis for the evolution of the fastest growing perturbation wavemode under surface diffusion. The initial condition is taken to be a slightly undulating surface with an amplitude-to-wavelength ratio around 1%. As illustrated in Figures 7a and 7b, the surface motion is quickly dominated at the vicinity of the valley, driven by the concentration of chemical potential $\chi(s)$ there. Figure 7c shows that the exponents of all coefficients a_n increase linearly with time over a significant portion of the evolution history, indicating exponential growth of all Fourier wavemodes. Beyond a characteristic time, these coefficients begin to accelerate, starting from
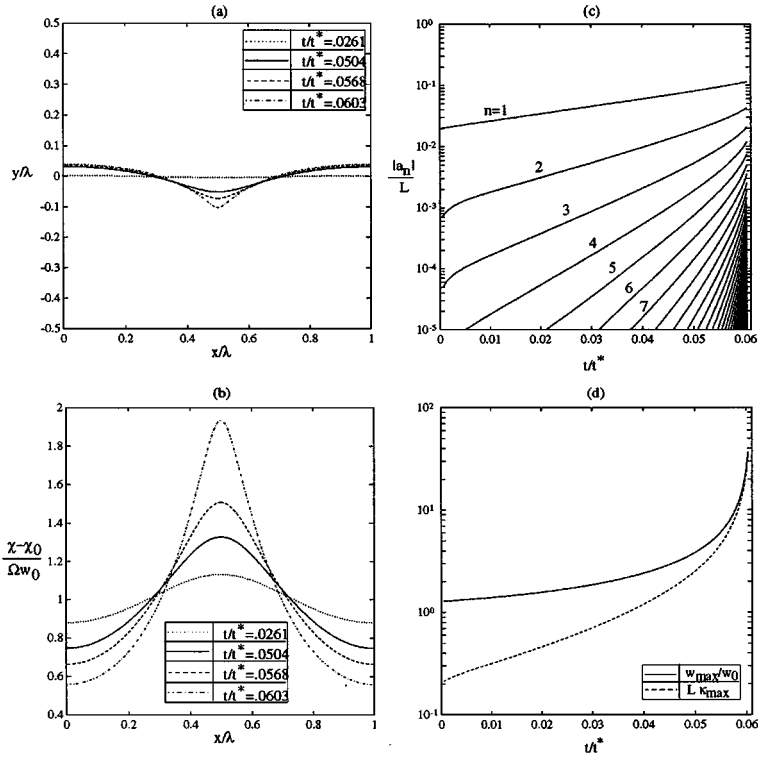
Figure 7 For the case of fastest-growing perturbation wavemode, the evolution of (*a*) the surface shape $h(x, t)$, (*b*) the chemical potential χ along the surface, (*c*) the Fourier coefficients a_n of the surface shape, and (*d*) the maximum strain energy density and the maximum curvature (12).

high-frequency modes, which indicates that the morphology is approaching a cusped profile rather than a steady state. The cusp formation is also demonstrated in Figure 7*d*, in which the strain energy density and curvature at the valley increase with time toward singular values.

The results in Figure 7 have been normalized for the following characteristic length, time, and energy density: $L = \gamma/w_0$, $t^* = k_B T \gamma^3 / D_s \delta_s \Omega^2 w_0^4$, $w_0 = (1 - v^2)\sigma_0^2/2E$. The time to reach cusp formation estimated from the numerical simulation is $t_{\text{cusp}} = 0.063 t^*$.

Effect of Substrate: A Boundary Layer Model

Owing to long-range atomic interactions, the surface energy of a heteroepitaxial film can be influenced by the substrate. This influence usually becomes insignificant as the distance between the film surface and interface exceeds a few atomic

monolayers, indicating that the surface energy of the film/substrate system, $\gamma_{f/s}$, should gradually take the bulk value γ_f with increasing film thickness. On the other hand, when the film thickness approaches zero in a continuum sense, the film surface energy can be equated to $\gamma_s - \gamma_{fs}$, where γ_s is the surface energy of the substrate and γ_{fs} is the interface energy assumed to be independent of the film thickness. A boundary layer model has been developed (13) to account for the effects of substrate, in which the surface energy of the film/substrate system is assumed to be $\gamma_{f/s} = \gamma_f + (\gamma_s - \gamma_f - \gamma_{fs})\exp(-y/l)$, where y is the distance to the interface and l is a characteristic length. The surface energy thus varies as $\gamma_{f/s}[f(x)]$ along a rough surface profile $y = f(x)$.

A linear stability analysis (13) based on the above boundary layer model of a heteroepitaxial film indicates that the effect of substrate on film evolution depends on the sign of $\gamma_s - \gamma_f - \gamma_{fs}$. If $\gamma_s - \gamma_f - \gamma_{fs} < 0$, a flat film is unstable for any thickness of the film. If $\gamma_s - \gamma_f - \gamma_{fs} > 0$, the growth mode is determined by whether the film is strained. For a lattice-matched structure, a flat film is stable. For a lattice-mismatched structure, there exists a critical thickness h_{cs} below which a flat film is stable. When the film thickness exceeds h_{cs}, a flat film becomes unstable and surface roughening occurs. The critical thickness h_{cs} is found to depend on the lattice mismatch as well as the surface energy difference $\gamma_s - \gamma_f - \gamma_{fs}$. Increasing the lattice mismatch will reduce h_{cs}; when the mismatch strain exceeds a critical value, h_{cs} becomes negative. This study suggests that the substrate effects may stabilize a heteroepitaxial film a few monolayers thick. However, it is impossible to maintain even a monolayer for arbitrarily large lattice mismatch. Reducing the lattice mismatch ε_0 in the heterostructure increases h_{cs} modestly, according to $\ln |\varepsilon_0^{-1}|$.

Summary of Critical Length and Time Scales for Surface Roughening

For the convenience of the reader, we summarize in Table 1 the critical length and time parameters for surface roughening.

Table 1   The critical length and time parameters of surface roughening

Lattice mismatch	$\varepsilon_0 = (a_f - a_s)/a_s$
Basic length scale	$L = \gamma/w_0 = \gamma(1 - v)/(E\varepsilon_0^2)$
Initial instability	$\lambda_{cr} = \pi L/(1 + v)$
Two-dimensional cusp formation	$\lambda_{lcr} = 4L/(1 + v)$
Three-dimensional cusp formation	$\Lambda_{lcr} = 4L$
Dominant wavelength	$\lambda_{\text{diffusion}} = 4\pi L/[3(1 + v)]$
Cusp formation time	$t_{\text{cusp}} = 0.063\, k_B T \gamma^3 / D_s \delta_s \Omega^2 w_0^4$

Table 2 Estimated value of the dominant wavelength of surface roughening under surface diffusion

$\lambda_{\text{diffusion}}$	15% Ge	18% Ge	22% Ge	25% Ge	35% Ge	50% Ge	75% Ge
Length (nm)	742	516	349	272	142	72	34

For estimates of the length scales of surface roughening, we take the biaxial plane modulus of Si(100) to be 180 GPa and that of Ge(100) to be 141 GPa, and we take the Poisson ratio $v = 0.22$ and the surface energy density $\gamma = 1.5$ J/m^2. Under these assumptions, numerical values of the dominant wavelength $\lambda_{\text{diffusion}}$ for Si$_{1-x}$Ge$_x$ films with different alloy fractions are tabulated in Table 2.

The surface diffusivity for (100)-type silicon surfaces has been reported to be $D_s = 8.45 \times 10^{-6} \exp(-0.83 \, \text{eV}/k_B T)$ cm^2/s for diffusion along $\langle 100 \rangle$ type directions (7). For estimates of the time scale of surface roughening, we take this as the surface diffusivity of Si$_{1-x}$Ge$_x$ alloys. The estimated time for cusp formation is tabulated in Table 3 (12, 24). For a 25% Ge film, it takes ~25 min to reach the cusping condition at 700°C and about 8 min at 800°C. For a 15% Ge film, it takes ~10 h at 800°C and 22 h at 700°C to reach the cusping condition. For even lower misfit strains and lower annealing temperatures, the time of evolution becomes astronomically large and surface roughening is kinetically suppressed. Note that the time to reach cusp formation depends strongly on the misfit strain (germanium fraction) and the evolution temperature.

