Materials Science and Engineering R 46 (2004) 127–176

R Reports: A Review Journal

www.elsevier.com/locate/mser

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Self-organized growth on GaAs surfaces

Bruce A. Joyce*, Dimitri D. Vvedensky

The Blackett Laboratory, Imperial College, London SW7 2BW, UK

Accepted 28 October 2004

Available online 8 December 2004

Abstract

GaAs(001) has been one of the most intensively studied surfaces for the past 30 years due both to its importance as a substrate for epitaxial growth and to the challenge its phase diagram of complex structures presents to computational methods. Yet despite substantial experimental and theoretical effort, a number of fundamental questions remain concerning growth kinetics and mechanisms on this surface, even for homoepitaxy, but more especially in the formation of heterostructures. These issues have acquired a renewed timeliness because the quantum dots that are formed during the Stranski–Krastanov (SK) growth of InAs on GaAs(001) can be used for optoelectronic applications and have potential in quantum dot-based architectures for quantum computing.

In this review we survey the current state of understanding of growth kinetics on GaAs surfaces, beginning with the simplest case, homoepitaxy on GaAs(001). We compare interpretations of recent reflection high energy electron diffraction measurements taken during the initial stages of growth with predictions of ab initio density functional calculations. We also consider the extent to which snapshot scanning tunnelling microscopy images from rapidly quenched samples truly reflect the growing surface structure as revealed by in situ real-time methods. We then examine the present experimental and theoretical status of the SK growth of InAs quantum dots on singular orientations of low-index GaAs surfaces, focussing on such issues as the importance of substrate orientation and surface reconstruction of the substrate, wetting layer formation, the nucleation kinetics of quantum dots, their size distributions and the role of strain. The systematics and anomalies of the phenomenology will be highlighted, as well as the current understanding of quantum dot formation.

PACS: 68.55.Ac; 68.55.Jk; 68.65.Hb

Keywords: GaAs surfaces; Self-organization; Quantum dots; Heteroepitaxy; Reconstructions; Stranski–Krastanov growth

1. Introduction

The observation of carrier confinement in quantum wells in the 1970s spawned an enormous effort aimed at fabricating heterostructures with greater dimensional confinement. This was driven by the realization that the attractive optical and transport properties of low-dimensional structures stem from the fundamental changes in the densities of states and the Coulomb interaction between carriers as their effective dimensionality is reduced. The confinement of carriers to one (quantum ''wires'') or zero (quantum ''dots'') dimensions is relatively easy to achieve when the lateral dimensions of the confining region is $\sim$500–1000Å. The techniques that permit the fabrication of such structures are based on lithographic processing. Unfortunately, the lithographic process often produces interfaces with a high defect density and even damage to the bulk material itself. Moreover, carrier confinement

* Corresponding author. Tel.: +44 20 7594 7573; fax: +44 20 7594 7604.

E-mail addresses: b.joyce@imperial.ac.uk (B.A. Joyce), d.vvedensky@imperial.ac.uk (D.D. Vvedensky).

0927-796X/$ – see front matter
doi:10.1016/j.mser.2004.10.001

within regions of such lateral dimensions leads to a limited separation of subband energies (typically a few meV), which in most cases is exceeded by thermal broadening. Thus, most of the interesting physical behaviour in such heterostructures is observed only at very low temperatures ($T \ll 4$K). This property distinguishes mesoscopic structures from true *quantum* heterostructures, in which the inter-subband separation is larger than the typical Coulomb interaction energy between carriers. For these reasons, fabricating quantum wires and quantum dots directly with in situ growth techniques provides an attractive alternative to ex situ processing with lithography.

Quantum heterostructures must aim to satisfy four basic criteria. (i) Their lateral size(s) (in relation to carrier effective masses) must be small enough to achieve genuine quantum confinement in the sense discussed above. (ii) Their size distribution must be relatively narrow. This is especially important for optical applications, where the inhomogeneous broadening of electronic transitions over an ensemble of heterostructures must be minimized. (iii) Their interfaces must have a high degree of definition and produce an abrupt carrier-confining potential. (iv) They must have structural and chemical stability, i.e. the interfaces must retain their chemical integrity and be stable against the formation of structural defects, such as dislocations, during all stages of fabrication. Attempts at producing heterostructures that satisfy these criteria have revolved largely around the deposition of atoms and molecules on ordered, usually single crystal substrates, and their subsequent incorporation to form thin films and other reduced dimensionality structures by processes such as molecular beam epitaxy (MBE), chemical vapour deposition (CVD) and metal-organic vapour phase epitaxy (MOVPE).

In this review we will restrict our attention mainly to the homoepitaxial growth of GaAs(001) and the growth of InAs on low-index oriented GaAs substrates, in each case using MBE growth. Despite this limitation on the choice of materials and method of deposition, much of the underlying physics, both experimental and theoretical, can be addressed in a reasonably comprehensive manner. Accordingly, although we focus on results that are particular to this specific materials system, we emphasize concepts that have a more general scope. We have separated our discussion into two parts. In the first, we review the experimental and theoretical work on the phase stability and growth kinetics on GaAs(001). This is an experimentally well-characterized surface whose theoretical understanding is a prerequisite for any discussion of heteroepitaxy. Several atomistic models have been proposed for reconstructions on GaAs(001) and for the growth kinetics on GaAs(001)–$\beta(2 \times 4)$, the most common substrate for homoepitaxy. The second part of this review is devoted to heteroepitaxy of InAs on GaAs. This presents substantial additional challenges because of the compositional and morphological consequences of the competing strain relaxation mechanisms on the low-index orientations of GaAs. We conclude with a summary of our current understanding of quantum dot formation in the InAs/GaAs(001) system and of the outstanding issues.

2. Part I: homoepitaxial growth of GaAs(001)

2.1. Background

GaAs(001) has become the archetype for the study of reaction kinetics, growth mechanisms and atomistics of III–V compounds. The earlier work has been extensively reviewed ([1,2] and references therein), so here we will concentrate on the present understanding of the atomistics of the growth process. Until comparatively recently the prevailing opinion had been that the morphology of the growing surface was dominated by the Ga and that the function of As (supplied as As_2 or As_4) was mainly to ensure stoichiometry and to establish or maintain a particular surface reconstruction. Support for this viewpoint came from a large body of experimental and theoretical work that provided

a quantitative interpretation of reflection high-energy electron-diffraction (RHEED) intensity oscillations [3,4] and scanning tunnelling microscopy (STM) measurements [5] without any explicit reference to the As flux. There are, however, several well-known effects which are directly attributable to As kinetics, including the influence of the As flux on the transition to the step flow growth mode on vicinal surfaces [6], As-induced RHEED oscillations on singular surfaces [7] and the sensitivity of Ga adatom mobility to variations in the As flux [8]. Understanding the atomistic origins of such kinetic phenomena, as well as changes of the equilibrium morphology of GaAs(001) surfaces as a function of the As (or Ga) chemical potential [9,10], requires treating Ga and As on an equivalent basis.

Farrell et al. [11] made a notable early attempt to describe the growth in atomistic terms based on an hypothesis known as the ''electron counting rule'', which stipulates that, under conditions of charge neutrality, all As dangling bonds are fully occupied, while all Ga dangling bonds are empty. The strict application of this rule limits the number of equilibrium reconstructions and provides some justification for the island structures seen during growth [12], but leaves open the question about the possible role of transient, but nevertheless essential, intermediate structures that do not obey electron counting. Madhukar and Ghaisas [13] performed kinetic Monte Carlo (KMC) simulations where the reaction kinetics of As were included explicitly, from which they were able to explain a number of trends in RHEED oscillation data. More recently, with the benefit of atomic-resolution scanning tunnelling microscopy (STM) measurements, Itoh et al. [14,15] used KMC simulations that included the (2×4) surface structure of the substrate and the kinetics of As_2 incorporation to reproduce several experimentally measured quantities associated with island kinetics, such as the evolution of islands and their size distribution. First-principles density functional calculations have been used to determine the energetics associated with several surface processes, such as Ga adatom diffusion [16,17], As incorporation [18,19] and the stability of Ga clusters [20,21]. These calculations have provided the basis for models of the initial steps in the homoepitaxy of GaAs(001)–(2×4)[21–23].

Although the most recent of these models can explain particular aspects of the process in considerable detail, a number of fundamental questions still remain. These include (i) the nature of Ga incorporation, specifically the mechanism by which Ga forms bonds with neighbouring, initially dimerized, As atoms. First principles calculations [17] have only examined pathways from an initial state in which neighbouring As dimers are first broken; (ii) the surface site(s) from which growth is initiated; (iii) the possible anisotropy of initial islands; (iv) the dependence of growth morphology on the As flux. It is also important to realize that while the $\beta 2(2 \times 4)$ structure (see Section 2.2) provides the template for growth, i.e. it exists in an As flux at typical growth temperatures ($\approx 580\,^{\circ}C$), it changes extremely rapidly to a substantially different structure once growth commences following the introduction of a Ga flux, as shown by in situ RHEED observations. We will discuss this effect in detail in Section 2.3, but we note here that it has not been considered in any available growth model.

The origin of these growth characteristics can be traced to the interaction between Ga adatoms and As-dimers and in the following sections we discuss a model that provides a way of understanding these phenomena. Before doing so, however, we consider to what extent it is possible to provide experimental justification for the basis of the model. It is unlikely to be direct in the sense of the observation of surface interactions on an atomistic scale, but the most relevant would in principle be from techniques that are sensitive to local bonding arrangements and are capable of acquiring information dynamically.

There are effectively only two methods that fulfil these criteria; reflection anisotropy (difference) spectroscopy (RA(D)S) and photoemission. Unfortunately, neither has been used in a manner that is able to provide unequivocal confirmation of the bonds present during growth. With RDS, it is essential

that the response be followed spectroscopically if such information is to be deduced, but measurements during growth have mostly concentrated on the periodic variation of the signal at some fixed energy [24,25]. This is then treated in an analogous manner to RHEED intensity oscillations [3,4], where the period corresponds to monolayer growth. It is an interesting phenomenon, but not directly relevant to the establishment of reaction pathways in the present context. A similar situation exists with regard to static reconstructions of GaAs(001) surfaces, which have been studied with RDS by correlating spectral lineshapes with RHEED patterns [26–28], but can only be related to growth in a somewhat indirect manner.

In photoemission, angle- and energy-integrated data have been collected in real time during growth [29–32], but electronic structure can only be determined from angle- and energy-resolved measurements, which have not yet been used dynamically. The measurements made were concerned with the oscillating intensity during growth, which was also found to have a period corresponding to the growth of a monolayer, analogous to RHEED, but the intensity value was very dependent on surface stoichiometry, being greatest for the highest arsenic concentration. It was also found that high step edge densities produced high photoelectron yields and lower ionisation thresholds, but it is not clear how these facts can be related to local bonding arrangements, since the state of the arsenic on the surface is not known.

Support must therefore be deduced from measurements that do not directly relate to bonding configurations, but which contain elements of that information within them. STM can provide an indication of growth sites with atomic resolution [14,15], at least in ''snap-shot'' mode, i.e. by rapid quenching following a growth sequence, but in general precise bonding details cannot easily be derived from the images. Nor is it clear whether quenching can ever be fast enough to maintain the structure present during growth, in view of the rate of change of RHEED patterns observed at the commencement of growth [Section 2.3.1]. RHEED can indicate long-range surface order (structure) and symmetry under dynamic (growth) conditions, but cannot identify specific bonds except by inference from symmetry. Finally, macroscopic kinetic measurements are relevant in the sense that any model must at least be consistent with overall rate processes [33].

An important component of the reaction pathway deduced from these kinetic measurements was that prior to any localized adsorption the incident As_2 molecules were present in a weakly bound, highly mobile precursor state, from which they could either desorb or become more tightly bound to the surface [33], although no specific site could be identified. Itoh et al. [14,15] showed from KMC simulations that without this concept unphysically high As_2/Ga flux ratios were needed to reproduce experimental results on island formation during growth. First principles calculations [19] also supported the idea of a precursor state to hold As_2 molecules close to the surface so that they could be funnelled into strongly bound sites during growth. We may reasonably conclude, therefore, that precursor state kinetics are an essential feature of the overall growth process.

2.2. The substrate

An atomistic description of the growth kinetics on GaAs(001) is an experimental and modelling challenge for two main reasons: the vastly different surface kinetics of the incident species (atomic Ga and either As_2 or As_4) and the presence of reconstructions with considerable atomic rearrangements whose structures are determined by the relative fluxes of these species and the substrate temperature. The (2×4) reconstruction occupies the largest region of the temperature-incident flux ''phase diagram'' [34] and is used for most studies of epitaxial growth on this orientation. The template for growth is normally produced by growing a so-called buffer layer by MBE, typically a few hundred nanometres thick, for which the equilibrium surface reconstruction is generally accepted to be the

$\beta 2(2 \times 4)$ structure. We illustrate the unit cell in Fig. 1, for which there is virtually full agreement based on RHEED, glancing angle X-ray scattering (GIXS) and STM experiments, together with several first-principles density functional calculations [35,36,37,38,39]. It is an As-terminated structure that substantially modifies the three topmost layers of the surface: alternate pairs of dimer rows along $[\bar{1}\,1\,0]$, together with their underlying Ga atoms are missing and the exposed As atoms in the third layer form dimers along this direction. It has been known for over 20 years, however, that the investigation of the long-range order, either by diffraction techniques or STM, indicated the presence of surface defects in every sample measured, but their influence on growth was, and remains, unknown.

The various surface reconstructions of GaAs(001) have different surface stoichiometries. The thermodynamic stability of these reconstructions as a function of the Ga chemical potential (μ_{Ga}) (the chemical potentials of Ga and As are constrained to be equal to the energy per surface unit cell of bulk GaAs) has been calculated by Northrup and Froyen [9,10], and the result is displayed in Fig. 2. The calculated phase diagram reproduces the trend seen in RHEED: a $c(4 \times 4)$ phase at high μ_{As} (low μ_{Ga}), $\beta 2(2 \times 4)$ phase at intermediate μ_{As}, and (4×2) at low μ_{As}. It must be borne in mind, however, that these are *zero temperature* calculations, so for example, the narrow chemical potential window for the $\alpha(2 \times 4)$ phase, might indicate that it is, in fact, a nonequilibrium structure [10].

The first report [40], before STM images were available, interpreted the curved streaks in RHEED patterns taken in the [010] azimuth as arising from anti-phase domains formed by tilted As–As dimer chains, and proposed several possible domain boundary configurations. This model, however, assumed a complete topmost layer of As dimers, but as soon as STM images were available [41], it became clear that this assumption could not be correct.

