# Origin of Apparent Critical Thickness for Island Formation in Heteroepitaxy

Yuhai Tu and J. Tersoff
*IBM Research Division, T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598, USA*

(Received 14 May 2004; published 16 November 2004)

We find that a continuum model of heteroepitaxy exhibits a sharp crossover with increasing coverage, from planar growth to island formation. The "critical thickness" at which this Stranski-Krastanov transition occurs depends sensitively on misfit strain, with a dependence strikingly similar to that seen experimentally. The initial planar growth occurs because of intermixing of deposited material with the substrate. While the transition is strictly kinetic in nature, it depends only weakly on growth rate. The role of surface segregation is also discussed.

DOI: 10.1103/PhysRevLett.93.216101 PACS numbers: 68.65.Hb, 68.35.Dv, 81.15.Aa, 81.16.Dn

In heteroepitaxial growth of strained layers, the growing layer often remains planar up to some "critical thickness," at which point three-dimensional (3D) islands form [1–6]. This Stranski-Krastanov transition is of great practical importance, since many device structures require smooth planar layers, while 3D islands hold promise as self-assembled quantum dots [7].

The transition also poses a long-standing puzzle in the understanding of heteroepitaxy. A *thick* planar layer under stress is known to be *unstable* [8,9] and ultimately forms 3D islands [4]. On the other hand, *atomically thin* layers can be stabilized by surface and interface energy. For the two most-studied systems, Ge on Si(001) and InAs on GaAs(001), islands appear only after deposition of a few atomic layers [1,3,4]. Calculations for Ge on Si(001) suggest that there are stabilizing finite-thickness effects and that these may be sufficiently large to explain the observed thickness [10].

The real puzzle is the behavior of alloys having lower strain—$In_xGa_{1-x}As$ on GaAs or $Ge_xSi_{1-x}$ on Si. As the composition $x$ and the corresponding misfit strain are reduced, the critical thickness is observed to increase dramatically. Transition thicknesses as large as 3–10 nm [2,6] are seen for moderately In-poor InGaAs on GaAs(001), corresponding to 10–30 monolayers (ML); and similar behavior is seen for GeSi on Si(001) [5]. In principle, this could be explained by the persistence of stabilizing surface-substrate interactions up to such large thicknesses [11,12], but there is no direct evidence for such a mechanism. At the same time, kinetic models might be expected to give a critical thickness that is sensitive to growth rate [13], while no dramatic dependence has been noted experimentally.

Here we report extensive simulations of strained-layer growth, using a continuum model that properly accounts for the alloy compositional degrees of freedom [14,15]. Under generic assumptions, we find behavior closely resembling that seen experimentally, as illustrated in Figs. 1 and 2. The surface remains rather flat up to some thickness at which it suddenly starts to roughen. This apparent critical thickness depends sensitively on composition, but only weakly on deposition rate. Because the behavior arises automatically in a published generic model that was not constructed with the critical-thickness problem in mind, we believe that we have identified a key mechanism underlying the apparent transition.

Our results confirm the insightful proposal of Cullis *et al.* [6] that the key factor controlling the transition is the continuous increase of surface composition during growth. Because of mixing between surface and subsurface layers, the surface composition is initially very dilute and only asympotically approaches the composition of the deposited material. The specific mechanism we find is

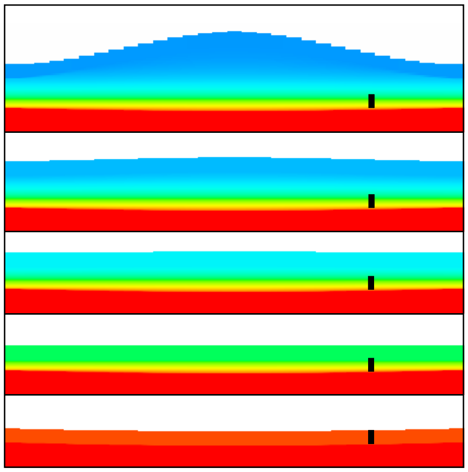

FIG. 1 (color online). Evolution of structure and composition during heteroepitaxy, for nominal $Si_{0.60}Ge_{0.40}$ on Si(001), at deposition rate of $10^4$ (arbitrary units). The onset of nonplanar morphology is more abrupt at lower growth rates. Colors indicate composition, from pure Si substrate (bottom) to $Si_{0.62}Ge_{0.38}$ (top). The bottom panel is the initial surface (slightly nonplanar), and subsequent panels are at equal time intervals. The figure shows one unit cell of periodic system; the lateral size is $640w_s$. Surface-layer thickness $w_s$ is indicated by a black rectangle in bottom panel; the vertical scale is greatly expanded to show the small perturbation. The rectangle is repeated in the same position in subsequent panels for reference. Surface steps are a graphical artifact.

