Direct Imaging of Surface Cusp Evolution During Strained-Layer Epitaxy and Implications for Strain Relaxation

D. E. Jesson and S. J. Pennycook

Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6030

J.-M. Baribeau and D. C. Houghton

Institute for Microstructural Sciences, National Research Council of Canada, Ottawa, Canada K1A OR6

We have directly imaged the evolution of surface cusps during strained-layer epitaxy. The cusps arise naturally as a result of gradients in the surface chemical potential. High stress concentrations at the cusp tip have important implications for strain relaxation in the film via dislocation nucleation.

The shape evolution of a thin overlayer film subjected to substrate induced misfit strain remains a central issue in strained-layer epitaxy. Although it is understood that flat surfaces are unstable to the formation of surface waves [1], little is known about the detailed shape or time evolution of surface features. In this Letter, we demonstrate the surprising result that strained overlayers are unstable to the formation of a surface cusplike morphology as a consequence of gradients in the surface chemical potential. The cusps are associated with high stress concentrations, and we consider the potential implications of this morphological development as a means of strain relaxation in the film via dislocation nucleation.

To study the surface profile evolution of a strained epitaxial layer, we have performed a $Si_{0.5}Ge_{0.5}/Si(001)$ molecular beam epitaxy (MBE) growth experiment in which 2 monolayer Ge marker layers were deposited at selected intervals. These marker layers were then imaged along [110] as bright lines in Fig. 1(a), using the Z-contrast technique [2], to provide a direct map of the evolving nonequilibrium surface morphology [3]. We believe that these data represent the first cross sectional morphological map detailing the coherent-to-incoherent transition of a strained layer under conditions of high supersaturation.

The morphological development in Fig. 1(a) can be broadly classified into three distinct regions. Initially, the surface profile is flat and develops into a wavelike modulation of amplitude 2 nm at a thickness of about 25 nm. Most surprising, however, is the development of cusplike features at about 40 nm. Similar images were obtained along the [100] projection, so cusps can be regarded as lines of intersection between coherent islands. The cusp shapes would seem to be directly related to strain relaxation in this system since the surface profile rapidly flattens after their appearance. A flat morphology then persists for the subsequent film growth.

To understand the evolution in Fig. 1, we consider the morphological development of the surface to be governed by gradients in the surface chemical potential [1,4],

$$\mu(x) = \mu^s + \gamma \Omega \kappa(x) + \sigma_\tau^2(x) \frac{\Omega}{2E} . \tag{1}$$

Here μ^s represents the chemical potential of the unstressed surface, Ω is the atomic volume, and γ is the surface free energy per unit area. The second term corresponds to the surface free energy contribution to the chemical potential and therefore involves the surface curvature $\kappa(x) = -h''[1 + (h')^2]^{-3/2}$, where $h(x)$ specifies the height of the surface at lateral position x (see Fig. 2). The third term involves the elastic energy contribution, where $\sigma_\tau(t)$ represents the local stress component tangential to the surface and E is the Young's modulus appropriate to the film surface. Physically, a gradient in μ along the surface will establish an atomic drift velocity leading to a rate of change in surface profile given by [1,4]
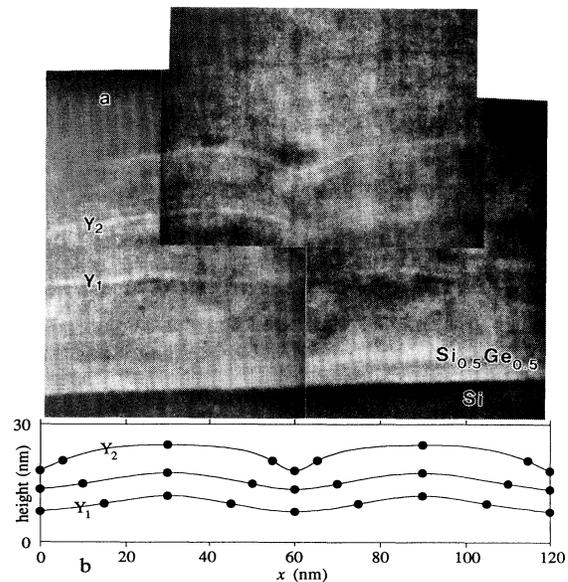

