Materials Science and Engineering, B14 (1992) 245–253

Theoretical modeling of heteroepitaxial growth initiation

Efthimios Kaxiras
Department of Physics and Division of Applied Science, Harvard University, Cambridge, MA 02138 (USA)
O. L. Alerhand
Bell Communications Research, Red Bank, NJ 07701 (USA)
J. Wang and J. D. Joannopoulos
Department of Physics, MIT Cambridge, MA 02139 (USA)

(Received January 16, 1992)

Abstract

We review recent theoretical investigations on the initiation of heteroepitaxial growth. The thermodynamic premises on which the various growth models are based are considered. A qualitatively new model which applies to polar-on-non-polar heteroepitaxial growth is discussed. Specific consequences of the new growth model are explored and shown to be consistent with experimental observations of GaAs growth on Si(100) vicinal surfaces. First-principles calculations of the total energy of competing structures provide support for the proposed model. The total-energy comparisons afford a resolution to a long-standing puzzle concerning the effect of temperature on the orientation of GaAs overlayers on vicinal silicon substrates.

1. Introduction

The process of heteroepitaxial growth of material A (the adsorbate) on material B (the substrate) has long been thought to proceed via one of three possible modes: (a) layer growth, also known as the Frank–van der Merwe mode, in which A grows on B layer by layer, usually by step flow (in this case the formation of layers of A on B is energetically favorable for arbitrary thickness); (b) island growth, also known as the Volmer–Weber mode, in which the formation of layers of A on B is strictly forbidden due to the surface and interface energy balance, and only islands of A can be formed; (c) layer-plus-island growth, also known as the Stranski–Krastanov (SK) mode, in which layers of A on B are permitted up to certain thickness, but they eventually become energetically unstable and island growth takes over. This transition from island to layer growth is a stress-driven transition.

In terms of the surface energies of the adsorbate A and the substrate B (denoted by σ_A and σ_B) and the interface energy between the two materials (denoted by σ_I) the above three modes of growth correspond to the following relations:

$\sigma_B \gg \sigma_A + \sigma_I$

$\sigma_B \ll \sigma_A + \sigma_I$

$\sigma_B \approx \sigma_A + \sigma_I$

These three modes exhaust the thermodynamic possibilities, assuming that: the substrate is homogeneous or, equivalently, that substrate inhomogeneities are unimportant in the growth process; global equilibrium between adsorbate and substrate is attainable and is maintained in the system during growth.

With these assumptions a satisfactory description of growth is obtained for most cases where the macroscopic observables are important. However, in cases where these assumptions break down, new modes of growth can appear with interesting consequences. We shall describe here such a growth mode, which may be the prevalent channel by which growth of GaAs proceeds on Si(100) vicinal substrates. This system is of great technological importance, due to its potential applications in optoelectronic devices, and has been studied intensively by a variety of experimental methods.

2. New mode of heteroepitaxial growth

Consider a system in which the following premises apply:

(1) The mode of growth is dependent on the initial stages of growth nucleation.

(2) Global equilibrium is not attainable, but local equilibrium is readily achieved.

0921-5107/92/$5.00

(3) The role of the substrate morphology is crucial and influences strongly the growth initiation.

(4) Factors other than surface stress can be equally important in the process of growth initiation. An example of such a factor is the induced dipole field at the interface between a non-polar and a polar semiconductor along a direction of alternating anion and cation planes.

(5) The kinetics of growth initiation are very important. In particular, kinetically allowed configurations of low energy are preferred during growth initiation.

With these conditions, it is possible to imagine alternative modes of heteroepitaxial growth, which are qualitatively different from the thermodynamic modes of growth discussed above. Since our ultimate goal is to relate such alternative modes of growth to a real system, we shall focus the discussion by considering GaAs as the adsorbate and Si(100) as the substrate. This is a case of polar-on-non-polar material growth, in a crystallographic direction along which a perfect interface would produce macroscopically large dipole fields [1]. The mode of growth that we shall discuss could be applicable to other systems with similar characteristics, and under the same growth conditions.

The alternative mode of growth that we propose consists of the following process [2]. The deposited material forms initially a thin film on substrate terraces, which is stable up to a certain thickness, but the presence of the film inhibits further growth on the areas that it covers. The stability of the film is related to elimination of the polarity associated with the ideal interface. Because this thin film eliminates the dipole fields, it cannot have the ideal bulk structure of the adsorbate. This leads to inhibition of further growth of the film, since the bulk structure of the adsorbate must dominate at higher coverage. However, nucleation of centers with the ideal bulk structure of the adsorbate may be favorable at surface imperfections (such as steps and ledges) where a non-planar arrangement of the atoms is possible, which may naturally eliminate or reduce the polarity problem. Such nucleation centers will then start to grow into islands concurrently with the thin film on the terraces. The growth of the nucleation centers can continue uninhibited, since their structure is that of the ideal bulk adsorbate. These centers will then quickly take over and dominate the growth process.

Let us consider the differences between this mode of growth and the three thermodynamic modes mentioned earlier. Obviously, our model is most similar to the layer-plus-island (SK) mode since it involves both the formation of a thin film and island nucleation. The similarity, however, is superficial and indeed many differences exist.