CONTROLLED ANNEALING EXPERIMENTS

Experiments concerned with surface roughening and dislocation formation in heteroepitaxial thin films have been conducted during film growth (15–22) and annealing (23, 24, 38, 44, 46–48). Observations during film growth are strongly

Table 3 Estimated annealing time for cusp formation as a function of germanium composition and temperature (12, 24)

Seconds (°C)	15% Ge	18% Ge	22% Ge	25% Ge	35% Ge	50% Ge	75% Ge
500	1.8×10^6	4.3×10^5	3.0×10^4	1.8×10^4	618	36	2.41
600	4.22×10^5	6.0×10^4	2.4×10^4	3620	210	11	0.75
700	1.2×10^5	2.4×10^4	7840	1510	60	4.10	0.22
750	6.2×10^4	1.3×10^4	2895	603	24	2.63	0.11
800	3.6×10^4	7805	1810	480	17	1.25	0.053
850	3.0×10^4	6060	1390	265	12	0.63	0.028
900	1.8×10^4	3015	660	175	7.5	0.36	0.016

influenced by the changing conditions during growth, the increasing film thickness, the variation in film strain as misfit dislocations are introduced, and the kinetics of interaction between vapor and solid phases. Annealing experiments in the UHV environment (44, 48) show surface features that are difficult to interpret owing to surface reconstruction at the atomic scale (48). Thus, for a controlled study on the phenomenon of surface roughening, we focus our discussions on a series of controlled annealing experiments conducted in a hydrogen environment on defect-free, flat as-grown $Si_{1-x}Ge_x$/Si heterostructures (23, 24, 38, 46–48). These films were grown by low-pressure chemical-vapor deposition (LPCVD) in an ASM Epsilon-1 single-wafer chamber reactor at Stanford University. The annealing was conducted in situ under hydrogen flow in the chemical vapor deposition (CVD) reactor after the growth process was terminated.

Surface Roughening in Subcritical Thin Films

To study surface roughening separately from dislocation generation, heteroepitaxial films with 15% and 18% in Ge fractions were grown epitaxially on (100) Si substrates to 100 Å in thickness. This thickness is below Matthews' critical thickness for dislocation formation. The as-grown films were then annealed at a range of temperatures between 700°C and 850°C. The process of surface roughening was studied for annealing periods ranging from minutes to hours.

The surface-roughening theory (34) has indicated that the surface evolution in a heteroepitaxial film should commence in the form of three-dimensional budding or small-island formation caused by independent undulations in different orientations along the surface. This prediction is tested by observing the initial stage of roughening via controlled annealing experiments. Figure 8 shows a representative image of the initial three-dimensional budding stage of surface roughening.

Because the fastest growing wavemode under surface diffusion has a wavelength larger than that required for cusp formation in two-dimensional morphology but smaller than that required for cusp formation in three-dimensional morphology, the surface-roughening theory (34) also predicted that the three-dimensional budding stage of surface roughening will be replaced by two-dimensional cusp formation as the roughening proceeds. This is indeed observed in the annealing experiments shown in Figure 9, which shows slope-sensitive AFM images of 100-Å-thick 18% Ge films subjected to annealing for 1, 5, 20, and 120 min, respectively. In all of these images, the surface features are magnified 20× to 30× in height, and the scan edge marks a ⟨110⟩-type crystallographic orientation. The as-grown state of these films was virtually flat where an average surface roughness of only 1.5 Å was measured by AFM (23). The sequence of changes in morphology shown in Figure 9 was obtained
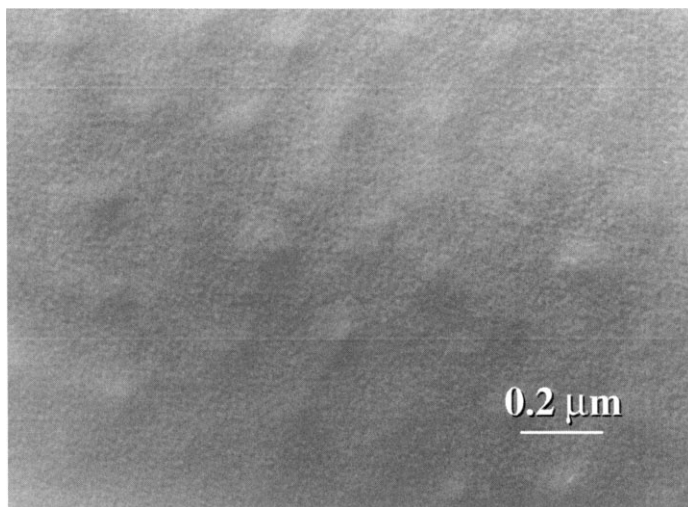
Figure 8   Bright-field plan view TEM image of a 100-Å-thick $Si_{0.82}Ge_{0.18}$/Si film annealed at 700°C for 1 h. This image reveals the three-dimensional surface pattern at the initial stage of surface roughening (initial budding) (24).

through a series of controlled annealing experiments. Observe the evolution of two-dimensional surface ridges, which shows that the film surface was being driven toward two-dimensional cusp formation. After 5 min of annealing, the average wavelength and amplitude of the surface features are around 4400 and 264 Å, respectively. Long ridges of several micrometers can form during the course of roughening, but these features themselves are unstable, resulting in further pinching (or roughening) of ridges into three-dimensional surface features. After 20 min, the average wavelength and amplitude of ridges reach 1.17 μm and 626 Å, respectively. The mode of roughening indicates a transition from two- into three-dimensional roughening or islanding. After 2 h of annealing, the roughening has progressed to a stage at which the substrate is exposed (possibly with a few monolayers of film coverage) between large islands, and the islands have an average wavelength and amplitude of 1.64 μm and 1600 Å, respectively.

The presence of misfit dislocations in heteroepitaxial films causes nonuniform stress distribution in a heteroepitaxial film and can also result in surface undulation (49). To understand whether the observed surface roughening is solely from the instability caused by lattice mismatch, transmission electron microscopy (TEM) imaging was conducted on the 18% Ge samples. Figure 10 shows (100) plan view images for the samples annealed at 850°C for 5 min and
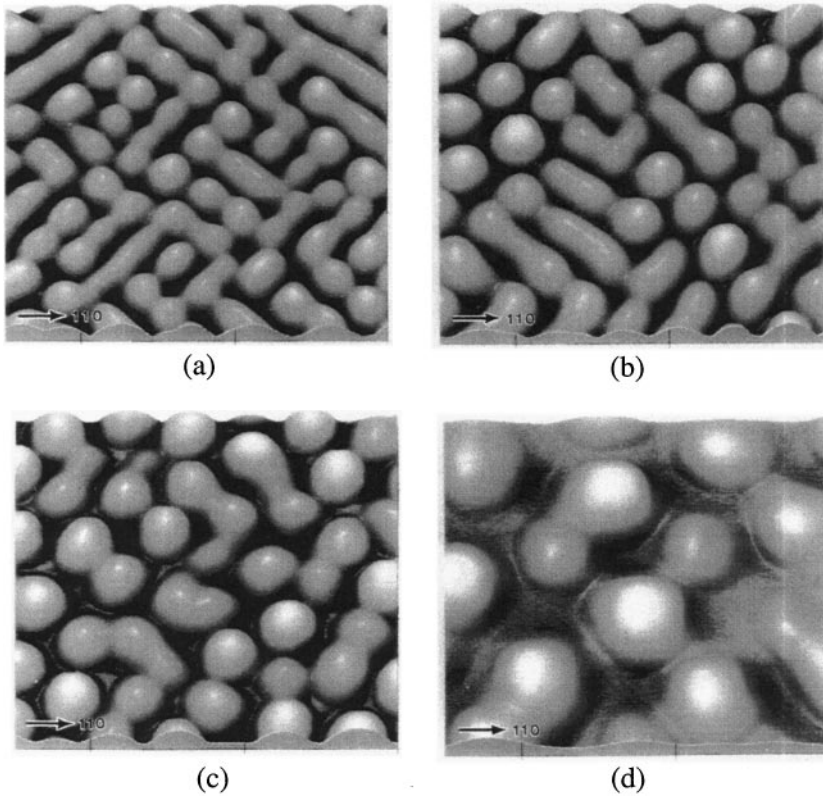
Figure 9   AFM contact mode images for 100-Å-thick $Si_{0.82}Ge_{0.18}$/Si films. The images reveal evolving surface patterns on samples annealed at 850°C for (*a*) 1 min, (*b*) 5 min, (*c*) 20 min, and (*d*) 2 h, respectively (23).