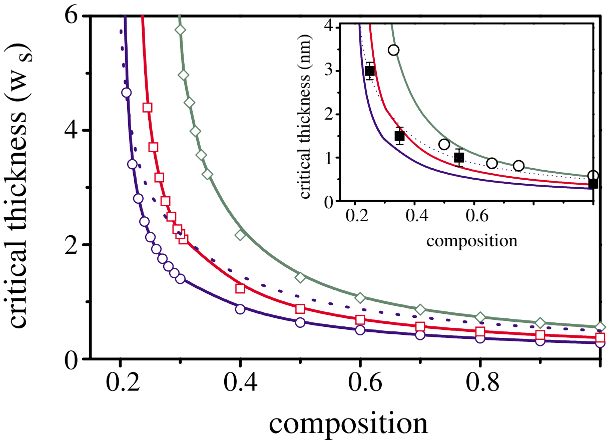

FIG. 2 (color online). Apparent critical thickness vs composition $c_{\mathrm{dep}}$ of the deposited material, for growth rates 1, $10^2$, and $10^4$ (in arbitrary units). The bottom curve corresponds to the slowest growth rate. (The dotted curve is the slowest growth rate, including surface segregation as described in the text.) The inset shows experimental results for $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{As}$ on GaAs(001), from Ref. [2] (open circles) and from Ref. [6] (solid squares). The lines in the inset are theoretical curves copied from the main panel, with arbitrary vertical scaling chosen to emphasize similarity to experiment.

surprisingly simple. Strained layers are unstable, and the growth rate of the instability is a very sensitive function of strain [8]. Thus, as the surface composition increases, the time scale for the instability decreases. The surface remains essentially planar until this time scale becomes comparable to the growth time. Surface segregation, paradoxically, can increase the critical thickness, because the surface-segregated layer does not contribute to the strained-layer instability.

The model used here is described elsewhere [14,15]. In brief, the system evolves by surface diffusion, while bulk diffusion is assumed to be negligible. Atoms within a few atomic layers of the surface are generally more mobile than in the bulk [16,17], so we assume that atoms within a depth $w_s$ (perhaps 2–4 ML) are in equilibrium with the surface. The free energy of this surface region, $g_s(\xi)$, is a function of $\xi$, which is the surface composition averaged over the depth $w_s$. In general, $g_s$ may be different than the bulk free energy function $g_b$, and any difference drives surface segregation. The composition and morphology evolve as

$$v = \sum_\nu [F_\nu + \nabla \cdot (D_\nu \nabla \mu_\nu)] \tag{1a}$$

$$w_s \frac{d\xi_\nu}{dt} = F_\nu + \nabla \cdot (D_\nu \nabla \mu_\nu) - \xi_\nu v. \tag{1b}$$

Here the subscript $\nu$ labels the two alloy components. $F_\nu$ is the incident flux of each component, $D_\nu$ is the diffusivity (taken as $\xi_\nu D_0$), $v$ is the local growth velocity of the surface normal to itself, and $\mu_\nu$ is the chemical potential for each species [14,15].

We choose parameters based on Ge/Si(001) at 600 °C, insofar as possible [18]. However, the results cannot be compared quantitatively with any specific system. Here, as in all treatments of stress-induced instability, a crucial factor is the "surface stiffness" term $\gamma\kappa$ in the chemical potential [14,15], where $\gamma$ is the stiffness and $\kappa$ is the local curvature of the surface. Yet while the surface *energy* $\Gamma$ is relatively well known for the Si(001) surface, little is known about the surface *stiffness* $\gamma = \Gamma + d^2\Gamma/d\theta^2$. In fact, only recently has there been clear evidence that the Ge/Si(001) surface is unfaceted (finite stiffness) under typical growth conditions [19]. The stiffness reflects the free energy of steps on the (001) surface, so it may be a sensitive function of temperature, composition, and strain. The composition-dependent diffusivity and the depth $w_s$ of interdiffusion are also poorly known. We therefore focus on the *qualitative* aspects of the behavior, which are insensitive to the specific parameters. We refer to "Ge" and "Si" for convenience in discussing the results, but we could as well refer to "InAs" and "GaAs," the system most-studied experimentally.