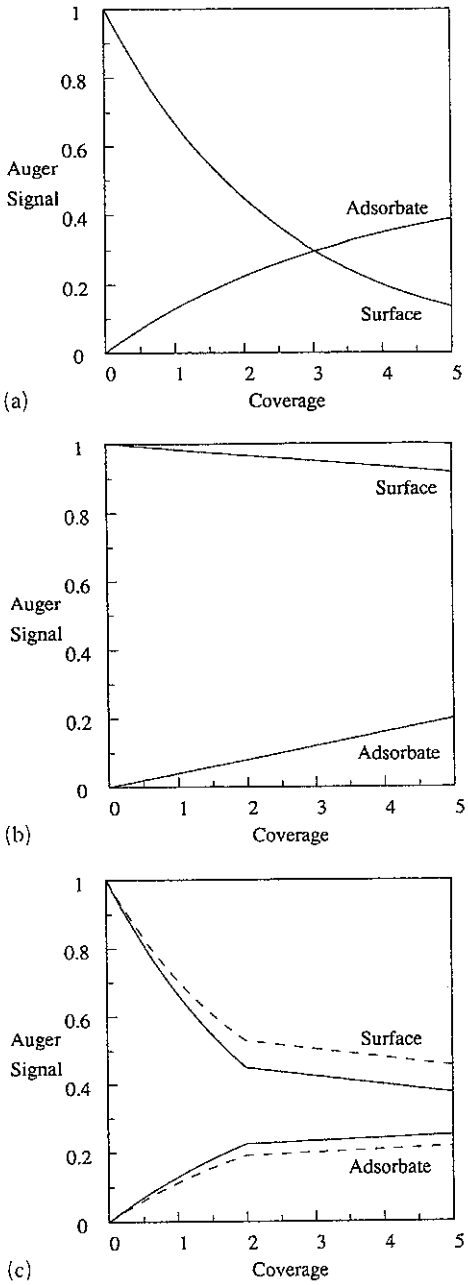

Fig. 1. Auger signal for the different modes of growth (the signal normalization is arbitrary): (a) layer-by-layer growth; (b) island growth; (c) layer-plus-island growth in the conventional (SK) thermodynamic model (——) and in the new model discussed here (– – –).

First, the islands and the film begin to grow together, rather than at different stages as in the SK mode. To emphasize this point, we present in Fig. 1 the expected behavior of the Auger signal for the various modes of growth. (This discussion of Auger signals is adapted from ref 3.) Since the Auger signal depends exponentially on the escape depth, layer growth gives an exponential decrease of the surface Auger signal as a

function of coverage and a corresponding increase of the adsorbate signal (Fig. 1(a)). Island growth gives a linear decrease of the surface Auger signal and a corresponding linear increase of the adsorbate signal (Fig. 1(b)), which reflects the total area covered by the adsorbate islands. In the SK mode of growth, the substrate signal shows an initial exponential decrease, which is followed by a linear regime (Fig. 1(c)). The point where the change from exponential to linear behavior occurs corresponds to the transition from layer to island growth. In the new model discussed above, the decrease of the substrate Auger signal will be less than exponential with coverage, since a substantial amount of the deposited material is incorporated in the nucleation centers, from the very beginning of growth. In fact, the deviation of the Auger signal from exponential behavior in our growth model provides a direct measure of the relative amount of material incorporated into the nucleation centers. This deviation will depend on surface preparation conditions (density of imperfections acting as nucleation centers) and could be used as evidence in support of the model.

The second difference concerns the nature of the transition from layer to island growth. In the traditional SK model, this transition is stress driven with the initial adsorbate layers being metastable on the surface. In the new model, the thin film is actually more stable on the surface than ideal layers of the adsorbate, since the thin film (by assumption) eliminates the polarity problem. Thus, there is no sharp stress-driven transition from layer to island growth, but the layer growth is inhibited and superseded by the successful island growth at nucleation centers.

A final difference is that the structure of the islands and their position on the surface are now intimately related to the structure and the position of the surface imperfections that can act as nucleation centers. None of these effects is present in SK growth, where the morphology of the substrate does not enter, and the position and size of islands formed after the stress-driven transition are statistically distributed.

3. Theoretical framework of calculations

Before discussing the specific application of our model to GaAs-on-Si(100) growth, we give a brief review of the theoretical framework used to perform total-energy comparisons. The total-energy comparisons will prove crucial in assessing the validity of the model. The calculations are based on an implementation of the Hohenberg–Kohn–Sham theorem [4], which asserts that the total energy of a system of interacting electrons and ions is a function of the electronic ground state density $n(\mathbf{r})$, which is determined by the ionic distribution $\{\mathbf{R}\}$:

$$E[n(\mathbf{r}),\{\mathbf{R}\}] = T[n(\mathbf{r})] + \int \mathrm{d}^3 r V_{\mathrm{ion}}(\mathbf{r})\, n(\mathbf{r}) + \frac{1}{2}\iint \mathrm{d}^3 r\, \mathrm{d}^3 r'\, \frac{n(\mathbf{r})\, n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\mathrm{xc}}[n(\mathbf{r})] \tag{1}$$