2 h. No dislocation activity is observed from the beginning of surface roughening with formed ridges oriented along ⟨100⟩-type directions to the later stage of isolated large islands.

The effect of surface roughening on strain relaxation of 100-Å-thick 18% Ge films is summarized in Figure 11, which shows X-ray rocking curve measurements for the 700°C and 850°C annealing series. The peak position for the as-grown film is within 0.3% of the value estimated for a dislocation-free, perfectly flat film. After 5 min of annealing at 850°C, at which point extensive two-dimensional ridge formation was observed, the nominal film strain was relaxed by 42%. After 2 h at 850°C, the film was almost fully relaxed, to about 93%.
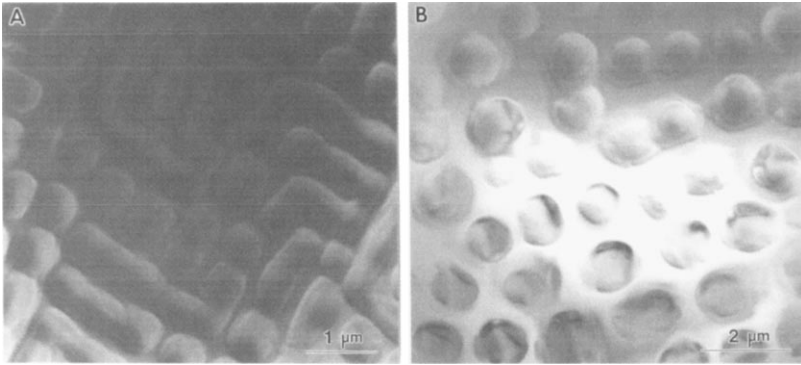
Figure 10   Bright-field plan view TEM image of a 100-Å-thick $Si_{0.82}Ge_{0.18}$/Si film annealed at 850°C for (*a*) 5 min and (*b*) 2 h. The images reveal that the surface-roughening proceeds with no dislocation formation in subcritical films (23).

At the temperature and time scales of the annealing experiments, the possibilities for germanium to diffuse out of the film into the underlying substrate or to agglomerate into nonuniform composition bands within the film (50) need to be considered. Energy-dispersive X-ray digital mapping on the (110) cross-section samples has been conducted (24), with results showing neither germanium interdiffusion nor phase separation within the film.
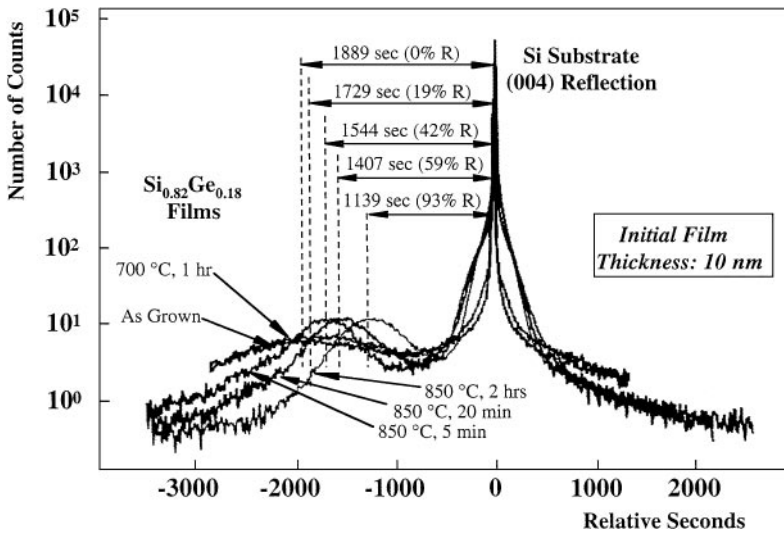
Figure 11   X-ray rocking curve measurements for 100-Å-thick 18% Ge films (23).

In summary, the controlled annealing experiments on subcritical $Si_{0.82}Ge_{0.18}$/ Si(100) films have shown that surface roughening occurs in dislocation-free films and proceeds in a sequence of three- $\rightarrow$ two- $\rightarrow$ three-dimensional patterns, as predicted by the surface-roughening theory (34). At the initial stage of roughening, undulations in different directions along the surface lead to three-dimensional budding of small islands along the film surface. As the roughness amplitude grows, the three-dimensional budding is replaced by the cusp formation process. The dominant wavelength selected by surface diffusion is sufficient for two-dimensional cusping morphology but is too small for three-dimensional cusping morphology. This causes the initial three-dimensional budding to evolve into two-dimensional ridge formation observed in experiments. The surface eventually evolves into three-dimensional isolated islands owing to further instability of the two-dimensional ridges.

Controlled Annealing of Supercritical Thin Films

Having established the surface-roughening process as a mechanism independent of dislocation activity by experimenting with subcritical films, it is now insightful to consider the interaction between surface roughening and defect formation in supercritical films. For this purpose, controlled annealing experiments at 700°C and 800°C were conducted on heteroepitaxial films 500 Å in thickness and containing 22% Ge. The annealing was done under hydrogen flow in the CVD reactor.

As shown in Figure 12*a*, the as-grown film is flat and dislocation free. The film thickness is about fivefold larger than the corresponding critical thickness for dislocation formation. In this case, kinetic limitations on surface roughening and dislocation formation have resulted in flat, defect-free as-grown films with thicknesses much larger than Matthews' critical thickness. Figure 12*b–d* shows a sequence of bright-field cross-section images for anneal series at 800°C for 5 min, 20 min, and 1 h, respectively. The contrast distribution in these TEM images was convoluted by the strain distribution in the film, the changing film thickness, and the bending of the sample. Figure 12*b* reveals shallow undulations with varying amplitude and wavelength, with various defects observed in the film. The film/substrate interface remains flat. Figure 12*c* reveals a sharp cycloid-like valley region, where the stress concentration is very high. Defects are expected to form in such regions. Figure 12*d* shows isolated islands after 1 h at 800°C. At this late stage of roughening, the tops of the islands are almost stress free and are becoming flat, as facets tend to form owing to surface energy minimization.

AFM images of surface evolution in supercritical films have revealed surface roughening along ⟨110⟩-type directions, in contrast to what was observed in

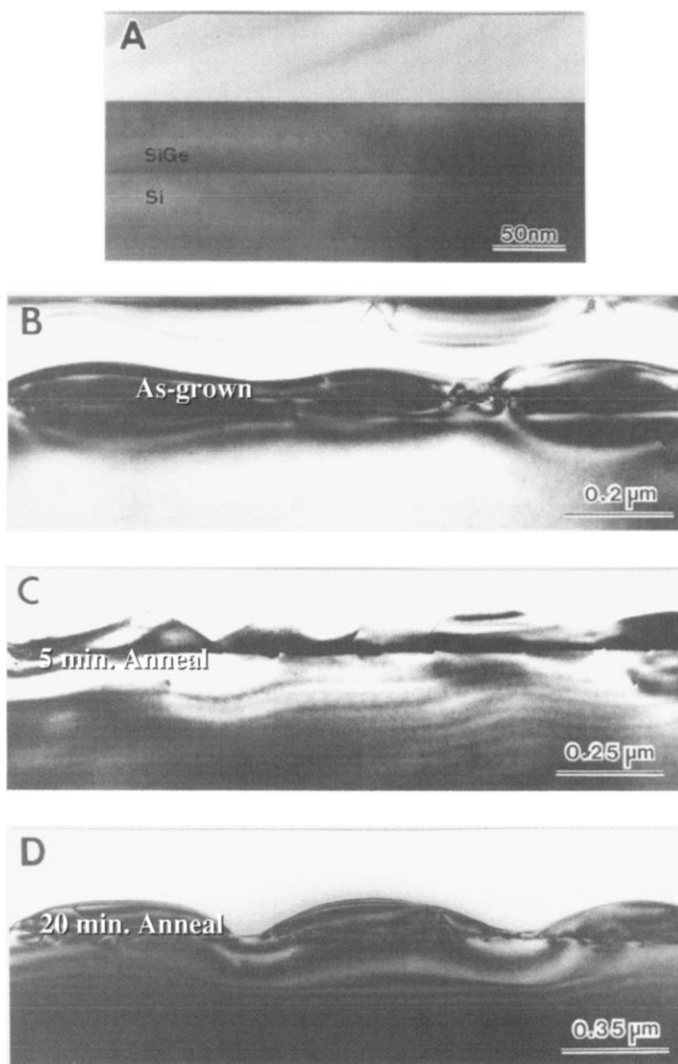*Figure 12* Cross-section TEM images of a 500-Å-thick $Si_{0.78}Ge_{0.22}/Si$ film in the (*a*) as-grown stage and (*b–d*) annealed at 800°C for 5 min, 20 min, and 1 h, respectively (23).