Larsen and Chadi [42] considered various possible alternative surface structures and concluded that a surface dimer vacancy model was the most probable, having either three dimers and one missing dimer or two dimers and two missing dimers in the (2×4) unit cell. The latter is effectively the $\beta 2(2 \times 4)$ unit cell, although it had not been established at the time. They then attempted to explain various anomalies in the RHEED patterns in the [110], [010] and $[1\,\bar{1}\,0]$ azimuths on the basis of defects which could be generated by an assembly of such unit cells. The important diffraction features were that all lattice points lay on a (1×4) reciprocal lattice, but with diffuse streaks in the half-order positions. They explained them with reference to the phasing of As–As dimers on either side of the missing rows of As atoms in the topmost layer. For a (2×4) unit cell, the phase is the same on both sides, but if it is opposite (in consecutive (2×4) unit cells along [110]), then the surface cell becomes $c(2 \times 8)$, as illustrated in Fig. 3. If the phasing is assumed to be random, one-dimensional disorder boundaries will exist and Larsen and Chadi attributed the diffuse half-order streaks to their presence. This type of structure has since been clearly identified in STM images by Pashley and co-workers [41].

Garreau et al. [37] used in situ grazing incidence X-ray diffraction of MBE grown surfaces vacuum quenched to room temperature to confirm the $\beta 2(2 \times 4)$ unit cell structure shown in Fig. 1 and to determine the atomic coordinates. They also observed a systematic broadening of half-order lines compared to the neighbouring quarter-order ones in a narrow azimuthal range, for which they claimed the only possible explanation was the presence of phase defects, due to the component of the fault vector along [110]. This means that the only defects they identified are cooperative shifts of whole columns of (2×4) units along the [110] direction on a "substrate template", i.e. the bulk (1×1) structure. The shifts are then $\frac{1}{4}a_0$, $\frac{1}{2}a_0$ and $\frac{3}{4}a_0$ with respect to the unfaulted reconstruction domain, where a_0 is the bulk lattice parameter. This type of defect is known as a kink and has been unequivocally identified in STM images [43], but in general kinks are present in extremely low numbers unless growth is carried out using very high As/Ga flux ratios and we will not discuss them further.

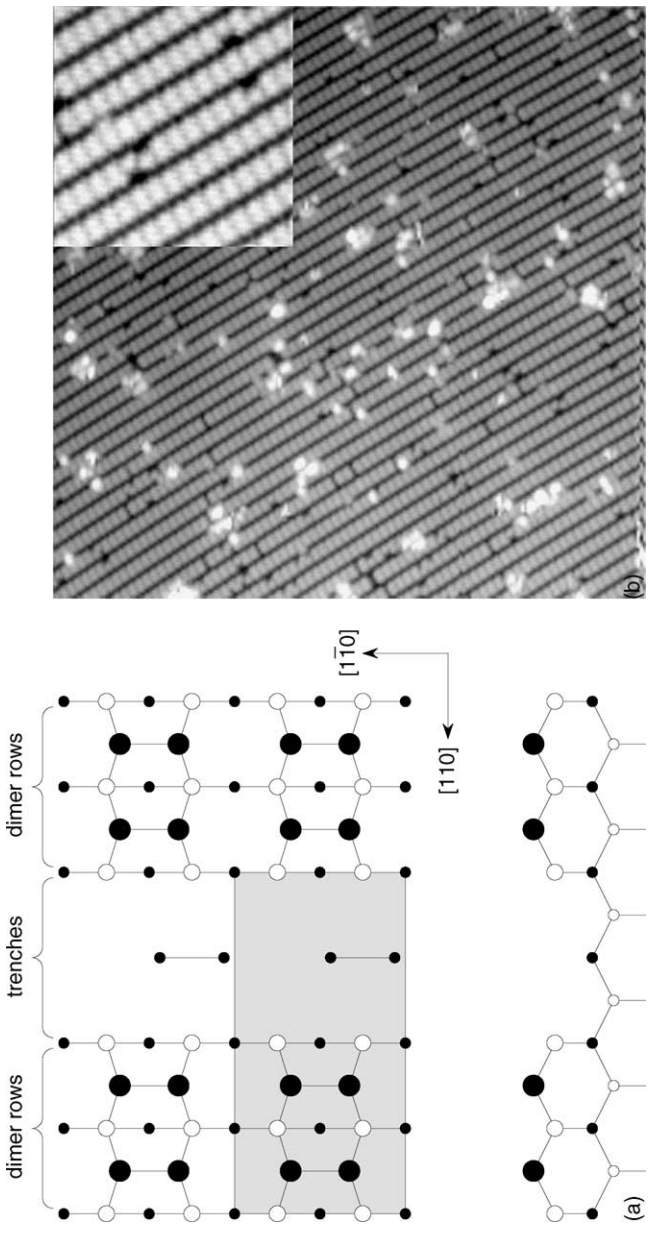

Fig. 1. (a) Side and plan views of the $\beta 2(2 \times 4)$ reconstruction of GaAs(001), showing alternate rows of dimer pairs and missing-dimer trenches. Filled and open circles represent As and Ga atoms, respectively, with the size indicating their proximity to the surface. The surface unit cell is indicated by the shaded region. (b) STM image (500Å $\times$ 500Å) of GaAs(001)–(2×4) after annealing and quenching. The inset shows a higher resolution (100Å $\times$ 100Å) image of the same surface. The dimer rows and missing dimer trenches appear as light and dark stripes, respectively, running diagonally across the images (courtesy: G.R. Bell).

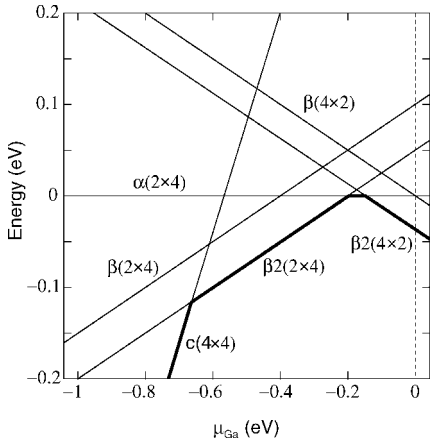

Fig. 2. Formation energies for a (1×1) unit cell for GaAs(001) surfaces as a function of μ_{Ga} over the thermodynamically range $-0.92 < \mu_{Ga} - \mu_{Ga(bulk)} < 0$ eV as determined by first-principles density functional calculations (adapter from [9]). The emboldened line indicates the equilibrium phases.

Hashizume et al. [36] analysed RHEED rocking curves from $\beta 2(2 \times 4)$ surfaces using dynamical theory [44] and apparently confirmed the unit cell structure illustrated in Fig. 1, as did a comparison of calculated and experimental RHEED spot intensity profiles in the $[1\ \bar{1}\ 0]$ azimuth, but the influence of defects on the RHEED patterns was not discussed. Subsequently, Ichimiya et al. [45] also concluded from RHEED rocking curves and integrated intensities in the $[1\ \bar{1}\ 0]$ azimuth that the 2-dimer, 2-missing dimer $\beta 2(2 \times 4)$ structure shown in Fig. 1 is the correct form. They also pointed out that for this azimuth the diffraction pattern consists of spots on the Laue zone circles, so the antiphase domains [relative areas of (2×4) and $c(2 \times 8)$] are large in this direction. The significant omission, however, is that there are no rocking curve data for fractional (half) order beams in the [110] azimuth. This reflects much earlier work

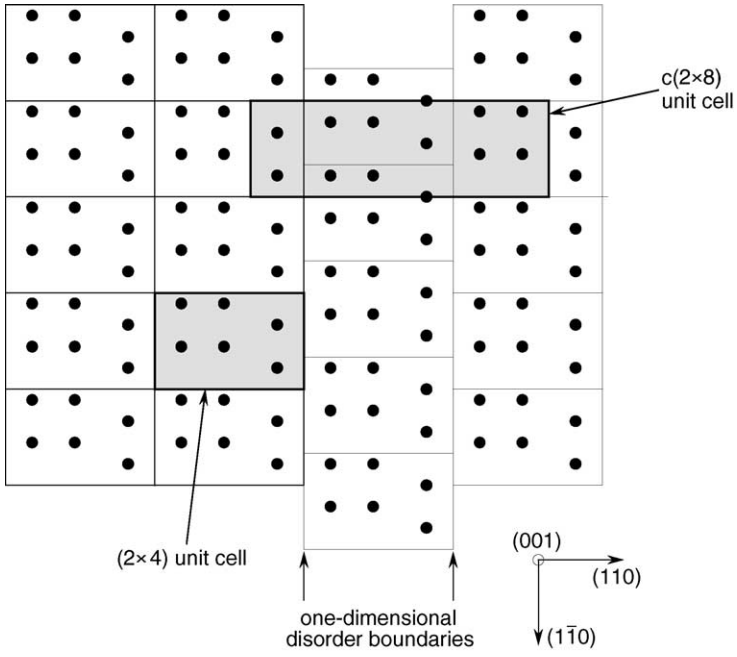

Fig. 3. Alignment of As-dimers showing the formation of (2×4) and $c(2 \times 8)$ unit cells.

by Larsen et al. [46], who reported that in this azimuth it was only possible to measure rocking curves for the (0, 0) and integral order beams because the elastically diffracted fractional order beams were very weak. We note that McCoy et al. [47] showed very good agreement with the data of Larsen et al. in their calculation of the atomic geometry of the $\beta 2(2 \times 4)$ surface using dynamical RHEED intensity analysis, but of course they were also restricted to [1 $\bar{1}$ 0] azimuthal results.

Very recently, Ohtake et al. [39] have analysed RHEED rocking curves and patterns, using dynamical theory, in both [1 $\bar{1}$ 0] and [110] azimuths from a $\beta 2(2 \times 4)$ surface, but again only integral order curves were available in the latter azimuth. They included the effects of antiphase domain regions introduced by Larsen and Chadi [42] and the agreement with the basic cell structure of Fig. 1 is good, but of course there are no results for the half-order feature in the [110] azimuth. Their calculation of atomic coordinates does, however, attempt to take account of the presence of the antiphase domains, despite this lack. The diffuse streaks in the [110] azimuth patterns are simply attributed to the domain structure based on the Larsen–Chadi model, so we must conclude that it has not yet been possible to take account of half-order beam effects in the [110] azimuth in any calculation of atomic positions from diffraction results.

Finally, LaBella et al. [38] used a combination of bias-dependent STM imaging and first principles electronic structure calculations to provide evidence for the atomic arrangement of the $\beta 2(2 \times 4)$ structure, but their images were very high magnification and resolution, from a very small area, and so were not appropriate for visualizing antiphase defects. Antiphase domains did not feature in the calculations.

It is quite clear that there is extensive agreement on the basic structure of the $\beta 2(2 \times 4)$ unit cell and even quite good agreement on the atomic coordinates within that cell, at least when the symmetry is treated as (2×4), and the possible implications of even a $c(2 \times 8)$ arrangement are ignored. An antiphase defect does, however, appear to be present on all surfaces displaying the $\beta 2(2 \times 4)$ structure and results from the phasing of As–As dimers on either side of the missing As-dimer rows along [110], to produce either (2×4) or $c(2 \times 8)$ cells and their associated antiphase domain boundaries, as shown by the STM image in Fig. 4, which also shows a roughly equal distribution of the two domains. As we have stressed though, there has been no significant attempt to take account of this structure in the determination of atomic coordinates, and furthermore, RHEED azimuthal information has been restricted to [110], [010] and [$\bar{1}$ 1 0].

From the point of view of understanding the atomistics of any subsequent growth process we can conclude that the substrate structure is apparently known at a very local level, although a multi-azimuthal RHEED investigation has thrown some possible doubt on the As atom positions in the third layer [48], because a problem concerning the intensity of certain reflections as the azimuth is rotated has emerged. Over areas realistic for growth studies, however, structural details are significantly less well defined.

2.3. Island nucleation

In this section we describe a step-by-step growth sequence for the nucleation and growth of islands on GaAs(001)–(2×4). Although the actual sequence has not been observed directly, we will support the plausibility of each step with the available theoretical and experimental evidence. Our model differs from previous descriptions in that the As dimers are not treated simply as ‘‘opportunistic spectators’’ waiting for an appropriate configuration of Ga atoms, but as active participants in Ga incorporation. We argue that excess As dimers perform an essential function in providing a pathway for the insertion of Ga into the top-layer surface As dimers. This provides a natural means of understanding where island nucleation is most favourable.

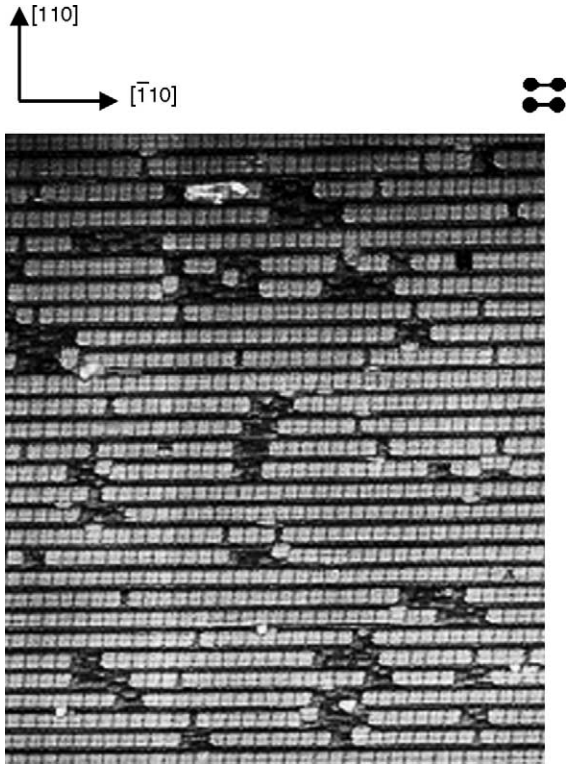

Fig. 4. STM image from a $\beta 2(2 \times 4)$ surface, showing the presence and distribution of antiphase domains.