where, on the right-hand side, the first term is the kinetic energy of a system of interacting electrons, the second term is the electron–ion interaction, the third term is the Coulomb electron–electron interaction, the fourth term is the exchange–correlation interaction, which contains all the many-body aspects of the electron wavefunction, and, for simplicity, we have neglected the electrostatic interaction of the ions which contributes a constant term. This system can be mapped onto a system of non-interacting particles with exactly the same density as the interacting electrons. The non-interacting particles experience an effective potential given by

$$V_{\mathrm{eff}}(\mathbf{r}) = V_{\mathrm{ion}}(\mathbf{r}) + \int \mathrm{d}^3 r' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + v_{\mathrm{xc}}[n(\mathbf{r})] \tag{2}$$

where the exchange–correlation potential v_{xc} is obtained from the local density approximation (LDA) to the exchange–correlation energy functional

$$E_{\mathrm{xc}}[n(\mathbf{r})] = \int \mathrm{d}^3 r\, \varepsilon_{\mathrm{xc}}[n(\mathbf{r})]\, n(\mathbf{r}) \tag{3}$$

The non-interacting single-particle equations are

$$-\frac{\hbar^2 \nabla^2}{2m}\psi_i(\mathbf{r}) + V_{\mathrm{eff}}(\mathbf{r})\,\psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \tag{4}$$

and the density is obtained from the occupied single-particle states

$$n(\mathbf{r}) = \sum_{i=\mathrm{occ}} |\psi_i(\mathbf{r})|^2 \tag{5}$$

This formulation involves only one approximation (LDA, eqn. (3)). The LDA exchange–correlation functional is usually obtained by fitting exact results for the uniform electron gas (see for example ref. 5), and is independent of the system under consideration. LDA makes it possible to obtain a self-consistent solution to the set of single-particle equations (2)–(5) efficiently while taking into account many-body effects in a satisfactory manner, for most practical purposes. Indeed, experience has shown that LDA works extremely well for comparing relative energies. Finally, we use the

pseudopotential formalism to restrict ourselves to the description of the valence electrons. This formalism allows one to freeze the core electrons but at the same time include their influence on the valence electrons through an effective potential (the pseudopotential). (For a recent review of pseudopotential applications see ref. 6.)

Up to this point, the discussion has been concerned with solving the electronic degrees of freedom for a given external potential $V_{\text{ion}}(\mathbf{r})$ which is determined by the distribution of ions $\{\mathbf{R}\}$. Unless the ionic distribution corresponds to a true minimum of the system's total energy, the ground state electronic distribution gives rise to forces on the ions, which are calculated through the Hellmann–Feynman theorem. Thus, the second step in obtaining the ground state of the system, with all the degrees of freedom optimized, involves relaxing the ions until the forces on them are vanishingly small. The separation of the two sets of degrees of freedom (ionic and electronic) is of course based on the Born–Oppenheimer approximation and derives from the very large mass difference between ions and electrons. Recently, Car and Parrinello have introduced a scheme for a simultaneous optimization of electronic and ionic degrees of freedom [7]. The time evolution of the system of electrons and ions is described by a lagrangian in which the electrons have fictitious masses:

$$\mathscr{L} = \sum_i \frac{1}{2}\mu \int \mathrm{d}^3 r |\psi_i'(\mathbf{r})|^2 + \sum_{\mathbf{R}} \frac{1}{2} M_{\mathbf{R}} |\mathbf{R}'|^2 - E[\{\psi_i\}, \{\mathbf{R}\}] \quad (6)$$

(the prime indicates differentiation with respect to time). The total energy of the system for a given ionic configuration $\{\mathbf{R}\}$ plays the role of the potential energy in the lagrangian. This formulation results in the following equations of motion for the electronic degrees of freedom:

$$\mu\psi_i'' = -\frac{\delta\langle\mathscr{H}\rangle}{\delta\psi_i^*} + \sum_j \Lambda_{ij}\psi_j \quad (7)$$

where $\mathscr{H}$ is the LDA hamiltonian and the lagrange multipliers Λ_{ij} ensure orthonormality of the electronic eigenfunctions. The coupling between the electron and ion degrees of freedom through the fictitious electron masses is adjusted to give optimal rates of relaxation, and the whole procedure can be much more efficient than keeping the relaxations separate. For the calculations reported here we use a variant of the Car–Parrinello scheme, which incorporates a conjugate gradient algorithm to solve for the electron wavefunctions and an integration over short time scales, which otherwise would dominate the evolution of the system toward an energy minimum [8]. This procedure has proven very successful in locating energy minima in systems with a large number of constituent ions. In our case, we compare the energy between carefully chosen representative configurations of a system, and use the above procedure to find the full local relaxation for a given configuration.