subcritical films, where the undulations were observed to align along $\langle 100 \rangle$-type directions. This apparent paradox has been resolved by observing the initial stage of surface evolution in defect-free, flat as-grown supercritical films. Figure 13a–d shows annealing experiments conducted on 22% Ge films with 500-Å as-grown thickness, at temperatures of 750 and 800°C for short periods of time. The surface of the sample shown in Figure 13a was annealed at 750°C for 10 s right after the growth process in the reactor. This image indeed reveals surface ridge formation along $\langle 100 \rangle$-type directions, which is consistent with the surface-roughening theory (6, 24, 38) and the observations on subcritical films. The corresponding plan view TEM image for this sample is shown in
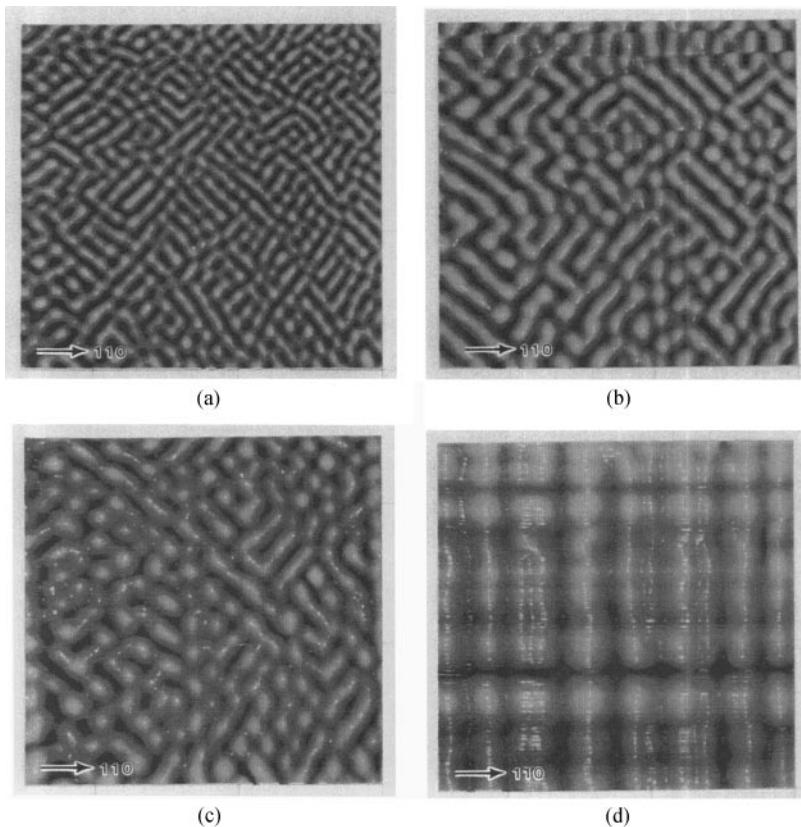
Figure 13   AFM images of a 500-Å-thick $Si_{0.78}Ge_{0.22}/Si$ film annealed at (*a*) 750°C for 10 s, (*b*) 800°C for 10 s, (*c*) 800°C for 1 min, and (*d*) 800°C for 5 min. These anneal images reveal the rotation of ridges from $\langle 100 \rangle$ to $\langle 110 \rangle$ type directions (38).
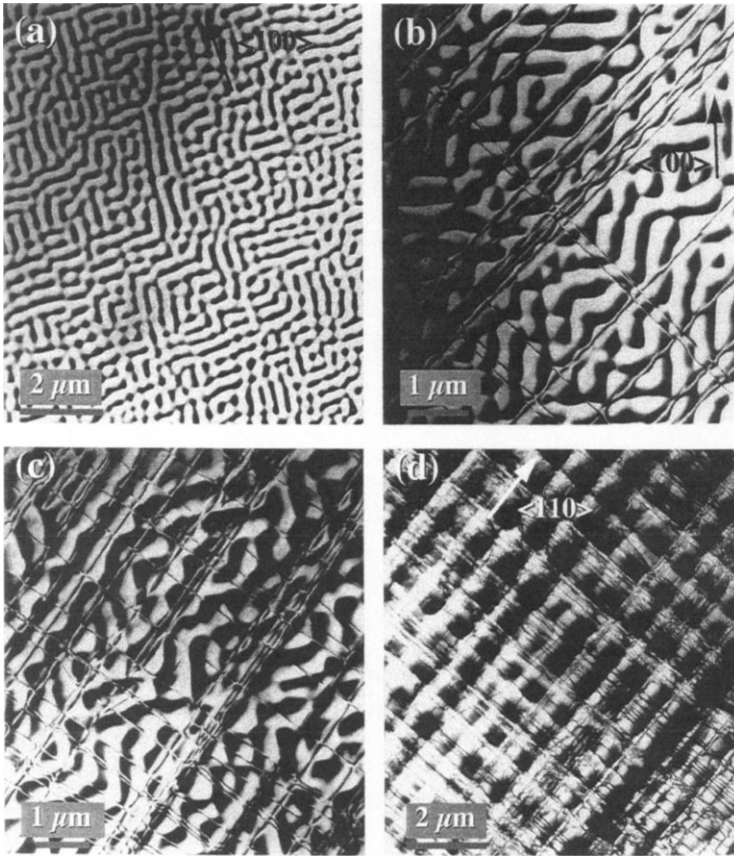
Figure 14   Plan view TEM images of a 500-Å-thick $Si_{0.78}Ge_{0.22}$/Si film annealed at (*a*) 750°C for
10 s, (*b*) 800°C for 10 s, (*c*) 800°C for 1 min, and (*d*) 800°C for 5 min.

Figure 14*a*, showing that no dislocations were formed in the film at this stage.
Therefore, the surface roughening precedes any dislocation formation. The
next image in Figure 13*b* was obtained after annealing at 800°C for 10 s, and it
reveals propagation of threading dislocations that shear through the $\langle 100 \rangle$ ridges
and cause them to rotate toward $\langle 110 \rangle$ directions. The corresponding plan view
TEM image in Figure 14*b* shows a cross-grid of misfit dislocations aligned along
$\langle 110 \rangle$ directions, although most of the ridges are still along $\langle 100 \rangle$ directions. It
has been shown (49) that misfit dislocations at the film/substrate interface can
cause the surface to undulate in the direction of dislocation alignment because
the dislocations change the strain distribution along the surface. Figures 13*c*

and 14*c* show that more surface ridges become aligned with the network of misfit dislocations as dislocation density increases, and Figures 13*d* and 14*d* indicate that the rotation process has reached near completion after annealing at 800°C for 5 min.

In summary, the annealing experiments on supercritical heteroepitaxial films have shown that the process of surface roughening and cusp formation plays an important role in the generation of misfit dislocations. Dislocation-free, flat as-grown films are observed to roughen into two-dimensional ridges along ⟨100⟩ directions before misfit dislocations are generated. This is consistent with the theoretical calculations on the anisotropic effects on surface roughening. The misfit dislocations change the distribution of strain and chemical potential along the film surface and force the surface ridges to coalign with them in ⟨110⟩ directions.