2.3.1. Step one: As-dimer breaking

The breaking of an As-dimer bond by Ga is not a facile process, but the adsorption of an As-dimer onto adjacent As-dimers in the top row of the (2×4) reconstruction, with the concomitant breaking of the two dimer bonds, has no barrier if the adsorbing dimer tilts so as to break those bonds successively [17]. The resulting configuration, which satisfies electron counting and is shown in Fig. 5(b), has a binding energy (which is also the desorption barrier) of 1.6eV. Under sufficiently As-rich conditions and/or low enough temperatures, the adsorption of As onto such sites leads to the formation of the $c(4 \times 4)$ surface reconstruction [17]. The ideal $c(4 \times 4)$ surface consists of rectangular blocks of top-layer As atoms, with six atoms per block and each block associated with a lattice point of the $c(4 \times 4)$ structure, so that all top layer As atoms are bonded to three other As atoms [49].

There is an assumption [49] that the structure of the growing surface is actually $\gamma(2 \times 4)$, because in situ RHEED observations show that in the $[\bar{1}\,1\,0]$ (four-fold symmetric) azimuth the $(0, \frac{2}{4})$ rod has a very low intensity. This feature is also observed when a $\beta 2(2 \times 4)$ surface is cooled in an arsenic flux to about 520 °C[36,49,50] and appears to be related to a structure that is intermediate between $\beta 2(2 \times 4)$ and $c(4 \times 4)$. STM images formed by cooling in this way show a high density of kink sites, which implies an excess of arsenic in the surface, consistent with the method of formation.

The $\gamma(2 \times 4)$ therefore corresponds to a higher As coverage than the $\beta 2(2 \times 4)$ and, according to recent molecular dynamics calculations [51], it has the lowest surface energy for the whole range of As coverage (chemical potential) relevant to growth. Very similar conclusions can be drawn from the first-principles total energy calculations of Northrup and Froyen [9,10], Ohno [52] and Schmidt and Bechstadt [53]. Northrup and Froyen additionally showed that a "uniform" $\gamma(2 \times 4)$ structure, i.e. $\beta 2(2 \times 4)$ with a random distribution of kinks, is unstable with respect to dissociation

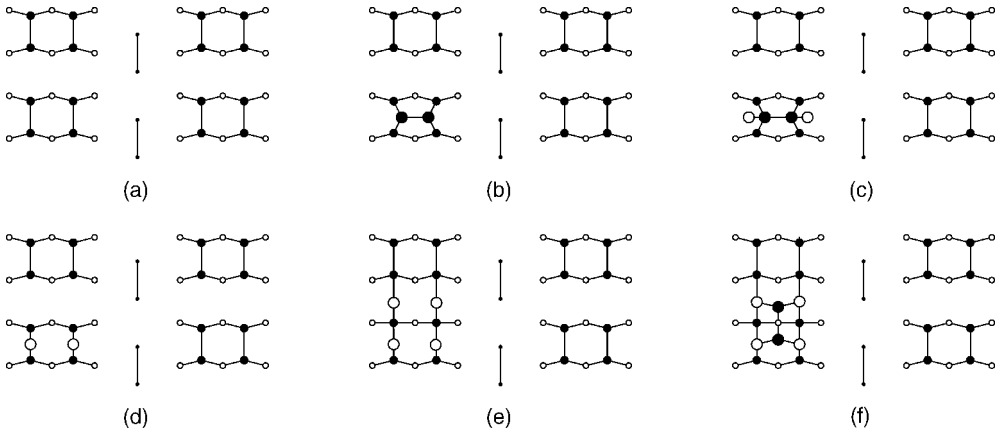

Fig. 5. Illustration of a possible attachment sequence of Ga and As atoms to the $\beta 2(2\times 4)$ surface. (a) Schematic diagram of topmost three layers of a $\beta 2(2\times 4)$ surface. (b) Adsorption of an As-dimer onto adjacent As-dimers in the topmost layer. (c) Ga atoms bond to dangling bonds of the adsorbed As_2. (d) Two Ga adatoms with adjacent As-dimer bonds. (e) Ga atoms in long-bridge sites, adjacent to the Ga dimer. (f) Formation of bonds with As_2. Filled and open circles represent As and Ga atoms, respectively.

into $\beta 2(2\times 4)$ and $c(4\times 4)$, such that $\gamma(2\times 4)\rightarrow \frac{3}{4}\beta 2(2\times 4)+\frac{1}{4}c(4\times 4)$, which is exothermic by 0.06eV per (1×1) cell. Zhang and Zunger [50], using the linear combination of structural motifs (LCSM) method (an approximate ab initio calculation for large unit cells), have also shown that the $\gamma(2\times 4)$ is not a single surface phase, but is a mixture of $\beta 2(2\times 4)$ and $c(4\times 4)$. All of these calculations are in accord with the original proposal by Hashizume et al. [36], based on RHEED and STM observations.

The question arises as to whether the comparison of a single RHEED feature is an adequate criterion on which to base the claim that the growing surface also has the $\gamma(2\times 4)$ structure. In fact it is not, as a more detailed examination of the relevant RHEED patterns (Fig. 6) shows that they are quite different from those formed by the γ-structure. The intensity distribution for all of the diffraction features from the growing surface is significantly different from that for the $\gamma(2\times 4)$ surface, and also from that of the static $\beta 2(2\times 4)$ structure. A full analysis of the recent RHEED results [48] is not yet available, but clearly the growing surface neither perpetuates the $\beta 2(2\times 4)$ starting template, nor is it represented by the $\gamma(2\times 4)$ model. It is also important to note that the pattern observed during growth appears within <40ms of the initial incidence of the Ga flux, but after termination of growth it takes almost 1h at 580 °C in an As_2 flux for the ideal $\beta 2(2\times 4)$ pattern to be fully restored. Despite these differences, it is probably safe to conclude, however, that the growing surface is more As-rich than the $\beta 2(2\times 4)$.

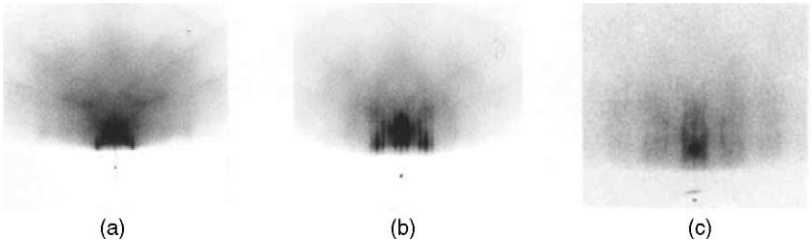

Fig. 6. RHEED patterns showing (a) static $\beta 2(2\times 4)$ surface (i.e. at 580 °C in an As_2 flux, no Ga flux); (b) growing surface at 580 °C; (c) $\beta(2\times 4)$ surface at 530 °C, all in the $[\bar{1}\,1\,0]$ azimuth.

2.3.2. Step two: insertion of Ga into dimers

Provided we assume that the growth surface contains As dimers, with the possibility of trimers also being present, which seems completely reasonable even though a detailed knowledge of the surface structure is not available, the breaking of adjacent As dimer bonds by an adsorbed dimer creates a pathway for the adsorption of Ga adatoms into the trimer bonds so-formed. Assuming an attempt frequency of $10^{13}\ s^{-1}$, the lifetime of As_2 in the precursor state is estimated to be 3×10^{-4} s at 580 °C. The hopping barriers of Ga adatoms have been calculated [17] as 1.5eV along [110] and 1.2eV along $[\bar{1}\,1\,0]$ for the $\beta 2(2 \times 4)$ surface. Since the process with the lower barrier (parallel to the dimer rows) will dominate migration, and again assuming an attempt frequency of $10^{13}\ s^{-1}$, we estimate the residence time per site of a Ga adatom to be 10^{-6}s, which is two orders of magnitude shorter than the lifetime of an As_2 molecule in the precursor state. The binding energy of the adsorbed As_2 is practically the same if either there are adsorbed As_2 dimers in adjacent (2×4) unit cells or if there are Ga adatoms in nearby trenches [19]. When combined with the fact that the reservoir of As_2 precursors is filled under typical (As-rich) conditions [14,15], the implication is that there is an appreciable number of adsorbed As-dimers having a trimer configuration, apparently providing ample opportunity for the attachment of a pair of migrating Ga adatoms to an adsorbed As_2. This produces the configuration shown in Fig. 5(c), in which the Ga atoms are bonded to the dangling bonds of the adsorbed As_2. The As_2 then desorbs into the reservoir, leaving two Ga adatoms with adjacent As-dimer bonds, as shown in Fig. 5(d). Each of these Ga atoms uses two electrons to form bonds with two As atoms. The remaining electrons are used to form a bond between the two Ga atoms, thus creating a stable surface complex [21] in which the Ga dangling bonds are empty, so satisfying the electron counting rule. Indeed this is the most stable of the nearest-neighbour pairs of Ga atoms calculated by Kratzer et al. [21].

We need to ask to what extent we are able to provide experimental evidence for these mechanisms. The (assumed) presence of As-trimer bonds is not particularly well supported by STM images produced by quenching the growing surface, which look much more $\beta 2(2 \times 4)$-like. It is probable, however, that there will be sufficient time at elevated temperatures (>580 °C) during quenching to desorb a substantial proportion of the excess arsenic, so that on subsequent examination by STM the surface will indeed appear to be $\beta 2(2 \times 4)$. An indication of the relative bond strengths of As dimers and trimers can be inferred from thermal desorption spectroscopy (TDS). Banse and Creighton [54] found that the lowest energy state was that of As_4 desorbing from a $c(4 \times 4)$ surface, with a desorption peak temperature of 440 °C. The fact that it is the tetramers which desorbs is interesting in itself, since it is only the dimer which has been identified desorbing from GaAs under both Langmuir and Knudsen conditions in the absence of excess arsenic [55]. TDS results for desorption from a (2×4) surface did indeed show that As_2 was desorbing with peaks at 480 and 570 °C[54], i.e. significantly higher than for $c(4 \times 4)$. This is also in keeping with the modulated molecular beam results of Foxon and Joyce [33], who showed that at comparatively low temperatures during growth from As_2, As_4 was a desorption product, formed either by association of dimers in the precursor state or by desorption from a $c(4 \times 4)$ surface phase. On the other hand, Sasaoka et al. [56] observed only the desorption of As_2 from the $c(4 \times 4)$ surface, although they did see As_4 desorbing from a (2×4) structure on which As_4 had previously been adsorbed, and they also found there to be a desorption peak corresponding to the excess arsenic on the $\gamma(2 \times 4)$ surface. The reason for this apparent discrepancy is not clear, but the important point is that both $c(4 \times 4)$ and $\gamma(2 \times 4)$ surfaces contain As which is more weakly bound than on the $\beta 2(2 \times 4)$.

No spectroscopic RDS measurements have been made during growth, but results obtained during adsorption or desorption of As_2 on or from variously reconstructed GaAs(001) surfaces may possibly

shed some light on the dynamics of high As-coverage structures. In a recent paper, Begarney et al. [57] reported on the spectral changes that occurred as the arsenic surface coverage was progressively reduced by desorption and compared them to the corresponding STM images. The effective starting surface was $c(4 \times 4)$, with an As coverage of ≈ 1.75 monolayers (ML), which ultimately transformed to (2×4) as arsenic was desorbed. The most interesting feature concerns the vertical arrangement of surface phases. After a ≈ 0.1ML reduction of the As coverage, pits one bilayer deep, but retaining the $c(4 \times 4)$ structure, began to form, but with some distortion of the blocks of three trimers that formed the initial surface. Since the pits have the same structure as the upper surface, they must have been formed by the exchange of As atoms in the second layer down with Ga atoms in the third. This surface retains the $c(4 \times 4)$ symmetry everywhere in the image, but a negative spectral feature at 2.8eV disappears during the desorption, so it cannot relate solely to the basic $c(4 \times 4)$ arrangement, contrary to previous suggestions [24,58].

With further As desorption, a positive peak emerges at 2.9eV, with a peak height proportional to the developing (2×4) coverage. During this surface "phase transition", three atomic bilayers are exposed, with (2×4)-like regions in the layer above the $c(4 \times 4)$ structure. This is in complete accord with previous results: Kamisawa and Yamaguchi [59] used STM to observe the transition in both directions [i.e. from $c(4 \times 4)$ to (2×4) by heating under an As_4 flux and in the reverse direction by cooling in an As_4 beam] and found that the (2×4) domains were above the $c(4 \times 4)$ regions irrespective of the direction of the transition. They believed, however, that the process involved Ga detachment only from steps, whereas it is clear that pit formation on terraces leads to the same morphology in a more general way. Bell et al. [60] also found the same result without the need to invoke steps.

An important feature of the transition region observed by Begarney et al. [57] is the appearance of a $(2 \times n)$ periodicity, where the two-fold direction is along $[\bar{1}\,1\,0]$, but the value of n along [110] can be 2, 3 or 4, corresponding to distances of 8, 12 and 16Å, respectively, together with a positive optical feature at 3.25eV. The $c(4 \times 4)$ and $(2 \times n)$ structures are at the same height, but are $0.35a$ (a is the GaAs lattice constant) below the (2×4). The uppermost ridge of the $(2 \times n)$ must therefore be due to As dimers along [110] bonded to As atoms in the second layer to form trimers, which evoked the suggestion from STM contrast that the ridge dimers are buckled, with the up-atom having a filled lone-pair and the down atom an empty or half-filled orbital. This is a transition phase that does not obey the electron counting rule, but it is significant that spectral features exist which are indicative of transition states and the bonding arrangement might correspond to the state during growth immediately prior to Ga insertion. At least there is definite evidence that excess As can exist in the surface, which would facilitate the insertion process.

2.3.3. Step three: the critical nucleus

The bonding of As_2 to the GaAs(001) surface requires 14 electrons: two in each of four covalent bonds to the surface, two in the dimer bond and two in each of the two dangling bonds. The As_2 molecule supplies ten electrons, so the surface must supply the remaining four. The Ga configuration shown in Fig. 5(d) provides only two electrons (those in the dimer bond), so the nucleation of the next layer of GaAs requires additional neighbouring Ga atoms. Steric considerations require these Ga atoms to reside in long-bridge sites adjacent to the Ga dimer, producing the configuration in Fig. 5(e), which first-principles calculations [21] show to be stable. Each of the two additional Ga atoms bond to As dangling bonds, leaving their three valence electrons in higher-energy dangling bond states. These states are the reason why this configuration is considerably less stable than that with four Ga atoms in Ga-dimers within adjacent top row As-dimers, which clearly does satisfy electron counting. The excess of electrons promotes the formation of bonds with As_2, forming the configuration shown in

Fig. 5(f). It is possible that this complex forms initially with only three Ga atoms (As_2 binding energy of 1.7eV [21]), after which the fourth Ga is captured, but As_2 bound to four Ga atoms is by far the most stable complex, with a binding energy of 2.7eV. This configuration has four electrons in each of two Ga dangling bonds, so it does not satisfy the electron counting rule, but these electrons will be required for growth across the trenches.