FIG. 1. (a) [110] Z-contrast image of a $Si_{0.5}Ge_{0.5}$ alloy grown by MBE with 0.25 nm Ge marker layers deposited at selected intervals. The growth temperature was 400 °C and the deposition rate 2 Å s^{-1}. The profile simulations in (b) correspond to the period of cusp formation in (a) between vertical ordinates Y_1 and Y_2. The solid circles denote the peak, valley, and inflection positions as determined from Eqs. (3), (4), and (5) (see text).
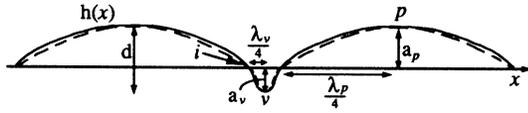
FIG. 2. The surface cusp profile $h(x)$ (solid line) is approximated at the peak p and valley v by the two sinusoidal half periods shown (dashed lines). The wavelengths λ_p,λ_v are related to the frequencies $u_p = 2\pi/\lambda_p$ and $u_v = 2\pi/\lambda_v$, respectively.

$$\frac{\partial h}{\partial t} = \frac{D_s\Omega}{kT}\eta[1+(h')^2]^{1/2}\frac{\partial^2\mu(x)}{\partial s^2} + \xi[1+(h')^2]^{-1/2}. \quad (2)$$

Here D_s is the surface diffusion coefficient, η the number of atoms per unit area, kT the thermal energy, and ξ the deposition rate. The derivatives with respect to s are along the surface.

To solve Eq. (2), we now make a number of simplifying assumptions consistent with our experimental data. This will lead to a qualitative but predictive model which embodies the essential physics governing the shape evolution of the surface. From Fig. 1(a) it can be seen that the surface profile is slowly varying and is essentially defined by three points: the profile peak p, the point of inflection i, and the valley v. If we approximate the profile extrema by two different sinusoidal half periods, as defined in Fig. 2, then it is possible to evaluate the curvature contributions to the chemical potential analytically. To model the elastic contribution, we note that for a sinusoid of wavelength λ, this is given to first order in h/λ^1 by $\sigma^2[1-(8\pi h/\lambda)\sin(2\pi x/\lambda)]\Omega/2E$. Modeling the initial roughness in Fig. 1(a) as a sine wave of amplitude 2 nm and wavelength 60 nm, this implies the elastic contribution at the peak is substantially reduced, and at the inflection point is close to $\sigma^2\Omega/2E$. In our model, we therefore assume that values of zero and $\sigma^2\Omega/2E$ are maintained at the peak and point of inflection, respectively, during growth, although the position of inflection will, of course, vary spatially as the profile evolves. The elastic contribution at the valley of depth d is similar to the compression of an elliptical crack producing an elastic contribution [5] of $\sigma^2[1+2(d/\rho_v)^{1/2}]^2\Omega/2E$. Here, ρ_v is the valley radius of curvature which in terms of the sinusoid parameters is given by $a_v^{-1}u_v^{-2}$. Within the scope of this model, it is then possible to solve Eq. (2) directly to obtain analytical expressions for the instantaneous growth velocities at the peak, valley, and point of inflection as

$$V_p = \xi - \frac{D_s\Omega^2\gamma\eta}{kT}[a_pu_p^4+3a_p^3u_p^6], \quad (3)$$

$$V_v = \xi + \frac{D_s\Omega^2\eta}{kT}\gamma a_vu_v^4+3\gamma a_v^3u_v^6 - \frac{\sigma^2}{E}a_v^{1/2}u_v^3d^{1/2}$$

$$- \frac{3\sigma^2}{E}a_v^{5/2}u_v^5d^{1/2} - \frac{2\sigma^2}{E}a_vu_v^4d - \frac{6\sigma^2}{E}a_v^3u_v^6d, \quad (4)$$

$$V_i = \xi[1+(h')^2]^{-1/2}, \quad (5)$$

with the symbols defined in Fig. 2. Equation (5) follows from the approximate antisymmetric nature of the profile

close to the point of inflection so that even derivatives vanish.