4. Growth of GaAs on Si(100)

The first consideration, according to our new model of heteroepitaxial growth, is to establish whether the adsorbate can grow in a two-dimensional layer-by-layer mode on the substrate. By analyzing a number of different possibilities, Kaxiras and Joannopoulos have shown that the layer-by-layer growth of GaAs on Si(100) is severely inhibited by the formation of a thin film of mixed-layer composition [9]. This film (shown in Fig. 2) eliminates the problems of polarity associated with the interface between bulk phases of silicon and GaAs in the (100) crystallographic direction. The stability of the mixed-layer film derives from the nonpolar character of the interface, as well as from the optimal relaxation of surface Ga and As atoms: surface As atoms are raised outward from the surface and assume a pyramidal p^3-bonded configuration whereas surface Ga atoms recede toward the bulk and assume a planar sp^2-bonded configuration (see Fig. 2(b)). However, the structure of the mixed layer film is not that of bulk GaAs, thus inhibiting further growth of the bulk adsorbate phase. The mixed-layer film need not have a perfect long-range periodicity as shown schematically in Fig. 2, but its average local structure is expected to be close to that of the mixed-layer configuration. This structure is energetically very stable, with an energy of -1.5 eV per unit cell (containing two

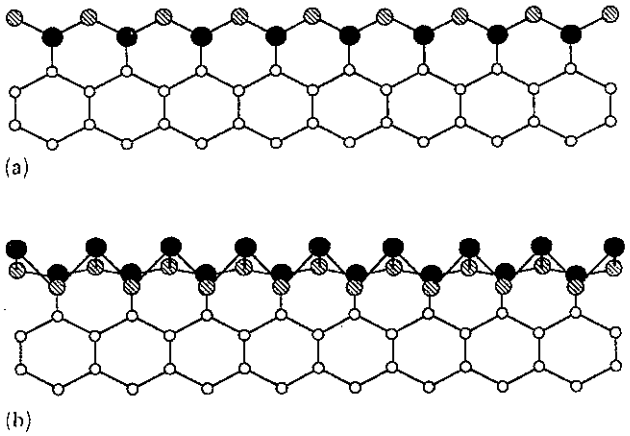

Fig. 2. The structure of (a) pure GaAs bilayers and (b) mixed Ga–As bilayers on terraces of Si(100): ●, arsenic; ⊛, gallium; ○, silicon. The surface structure in (b) is a schematic representation of the calculated atomic relaxation (see text).

Ga and two As atoms on top of silicon) relative to films with the bulk structure of GaAs on Si(100). These comparisons strongly indicate that layered growth will be stunted by the formation of mixed-layer films on flat regions (terraces) of the silicon substrate. This is consistent with experimental observations which indicate that, from the very beginning of growth on vicinal substrates, three-dimensional islands begin to appear [10–12].

The natural candidates for the nucleation of islands are imperfections on the silicon substrate, which are typically steps one or two layers high. Here we concentrate on a particular type of step, a double-layer step with the silicon dimers on either side of the step oriented parallel to the step edge (called DLB). Several other step configurations can be observed on vicinal silicon surfaces (for a review of experiments see ref. 13). We choose to concentrate on DLB steps for two reasons. First, these are the lowest-energy steps on the Si(100) surface [14], and surfaces with almost exclusively this type of step can be produced after long annealing [15]. Second, this type of step naturally eliminates the creation of anti-phase boundaries in the GaAs overlayers, a defect associated with nucleation at single-layer steps of the substrate [16]. The choice to concentrate on DLB steps as potential nucleation centers neglects other possibilities, but it does provide interesting insight to the mechanism of growth initiation.

We consider first the incorporation of Ga and As atoms in the neighborhood of a DLB step, and the difference with the layered growth. After examining several possibilities, we have found that there exists at least one way of incorporating Ga and As atoms in the neighborhood of a DLB step which is compatible with the bulk structure of GaAs, shown in Fig. 3. This configuration has several advantageous features, as follows. It eliminates the tendency for mixed layers which are incompatible with the bulk structure of GaAs. This is due to the bonding arrangement near the DLB step edge, which would lead to bonds between like atoms (Ga–Ga or As–As) upon mixing. In general, like-atom bonds in a polar material are energetically unfavorable. The second advantage is a reduction in the polarity, since now the sharp planar interface that was unavoidable at the terraces is replaced by a contoured interface which follows the contour of the DLB step. The exposed surfaces of the GaAs seed created around the DLB step begin to resemble the {111} surfaces of GaAs, as indicated in Fig. 3 (whether it is the arsenic-terminated or the gallium-terminated (111) surface depends on the details of substrate preparations, as will be discussed below). Finally, the seed described above and shown in Fig. 3 has lower energy than the optimal planar structure on a terrace. Our calculations for an infinitely long seed along the step direction give an energy lower than the mixed layer by 0.5 eV per unit length, which in the seed of Fig. 3 contains 4 Ga and 6 As atoms.

All these observations argue that the seeds nucleated at DLB steps are very stable local structures, likely to be formed at the initial stages of growth. Once such a seed is formed, and assuming that local equilibrium conditions prevail, it will be energetically favorable to incorporate additional material in this structure, which corresponds to expanding the bulk phase of GaAs, rather than on the nearby terraces, where the mixed layer structure is formed.