2.3.4. Quantitative considerations

STM experiments have been combined with modelling to characterize more quantitatively the atomistic nucleation, growth and structural transformation kinetics, using a mean field rate equation as well as KMC simulation [14,15]. We therefore need to resolve how accurately such data reflect real-time surface processes, bearing in mind that the actual images are only snapshots of quenched surfaces.

One of the most direct measures of surface nucleation and growth kinetics is obtained from the submonolayer regime, prior to any significant coalescence, by plotting the sizes of islands, as a fraction of the average island size, against their relative frequencies. The distributions of island sizes so obtained have been studied for a variety of systems and their gross shapes shown to be sensitive to processes such as attachment and detachment rates [61,62], barriers to attachment [63], and evaporation [64,65]. The submonolayer island morphology also provides important signatures about processes that are operative in the multilayer regime. Apart from processes that are intrinsic to a system, such as those cited above, there is the effect of various surface impurities that are introduced either deliberately (''surfactants'') or are unavoidably present because of the polyatomic molecules used to deliver the atoms of the growing material, as in MOVPE. Surfactants improve film morphology by promoting layer-by-layer growth while reaction products can inhibit the lateral growth of islands [66,67] by attaching to island edges and thus forming a (physical and free energy) barrier to attachment by adatoms [63,68,69]. Strain can also affect the morphologies of heteroepitaxial islands by causing island-size- and island-shape-dependent changes to both attachment and detachment barriers at island edges [70,71]. An important practical application of these ideas is the growth of three-dimensional (3D) islands during Stranski–Krastanov (SK) growth, as will be discussed in the second part of this review.

Fig. 7(a) shows the island-size distributions obtained from STM images compared with distributions obtained from kinetic Monte Carlo simulations. The distributions for the (110) and (111)A surfaces are qualitatively similar and have the general form expected for critical island sizes $i^* > 1$ and single adatom capture [62,72,73]. The corresponding simulated distributions are labelled by the ratio λ of the detachment rate of single atoms from island edges to the capture rate of adatoms by all islands [62], with the value for the (111)A surface ($\lambda \approx 10$) being discernibly larger than that for the (110) surface ($\lambda \approx 5$). The overall level of agreement between the simulated and measured distributions allows us to conclude that for these two surfaces, islands nucleate by the binding of two Ga adatoms and grow by the net capture of single Ga adatoms. The distribution for GaAs(001)–(2 ×4) suggests an altogether different growth sequence. One of the main aspects of the KMC simulations for growth on this surface was that islands can grow only by the simultaneous capture of two migrating Ga adatoms. This produces a scenario whereby the nucleation of islands is a more facile process than growth, resulting in a large number of small islands.

STM images of islands on the low-index surfaces of GaAs, which are displayed in Fig. 7(b), confirm the basic conclusions deduced from the island-size distributions. The images of GaAs(001)–(2×4), GaAs(110), and GaAs(111)A–(2×2) were taken after the deposition of 0.2ML of Ga. The growth temperatures were 580 °C for the (001) and 480 °C for the (110) and (111)A surfaces, with

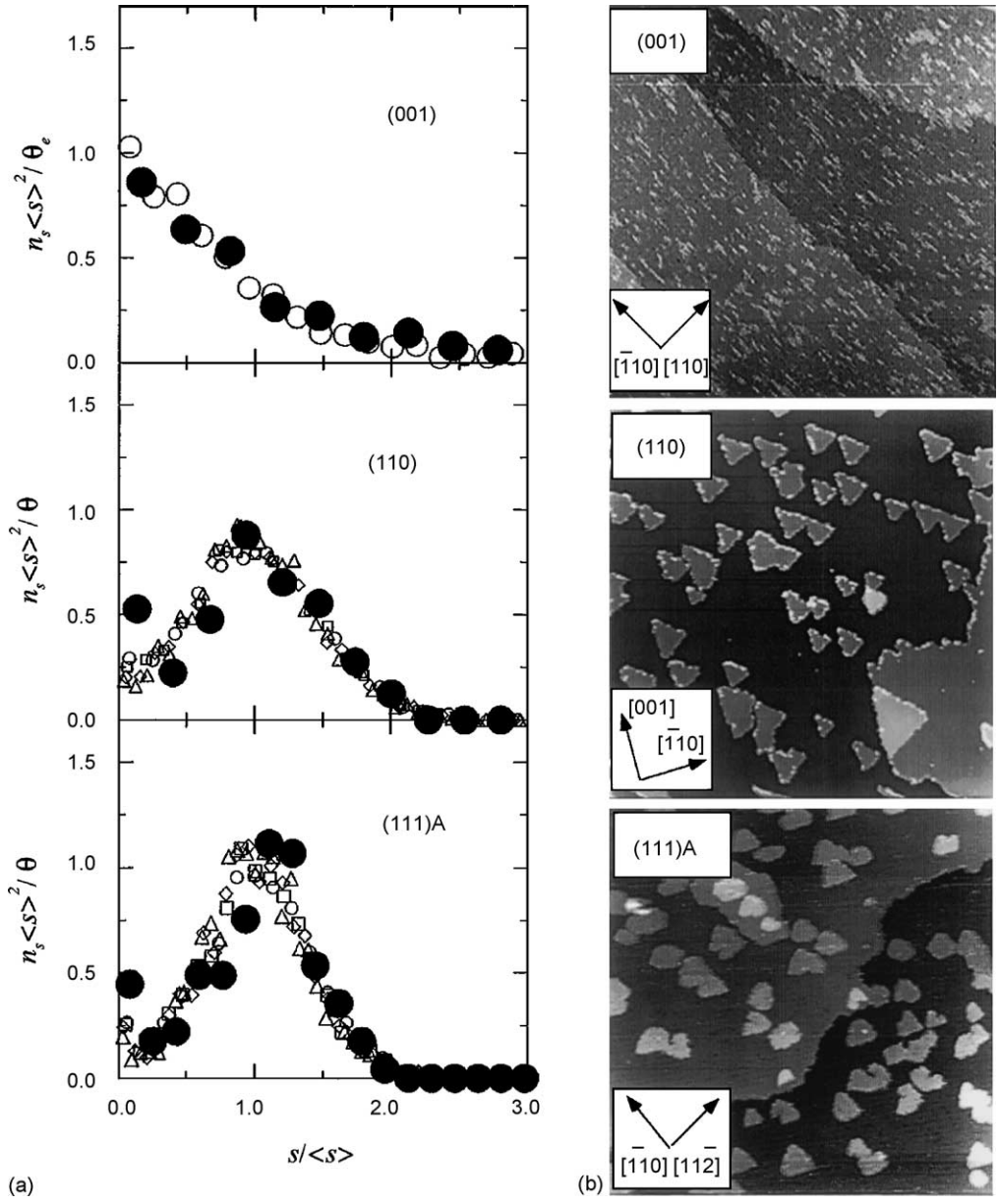

Fig. 7. (a) Island-size distributions obtained from STM images (filled circles) and from simulations (open symbols) at 0.2ML. The simulations for (001)–(2×4) are described in the main text; those for (110) and (111)A–(2×2) are taken from [61]. (b) Filled-state STM images of the low-index GaAs surfaces taken after the deposition of 0.2ML of Ga. The scan area for each image is 2000Å $\times$ 2000Å. Reprinted with permission from A.R. Avery et al., Phys. Rev. Lett. 79 (1997) 3938. Copyright 1997 by the American Physical Society.

As:Ga = 6:1 and a growth rate of 0.05ML/s for all surfaces. The (001) and (110) surfaces were grown with As_4 and the (111)A was grown with As_2. The islands on the (110) and (111) surfaces are similar both in density and morphology, but the morphology of the (001) surface is strikingly different from these, having a much higher density of smaller islands. An analysis of several images yields island densities of 17,000, 1700 and 1900 μm^{-2} for the (001), (110), and (111)A surfaces, respectively, i.e.

an order of magnitude difference between the island density on the (001) surface and those on the other two surfaces at the same nominal coverage.

3. Part II: heteroepitaxial growth: InAs on GaAs

3.1. Effects of strain on growth

We now turn our attention to the heteroepitaxial system in which GaAs is the substrate material for the deposition of InAs, where there is a lattice mismatch of $\approx 7\%$. This has important consequences for the growth mode, the mechanisms by which the strain is relaxed during growth and the processes of alloying and segregation. Classically [74,75,76], three basic growth modes have been identified, referred to as Frank–van der Merwe (FM) or layer-by layer growth, Volmer–Weber (VW), or the growth of three-dimensional islands which leave part of the substrate uncovered, and Stranski–Krastanov, where 3D islands form on top of a complete, flat pseudomorphic film one or two monolayers thick. The FM and VW growth modes can be understood on the basis of thermodynamic wetting arguments. FM growth is favoured if

$$\gamma_e + \gamma_i < \gamma_s, \tag{1}$$

where γ_e, γ_i and γ_s are the free energies of the vacuum-layer interface, the layer-substrate interface and the substrate-vacuum interface, respectively. The VW mode occurs when

$$\gamma_e + \gamma_i > \gamma_s, \tag{2}$$

The SK mode, however, is generally accepted to be strain related, because the accommodation of misfit strain between layer and substrate requires that the strain energy increases with film thickness, which means that while the balance is initially in favour of two-dimensional (2D) growth, it subsequently becomes unstable and 3D islands are formed. The strain in these islands is then relaxed by the formation of misfit dislocations.

There is, though, a second version of SK growth, whereby coherent (i.e. dislocation free) 3D islands form on top (or from within) the WL. These islands have become known as self-assembled quantum dots (QDs) and not only do they have significant technological potential, ranging from optoelectronic devices to possible architectures for quantum computing, but they also provide an excellent basis for studies of fundamental issues in heteroepitaxial growth. It is only the latter aspect we will deal with here and we begin by discussing the mechanisms that are available for the relaxation of the misfit strain. We do this by considering the behaviour on the three low index orientations and extend it to the influence of surface reconstruction on (001) oriented substrates. This leads naturally into a treatment of WL formation, i.e. the development of an alloy layer prior to the creation of misfit dislocations or of any form of 3D entities. We follow this by examining the nucleation kinetics of QDs, their size distribution and scaling behaviour, and compositional and crystallographic properties in relation to the SK growth mode. Several theoretical approaches, both kinetic and equilibrium models, have been developed in an attempt to account for QD formation, but none is so far capable of explaining self-consistently all of the experimental observations. The methods adopted include the thermodynamics of WL formation and of the 2D to 3D transition, equilibrium theories of 3D island formation and kinetic models for 3D island nucleation. We will discuss their various strengths and shortcomings.

3.2. Growth modes and strain relaxation on low index surfaces

Although we will be principally concerned with the evolution of QDs, we need first to demonstrate that this mode of strain relaxation and growth is the exception in the InAs–GaAs system, and is in fact limited to one orientation and probably also to one surface reconstruction of that orientation. While the subject matter of this article does not allow us to treat materials generally, this behaviour appears to be typical of many other semiconductor systems.

If there were a simple relationship between strain relaxation and growth mode they should be independent of orientation since the misfit is constant. That this is clearly not the case can be seen by examining the relative effects on the three low index orientations. We begin with growth on GaAs(110), because here the additional complexity of different surface reconstructions does not occur; the surface retains a (1×1) periodicity with respect to the bulk lattice. There is only relaxation brought about by a rotation of As atoms outwards and Ga atoms inwards by $\approx 30^\circ$ about the surface plane, with no significant change of bond length.

3.2.1. Growth on GaAs(110)

This system has been examined in considerable detail using RHEED, TEM, TED and STM for a wide range of film thicknesses, using growth by MBE (beams of In and As_2). The important features are (i) growth follows a layer-by-layer (FM) mode, irrespective of thickness; (ii) the extent of formation of a WL is temperature dependent and appears to involve only cation exchange in the topmost ML, rather than extensive alloying; and (iii) strain relaxation occurs solely by the formation of misfit dislocations.

3.2.1.1. Growth mode. It has been shown, using a combination of RHEED [77] and STM [78], that growth occurs by a process of 2D island nucleation and coalescence. The RHEED patterns in the orthogonal [1 $\bar{1}$ 0] and [001] azimuths remain streaked at all thicknesses in the range investigated, as illustrated in Fig. 8 for the substrate and a 60Å thick film. As expected, there is no evidence of reconstruction and the intensity modulation along the rods arises mainly from intersection with Kikuchi lines; with growth the rods become somewhat broader and more uniform in intensity, but importantly there is no indication of a spot pattern or the presence of facets, either or both of which would occur if growth became 3D. STM images of sub-ML coverages show a number density of 2D nuclei comparable to that for GaAs(110) homoepitaxy ($1.0 \times 10^{11}\text{cm}^{-2}$, cf. $2.7 \times 10^{11}\text{cm}^{-2}$) for very similar growth conditions. The nuclei are either single ML or double ML (a bilayer) in height, which is also a common feature of GaAs(110) homoepitaxy for a wide range of growth conditions

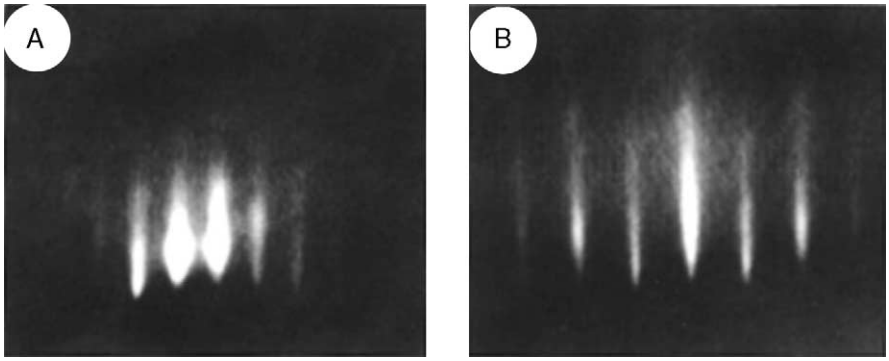

Fig. 8. RHEED patterns for growth of InAs on GaAs(110) in (A) [1 $\bar{1}$ 0] and (B) [001] azimuths.

[79]. For un-annealed substrates with a high step density, growth occurs predominantly by propagation from these steps, rather than by the formation of separate 2D nuclei, which is the expected behaviour for this growth mode.