Formation of Defects in Thin Films

A most remarkable consequence of surface roughening is that it results in the formation of cycloid-like surface valleys with very large stress concentrations (9, 14, 24), which can significantly reduce the activation energy for defect formation (24). Figure 15*a* shows a cross-sectional bright-field TEM image of such a region in a 500-Å-thick, 22% Ge film annealed at 800°C for 20 min. This image shows a very sharp cycloid-like surface valley, which is a potential site for defect nucleation. The generation of misfit dislocations during two-dimensional cusp formation has been shown in Figures 13 and 14. Calculation of the activation energy (24) shows that dislocation half loops can readily nucleate at surface cusps with little or no activation energy barrier. Expansion of such loops leads to propagating threading dislocations which create misfit dislocations along the film/substrate interface, as illustrated in Figure 5*b*.

Other types of defects are also observed to form near the cycloid-like valleys. Figure 15*b* shows images of two types of nanoscale defects observed at the tip of the cusp valley shown in Figure 15*a*. The first defect is called a V-shaped defect (in this case a double defect), and the second defect is called a wedge-shaped defect. This clearly shows the role of surface roughening in nucleating defects at sharp valley regions with large stress concentrations. The V-shaped defect at cusps (51) contains a 90° stair rod dislocation at its tip and two stacking faults on the side, which extend to the surface (24, 47). The nucleation of this defect has been discussed via continuum dislocation theory (47) and atomistic mechanisms near surface ledges along a diamond cubic crystal (52). The wedge-shaped defect can be modeled as wedge disclinations, which form through a sequence of successive steps of nucleation and growth (53). The wedge-shaped defect shown in Figure 15*b* is in the final stage of the growth process, and it was observed after an annealing experiment at 800°C for 5 min
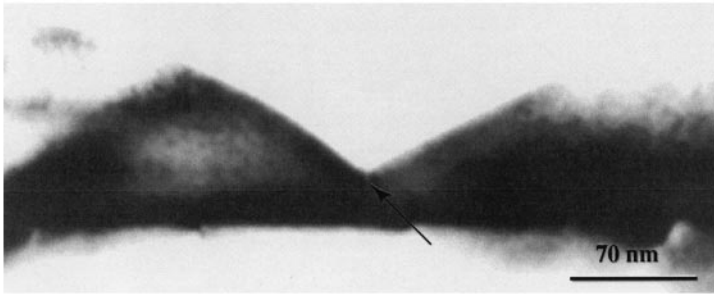
(a)
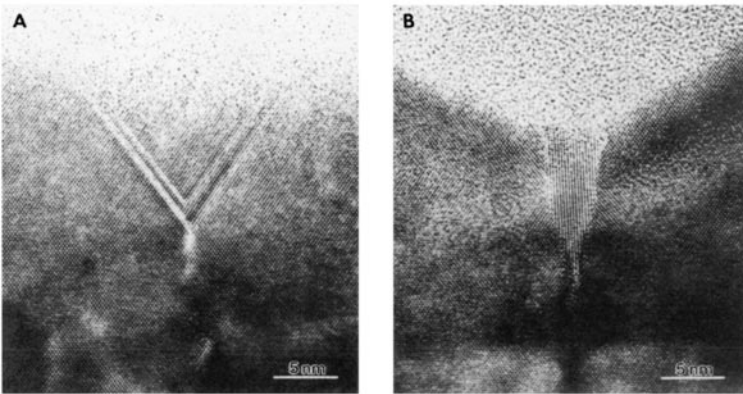
(b)

Figure 15 (*a*) Cross-section TEM image of a 500-Å-thick $Si_{0.78}Ge_{0.22}$/Si film annealed at 800°C for 20 min, showing the formation of a cycloid-like valley with high stress concentration. (*b*) High-resolution cross-section TEM images of a V-shaped defect (*A*) and a wedge-shaped defect (*B*) (23).

(23). Similar defects that resemble the earlier stages of the nucleation process were also observed in pure Ge films (54, 55).

Surface Roughening in Tensile Films

It has been questioned (56) whether surface roughening also occurs in films with a positive mismatch strain, even though the theory described earlier predicts roughening regardless of the sign of the stress in the film. Experiments with tensile Si-Ge alloy films grown on higher composition graded Si-Ge alloy

substrates in the UHV environment of an MBE reactor did not reveal surface roughening (56). In view of this controversy, annealing experiments with $Si_{0.5}Ge_{0.5}$/Ge films 150 Å in thickness were conducted in the hydrogen environment and compared with similar experiments with $Si_{0.5}Ge_{0.5}$/Si films (24). For a germanium substrate, where the film is under a tensile mismatch strain, surface roughening was indeed observed. It was observed that the average amplitude and wavelength of the surface islands on $Si_{0.5}Ge_{0.5}$/Si(100) films are 940 Å and 4680 Å, respectively, whereas they are 0.24 μm and 1.4 μm on $Si_{0.5}Ge_{0.5}$/Ge(100) films. From the surface-roughening theory, the wavelength of surface features is inversely proportional to the strain energy density at the surface, hence inversely proportional to the elastic stiffness of the film at a given mismatch strain. A possible explanation for the observed difference in size scales of surface features on $Si_{0.5}Ge_{0.5}$/Si(100) and $Si_{0.5}Ge_{0.5}$/Ge(100) films is that materials tend to elastically "soften" under large tensile stresses and "stiffen" under large compressive stresses (24).

Effect of Capping Layer on Surface Roughening

A series of experiments were conducted to investigate 100-Å-thick $Si_{0.82}Ge_{0.18}$/ Si(100) films with a capping layer of various thickness, subjected to a controlled annealing process at 850°C for 5 min (48). These films are subcritical with respect to dislocation formation, and their surfaces are flat in the as-grown state. The capping layer thickness ranges from 0 to 100 Å. The experiments show that the surface roughness becomes smaller as the thickness of the capping layer increases. The average amplitude of surface ridges drops from 510 Å in the absence of a capping layer to 220 Å with a capping layer of 30 Å. This may be intriguing because one may think that a stress-free capping layer with a few monolayers should be sufficient in suppressing the surface roughening. A possible explanation of this behavior is that germanium may be diffusing tens of angstroms into the capping layer, which will cause a stress to develop within the capping layer. The surface roughening is completely suppressed by a capping layer of 100 Å on the film subjected to annealing at 850°C for $\leq$1 h.

DIFFUSION TRAPPING OF DISLOCATIONS IN HETEROEPITAXIAL FILMS

Why do multiple types of defects occur in heteroepitaxial films? What are the mechanisms by which each type of these defects is nucleated? Dislocation nucleation in thin films has been traditionally modeled as a thermally activated gliding process (26, 27, 44, 57, 58) in which the location of the free surface is assumed to be fixed in material coordinates, and a dislocation loop expands on a preferred glide plane after nucleation. However, if mass transport along the

surface is taken into account, as it occurs in the evolution of surface morphology in unstable strained layer systems, dislocations can also be formed without glide through rearrangement of the material. Mass transport occurs through motion of atomic ledges or steps at the crystal surface. A scenario may be envisaged in which a dislocation is generated at a surface feature such as a ledge. Subsequently, this dislocation is trapped behind the free surface, not by glide of the dislocation but by motion of the surface caused by surface diffusion and/or deposition. Therefore, two classes of dislocation motion at rough surfaces of a heteroepitaxial film should be considered (34). The first class consists of dislocations that form by a glide mechanism and are able to move away from the surface on a glide plane under the action of the internal stress associated with the mismatch strain. The second class consists of sessile dislocations that can form only through the addition or rearrangement of mass on the surface.