We can now follow the evolution of the surface profile away from an initial perturbation by applying Eqs. (3)–(5) for a short duration Δt. Using the experimental evidence that the profiles are smooth, it is then possible to fit new sinusoids, reapply Eqs. (3)–(5) for a further duration Δt, and hence track the evolving surface morphology as shown by the simulations in Fig. 1(b) [6]. It can be seen that our model qualitatively reproduces all of the features observed experimentally. This confirms that cusp formation is physically reasonable and that our images could form the basis for quantitative theoretical studies of nonlinear surface instabilities.

After the initial wetting layer, Fig. 1(a) indicates that the alloy grows with a metastable flat surface morphology up to a thickness of 25 nm at which a wavelike morphology develops. This observation is consistent with the far-from-equilibrium nature of MBE growth, in which the deposition term in Eq. (2) retards the morphological evolution defined by the gradient in surface chemical potential. As shown in Fig. 1(b), our model predicts that a sinusoidal profile of small amplitude will gradually evolve into a cusplike morphology. Although the curvature terms in Eqs. (3) and (4) tend to flatten the profile, the strain terms in Eq. (4) result in a net flux of atoms away from the valley towards the peak. The stress concentration at the valley is then further increased as the cusp sharpens which accentuates the migration. At a critical stage of the evolution, our model indicates that the cusp curvature and concomitant stress concentration accelerate rapidly.

We now consider the important implications of such stress concentration as a means of strain relaxation in the film via dislocation nucleation. The process by which dislocations nucleate in defect-free semiconductor epilayers is a fundamental and yet unresolved issue in strained-layer epitaxy. The basic conceptual difficulty arises because appreciable local stress concentrations are required to overcome the dislocation nucleation barrier [7–11]. In particular, for surface half-loop nucleation, the stress field must be sufficiently large to expand the loop beyond its critical radius. The appreciable stress concentrations associated with surface cusps, therefore, provide a natural explanation for how dislocations can nucleate in defect-free films. Consider, for example, a 60° half loop nucleating at the tip of the cusp and gliding along a (111) plane. For simplicity, we assume that the stress field will be dominated by σ_{xx} [12] so that using the stress solution of Sneddon [13], the total energy as a function of loop radius R_L is given by [14]

$$U(R_L) = R_L\frac{b^2\mu}{8}\frac{2-\nu}{1-\nu}\ln\left[\frac{\alpha R_L}{b}\right] - \frac{R_L\mu b^2}{4}\cos\beta$$

$$- \frac{C(R_L)}{4}\left[5\cos\frac{\phi}{2} - \cos\frac{5\phi}{2}\right]\cos\lambda\cos\phi. \quad (6)$$

Here, the first term specifies the energy cost of a dislocation of core parameter a and Burgers vector b in a medium of shear modulus μ and Poisson's ratio v. The second term specifies the energy gained by the removal of a step, β being the angle between the Burgers vector and the direction in the slip plane which is perpendicular to the intersection of the slip plane with the surface. The last term defines the elastic energy released by the loop, where λ and ϕ are the angles defined by Matthews [7]. Note in this case ϕ also appears as a polar coordinate in the crack stress field. The factor $C(R_L)$ depends upon

the model used for the radial (R) dependence of the stress field away from the crack tip. In the case of the stress solution [13] $\sigma_{xx} \propto (d/2R)^{1/2}$, then the calculation can be performed analytically to give

$$C^A(R_L) = 4.94 d^{1/2} R_L^{3/2} \mu \frac{1+v}{1-v} \varepsilon b . \tag{7}$$

Likewise, modeling the radial stress field of a "blunt" crack [15] as $\sigma_{xx} \propto \rho_v^{1/2}[1 + 2(d/\rho_v)^{1/2}]/(\rho_v + 4R)^{1/2}$ but retaining the angular (ϕ) dependence of the Sneddon solution in Eq. (6) yields

$$C^B(R_L) = \mu \frac{1+v}{1-v} \varepsilon b \rho_v^{1/2} [1 + 2(d/\rho_v)^{1/2}] \int_{-R_L}^{R_L} [\rho_v + 4(R_L^2 - z^2)^{1/2}]^{1/2} - \rho_v^{1/2} dz . \tag{8}$$