A cross-section of a small island grown in a self-similar way around the DLB seed is shown in Fig. 4. This island clearly exhibits the (111) exposed surface (in this case the arsenic-terminated one). Experimental evidence in support of this growth model comes from transmission electron microscopy (TEM) or scanning electron microscopy (SEM) experiments. High resolution TEM pictures of GaAs islands [17] grown on Si(100) exhibit a cross-section along their long axis which is exactly similar to that of the island shown in Fig. 4. Cross-sections of the islands along different directions clearly indicate that the islands are elongated, with a pyramidal shape. From the preceding discussion, this is precisely the shape one would expect

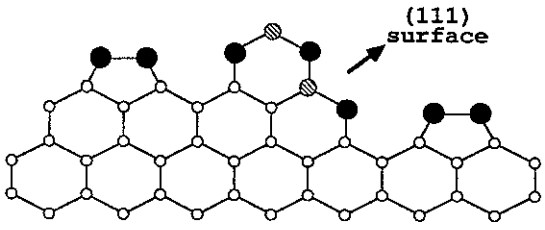

Fig. 3. GaAs seed nucleated at a silicon DLB step: ●, arsenic; ⊗, gallium; ○, silicon. The beginning of formation of (111) facets (in this case arsenic terminated) is indicated.

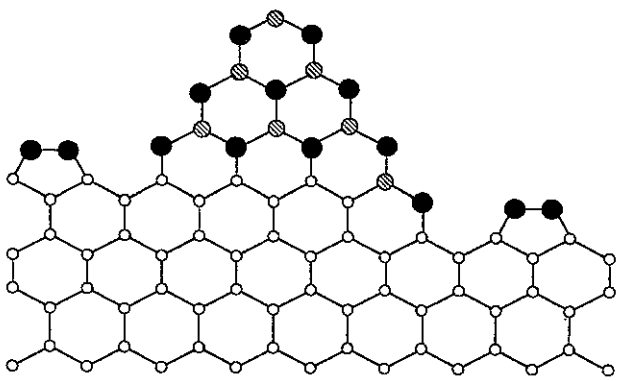

Fig. 4. GaAs island obtained through self-similar growth of the seed of Fig. 3: ●, arsenic; ⊗, gallium; ○, silicon.

from islands nucleated at steps. Since the steps are usually finite in length, one expects islands of finite length as well. Indeed, SEM pictures [18] of GaAs on annealed vicinal Si(100) show that the majority of the islands are elongated with their long axes parallel to one another, as expected for DLB nucleated seeds. The elongated shape of the islands could in principle be attributed to a preference toward exposing one type of surface. If this were the only driving force behind island formation, the distribution and orientation of the islands on the substrate would be random. The fact that all islands lie parallel to one another on annealed substrates strongly indicates that they are also nucleated by the same process and at similar substrate imperfections. Since DLB steps are predominant on annealed silicon substrates and they lie parallel to one another, the nucleation process described above seems a plausible model for describing the growth initiation.

It appears then that carefully prepared silicon substrates with regularly spaced DLB steps will be the most favorable ground for initiation of GaAs island growth. Once islands have been formed around DLB steps, their coalescence can produce GaAs overlayers that have fewer defects than usual growth on substrates without any preparation. The density of DLB steps on the substrate will determine how the islands start coalescing. The regions where the islands meet one another can act as nucleation centers for dislocations that do not thread through the GaAs overlayers. Such defects are needed to accommodate the lattice mismatch between silicon and GaAs. At the level of growth discussed here (the equivalent of few monolayers) most of the lattice mismatch can be relieved by outward relaxation of the GaAs islands. This relaxation cannot be effective for large islands, since the core of the islands becomes essentially bulk material and its straining is energetically costly. The details of island coalescence and dislocation nucleation within the present model deserve further examination.

5. The sublattice orientation dilemma

The nucleation model of the previous section affords a direct and natural explanation of the long-standing sublattice orientation dilemma. It has been reported that the orientation of GaAs grown on identical Si(100) substrates depends sensitively on the temperature conditions during growth initiation. In particular, GaAs overlayers grown at low temperature (below 700 K) compared with those grown at high temperature (900 K) exhibit a 90° crystallographic rotation [19–22]. Naively, this can be explained by the following argument. At low temperature one kind of atom (say, arsenic) forms the initial layer in contact with the silicon substrate, whereas at high temperature the other kind of atom (gallium) forms the contact layer.

This explanation is untenable for several reasons. First, it is well known that arsenic, due to its high volatility, always forms the initial layer on silicon under epitaxial growth conditions. The above explanation of the sublattice rotation would necessitate that the entire first layer of arsenic is removed and replaced by gallium upon increasing the temperature by 200 K. It is unlikely that under the higher temperature conditions the gallium-for-arsenic substitution of the initial layer would be as complete and perfect as is necessary for a 90° crystallographic reorientation of the GaAs crystal. Moreover, the gallium-for-arsenic substitution does not address any of the usual problems of polarity at the planar interface, substrate morphology *etc.* These problems will also hinder the complete substitution. Lastly, as we have argued above, the formation of pure monolayers of either type (gallium or arsenic) in contact with the substrate is energetically unfavorable, and instead mixed layers are much more stable on the substrate terraces. The gallium-for-arsenic substitution assumes that motion of atoms is easier under the higher temperature conditions, whereas at low temperature the arsenic overlayer is frozen in place. Given this assumption, however, it is likely that, once atoms are allowed to interchange positions, the more stable mixed-layer configuration will take over. This would certainly make it harder to explain a complete 90° reorientation of the GaAs overlayers, since mixed layers would not give a preference in overall orientation.