A few interesting ideas have emerged from studies of dislocation formation by surface diffusion. One idea (59) involves an evolving surface joining behind a deepening valley and trapping the valley region as a hollow dislocation core with a super Burgers vector, somewhat like water bubbles trapped at an agitated surface. Another suggestion (44) is that a dislocation is generated by shear of the lattice in the vicinity of an atomistically sharp cusp tip; the dislocation is punched out of the cusp by an appropriately resolved shear stress. These mechanisms are all feasible from a theoretical point of view. However, there has been neither experimental evidence nor quantitative theoretical calculations to support the suggestion that an evolving surface will ever tend to trap a deepening valley as a super hollow dislocation. Furthermore, experimental observations have indicated that dislocation nucleation occurs long before the cusp becomes atomistically sharp. An explanation of the experimental observations on defect formation in heteroepitaxial films is provided by the mechanism of dislocation trapping (52) via diffusional mass transport. This is motivated by the following considerations. During stress-controlled mass transport, atoms move away from surface valleys toward surface peaks. The local elastic strain near a surface valley continues to increase in magnitude as the local curvature of the surface increases. Figure 16a shows a surface evolving toward a cusp-like (cycloid-like) morphology. At a critical level of mean strain, bifurcation of the surface evolution becomes possible with evolution continuing toward a cusped morphology along one branch and with this process interrupted by dislocation formation along the other branch. From an atomistic point of view, as the surface valley deepens (and sharpens), atoms leave behind double-ledge-like structures shown in Figure 16b. If such double ledges are stable, the process leading to cusp formation continues. On the other hand, when the strain becomes sufficiently large owing to geometrical amplification, the double-ledge
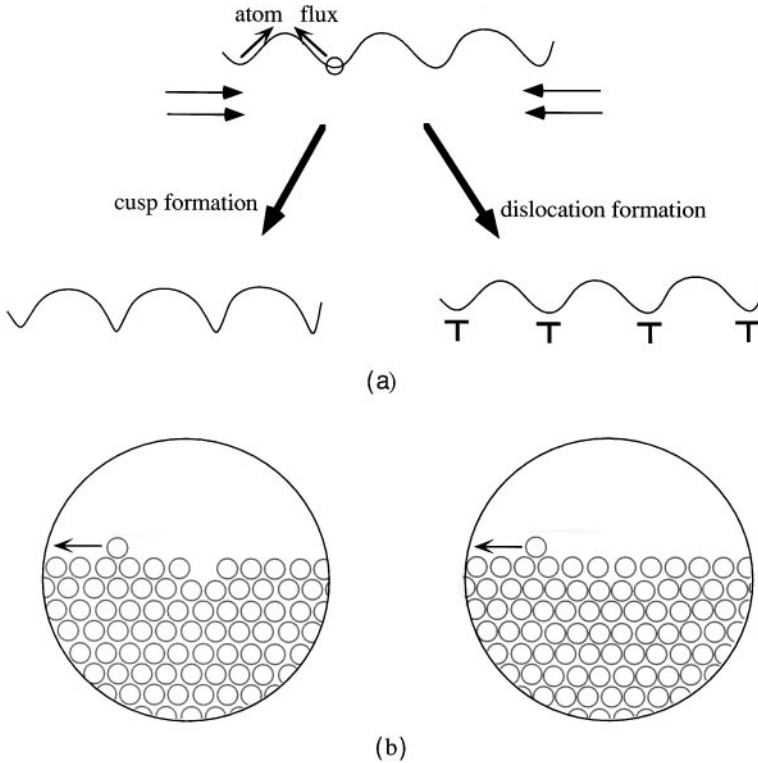
Figure 16   (*a*) Bifurcation of a cusp-forming surface instability into dislocation trapping. (*b*) The stability of an atomic double ledge at a surface valley (52).

structures may close or collapse to form dislocations. Dislocation formation causes the local strain energy in the surface layer to be reduced, thus reducing the driving force for cusp formation. Subsequently, atomic flux is reversed so that atoms start to accumulate at the valley, and the dislocation just formed becomes trapped as a bulk defect. In this mechanism of dislocation formation, the stability of atomic double-ledges becomes a fundamental criterion for determining whether the surface should continue evolving along a path toward cusp formation or it should select the alternate path involving dislocation trapping.

Dislocation Formation at a Double Ledge: the V-Shaped Defect

Dislocation formation at an atomic double-ledge structure has been studied, using molecular dynamics simulation on a model Lennard-Jones crystal (60). It

was found that the double ledge unstably collapses to form a one-dimensional Frenkel-Kontorova dislocation (61) at 6–7% strain, representing a misfit between the surface layer and the underlying substrate. This study suggested the following mechanism for the formation of the V-shaped defect in diamond cubic crystal: First a sessile 90° Lomer-Cottrell dislocation is introduced with a Burgers vector of $a/2\langle 110 \rangle$ owing to collapse of a double ledge as shown in Figure 17. During the surface roughening, atoms move away from surface
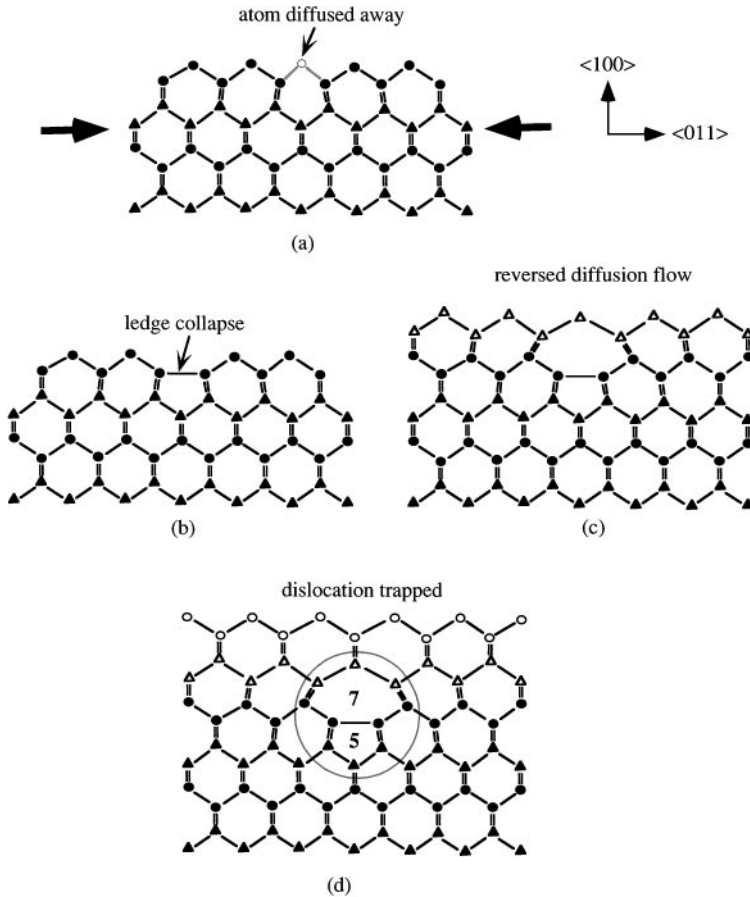
Figure 17   Atomistic model of formation of a perfect Lomer-Cottrell dislocation near a double-ledge configuration. (*a*) Outgoing atomic flux creates an atomic gap (double ledge) at a surface valley; (*b*) the double ledge collapses owing to the background compressive stress; (*c*) the reversed atomic flux traps the collapsed double ledge and begins to form a vertical 5–7 ring structure; (*d*) the 5–7 ring is trapped as a perfect Lomer-Cottrell dislocation by reverse diffusional flow (52).

valleys toward surface peaks. The local elastic strain near the valley continues to increase in magnitude as the local curvature of the surface increases, and atoms leave behind double-ledge-like structures shown in Figure 17a. This corresponds to a missing row of atoms in the surface layer (a double ledge) which are assumed to have diffused away toward a neighboring peak region. If such double ledges are stable, the process leading to cusp formation continues. On the other hand, when the strain becomes sufficiently large owing to geometrical amplification, the double-ledge structures may close or collapse. This is shown in Figure 17b. The ledge collapse causes the local strain energy in the surface layer to be reduced, thus reducing the driving force for cusp formation. Subsequently, atomic flux is reversed so that atoms start to accumulate at the valley, as shown in Figure 17c. Further reversed diffusional flow traps a vertical 5- to 7-ring structure, as in Figure 17d. This 5–7 ring characterizes the core structure of a Lomer-Cottrell dislocation (62). In the process described, the Lomer-Cottrell dislocation just formed becomes trapped as a bulk defect.

The experimentally observed V-shaped defect consists of an a/6 ⟨110⟩ stair-rod dislocation and two a/6 ⟨112⟩ Shockley partial dislocations inclined at 60°, which are separated from the stair-rod dislocation by stacking faults (47). This is caused by an unstable transformation of the Lomer-Cottrell dislocation into stair-rod and Shockley partial dislocations with stacking faults. Because the Lomer dislocation forms near the surface layer, the free surface exerts a strong image force on the core and draws out a pair of Shockley partial dislocations to the surface, leaving stacking faults behind. An analysis of the stability of a Lomer-Cottrell dislocation near a free surface shows that there is a critical compressive strain beyond which a perfect Lomer-Cottrell dislocation is stable against the free-surface–induced emission of Shockley partials (47). It is also confirmed that the stresses in the films in the annealing experiments are below this critical level so that the formation of V-shaped defects is predicted.