These two expressions should be compared with the expression appropriate for a flat surface [7]

$$C^C(R_L) = \pi R_L^2 \mu \frac{1+v}{1-v} \varepsilon b , \tag{9}$$

which, of course, requires $\phi = 0$ in Eq. (6). In Fig. 3, we have plotted the total energy [Eq. (6)] as a function of loop radius using the above expressions for $C(R)$ and a cusp geometry appropriate to the experimental data in Fig. 1. It can be seen that although the nucleation barrier at a flat surface is around 70 eV, nucleation at the cusp is effectively barrierless. These trends appear to be relatively insensitive to the crack model used, which suggests a further simplification in our interpretation. Since the critical loop radius R_c is appreciably reduced by the cusp stress field, then for $\rho_v \gg R_c$ in Eq. (8) we can consider nucleation at a cusp to be similar to dislocation nucleation at the flat surface of a film which is uniaxially stressed to the tip stress. If $\varepsilon^*(T)$ is the system dependent flat film strain at which the nucleation barrier becomes 37 kT, the estimated available nucleation energy [16], then the critical cusp geometry $G_c(T) = d/\rho_v$ giving rise to barrierless nucleation is
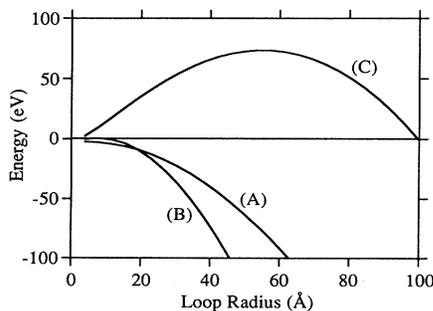
$$G_c(T) = \frac{1}{4} \left[\frac{\varepsilon^*(T)}{\varepsilon} - 1 \right]^2 . \tag{10}$$

$G_c(T)$ is plotted as a function of misfit ε in Fig. 4 for the Si-Ge system at 400 °C and 750 °C. The experimental critical geometry observed in Fig. 1 lies very close to the G_c (400 °C) curve for $\varepsilon = 0.02$. Images containing cusps can also be seen in Ref. [17] for $\varepsilon = 8.4 \times 10^{-3}$ but were interpreted as sinusoids by the authors. Estimating $d = 38$ nm and $\rho_v = 24$ nm, we find that in this case the cusp geometries are subcritical, lying below the 750 °C critical geometry curve in Fig. 4 as we would expect for a coherently strained film.

Cusp formation is clearly very sensitive to growth conditions [Eq. (2)]. A more appropriate definition of critical thickness in strained-layer epitaxy is that film thickness at which a critical cusp develops, having a surface stress concentration capable of nucleating a dislocation. In the case of large misfits, large stress concentrations can develop as a result of sharp cusps (small ρ_v) without the need for very large islands (large d), so that these features represent likely sources for the introduction of the first dislocations. It is also clear from Fig. 1 that a cusp can act as a multiple source over a distance of
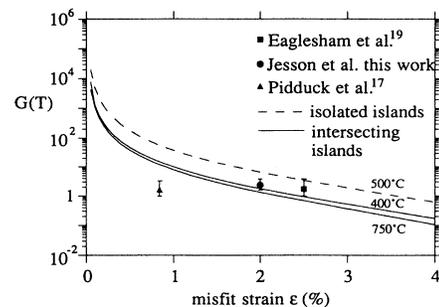
FIG. 3. Total energy curves for nucleation of a 60° half loop in a $Si_{0.5}Ge_{0.5}$ alloy as a function of loop radius R_L. The curves A and B correspond to the cusp functions $C^A(R_L)$ and $C^B(R_L)$, and C is the flat surface solution corresponding to $C^C(R_L)$ (see text). The cusp geometry ($d = 7$ nm, $\rho_v = 3$ nm) is appropriate to Fig. 1. The calculation assumes $a = 4$.