This puzzling situation can be explained in a very natural manner, if one assumes that growth initiation is directly related to substrate imperfections. Let us again concentrate on double-layer steps of Si(100). A number of experimental studies [23–27] of arsenic-covered vicinal Si(100) surfaces have reported that, upon changing the temperature or the deposition conditions, the orientation of arsenic dimers rotates by 90°. This observation depends crucially on the existence of steps on the substrate, since on a perfectly flat surface there exists only one possible orientation of the surface-layer dimers. The existence of steps introduces a second possibility, with the direction of a step providing an alignment axis. In the case of DLB steps on the clean silicon surface the lowest energy configuration corresponds to silicon dimers oriented parallel to the step direction. When arsenic is deposited on this surface, the kinetically allowed configuration corresponds to arsenic dimers perpendicular to the step direction [28], as shown in Fig. 5(a) (this configuration will be denoted as Si(100):As 1×2, or λ_{DA}). This is a result of the topology of the (100) surface of the diamond structure, in which successive layers along the

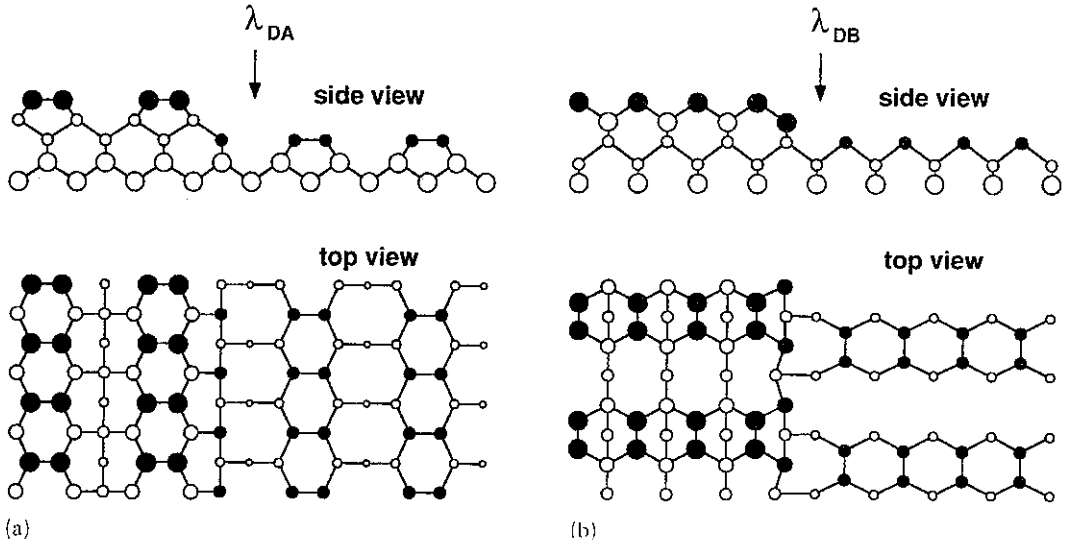

Fig. 5. Side and top views of (a) the metastable (λ_{DA}) (Si(100): As 1×2 and (b) the equilibrium (λ_{DB}) (Si(100): As (2×1) configurations of arsenic-covered double-layer steps on Si(100): ●, arsenic; ○, silicon.

(100) direction are rotated by 90° with respect to one another.

However, the kinetically allowed structure is not necessarily the lowest-energy, equilibrium, configuration. We have performed total-energy calculations to investigate this point, and we find that a structure with arsenic dimers oriented parallel to the step direction, shown in Fig. 5(b), has lower energy by 0.2 eV per unit length along the step (this configuration will be denoted as Si(100):As 2×1, or λ_{DB}) [29]. This structure will encounter an energy barrier to its formation, since it involves a 90° rotation of the arsenic dimers relative to the kinetically allowed structure. It is natural then to expect that the equilibrium structure will be unreachable at low temperature, but will be formed upon raising the temperature beyond a certain critical value. The reason behind the lower energy of the λ_{DB} configuration compared with the λ_{DA} configuration can be seen by examining the surface strain corresponding to the two structures. Top and side views of the two configurations are compared in Fig. 5, and the calculated strains are shown in Fig. 6. From our calculations we find that λ_{DB} produces a non-uniform relaxation along the step direction, which relieves surface stress without introducing substrate strain, whereas λ_{DA} produces a uniform contraction of the upper terrace, which results in substrate strain and correspondingly higher energy [29].

The picture that emerges from the above discussion is the following. Low-temperature growth initiation begins from an arsenic dimer layer, in which the arsenic dimers are oriented perpendicular to the step direction. This is a kinetically allowed, but metastable, structure.