Dislocation Formation at a Two-Atom-Gap Double Ledge: the Wedge Defect

The collapse of a double ledge near a cycloid-like surface valley has provided an explanation of the V-shaped defect. A remaining question is why the wedge-shaped defect shown in Figure 15b is also observed. The wedge-shaped defect has been analyzed by using the continuum concept of disclinations (53). The atomic structure of the early stage of this defect has been investigated in connection with experimental observations during deposition of Ge on Si (54, 55). A possible atomistic sequence that distinguishes the wedge-shaped defect from the V-shaped defect is shown in Figure 18. Suppose, instead of one missing row of atoms near a valley region of high stress concentration, two rows of atoms are missing near such a region, as in Figure 18a. Under the combined effects
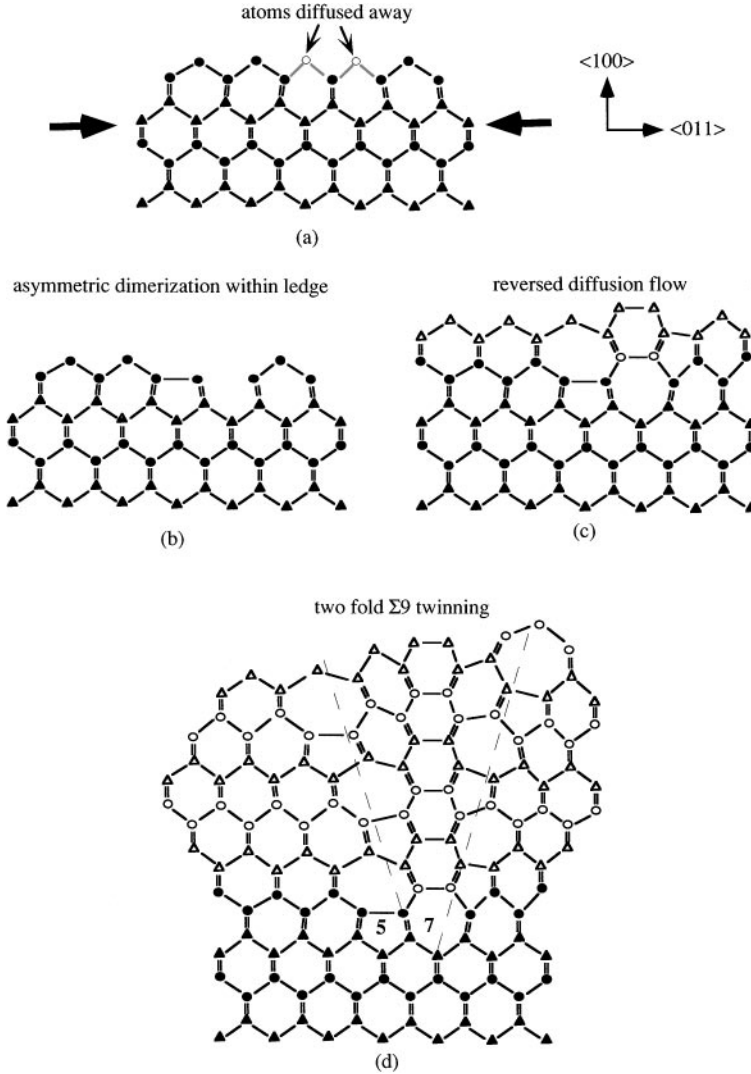
atoms diffused away

<100>

<011>

(a)

asymmetric dimerization within ledge

reversed diffusion flow

(b)

(c)

two fold Σ9 twinning

5   7

(d)

Figure 18   Atomistic model of formation of a twofold Σ9 twinning wedge near a two-atom–gap double-ledge configuration. (*a*) Outgoing atomic flux creates a two-atom–gap double ledge (at a surface valley); (*b*) part of the ledge collapses owing to the background compressive stress; (*c*) the reversed atomic flux traps the collapsed ledge and begins to form a horizontal 5–7 ring structure along the surface; (*d*) diffusive trapping of the horizontal 5–7 ring leads to the twofold Σ9 twinning wedge (52).

of high stress and high temperature, a dimer-like collapse occurs between the protruding row of atoms in the center of the gap and one on the ledge, which is taken to be the left one in Figure 18*b*. The entrapment of this event by reversed diffusional flow is the starting point at which the wedge defect differs from the steps leading to the Lomer dislocation in Figure 17. The collapsed portion of the ledge causes a widening of the rest of the original gap so that, instead of accommodating one atom as in the formation of a Lomer-Cottrell dislocation, it is wide enough to accommodate two atoms during reversed diffusional growth, as illustrated in Figure 18*c*. In contrast to the vertical 5- to 7-ring structure that leads to the Lomer-Cottrell dislocation, a horizontal 5–7 ring is formed in the situation and subsequent events lead to a twofold $\Sigma 9$ twinning structure.

CONCLUSIONS

We have discussed some of the recent theoretical and experimental studies on surface roughening in lattice-mismatched heteroepitaxial films. On the theoretical side, the surface roughening is understood as a consequence of the competition between strain energy and surface energy in heterostructures. The basic length scale of roughening is the ratio between the surface energy density and the strain energy density of the film. The time scale of roughening is associated with the kinetics of surface diffusion, which are very fast owing to the conditions of high temperature, nanometer-scale thicknesses, and large lattice mismatch in heteroepitaxial films. A most remarkable feature of the surface roughening is the formation of cycloid-like cusps at surface valleys with singular stress concentration at the cusp tip. The essential features of surface roughening are described by a cycloid model based on closed-form solutions to the elasticity problem of a cycloid surface.

On the experimental side, we have focused our discussion on a series of controlled annealing experiments on Si-Ge/Si heterostructures, which were conducted under the guidance of the surface-roughening theory. Annealing experiments on films with thicknesses below Matthews' critical thickness for dislocation formation have shown that surface roughening occurs in dislocation-free films and that the surface roughness proceeds in a three-dimensional (budding) → two-dimensional (ridges) → three-dimensional (islands) pattern. The latter observation is explained by the theoretical result that the dominant wavelength of surface roughening under surface diffusion is sufficient for two-dimensional cusp formation but too small for three-dimensional cusp formation. Annealing experiments on defect-free, flat as-grown films with thicknesses larger than Matthews' critical thickness have shown that surface roughening precedes dislocation formation, that formation of misfit dislocations causes the

surface ridges to rotate from ⟨100⟩ directions to ⟨110⟩ directions, and that multiple types of defects form at cycloid-like surface valleys. This study is consistent with the anisotropic theory of surface roughening that predicts that the surface ridges in a dislocation-free film should align in ⟨100⟩ type directions. It also indicates that the later stage of surface roughening in heteroepitaxial films is also affected by the presence of misfit dislocations.

Surface roughening plays an important role in the nucleation of various defects in heteroepitaxial films. Dislocation theory has been used to understand the processes by which various observed defects form. The formation of sessile dislocations at surface valleys is explained by diffusional entrapment of dislocations at evolving surfaces. In this process, an atomic-scale dislocation embryo is generated as a surface ledge collapses near regions of large stress concentration. The local stress relaxation from dislocation formation causes reversed diffusional flow that traps the dislocation behind the free surface. Note that diffusive trapping of dislocations does not require glide dislocations to be nucleated on a preferred slip plane. We have discussed atomistic models for the entrapment of sessile defects observed near cycloid-like surface valleys in heteroepitaxial films.

FUTURE RESEARCH

Although the basic mechanism of surface roughening in heteroepitaxial films is now reasonably well understood, there are several important aspects that are worth further investigation.

1. Annealing experiments conducted under the UHV conditions in an MBE reactor (44) exhibit very different surface features compared with those under the hydrogen environment in a CVD reactor. This suggests that surface reconstruction along semiconductor surfaces can dramatically affect the surface-roughening process. We think that hydrogen saturation of dangling surface bonds has prevented such reconstruction from occurring in experiments conducted under the hydrogen environment, which explains why these experiments are consistent with the surface-roughening theory. However, since many nanoscale devices are grown by MBE, it is important to conduct more theoretical and experimental studies on surface roughening in a UHV environment.