3.2.1.2. WL formation. Since the growth mode does not involve the SK mechanism, it could be anticipated that a WL will not form, and this is effectively the case in practice. The coverage is virtually identical to the amount of deposit and at 420 °C only 2% of In atoms occupy Ga sites in the topmost substrate layer, the rest form ML InAs islands with no alloying. The extent of alloying is somewhat greater at 480 °C, but still appears to be limited to the topmost substrate layer, since the coverage does not exceed that expected from the amount of InAs deposited.

3.2.1.3. Strain relaxation. In this system, strain relaxation, even in films a few ML thick, is entirely by the formation of misfit dislocations and occurs via two separate pathways [78,80,81,82,83,84]. The first involves the direct nucleation of pure edge-type dislocations, where the network is confined to the plane of the interface, with no threading component, and is fully established in the thickness range 3–5ML. The elastic distortion extends from the dislocations up to the surface. The complete process can be followed in situ by STM, as illustrated in Fig. 9. It develops from the formation, at $\approx$2ML, of a network of very small InAs islands separated by 1ML deep trenches. These islands are extremely uniform in size (mean area 495Å^2, standard deviation 310Å^2, height 2Å, or 1ML). With further deposition from 3 to 5ML, larger islands form which do not break up into a network, but which expand and coalesce to form an array of dark bands lying parallel to [001] and separated by $\sim$60Å. The dark bands are the manifestation at the surface of the interfacial misfit dislocations, having Burgers vector $\boldsymbol{b} = \frac{1}{2}a_0[1\,\bar{1}\,0]$ lying in the plane of the interface, where $a_0 =$ 6.054 Å for InAs. The island size uniformity is an effect that best accommodates the elastic strain in the film and their coalescence leads to the incorporation of short misfit segments. The absence of any threading component means that the dislocations are not located in the first InAs layer, but in the second or third layers, since they are first observed at 3–4ML. The misfit dislocations are visible at the surface because of the lattice distortion they induce. Plan view dark field TEM images, exemplified in Fig. 10, from the same layer (but not necessarily the identical region) confirm the regular array of misfit dislocations with the same Burgers vector, line direction and spacing, which is that predicted for the system misfit.

The elastic displacement of atoms due to the edge dislocation strain field has been determined from the STM measurements of the depth and lateral dimensions of the surface depressions and compared with various theoretical treatments by Belk et al. [82] for films <30ML thick. It was found that for satisfactory agreement the models must contain a free epitaxial layer surface [85,86]. The observed decrease in surface vertical displacement with increasing film thickness was attributed mainly to an increase in the full width at half maximum (FWHM) in a closely packed dislocation array, where it rapidly attains a limiting magnitude corresponding to approximately half the dislocation spacing. For films <20Å thick there are still significant deviations between theory and experiment, reflecting an enhanced stiffness in these thin films, due to the high lateral atomic displacements near the dislocation cores which limit the capacity of the film to deform further.

Oyama et al. [87] used first principles calculations to determine the core structure of the dislocations and found it to be based on an asymmetric five-fold coordinated In atom, which is maintained as the film thickness increases. There is as yet no direct experimental evidence for this structure, but the vertical displacement calculated for a 4ML thick film was in reasonable agreement with experiment, so the calculated core structure may well be related to the enhanced stiffness.

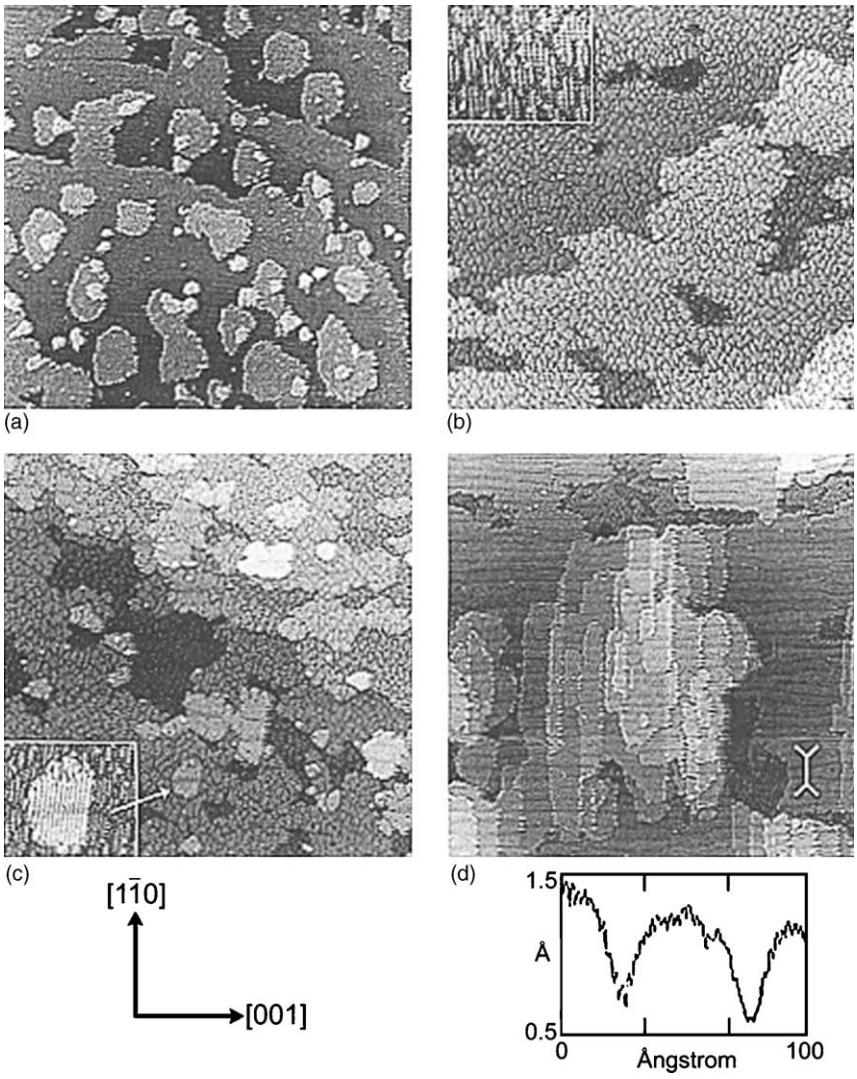

Fig. 9. STM images showing development of a misfit dislocation network during the growth of InAs on GaAs(110).

Maroudas et al. [83,84] have used a continuum elasticity theory combined with atomistic simulations to calculate surface profiles as a function of film thickness by introducing the concept of a thin compliant substrate. This is a substrate not constrained at its base, unlike the conventional case, and would correspond in practice to the growth of the GaAs buffer layer prior to the deposition of the InAs layer. The calculated profiles are in good agreement with those measured experimentally [84] for films $<$10ML thick, but do not realistically reproduce the rapid decrease in the size of the depression as the thickness increases, possibly because the density of the dislocation array was not treated adequately.

RHEED patterns show that the initial growth is pseudomorphic, i.e. the layer has the same in-plane lattice parameter as the GaAs substrate, but the strain begins to relax rapidly after $\approx$2ML and by 5ML it is $\approx$90% complete in the $[1\ \bar{1}\ 0]$ direction. This relaxation is not isotropic, however, because in the orthogonal [001] direction the strain relaxation process and the imaging by STM are quite different. Once the edge dislocation network is established, slip systems which relieve strain along

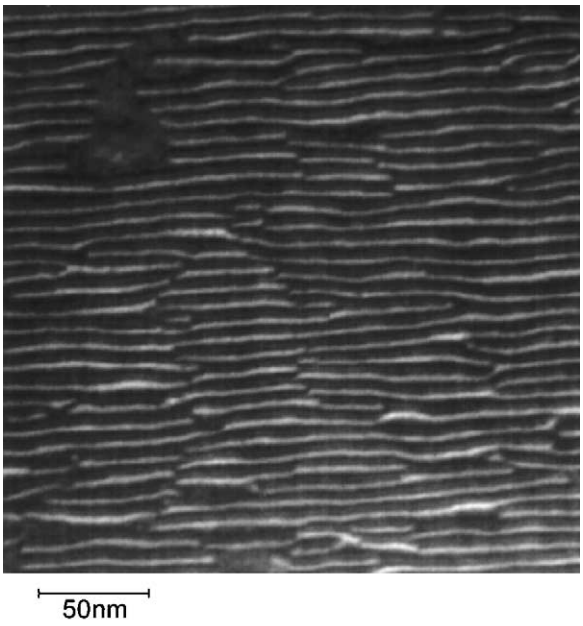

Fig. 10. Plan-view dark field TEM image ($g = [2\,\bar{2}\,0]$) confirming the array of misfit dislocations imaged by STM (Fig. 9) for 5ML of InAs on GaAs(110).

[001] become active, and although they relax strain much more slowly than the edge dislocations, they dominate the subsequent surface morphology. A discrete surface step is left behind by the motion of a dislocation towards the interface, followed by its expansion as the misfit segment increases in length. The morphology of the surface after deposition of 10ML of InAs is shown in Fig. 11, and shows the steps to be predominantly aligned along $[1\,\bar{1}\,0]$, with the buried network of pure edge dislocations still

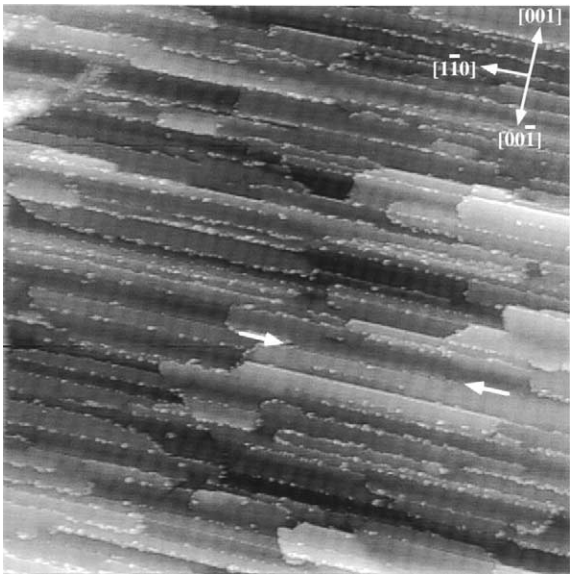

Fig. 11. STM image of 10ML of InAs on GaAs(110) showing slip steps, but with surface atom displacements due to pure edge dislocations still visible in the orthogonal direction.

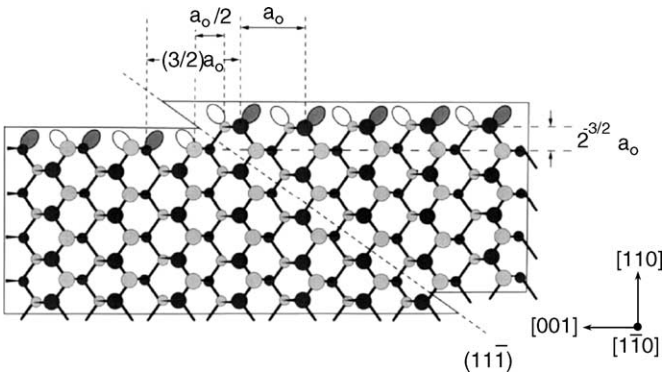

Fig. 12. Schematic diagram, viewed in the $[1\ \bar{1}\ 0]$ side projection, showing the structures (without surface reconstruction) for slip steps formed on the (110) surface after slip on the $(1\ 1\ \bar{1})$ plane for perfect 60° type, $\boldsymbol{b} = \frac{1}{2}a_0\,[1\,0\,1]$.

visible in the orthogonal [001] direction. The linear density of the steps is $\approx 125\,\mu\,m^{-1}$, giving an average separation of ≈ 80Å.

The orientation and compressive nature of the strain limit the number of possible slip steps to six, two due to partial dislocations and four linked to perfect dislocations. They can be identified from their associated Burgers vectors using TEM [88]. Of the perfect 60° dislocations, $\boldsymbol{b} = \frac{1}{2}a_0[1\,0\,\bar{1}]$ and $\boldsymbol{b} = \frac{1}{2}a_0[0\,1\,\bar{1}]$ on (111), and $\boldsymbol{b} = \frac{1}{2}[1\,0\,1]$ and $\boldsymbol{b} = \frac{1}{2}a_0[0\,1\,1]$ on $(1\,1\,\bar{1})$ are observed following "shuffle"-type slip in between the wider spaced {111} planes, as shown in Fig. 12. Both of these produce 2Å high steps shifted by half the unit cell dimensions parallel to the surface. Only the 90° set of partial dislocations is allowed for the relief of compressive strain, i.e. $\boldsymbol{b} = \frac{1}{6}a_0[1\,1\,\bar{2}]$ on (111) and $\boldsymbol{b} = \frac{1}{6}a_0[1\,1\,2]$ on $(1\,1\,\bar{1})$. Partial dislocation slip ("glide") is confined within the narrower-spaced {111} planes and the step displacement is two-thirds the (220) spacing normal to the surface (≈ 1.3Å). Both types have been identified in this system, and their combinations and directions of propagation, which lead to a complex morphology, have been analysed in detail by Belk et al. [80], and will not be discussed further here.

There is a third process which can relax strain in the [001] direction. It has been observed by X-ray diffraction space mapping [89] and transmission electron diffraction [88] that tilting occurs (i.e. the (110) planes of the layer and substrate are not precisely parallel). There is a divergence in the [001] direction of up to 2°, but any tilt in the orthogonal $[1\ \bar{1}\ 0]$ direction is negligible. It probably arises from differences in the proportions of the dislocations which slip down the two non-equivalent {111} planes.

3.2.2. Growth on GaAs(111)A

Results for this system are very similar to those for growth on (110) oriented substrates, although, of course, the strain-relaxing dislocation structures are significantly different. The GaAs (111)A surface also displays a (2×2) reconstruction, but the corrugation is only 0.2Å. RHEED results in both the $[1\ \bar{1}\ 0]$ and $[1\ 1\ \bar{2}]$ azimuths [90] first established that growth followed a 2D layer-by-layer (FM) mode for at least 30ML, over a significant range of growth conditions, and that substantial strain relaxation had occurred by 3ML. The (2×2) RHEED pattern is maintained throughout growth and coverages up to 1ML correspond to the amount of InAs deposited, so there is no substantial alloying or WL formation.