FIG. 4. The critical geometry $G(T)$ of intersecting (lower curves) and isolated islands (upper curve) required for the nucleation of 60° half loops. Relevant experimental data points are superimposed on the curves (see text).

perhaps 20 nm in the growth direction. Once the critical geometry is achieved, dislocations will be generated on closely bunched (111) planes, the rate of nucleation limited by glide away from the tip. For smaller misfits, it will take longer to reach the critical cusp geometry, so that heterogeneous sources will assume a greater importance [18].

Although it is clear from Fig. 1 that surface cusps arise as the intersection lines between coherent islands, we might also expect isolated islands to evolve in a similar way, governed by Eq. (2), so that strain accommodation could proceed as follows. Initially, cusps or isolated islands (half cusps) will develop as a result of gradients in the surface chemical potential. As discussed earlier, this evolution is sensitive to growth temperature, deposition rate, and misfit ε and will proceed to accentuate the stress concentration. If the growth experiment, including details of post growth quenching, is such that a stable cusp or island geometry is obtained at the wetting layer before the critical geometry is reached (Fig. 4), then coherent islands will be formed.

For isolated islands, stress concentrations will develop close to the island periphery where the surface radius of curvature ρ_I is at a minimum. The relevant stress concentration for an island of height d is $1 + G_I(T)^{1/2}$ where $G_I(T) = d/\rho_I$. In Fig. 4, we plot as a function of misfit the critical island geometry $G_I(T)$ [$= 4G_c(T)$] at which we might anticipate the introduction of dislocations. Large coherent Ge island geometries, as observed by Eaglesham and Cerullo [19] ($d = 50$ nm, $\rho_I = 30$ nm), lie below the 500 °C isolated island envelope for $\varepsilon < 0.03$ indicating that at least 1% of the misfit is accommodated by substrate distortion, which is in good agreement with the 1.5% estimation in Ref. [19]. Dislocations are presumably introduced into such large islands when the strain fields of neighboring islands appreciably interact. In the presence of a greater density of small islands (high supersaturations), it is interesting to postulate the spontaneous nucleation of perfect edge dislocations at the intersection cusps between islands, analogous to the island edge nucleation mechanism of Vincent for the case of nonwetting layers [20]. This avoids the difficulty of nucleating lower energy half loops into shallow islands and explains why perfect edge dislocations are observed when islanding occurs near to the substrate at high misfits. This contrasts with the observation of closely bunched 60° dislocations when critical cusps develop far from the surface. At very low misfits, presumably heterogeneous sources take over as the first active sources before the critical cusp geometry develops.

Finally, it is well known that surfactants suppress the formation of islands during growth due to a reduction in surface diffusion [21]. This is entirely consistent with our surface chemical potential model and would prevent the formation of critical stress concentrations at surface cusps or isolated islands. It is not therefore surprising that a completely different strain relaxation mechanism is observed in the case of surfactant-controlled growth [22].

[1] D. J. Srolovitz, Acta Metall. **37**, 621 (1989).

[2] S. J. Pennycook and D. E. Jesson, Phys. Rev. Lett. **64**, 938 (1990).

[3] For a different application of the Ge marker layer technique, see D. J. Eaglesham, H.-J. Gossmann, and M. Cerullo, Phys. Rev. Lett. **65**, 1227 (1990). We assume that the individual marker layers and the 2.5% increase in average Ge content after 40 nm of alloy growth does not appreciably influence the morphological development.

[4] W. W. Mullins, J. Appl. Phys. **28**, 333 (1957).

[5] C. G. Inglis, Trans. Inst. Naval Archit. **55**, 219 (1913).