2. Surface roughening is a promising process for fabricating quantum dot devices. It is possible to fabricate these structures by controlled annealing of very thin, defect-free as-grown films. Further studies on island formation are necessary (63, 64). Development of numerical simulation tools with three-dimensional simulation capabilities will be important in optimizing

the fabrication processes. One of the most important factors to investigate is the size distribution of islands as a function of annealing parameters. Another important task is to investigate the diffusion of Ge at the nanoscale level.

3. More controlled annealing experiments are necessary on heteroepitaxial films with a positive lattice mismatch, both in hydrogen and in UHV environments. It will be important to investigate the mechanisms of nanoscale defect formation in the presence of a tensile biaxial stress. It seems that diffusive trapping of dislocations can also occur in tensile films.

4. It is of interest to investigate methods to suppress the surface roughening. The use of capping layers (48) is one example of such methods. It is also interesting to explore the possibilities of using other mechanisms such as piezoelectric coupling (65) and substrate effects (13) to stabilize a flat surface.

**Visit the *Annual Reviews* home page at
http://www.AnnualReviews.org**

Literature Cited

1. Kordos P, Novak J, eds. 1998. *Heterostructure Epitaxy and Devices*. Boston: Kluwer
2. Pearsall TP, ed. 1990. *Strain Layer Superlattices*. Boston: Academic
3. Asaro RJ, Tiller WA. 1972. *Metall. Trans.* 3:1789
4. Grinfeld MA. 1986. *Sov. Phys. Dokl.* 31:831
5. Srolovitz DJ. 1989. *Acta Metall.* 37:621
6. Gao H. 1991. In *Modern Theory of Anisotropic Elasticity and Applications*, ed. JJ Wu, TCT Ting, DM Barnett. p. 139. Philadelphia: SIAM
7. Spencer BJ, Voorhees PW, Davis SH. 1991. *Phys. Rev. Lett.* 67:3696
8. Freund LB, Jonsdottir F. 1993. *J. Mech. Phys. Solids* 41:1245
9. Chiu Ch, Gao H. 1993. *Int. J. Solids Struct.* 30:2983
10. Yang WH, Srolovitz SJ. 1993. *Phys. Rev. Lett.* 71:1593
11. Spencer BJ, Meiron DI. 1994. *Acta Metall.* 42:3629
12. Chiu Ch, Gao H. 1994. *MRS Symp. Proc.* 317:369
13. Chiu Ch, Gao H. 1995. *MRS Symp. Proc.* 356:33
14. Chiu Ch. 1995. *Diffusive cusp-crack initiation on a stressed solid.* PhD thesis, Stanford Univ.
15. Eaglesham DJ, Cerullo M. 1990. *Phys. Rev. Lett.* 64:1943
16. Guha S, Madhukar A, Rajkumar KC. 1990. *Appl. Phys. Lett.* 57:2110
17. Snyder CW, Orr BG, Kessler D, Sander LM. 1991. *Phys. Rev. Lett.* 66:3032
18. Tersoff J, LeGoues FK. 1994. *Phys. Rev. Lett.* 72:3570
19. Yao JY, Andersson TG, Dunlop GL. 1988. *Appl. Phys. Lett.* 53:1420
20. Cullis AG, Robbins DJ, Pidduck AJ, Smith PW. 1991. *J. Cryst. Growth* 123: 333
21. Pidduck AJ, Robbins DJ, Cullis AG, Leong WY, Pitt AM. 1992. *Thin Solid Films* 222:78
22. Jesson DE, Pennycook SJ, Baribeau JM, Houghton DC. 1993. *Phys. Rev. Lett.* 71: 1744
23. Ozkan CS, Nix WD, Gao H. 1997. *Appl. Phys. Lett.* 70:2247
24. Ozkan CS. 1997. *Studies of morphological instability and defect formation in heteroepitaxial Si(1 − x)Ge(x) thin films via controlled annealing experiments.* PhD thesis, Stanford Univ.
25. Matthews JW, Blakeslee AE. 1974. *J. Cryst. Growth* 27:118
26. Freund LB. 1987. *J. Appl. Mech.* 54: 553
27. Nix WD. 1989. *Metall. Trans.* 20A:2217

28. Nix WD, Noble DB, Turlo JF. 1990. *MRS Symp. Proc.* 188:315

29. Gillard VT, Noble DB, Nix WD. 1992. *MRS Symp. Proc.* 239:395

30. Gillard VT. 1995. *Kinetics of strain relaxation through the formation of misfit dislocations in $Si_{1-x}Ge_x$ heteroepitaxial films.* PhD thesis, Stanford Univ.

31. Alivisatos AP. 1998. *MRS Bull.* 23:18

32. Zunger A. 1998. *MRS Bull.* 23:15

33. Gao H. 1991. *J. Mech. Phys. Solids* 39:443

34. Gao H. 1994. *J. Mech. Phys. Solids* 42:741

35. Grinfeld MA. 1993. *J. Nonlinear Sci.* 3:35

36. Gao H. 1991. *Int. J. Solids Struct.* 28:703

37. Tersoff J. 1995. *Phys. Rev. Lett.* 74:4962

38. Ozkan CS, Nix WD, Gao H. 1997. *MRS Symp. Proc.* 436:487

39. Rice JR. 1968. In *Fracture*, ed. H Liebowitz. 2:191. New York: Academic

40. Gao H. 1995. *Proc. R. Soc. London Ser. A* 448:65

40a. Gao H. 1995. *Proc. R. Soc. London Ser. A* 450:732

41. Mullins WW. 1957. *J. Appl. Phys.* 28:333

42. Mullins WW, Sekerka RF. 1964. *J. Appl. Phys.* 35:444

43. Williams ML. 1952. *J. Appl. Mech.* 74:526

44. Jesson DE, Chen KM, Pennycook SJ. 1996. *MRS Bull.* 21:31

45. Deleted in proof

46. Ozkan CS, Nix WD, Gao H. 1996. *MRS Symp. Proc.* 399:407

47. Ozkan CS, Nix WD, Gao H. 1997. *MRS Symp. Proc.* 442:373

48. Ozkan CS, Nix WD, Gao H. 1997. *MRS Symp. Proc.* 440:323

49. Jonsdottir F, Freund LB. 1995. *Mech. Mater.* 20:337

50. Guyer JE, Voorhees PW. 1998. *J. Cryst. Growth* 187:150

51. Cullis AG, Pidduck AJ, Emeny MT. 1995. *Phys. Rev. Lett.* 75:2368

52. Gao H, Ozkan CS, Nix WD, Zimmerman JA, Freund LB. 1999. *Phil. Mag. A* 79:349–70

53. Muellner P, Gao H, Ozkan CS. 1997. *Phil. Mag. A* 75:925

54. LeGoues F, Copel M, Tromp R. 1989. *Phys. Rev. Lett.* 63:826

55. LeGoues F, Copel M, Tromp R. 1990. *Phys. Rev. B* 42:11690

56. Xie YH, Gilmer GH, Roland C, Silverman PJ, Buratt SK. 1994. *Phys. Rev. Lett.* 73:3006

57. Beltz GE, Freund LB. 1993. *Phys. Stat. Solidi B* 180:303

58. Li J, Chiu Ch, Gao H. 1994. *MRS Symp. Proc.* 317:303

59. Grilhe J. 1993. *Europhys. Lett.* 23:141

60. Zimmerman J, Gao H. 1996. *MRS Symp. Proc.* 399:401

61. Frenkel J, Kontorova T. 1938. *Phys. Z. SowjUn.* 13:1

62. Nandedkar AS, Narayan J. 1989. *Mater. Sci. Eng. A* 113:51

63. Spencer BJ, Tersoff J. 1997. *Phys. Rev. Lett.* 79:4858

64. Floro JA, Chason E, Twesten RD, Hwang RQ, Freund LB. 1997. *Phys. Rev. Lett.* 79:3946

65. Chien NY, Gao H, Herrmann G, Barnett DM. 1996. *Proc. R. Soc. London Ser. A* 452:527

CONTENTS