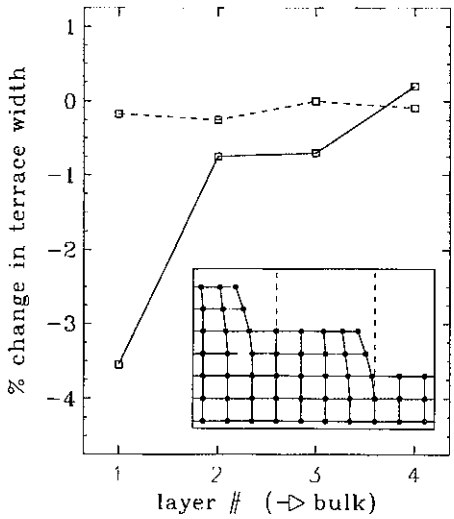

Fig. 6. Calculated surface strain (as percentage change in terrace width) resulting from the λ_{DA} (——) and λ_{DB} (– – –) structures.

At higher temperature, the stable equilibrium configuration of the arsenic dimers is reached, which involves a 90° rotation of the dimers so that they are parallel to the step direction. The ensuing overlayers of GaAs exhibit a 90° crystallographic rotation. These effects are shown schematically in Fig. 7. Notice that the low-temperature islands have the arsenic-terminated (111) surfaces exposed, whereas the high-temperature islands have the gallium-terminated (111) surfaces exposed.

The model described above is directly related to the presence and the effect of steps on the silicon surface, and provides a natural explanation of the experimental

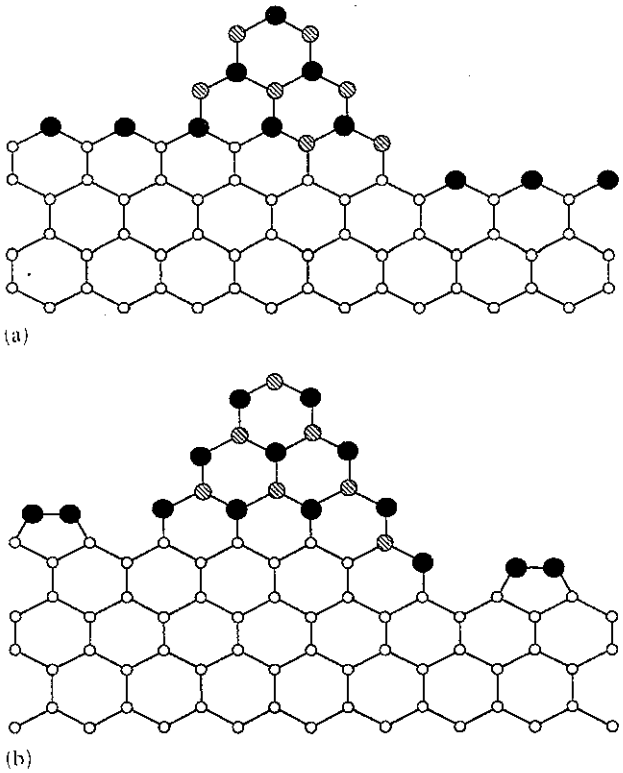

Fig. 7. GaAs islands grown at (a) high and (b) low temperatures at double-layer steps: ●, arsenic; ⊗, gallium; ○, silicon. The islands have different (1 1 1) surfaces exposed (see text).

observations concerning the orientation of GaAs overlayers as a function of growth temperature. The change in orientation emerges as a direct consequence of surface stress relief, by rotation of the arsenic dimers in the neighborhood of the steps. One aspect of this approach that remains unresolved is the precise way by which the arsenic dimers are rotated to reach the equilibrium structure. It appears that this rotation involves the replacement of the full top layer of silicon by arsenic. This, however, may not be as difficult to achieve as it seems at first sight. The presence of steps can facilitate the process considerably. It is possible that, once the substitution begins at some step edge, Si atoms move to the next-nearest step edge, which in a vicinal surface is only a few lattice constants away [27]. The presence of the extra Si atoms can precipitate the rotation of arsenic dimers in the next step. The details of this process such as diffusion and substitution pathways and the relevant energy barriers need to be examined further.

6. Summary and conclusions

We have discussed a qualitatively new model of heteroepitaxial growth, which takes into account explicitly the substrate morphology and the interplay of kinetics and thermodynamics during growth initiation. This model was compared with the generally accepted macroscopic thermodynamic modes of growth and in particular to the SK growth mode to which it bears some resemblance. The similarities and differences between the thermodynamic growth modes and our approach were pointed out in detail. The new model was then applied to growth of GaAs on vicinal Si(100) surfaces. Through first-principles total-energy calculations, it was shown that layered growth of GaAs on terraces of the substrate is inhibited by the formation of a mixed-layer structure that does not correspond to the bulk structure of GaAs. Instead, we find that nucleation at double-layer steps is energetically favorable and can promote bulk-like growth of GaAs. The coalescence of islands nucleated in well-prepared substrates can provide overlayers of GaAs with fewer defects than uncontrolled growth.