We begin for simplicity with the case where $\gamma$ is independent of composition, and there is no thermodynamic driving force for surface segregation ($g_s = g_b$). Figure 1 illustrates the evolution as $\mathrm{Ge}_{0.4}\mathrm{Si}_{0.6}$ is deposited on a Si surface. The slight nonplanarity of the initial surface, visible in Fig. 1, serves as the initial perturbation which eventually triggers the instability. The deposited material intermixes with a region of thickness $w_s$, so as long as the thickness of deposited material is much less than $w_s$, the composition at the surface is quite dilute. As a result, the planar geometry remains stable, except at wavelengths too long to be kinetically relevant; and the initial perturbation actually decays. As more material is deposited, the composition at the surface continuously increases. The growth rate of a perturbation is a highly nonlinear function of the composition, leading to the rather abrupt onset of the instability.

To explore the behavior systematically, we use an efficient linearized treatment, which is exact in the limit of small perturbations. We consider equal initial height perturbations $h_{q0}$ of the substrate at all $q$ (white noise) and calculate the time evolution of each $h_q$. We take the "apparent critical thickness" to be the thickness at which the largest $|h_q|$ is 10 times the original perturbation amplitude $h_{q0}$. (This criterion is somewhat arbitrary, but because of the suddenness of growth, the choice of criterion does not affect our conclusions.) It is already understood that a rapidly growing instability will cause the layer to break up into islands, so there is no need to explicitly simulate this highly nonlinear process.

In Fig. 2, we show this calculated critical thickness vs the composition $c_{\mathrm{dep}}$ of the deposited material. The critical thickness depends rather weakly on the deposition rate, considering the large range of rates shown. The overall shape of the curve is strikingly similar to what is seen experimentally (Fig. 2 inset).

In Fig. 3, we show in a more quantitative way the evolution of surface morphology and composition for an initial small sinusoidal perturbation of the substrate. As in Fig. 1, the perturbation initially decays, because the system is nearly unstrained. However, when the composition builds up sufficiently, the perturbation amplitude begins to grow rapidly, leading to the appearance of an abrupt transition.

The two solid lines in Fig. 3 correspond to different deposition compositions, giving different critical thicknesses. However, in both cases, the "transition" occurs when the surface composition reaches ~23% Ge. This supports a key conclusion of Cullis *et al.* [6] that the critical thickness corresponds to the thickness at which the surface composition reaches some threshold value.

Roughly speaking, we interpret this value of composition as follows. For a given growth rate $F$, there is a characteristic time scale for growth which is inversely proportional to $F$; for simplicity, we can take this time scale as $w_s/F$, the time to grow a layer of thickness $w_s$. There is a competing time scale, the time scale of the instability, which depends on the surface composition [8], decreasing with increasing misfit. Initially the composition is dilute (low strain) due to intermixing, and the instability time scale is long compared to the time scale of growth, so the planar morphology appears stable. With further deposition, the composition increases and the instability time scale decreases. When the instability time scale becomes comparable to the time scale of growth, the instability of the planar morphology becomes apparent. The critical composition thus depends on growth rate. But the instability time scale is extremely sensitive to composition (varying, in the simpler case of a uniformly strained single component, as $c_{\mathrm{dep}}^{8}$ [8,20]), and as a result, the critical composition and critical thickness depend only relatively weakly on the growth rate. By analogy with the single-component case, one might expect the apparent critical composition to vary as $F^{1/8}$.

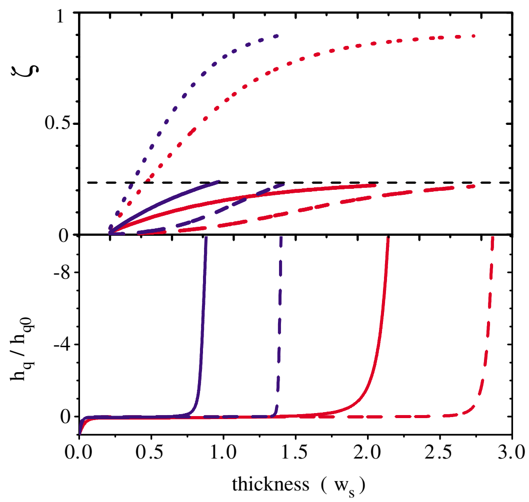

FIG. 3 (color online). Evolution of perturbation amplitude $h_q$ and surface composition $\xi$ with deposition thickness, at fixed growth rate (1 in arbitrary units). The left (blue) curve of each type (solid, dashed, and dotted) is for the deposition of $\mathrm{Ge}_{0.40}\mathrm{Si}_{0.60}$; the right (red) curves are for $\mathrm{Ge}_{0.25}\mathrm{Si}_{0.75}$. In each case we use the value of $q$ that determines the critical thickness in the case. The solid curves are results without surface segregation. For results with surface segregation, we show separately the composition $\xi^{(1)}$ of the surface monolayer (upper dotted curves) and $\xi^{(2)}$ of the rest of $w_s$ (dashed curves). The horizontal dashed line illustrates "critical composition."