The growth behaviour has been followed directly by in situ STM observations [91], which also showed that there were ML high islands on the substrate surface. As a consequence, the initial growth mode involved both step propagation from these existing islands and 2D island nucleation in regions

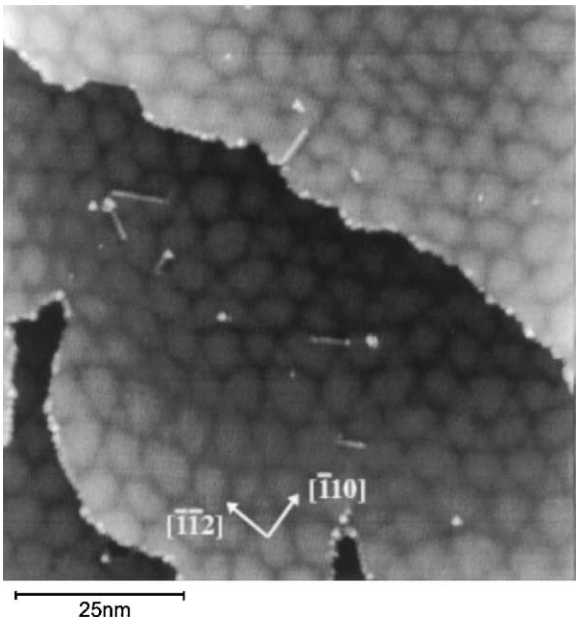

Fig. 13. STM image of a 5ML thick InAs film on a GaAs(111)A substrate showing a periodic network of misfit dislocations.

away from the steps. The islands coalesce as the amount of deposition is increased and the STM image from the boundary between any two coalesced islands displays a shallow (≈ 0.5Å) dark contrast band associated with the strain field induced by misfit dislocations at the interface. These misfit dislocations are in $\langle 1\,1\,2\rangle$ directions and can therefore form two differently directed junctions (see below). Fig. 13 shows a typical STM image of a 5ML InAs film, by which point the misfit dislocations form a periodic network, which is not, though, a simple hexagonal structure. The interface must therefore contain some partial dislocations at the boundaries between faulted and unfaulted domains, as well as stacking faults.

A possible model of the atomic arrangement for the InAs/GaAs(111)A interface based on these observations is shown in Fig. 14. Three partial dislocations form triangles referred to as types I and II, where the line density of dangling bonds is 1.5 times greater in type II than in type I, so it is likely that the type II dislocation has the higher formation energy. For complete relaxation the period of the network (labelled d in Fig. 14) has to be ≈ 6.0nm. Fig. 15 shows the extent of relaxation as a function of layer thickness, measured from STM images and also by TED for comparison. The level of agreement for the layer measured by both methods is clearly good and it can be seen that over 80% of the strain has been relaxed in layers ≥ 10ML thick. The threading segments of the misfit dislocations have been identified from high resolution STM images as having a Burgers vector of $\frac{1}{2}a_0[\bar{1}\,1\,0]$, which is in agreement with that for one of the perfect dislocations in the model shown in Fig. 14.

3.2.3. Growth on GaAs(001)

It is quite clear from the first part of this review on GaAs homoepitaxy that the GaAs(001) surface displays a very wide range of stoichiometry-dependent reconstructions, which can also show surface defect dependent variations. In this section we will concentrate on three stoichiometry-based reconstructions and largely ignore defect related issues, because any impact they might have on heteroepitaxy has not yet been treated.

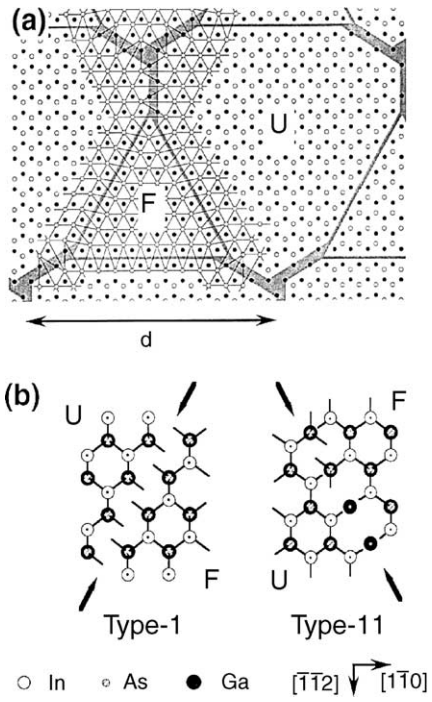

Fig. 14. (a) Model of the atomic arrangements for the InAs/GaAs(111)A interface. Open (closed) circles show the position of Ga (In) atoms just below (above) the interface. ''F'' and ''U'' indicate faulted and unfaulted interface regions, respectively. (b) The atomic structure of two possible partial dislocations.

3.2.3.1. Cation-stable surfaces. The Ga-stable GaAs(001) surface reconstructs with a (4×2) symmetry, but it is not possible to maintain such a structure in the temperature range available for InAs growth, which is typically $<520\ ^{\circ}C$. It is, however, possible to grow under strongly In-rich conditions, which results in a 2D layer-by-layer growth mode that has been called ''virtual surfactant-mediated epitaxy'' by Tournié et al. [92]. Experimentally, there seems general agreement [92,93,94,95] that this layer-by-layer mode operates provided that a (4×2) surface symmetry is maintained. This can either be established by the initial deposition of a sub-ML quantity of In at low temperature on a GaAs(001) (2×4) surface, followed by annealing at a higher temperature immediately prior to growth, or simply by growing with a very high In:As flux ratio at $\approx 450\ ^{\circ}C$, again starting from a (2×4) substrate surface. The reason for this behaviour is less clear-cut and

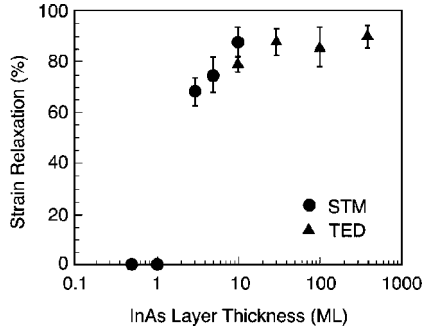

Fig. 15. Strain relaxation calculated from STM images and TED patterns as a function of InAs layer thickness on GaAs(111)A substrates.

conflicting arguments have been advanced, based on the assumption that it is indeed a surfactant mediated process. Snyder and Orr [93] correctly pointed out that lowering the surface free energy facilitates island formation, contrary to what is observed, but Grandjean and Massies [94] countered that this is an equilibrium concept and therefore does not apply to MBE growth. The issue therefore reduces to whether kinetics or energetics is the determining factor if the surfactant-driven hypothesis is correct. It is probably more realistic, however, to consider the growth mode displayed by this particular surface in the wider context of behaviour on low index GaAs surfaces in general, where the usual mode is 2D, so apparently this case is not exceptional.

We can do this by first considering the structure of the interface and how this translates into the growing film, and then dealing with the strain relaxation process. From a combination of STM, surface X-ray diffraction (SXRD) and X-ray standing waves (XSW), it appears that the initial stage of growth (<0.25ML) on a (2×4) substrate surface involves In adatom incorporation (as monomers) in the unoccupied hollow sites to form In rows along [110], to display a $(4 \times 2)/c(8 \times 2)$ symmetry. The In atoms are separated by 16Å between rows and by 4Å along them [95,96]. With increasing coverage up to 0.5ML, additional protrusions 8Å apart along [110] gradually fill the trenches to form what appears as a ladder-like structure in STM images. In subsequent growth, In adatoms fill the trenches and replace the surface Ga atoms to produce a structure essentially identical to the clean InAs–(4×2) surface, apart from relative In atom height differences produced by the mismatch strain. The mode remains 2D. It must therefore be concluded that whatever other factors might be important, the surface structure during growth is crucial for the 2D mode to propagate. It is also worth pointing out that the presently accepted model of the $4 \times 2/c(8 \times 2)$ reconstruction has moved away from the earlier concept of cation dimers in the outermost layer [95] to a structure in which there is subsurface dimerization of the Group III atoms in the second bilayer, accompanied by the formation of linear chains of the Group III atoms in the topmost layer [97], consistent with the observed growth behaviour.

The next question is how is strain relaxed in this system, and it appears to follow the idealized model, in that a periodic array of pure edge misfit dislocations is formed directly at the interface, with a spacing of $\approx$8nm, observed by TEM [98]. There is a kinetic restriction at the growing surface to surface roughening or island formation, so the strain remains in the layer until it starts to be relaxed by misfit dislocation formation at about 5ML. It has been suggested [98], although not confirmed, that the dislocations are formed by a homogeneous surface half-loop nucleation of pure edge dislocations via climb during growth, but no evidence for this mechanism has been found in STM images of layers slightly thicker than the minimum for relaxation (cf. growth on GaAs(110) substrates).

3.2.3.2. Anion-stable surfaces. In simple terms there are two As-rich reconstructions of the GaAs(001) surface, with symmetries (2×4) and $c(4 \times 4)$, respectively. In the first part of this review we discussed the (2×4) structure in considerable detail, and although it is the most stable, it is difficult to maintain it in an As flux adequate to avoid substantial In desorption at the growth temperatures appropriate for InAs growth (<520 °C). At these temperatures, in the presence of an As flux, there is a strong tendency for the GaAs surface to become $c(4 \times 4)$. As a consequence, most growth experiments have involved this surface, but we will deal first with the limited number where it was claimed that a (2×4) structure was present when growth commenced.

3.2.3.3. (2×4) Surfaces. Although Bressler-Hill et al. [99] maintained a (2×4) substrate surface, it was only at the expense of having no direct arsenic flux, but instead relying on an (unstated) arsenic background pressure for an In deposition rate of 0.25ML/s at 450 °C. They also used vicinal surfaces misoriented by $\approx 1^\circ$ in the [110] and $[\bar{1}\,1\,0]$ directions. The STM images they obtained suggested that this was not equivalent to the deposition of InAs, but followed more closely the behaviour of In alone

on a (2×4) surface kept stable at the low temperature used by a very small arsenic flux. In this case, with a (2×4) substrate surface, it has been calculated [100] that the energetically stable sites for In adatoms are along the missing dimer rows between As-trench dimers. Bell et al. [101] overcame the problem by using a valved Knudsen cell for their As supply, which enabled the (2×4) reconstruction to be maintained down to 470 °C. InAs films up to 1ML thick were deposited very slowly ($\approx 0.016 \text{MLs}^{-1}$) at this temperature, and at least in the sub-ML range, the InAs islands also displayed a (2×4) reconstruction. They pointed out, however, that even with this technique they were operating very close to the $(2 \times 4) \rightarrow c(4 \times 4)$ surface phase transition and the substrate often showed both reconstructions to be present concomitantly. At this early stage of growth, though, no 3D islands were formed, the coverage corresponded to the amount of InAs deposited, there was no discernible alloying and the island size distributions were similar to those observed in the homoepitaxy case [15], but with the same reservations about the STM snapshot technique. They therefore concluded that there was no apparent effect of strain on island formation on a (2×4) surface and any island anisotropy derived from the structure of the surface template on which growth occurred. This result is apparently at variance with earlier work by Belk et al. [102], which indicated that very rapid alloying occurred, but this is almost certainly due to the difficulty of maintaining the (2×4) reconstruction at these comparatively low temperatures in the presence of an arsenic flux. While the starting surface in [102] was indeed (2×4), it would have transformed almost immediately upon commencement of InAs deposition into a two- or three-phase structure, consistent with the published STM images.

In a subsequent paper [103], Bell and co-workers claim that the reconstruction of the substrate surface [$c(4 \times 4)$ or (2×4)] is only relevant at coverages $\leq$ 1ML. On the (2×4) surface growth occurs on top of the existing reconstruction, whereas on the $c(4 \times 4)$ the InAs incorporates directly into the surface layer in such a way that there is alloy formation from the outset. It is therefore only at very low coverages (<0.5ML) that the effect of the total strain is insignificant, even though individual islands may be pseudomorphic. At higher coverages they conclude that strain becomes the dominant issue and the behaviour is independent of the starting surface. This is superficially correct, but ignores any effects that the original surface structure may have on the extent of any alloying, alloy composition and composition gradients as the film develops. The STM images only show the final state morphology, not the surface symmetry which would have been shown by RHEED and which might have revealed details of alloying.

We must conclude that at the present time it is not possible to comment on strain relaxation processes which might occur for growth on the (2×4) surface per se, since the limited information available implies that after $\approx$0.5ML deposition it cannot be treated differently from the $c(4 \times 4)$ surface.

3.2.3.4. $c(4 \times 4)$ Surfaces. Since this is the naturally occurring reconstruction in the temperature range used for InAs growth, at least 90% of all reported work on QD formation in this system has been carried out with this starting surface. We will see that a very different pattern of growth related effects is observed and that the growth mode and strain relaxation behaviour is unique to this reconstruction. We will deal first with the formation of the WL, which occurs for coverages below about 2ML, and then discuss the processes which lead to the appearance of QDs as more material is deposited. Even at this stage, though, we should emphasize that QD formation in this material system is the exception, not the rule, and that the (001)–$c(4 \times 4)$ is the only low index surface on which the phenomenon is unequivocally observed.

3.2.3.5. Wetting layer formation. The initial growth of InAs on a GaAs(001)–$c(4 \times 4)$ surface is highly interactive and always results in the formation of a ternary alloy. Early work [102–104] was

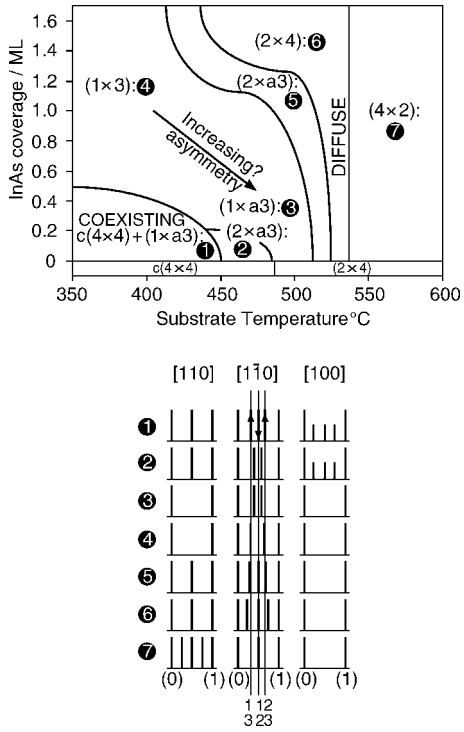

Fig. 16. A schematic illustration of the surface reconstructions observed by RHEED during the WL formation of InAs on GaAs(001) as a function of nominal InAs coverage and substrate temperature: (a) indicates an asymmetric pattern.

concerned with establishing a ‘‘surface phase diagram’’ of the structures present as a function of growth temperature and the amount of InAs deposited. The results are summarized in Fig. 16. Subsequent studies have largely confirmed the basic validity of this diagram, but of course details can vary as growth conditions, such as deposition rate, are changed. The essential point, however, is that alloy formation is the paramount effect in the ML growth regime.