Finally, the step-nucleated growth gives a natural explanation to the long-standing puzzle of a 90° crystallographic rotation of the GaAs overlayers when the growth temperature is raised by 200 K. This change in orientation can be traced to the orientation of arsenic dimers in the vicinity of steps, since the kinetically allowed structure is metastable and the equilibrium structure can only be reached by raising the temperature. The two structures differ by a 90° rotation of the arsenic dimers. The driving force behind the dimer rotation is a stress relief mechanism.

Unresolved issues that deserve further study include the accommodation of lattice mismatch and coalescence of large GaAs islands and the mass transfer implied by the rotation of arsenic dimers at elevated temperatures. Nevertheless, the growth model we presented is consistent with experimental observations and provides a plausible framework for understanding the complicated process of heteroepitaxial growth from a microscopic point of view.

Acknowledgments

During the course of this work we have benefited from useful discussions with G. W. Turner and R. S. Becker.

References

1 W. A. Harrison, E. A. Kraut, J. R. Waldrop and R. W. Grant, *Phys. Rev. B, 18* (1978) 4402.
2 E. Kaxiras, O. L. Alerhand, J. D. Joannopoulos and G. W. Turner, *Phys. Rev. Lett., 62* (1989) 2484.
3 J. A. Venables, G. D. T. Spiller and M. Hanbucken, *Rep. Prog. Phys., 47* (1984) 399.

4 P. Hohenberg and W. Kohn, *Phys. Rev. B, 136* (1964) 864.
W. Kohn and L. J. Sham, *Phys. Rev. A, 140* (1965) 1133.
5 J. P. Perdew and A. Zunger, *Phys. Rev. B, 23* (1981) 5048.
6 W. E. Pickett, *Comput. Phys. Rep., 9* (1989) 117.
7 R. Car and M. Parrinello, *Phys. Rev. Lett., 55* (1985) 2471.
8 M. C. Payne, J. D. Joannopoulos, D. C. Allan, M. P. Teter and D. H. Vanderbilt, *Phys. Rev. Lett., 56* (1986) 2656.
T. Arias, M. C. Payne and J. D. Joannopoulos, *Phys. Rev. B, 45* (1992) 1538.
9 E. Kaxiras and J. D. Joannopoulos, *Surf. Sci., 224* (1989) 515.
10 S. Nishi, H. Inomata, M. Akiyama and K. Kaminishi, *Jpn. J. Appl. Phys., 24* (1985) L391.
11 R. Hull and A. Fischer-Colbrie, *Appl. Phys. Lett., 50* (1987) 851.
12 D. K. Biegelsen, F. A. Ponce, A. J. Smith and J. C. Tramontana, *J. Appl. Phys., 61* (1987) 1856.
13 J. E. Griffith and G. P. Kochanski, *Crit. Rev. Solid State Mater. Sci., 16* (1990) 255.
14 D. J. Chadi, *Phys. Rev. Lett., 59* (1987) 1691.
15 T. Sakamoto and G. Hashiguchi, *Jpn. J. Appl. Phys., 25* (1986) L57.
16 H. Kroemer, *J. Cryst. Growth, 81* (1987) 193.
17 F. A. Ponce and C. J. D. Hetherington, in G. W. Bailey (ed.), *Proc. 47th Annu. Meet. of the Electron Microscopy Society of America*, 1989, p. 586.
18 M. Akiyama, T. Ueda and S. Onozawa, *Mater. Res. Soc. Symp. Proc., 116* (1988) 79.
19 M. Kawanabe, T. Ueda and H. Takasugi, *Jpn. J. Appl. Phys., 26* (1987) L114.
20 P. R. Pukite and P. I. Cohen, *Appl. Phys. Lett., 50* (1987) 1739.
21 R. Fischer, H. Morkoc, D. A. Neumann, H. Zabel, C. Choi, N. Otsuka, M. Longerbone and L. P. Erikson, *J. Appl. Phys., 60* (1986) 1640.
22 R. D. Bringans, M. A. Olmstead, R. I. G. Uhrberg and R. Z. Bachrach, *Appl. Phys. Lett., 51* (1987) 523; *Phys. Rev. B, 36* (1987) 9569.
23 R. S. Becker, T. Klitsner and J. S. Vickers, *J. Microsc. (Oxford), 152* (1988) 157.
24 J. Varrio, H. Asonen, J. Lammasniemi, K. Rakennus and M. Pesa, *Appl. Phys. Lett., 55* (1989) 1987.
25 T. R. Ohno and E. D. Williams, *J. Vac. Sci. Technol. B, 8* (1990) 874.
26 M. Uneta, Y. Watanabe, Y. Fukuda and Y. Ohmachi, *Jpn. J. Appl. Phys., 29* (1990) L17.
27 R. D. Bringans, D. K. Biegelsen and L.-E. Swartz, *Phys. Rev. B, 44* (1991) 3054.
28 R. D. Bringans, R. I. G. Uhrberg, M. A. Olmstead and R. Z. Bachrach, *Phys. Rev. B, 34* (1986) 7447.
29 O. L. Alerhand, J. Wang, J. D. Joannopoulos, E. Kaxiras and R. S. Becker, *Phys. Rev. B, 44* (1991) 6534.