If there were strictly a critical composition $c_{\mathrm{crit}}$ below which the planar surface is stable, the critical thickness in Fig. 2 would diverge as $c_{\mathrm{dep}}$ approaches $c_{\mathrm{crit}}$ [6]. Although there is no divergence in our simulations, and no strict $c_{\mathrm{crit}}$, this simplified picture nevertheless provides useful insight into the shape of the curves in Fig. 2, explaining why the critical- thickness varies so much more rapidly with $c_{\mathrm{dep}}$ at lower $c_{\mathrm{dep}}$.

Note also that, because the surface composition is continuously increasing, the instability time scale is continuously decreasing. Thus the growth of the perturbation is faster than exponential, contributing to the suddenness of the apparent transition.

A modest decrease in temperature can greatly decrease the diffusivity, typically by a factor of $\exp(-E/kT)$, where the activation energy $E$ is nearly 2 eV for Si. Within our model, a decrease in diffusivity (with no other changes) is equivalent to an increase in growth rate in Fig. 2 by the same factor. Thus from Fig. 2, we see that in the regime where the critical thickness is large, a small decrease in temperature may have an anomalously large effect in increasing the critical thickness. The same is true for a reduction in composition $c_{\mathrm{dep}}$.

Both InGaAs and SiGe exhibit strong, thermodynamically driven surface segregation. To model such surface segregation, we take $w_s$ to be 3 ML and divide it into a surface ML of composition $\xi^{(1)}$ and two subsurface MLs of composition $\xi^{(2)}$, where $\xi = \xi^{(1)}/3 + 2\xi^{(2)}/3$. We add to the surface-layer free energy a term driving segregation, $g_s \rightarrow g_s + U\xi^{(1)}/3$, and we take $U = 0.2$ eV/atom. For a given surface composition $\xi$, the decomposition into $\xi^{(1)}$ and $\xi^{(2)}$ is determined by minimization of free energy. Thus internal equilibrium over the thickness $w_s$ corresponds to strong enrichment of Ge in the top ML.

Figure 3 compares the evolution with and without surface segregation. Since segregation increases the fraction of strained material at the surface, it might be expected to have a destabilizing effect [6]. However, we find just the opposite, that if other parameters are unchanged, surface segregation increases the critical thickness. Moreover, the transition occurs at almost exactly the same surface composition with or without surface segregation, if we consider not the surface-segregated ML $\xi^{(1)}$, but rather the material $\xi^{(2)}$ immediately below (i.e., comparing solid and dashed lines in Fig. 3). The effect on critical thickness is also shown in Fig. 2 for the slowest growth rate.

Over a wide range of composition, the increase in critical thickness $h_c$ due to surface segregation is well described by $\Delta h_c \approx 0.64 c_{\mathrm{dep}}^{-1}$ ML. This can be understood by referring to Fig. 3 and considering how it would look for an idealized situation of perfect surface segregation. The surface composition $\xi^{(1)}$would rise linearly to 1, and the composition for the rest of $w_s$ would remain zero until the surface monolayer filled in. But the completed surface monolayer gives a stress that is independent of the local thickness; so it acts as a surface stress, not a bulk stress, and does not contribute to the linear instability [21]. In this case, we expect that surface segregation would increase the critical thickness by $\Delta h_c \approx c_{\mathrm{dep}}^{-1}$ ML.

In the actual simulations, the surface-monolayer composition is only ~0.90 by the time the layer reaches its critical thickness, and the composition just below has been rising continuously, so the effect on critical thickness is quantitatively reduced to $\Delta h_c \approx 0.64 c_{\mathrm{dep}}^{-1}$ ML. As seen in Fig. 2, this simple behavior breaks down for deposition composition below 30% Ge (perhaps because the surface layer is not saturated with Ge).

For a fully faceted system, island formation is expected to be a nucleated phenomenon [9], unless aided by kinetic roughening [22]. In either case, the rate for island formation again depends extremely sensitively on the alloy composition at the surface, so the behavior would be similar to that described here (except that kinetic roughening would introduce an additional dependence on growth rate [22]).