[6] Relevant simulation parameters for this and other calculations are $E = 107$ GN m^{-2}, $v = 0.27$, $\Omega = 21.26$ Å^3, $\eta = 6.51 \times 10^{-2}$ Å^{-2}, and $\gamma = 7.7 \times 10^{-2}$ eV Å^{-2}. The iteration time increments varied between 10 and 20 s. The major features of the simulations are relatively insensitive to the input parameters, including the expected 10% variation of γ with surface orientation; see D. J. Eaglesham et al., Phys. Rev. Lett. **70**, 1643 (1993). The rate of cusp development is, however, sensitive to the surface diffusion activation barrier. A best fit was achieved for a prefactor of 10^{-3} cm^{2}s^{-1} and 1.2 eV activation barrier.

[7] J. W. Matthews, J. Vac. Sci. Technol. **12**, 126 (1975).

[8] D. J. Eaglesham, E. P. Kvam, D. M. Maher, C. J. Humphreys, and J. C. Bean, Philos. Mag. A **59**, 1059 (1989).

[9] R. Hull, J. C. Bean, D. J. Werder, and R. E. Leibenguth, Appl. Phys. Lett. **52**, 1605 (1988).

[10] E. A. Fitzgerald et al., J. Appl. Phys. **65**, 2220 (1989).

[11] D. D. Perovic and D. C. Houghton, Mater. Res. Soc. Proc. **263**, 391 (1992).

[12] J. F. Knott, Fundamentals of Fracture Mechanics (Wiley, New York, 1973).

[13] I. N. Sneddon, Proc. R. Soc. London A **187**, 229 (1946).

[14] D. E. Jesson, S. J. Pennycook, J.-M. Baribeau, and D. C. Houghton (to be published).

[15] A. S. Tetelman and A. J. McEvily, Fracture of Structural Materials (Wiley, New York, 1967), p. 49.

[16] S. V. Kamat and J. P. Hirth, J. Appl. Phys. **67**, 6844 (1990).

[17] A. J. Pidduck et al., Thin Solid Films **222**, 78 (1992).

[18] See, for example, D. D. Perovic, G. C. Weatherly, J.-M. Baribeau, and D. C. Houghton, Thin Solid Films **183**, 141 (1989).

[19] D. J. Eaglesham and M. Cerullo, Phys. Rev. Lett. **64**, 1943 (1990).

[20] R. Vincent, Philos. Mag. **19**, 1127 (1969).

[21] R. M. Tromp and M. C. Reuter, Phys. Rev. Lett. **68**, 954 (1992).

[22] F. K. LeGoues, M. Copel, and R. M. Tromp, Phys. Rev. Lett. **63**, 1826 (1989).

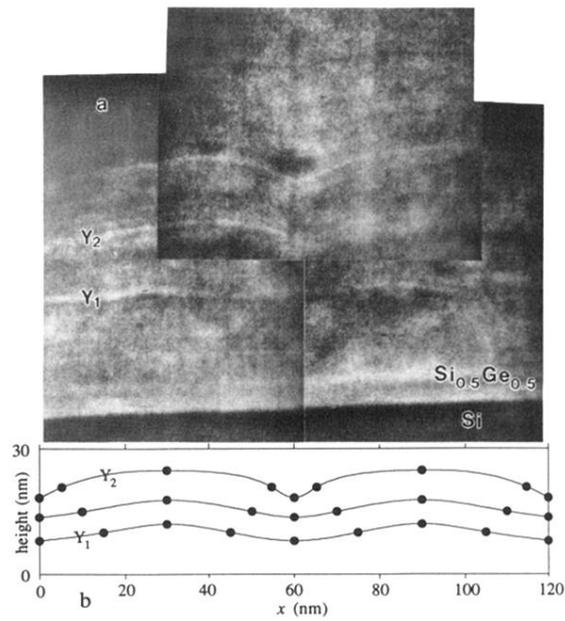FIG. 1. (a) [110] Z-contrast image of a $Si_{0.5}Ge_{0.5}$ alloy grown by MBE with 0.25 nm Ge marker layers deposited at selected intervals. The growth temperature was $400\,^\circ C$ and the deposition rate 2 Å s^{-1}. The profile simulations in (b) correspond to the period of cusp formation in (a) between vertical ordinates Y_1 and Y_2. The solid circles denote the peak, valley, and inflection positions as determined from Eqs. (3), (4), and (5) (see text).