Although growth is pseudomorphic, the $\approx 7\%$ misfit strain is mediated by this alloy formation, so the in-plane lattice parameter distortion is less than that which would be required to accommodate the total binary compound misfit. The In–Ga–As phase diagram shows complete miscibility, so forming an alloy is the expected result. This is confirmed by the STM-derived results shown in Fig. 17, which illustrate how deposition of sub-ML quantities of InAs can produce complete substrate coverage by a homogeneous layer. The temperature dependence of the extent (rate) of alloy formation shows it to be an activated process, for which the most probable barrier is either the surface diffusion or lattice incorporation of In. The question then arises as to why this process is effectively limited to this one type of surface.

We can begin by considering the surface structure of the alloy as a function of coverage and temperature and it is apparent from Fig. 16 that in general the RHEED patterns appear to indicate a complex range of structures, but in fact they can be explained reasonably systematically, provided the analysis is restricted to a $c(4 \times 4)$ starting surface. It then becomes apparent that the alloy phase is based on a (1×3) surface unit cell, the same as that found in non-dilute bulk alloys. There are, in addition, small regions of the associated (2×3) structure, a (2×4) phase found following significant InAs deposition (>1.2ML), and eventually a (4×2) structure formed at high temperature, which is associated with In segregation and is probably an alloy very dilute in Ga.

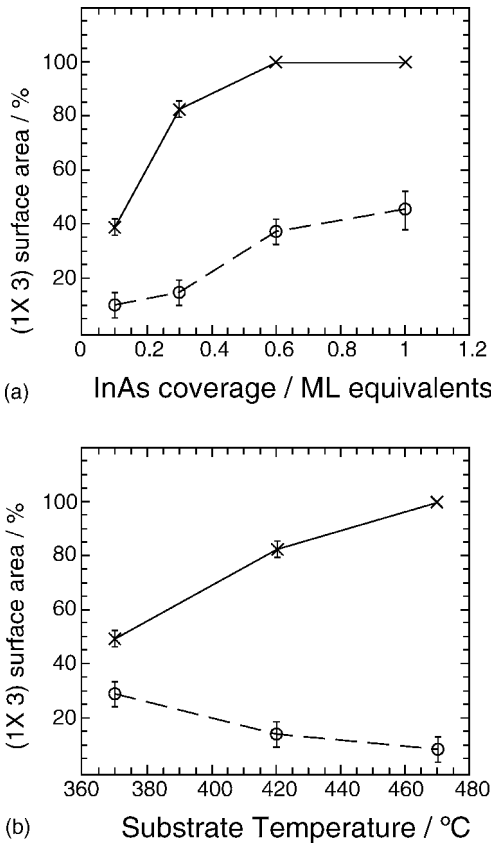

Fig. 17. The percentage of surface area covered by the (1×3) reconstruction as a function of (a) the amount of InAs deposited (in ML equivalents) at a substrate temperature of 420 °C; (b) the substrate temperature for a nominal 0.3ML deposition.

Structural models of the (1×3) and (2×3) surfaces, based on an alloy composition of InGaAs are characterized by continuous top-layer rows of As dimers along $[\bar{1}\,1\,0]$, with two cation positions in the third layer occupied by In and one by Ga. The doubled periodicity along $[\bar{1}\,1\,0]$ for the (2×3) reconstruction arises from an As dimer back-bonded to four third layer In atoms. These structures are illustrated in Fig. 18(a) and (b), but it must be emphasized that they only apply to layers of this fixed composition, for which thermodynamic and first principles calculations show the (2×3) to be the

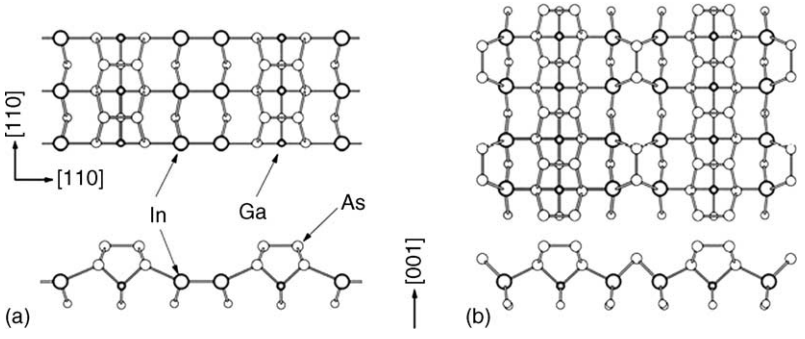

Fig. 18. Structural models for (a), (1×3) and (b), (2×3) reconstructed WL surfaces of InAs on GaAs(001).

more stable. In most cases, of course, the composition will deviate from this fixed ratio, and this results in so-called incommensurate phases [105,106,107].

Two questions then remain: why do alloys form so readily on the $c(4 \times 4)$ structure and what is the nature of the incommensurate phases? With regard to the ease of alloy formation, it has been shown using a combination of RHEED and STM that at low coverages InAs incorporates directly into the $c(4 \times 4)$ surface and alloyed domains form in the same layer as the remaining uncovered regions [60,103], whereas on the (2×4) reconstruction islands only nucleate on top of it and a uniform WL is not formed until almost a complete ML has been deposited [103]. For thicknesses >1ML there is apparently no influence of the starting reconstruction on the WL morphology [103], but this is hardly surprising given the difficulty of maintaining the (2×4) structure for any length of time under the conditions used for InAs deposition. We can therefore look to the structure of the $c(4 \times 4)$ surface [95] to account for the alloying behaviour, and two features seem to be important. One is its multi-layer construction, which means that in any defined area there is a high effective step density and the other is the outermost double layer of As atoms, which allows any excess cations to be accommodated in the second layer.

We consider next the incommensurate phases, which produce asymmetric diffraction patterns in the form of unequal spacing between the $\frac{1}{3}$-order features in the $[\bar{1}\,1\,0]$ azimuth, even at complete ML coverage, when there is no possible contribution from exposed regions of the substrate. These phases are present under almost all conditions of substrate temperature and coverage, compared to the very limited regions where the fixed composition (2×3) phase exists. This asymmetry effect is quite widespread and has been reported for several semiconductor systems, including Si/GaAs, InAs, InSb and GaSb, [108,109], as well as O_2/Ni (110), for which it was first reported by Germer et al. [110], with a subsequent generalization of the analysis by Houston and Park [111]. The origin is scattering from uniform sequences of different sized surface units formed from the same basic building block, in this case units of 2 and 3. The proportion of each defines the splitting of the $\frac{1}{2}$-order rods towards the $\frac{1}{3}$ and $\frac{2}{3}$ positions, but in addition, the domain sizes need to be small because of the limited coherence width of the incident electron beam. Here the domains were ~ 45Å and they are shown in the STM image of Fig. 19. Confirmation of this explanation has been obtained by Fourier transforming a computer-generated array of 2 and 3 block units with appropriate domain sizes, which produced a well-defined asymmetric (1×3) "diffraction pattern". The actual STM image was too noisy for the available computing power [112].

Overall, the evidence is for a reasonably well-ordered surface structure of the WL, but it is not fixed with respect either to detailed structure or composition. An important point is that the RHEED data are dynamic and represent the actual state of the surface prior to QD formation. In the light of the available measurements it seems reasonable to claim that the surface structure of the substrate dominates WL alloying behaviour, but a range of alloys is formed. An alloy WL appears to be a necessary precursor to QD formation, but its exact form is not a critical factor. In the following section we discuss the nature of the 2D–3D transition and the possible role the WL plays in this.

3.2.3.6. The 2D–3D growth mode transition. The transition in the growth mode from the WL growing two-dimensionally to the formation of 3D islands is the most distinctive aspect of this reconstruction and is, of course, the basis of the formation of self-assembled quantum dots. There are, however, several important features of this process which require comment before any attempt can be made to formulate a comprehensive mechanism:

- The 3D islands grow only after the formation of an alloy WL and are coherent, i.e. dislocation free and therefore strained.

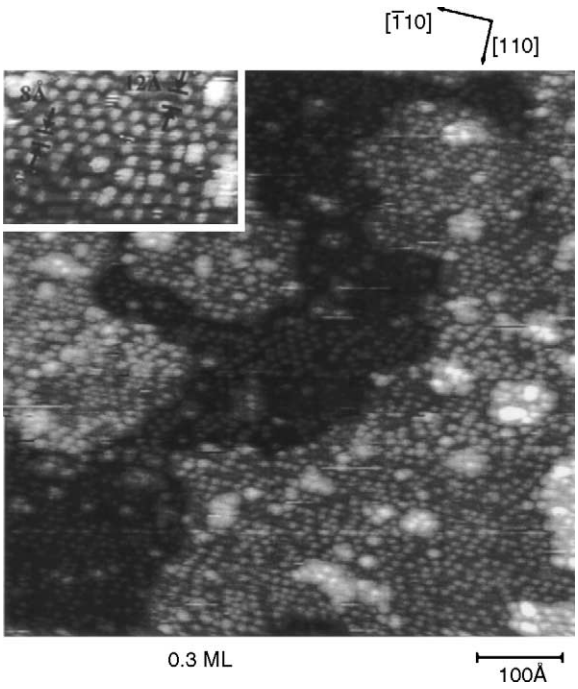

Fig. 19. STM image of "domains" in the $(1 \times a3)$ reconstructed WL surface, with 0.3ML InAs deposited on a GaAs(001)–$c(4 \times 4)$ reconstructed surface at 420 °C. The inset shows two- and three-fold periodicities that result in an asymmetric (1×3) RHEED pattern.

- The 2D to 3D transition is very rapid, occurring over <0.1ML of incremental deposition, with very mixed evidence for the existence of a possible precursor state.
- There is rather weak temperature and In flux dependence of the critical thickness for the transition.
- The QDs, once formed, are not InAs, even though no other material is being deposited, but are an (In, Ga)As alloy, whose In concentration increases with decreasing In flux. Clearly, Ga is being incorporated from the WL and/or the substrate.
- The final state of the QD size distribution appears to obey scaling laws derived for homoepitaxial growth, i.e. with no influence of strain, although earlier stages of growth appear to behave differently.

Dealing with these points in turn, we have treated WL formation in the previous section and in addition, there is ample evidence to show that the QDs are coherent, at least prior to the stage where they begin to coalesce. This evidence has been obtained in very large part by TEM examination, both plan-view and cross-section [113,114]. In Fig. 20 we show a typical plan-view TEM image in which the contrast is mainly due to the strain state of the QDs.

The rate of the transition from 2D to 3D growth has in almost all cases been monitored in situ by RHEED, so the information is in real time at the growth temperature. The technique involves monitoring the change from a diffraction pattern with streaks normal to the shadow edge, which is characteristic of 2D growth, to a spot pattern formed by electron transmission through 3D islands. Due to the shape of the islands (vide infra, Section 3.3), this change to a spot pattern is observed in the [110] azimuth, whereas in the orthogonal $[\bar{1}\,1\,0]$ azimuth the pattern displays chevron features as the result of refraction related effects [115,116]. These changes are illustrated in Fig. 21. There is, though, one other important effect which is seldom mentioned: immediately prior to the formation of 3D features,

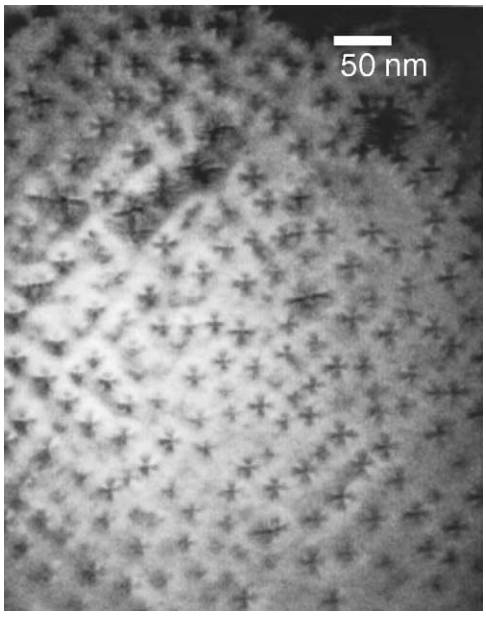

Fig. 20. Plan-view TEM image of QDs showing most of them to be coherent, and the contrast arising from their strain state.

the pattern virtually disappears, indicative of a very disordered transient state. The incremental amount of deposition required to bring about the complete change from 2D to 3D diffraction features is at most 0.1ML, which is about the resolution limit of the flux measurement. It corresponds to $\approx 5 \times 10^{13}$atomscm^{-2}, so for an island number density of 5×10^{10}cm^{-2}, which is fairly typical at conventional growth rates of ≈ 0.1MLs^{-1}, there would be $\approx 10^3$atoms per island. The fact that the first observable real-time feature of the transition is the presence of transmission spots provides a very strong indication of the early shape of the dots, in that they must be in the form of surface asperities, high enough and narrow enough for the electron beam to penetrate without significant interaction with

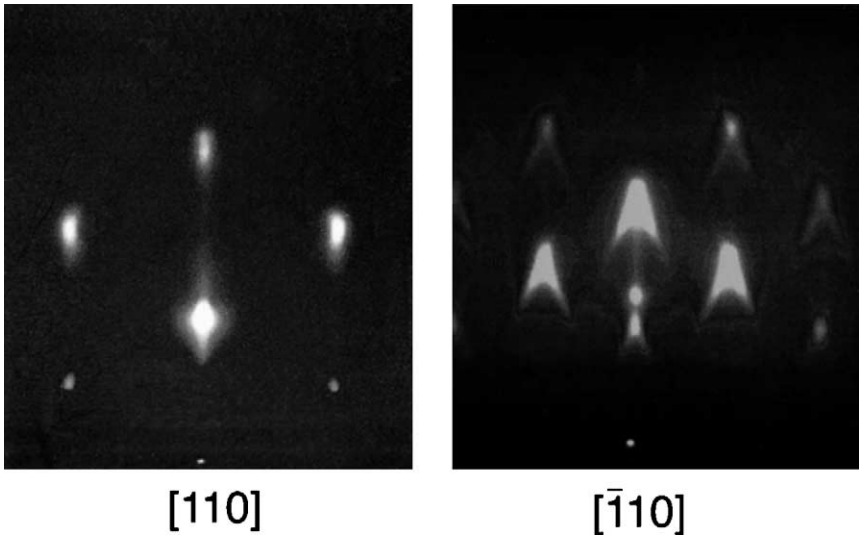

Fig. 21. RHEED patterns from ≈ 2ML of InAs on GaAs(001)–$c(4 \times 4)$, i.e. in the QD regime, in the [110] azimuth, showing transmission spots, and in the $[\bar{1}\,1\,0]$ azimuth, showing chevrons.

the surface. Given that they may contain as few as 10^3atoms, the lateral spread must be very small for the shape to be consistent with the transmission of $\approx$15keV electrons, unless, of course, additional atoms are incorporated from the WL at the moment of transition, but that would imply very rapid dissociation and subsequent mass transport (in relation to the flux the time scale would have to be $<$0.1s).