Finally, we note that there are other factors that may be important in specific systems, that are not included here. Most importantly, the surface stiffness $\gamma$ could depend sensitively on surface composition. (This also introduces another important and competing role for surface segregation.) In the most extreme case, the surface could change from faceted to unfaceted (infinite to finite stiffness $\gamma$) with changing composition at fixed temperature. The surface diffusivity could also be different for the two alloy components and could have a nontrivial dependence on composition. Also, we assume that the surface can fully equilibrate over a depth $w_s$ during growth. But this assumption could break down, especially for rapid growth at low $T$; and $w_s$ could vary with $T$.

In conclusion, we find that a simple continuum model of heteroepitaxy leads automatically to an apparent critical thickness for island formation. Moreover, the dependence on composition is strikingly similar to what is seen experimentally. The physics is essentially the same as in classic treatments of the strained-layer instability [8], except that we include intermixing at the surface. Within this model, there is no true transition, but rather a kinetic crossover due to the continuous increase of composition (and hence of stress) at the surface.

[1] D. Leonard, K. Pond, and P. M. Petroff, Phys. Rev. B **50**, 11687 (1994).

[2] P. M. Petroff and S. P. Denbaars, Superlattices Microstruct. **15**, 15 (1994).

[3] M. Asai, H. Ueba, and C. Tatsuyama, J. Appl. Phys. **58**, 2577 (1985).

[4] See D. J. Eaglesham and M. Cerullo, Phys. Rev. Lett. **64**, 1943 (1990), and references therein.

[5] J. A. Floro *et al.*, Phys. Rev. Lett. **79**, 3946 (1997).

[6] A. G. Cullis *et al.*, Phys. Rev. B **66**, 081305(R) (2002); T. Walther *et al.*, Phys. Rev. Lett. **86**, 2381 (2001).

[7] V. A. Shchukin and D. Bimberg, Rev. Mod. Phys. **71**, 1125 (1999).

[8] R. J. Asaro and W. A. Tiller, Metall. Trans. **3**, 1789 (1972); M. A. Grinfeld, Sov. Phys. Dokl. **31**, 831 (1986); D. J. Srolovitz, Acta Metall. **37**, 621 (1989); B. J. Spencer, P. W. Voorhees, and S. H. Davis, Phys. Rev. Lett. **67**, 3696 (1991).

[9] If the surface is faceted, it becomes metastable rather than unstable; see J. Tersoff and F. K. LeGoues, Phys. Rev. Lett. **72**, 3570 (1994).

[10] J. Tersoff, Phys. Rev. B **43**, 9377 (1991); K. Varga *et al.*, Surf. Sci. **562**, L225 (2004).

[11] I. Daruka and A.-L. Barabasi, Phys. Rev. Lett. **79**, 3708 (1997).

[12] H. R. Eisenberg and D. Kandel, Phys. Rev. Lett. **85**, 1286 (2000); Phys. Rev. B **66**, 155429 (2002).

[13] I. Daruka and A.-L. Barabasi, Phys. Rev. Lett. **78**, 3027 (1997).

[14] B. J. Spencer, P. W. Voorhees, and J. Tersoff, Phys. Rev. B **64**, 235318 (2001).

[15] J. Tersoff, Appl. Phys. Lett. **83**, 353 (2003); (unpublished).

[16] J. Tersoff, Phys. Rev. Lett. **74**, 5080 (1995).

[17] B. P. Uberuaga *et al.*, Phys. Rev. Lett. **84**, 2441 (2000).

[18] For these simulations we use the following values which, though somewhat arbitrary, are motivated by typical semiconductors, in particular, GeSi on Si(001): surface stiffness $\gamma = 20$ eV/nm$^2$; $w_s = 0.4$ nm; atomic volume 0.02 nm$^3$; $k_B T = 0.0776$ eV; elastic constants and misfit as for GeSi on Si(001). We assume equal diffusivities for the two components, for simplicity.

[19] J. Tersoff, B. J. Spencer, A. Rastelli, and H. von Känel, Phys. Rev. Lett. **89**, 196104 (2002), and references therein.

[20] J. Tersoff, Phys. Rev. Lett. **85**, 2843 (2000).

[21] J. Tersoff, Phys. Rev. Lett. **80**, 2018 (1998).

[22] J. Tersoff, Phys. Rev. Lett. **87**, 156101 (2001).