The problem with RHEED, however, is that although it provides dynamic, real-time information, it is not very sensitive to changes in 2D surface morphology, which might well occur if there is a significant precursor state to 3D island formation. This has indeed been proposed on the basis of snapshot STM images of quenched surfaces [117–120]. There is no corresponding sensitivity problem with STM, which is easily able to resolve the evolution of 2D (or low-profile 3D) precursors, even if their number density is small. The difficulty with this technique is, however, that it is not dynamic and the images are generally acquired at close to room temperature after quenching. This could in principle lead to the condensation of any mobile surface species, so features observed in STM images may not actually be present during growth. The effective transient disappearance of the RHEED pattern immediately prior to the 2D to 3D transition may indeed suggest the presence of just such a mobile adatom population at the crucial stage.

STM results relating to precursors started with the observations of Heitz et al. [117] and Ramachandran et al. [118], who described a so-called re-entrant 2D–3D morphology change in which quasi-3D clusters first appeared, then disappeared, only to reappear with further deposition, prior to the formation of 3D islands, for which the clusters acted as precursors. In a recent paper, however, Patella et al. [121] claimed that on volumetric grounds QDs could not be formed from these quasi-3D precursors. Instead, they attributed the transition to the initial formation of an intermixed $In_xGa_{1-x}As$ ($x \approx 0.82$) layer after deposition of 1ML, followed by a ''floating'' In surface population as deposition continued, from which In surface mass transport would lead to the rapid volume increase required to form QDs. We will discuss In surface segregation below. These authors also showed AFM images of InAs deposits at $\approx$0.15ML below the critical coverage for QD formation, which revealed a roughened surface with mounds about 3nm high and of lateral dimensions $\approx 1.2 \times 0.3\,\mu$m, elongated along $[1\,\bar{1}\,0]$, with some evidence of step bunching and step-edge meandering. It seems likely, though, that this behaviour is rather specific to the growth conditions used, since the composition of the WL is known to be quite variable and other investigators have not observed the type of morphological instability reported here. The concept of a mobile adatom population, not necessarily of In alone, may well be more general, however, and we will return to this point later.

Two papers by Krzyzewski et al. [119,120] come down very heavily in favour of an alternative model for the transition, in which large 2D islands form, but do not transform into 3D islands. Instead they co-exist with small, irregular 3D entities, between 2 and 4ML high, each containing about 150 atoms, which with an incremental 0.05ML deposition transform rapidly into ''mature'' QDs containing more than 10^4 atoms at a number density $>10^{10}$cm^{-2}. No explanation was given for this, but the numerical values suggest that it may be necessary for material from the WL to be incorporated into the mature QDs, although at the very low deposition rates used (0.017ML/s), the same group reported that the QDs were essentially 100% InAs [122]. Nor is it certain that the very small 3D entities can be associated unequivocally with processes occurring at the growth temperature, since the possibility that they form during the quenching and cooling stages prior to recording the STM images cannot be ruled out. Scaling analysis of the QD size distribution from this work implies that strain is important in the first stages of QD formation, but is not involved in their subsequent development. This latter conclusion is in agreement with earlier work by Ebiko et al. [123,124] and it is possible that elastic interactions in a dense array of islands could contribute to their growth kinetics, as originally suggested by Krishnamurthy et al. [125]. We will discuss the scaling concept in more detail in Section 3.3.2.3.

There is a comparatively weak temperature dependence of the critical thickness for QD formation, varying from ≈ 1.4ML at 350 °C to ≈ 1.8ML at 500 °C [126], which is consistent with the QDs having a higher Ga content at higher temperatures (see below). This is because the higher the Ga content, the greater the strain mediation between substrate and overgrowth, which would increase the amount of material that could be deposited before some of the strain is relaxed by the 2D–3D growth mode transition. There is an Arrhenius dependence of the number density of QDs on temperature [127], which might simplistically be attributed to shorter adatom migration distances at lower temperatures, but this would imply that conventional nucleation kinetics were being followed. The transition thickness has virtually no In flux dependence over the range 0.01–0.5MLs^{-1}, although the number density of dots decreases strongly with decreasing flux [116] and the dots are correspondingly larger.

There is ample evidence [122,128,129] to show that the amount of material in the dots at all substrate temperatures ≥ 400 °C is greater than the amount of material deposited beyond the 2D to 3D transition, and by ≈ 500 °C it is greater than the total amount of InAs deposited, as illustrated in Fig. 22. The somewhat surprising exception to this effect occurs at very low deposition rates (≤ 0.02MLs^{-1}), where the dot volume is virtually equal to the amount of InAs deposited after the transition [122]. Other than at these very low deposition rates, the QDs must therefore have the composition of an (In, Ga)As alloy. In general it has been found that the alloy composition is not homogeneous, but there is an In composition gradient from the apex of the dot to its base [130–134]. Although there is no direct proof, it appears certain that at the transition point material is transferred from the alloy WL (rather than directly from the GaAs substrate) into the dot by an activated process, since the amount transferred increases strongly with temperature. There appears to be no dependence on the exact surface structure of the WL, at least qualitatively, but it could influence the precise amount. The basic anomaly is why QD alloying is not observed at very low deposition rates. It clearly reflects a strong kinetic influence, since were the process to be close to equilibrium, alloy formation would be favoured at low rates. Finally, we need to discuss possible effects of the surface segregation of In on QD composition in this context.

Cullis and co-workers [135,136] have in fact shown that the segregation of In could play a very significant role in the 2D–3D transition. They used energy-selected imaging in the TEM to show that there is substantial lateral, but especially vertical (surface) segregation of In within the WL. This increases the surface strain, which inhibits the incorporation of In (and Ga) on lattice sites, thereby allowing the mobile surface adatom population to increase. As the transition is approached, the In concentration in the surface layer increases still further, allowing critical clusters for nucleation to be formed. These 3D islands are stabilized by stress relief due to the expansion of laterally unconstrained

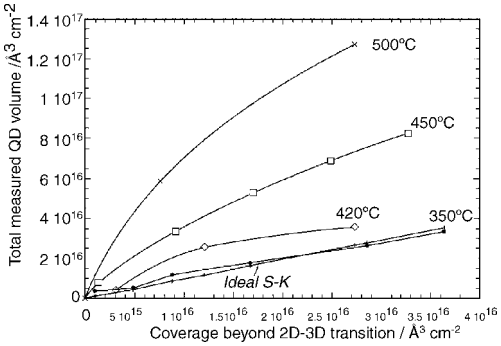

Fig. 22. Total measured QD volume as a function of the amount (effective coverage) of InAs deposited beyond the 2D–3D transition at different substrate temperatures for InAs on GaAs(001)–$c(4 \times 4)$. Ideal SK growth is indicated by (+).

lattice planes, and this stability allows them to act as sinks for adatoms, permitting them to grow. The plausibility of this model was further demonstrated by measurement of the inhomogeneous distribution of In in the QDs, which contain excess near their apices, a result demonstrated previously by several groups [131–134]. A more quantitative substantiation was provided by using the two-site exchange model of surface segregation with appropriate choice of parameters for this material system [137]. This considers the exchange of the Group III species between the top two layers during growth, and showed there to be a very substantial deviation in surface layer composition from the deposition flux. This model appears to be fully consistent with all of the RHEED data, and the large mobile adatom population immediately prior to 3D island formation would account for the transient loss of the pattern at this stage. STM images unfortunately cannot determine the presence or otherwise of a mobile adatom population prior to quenching.

3.3. Theories of Stranski–Krastanov growth

Several theoretical approaches have been developed to explain various features of QD formation, but none is so far capable of explaining all of the experimental observations. This is due, in large part, to the multiscale nature of this phenomenon. At the most fundamental level, first-principles calculations provide information about In and Ga adatom diffusion on GaAs(001) and on the homogeneously strained wetting layer. But the formation of 3D islands requires theoretical methods such as kinetic Monte Carlo simulations and rate equations that are capable of describing the statistical properties of quantum dot ensembles, including analyses based on the scaling theory developed for two-dimensional homoepitaxial islands. At the most coarse-grained level of description, continuum equations and thermodynamics enable connections to be established with classical elasticity theory. In this section, we review several calculations that have addressed various aspects of the energetics, kinetics, and thermodynamics of quantum dot formation.

3.3.1. First-principles studies of InAs/GaAs(001) heteroepitaxy

Although the equilibrium surface diagram (at $T = 0$) of GaAs(001) is well established (Fig. 2), there is comparatively less information available on the effect of strain on the stability of the reconstructions. First-principles calculations enable the effect of strain on the phase diagrams of GaAs(001) and InAs(001) to be examined and the influence of strain on the diffusion of In and Ga on the substrate and on the wetting layer to be determined.

Theoretical work [138,139,140] has established that, for growth conditions ranging from As-rich to Ga-rich, the stable reconstructions of the GaAs(001) and InAs(001) surfaces follow the sequence

$$c(4\times 4)\rightarrow \beta 2(2\times 4)\rightarrow \alpha 2(2\times 4)\rightarrow \zeta(4\times 2). \tag{3}$$

Ratsch [141] and Penev and Kratzer [142] have performed calculations of the surface free energy for these reconstructions for isotropically strained (001) substrates: their regions of stability in the $(\mu_{\mathrm{As}}, \varepsilon)$ plane are shown in Fig. 23. At positive ε the ''heterodimer'' $c(4\times 4)$ reconstruction of GaAs(001) appears as the stable reconstruction between the ''conventional'' $c(4\times 4)$ and $\beta 2(2\times 4)$ reconstructions. Since heteroepitaxy of InAs on GaAs is carried out typically under As-rich conditions, the GaAs(001) surface displays the $c(4\times 4)$ reconstruction, which is characterized by blocks of three As dimers, or three Ga–As heterodimers, sitting on top of a complete As layer (Fig. 24).

The detailed real-space structural information of equilibrated substrates is not available once growth is initiated. This is somewhat problematic for first-principles calculations, which rely on a basic structural model that can be relaxed to its minimum energy configuration. Test models are typically derived from a combination of the symmetry of the RHEED diffraction pattern and basic

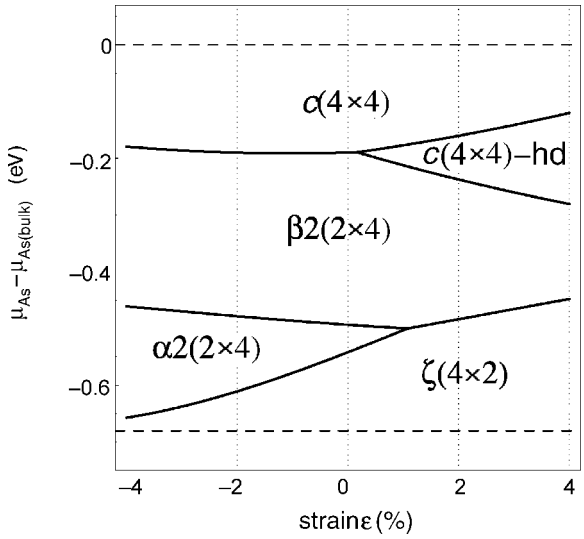

Fig. 23. Diagrams of surface phases (reconstructions) of GaAs(001) as a function of μ_{As} and isotropic strain ε. The broken lines mark the physically allowed range of μ_{As} (after Penev and Kratzer [142]).

stoichiometry. For example, after the deposition of $\sim \frac{1}{3}\frac{2}{3}$ML, the RHEED pattern indicates the appearance of a structure with a period of three lattice constants in the [110] direction. There is also evidence for some asymmetry and disorder in the diffraction patterns, which has been ascribed to the formation of domains [108]. For the deposition of $\frac{2}{3}$ML of InAs, a structural model with a (2×3) unit cell (Fig. 25(b)) has been proposed on the basis of X-ray diffraction data [105] that consists of rows of As dimers running along the $[\bar{1}\,1\,0]$ direction. The indium atoms are located only on lattice sites in the ''trenches'' between the chains of As dimers, avoiding the lattice sites directly below the As dimers, so this film may be regarded as a single ML of $In_{2/3}Ga_{1/3}As$ alloy. First-principles calculations [142,143] find that the surface energy of this (2×3) structural model is close to the that of the $c(4 \times 4)$ substrate, over a wide range of μ_{As} and, for $-0.2\,\text{eV} < \mu_{As} - \mu_{As(bulk)} < -0.1\,\text{eV}$, is slightly the more stable.

The investigation of In migration on the $In_{2/3}Ga_{1/3}As$ alloy surface can provide insight into adatom migration on the wetting layer, despite not necessarily corresponding to the actual structure of

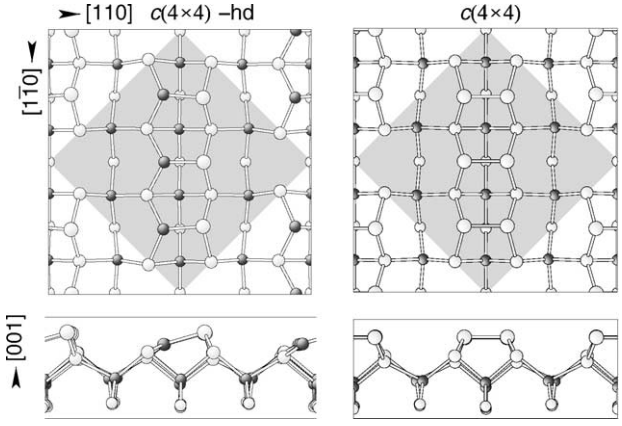

Fig. 24. Structural models of the conventional $c(4 \times 4)$ reconstruction and the $c(4 \times 4)$-hd heterodimer model. Ga atoms are shown darkened, As atoms are white). The shaded regions represent the surface unit cell (after Penev and Kratzer [142]).

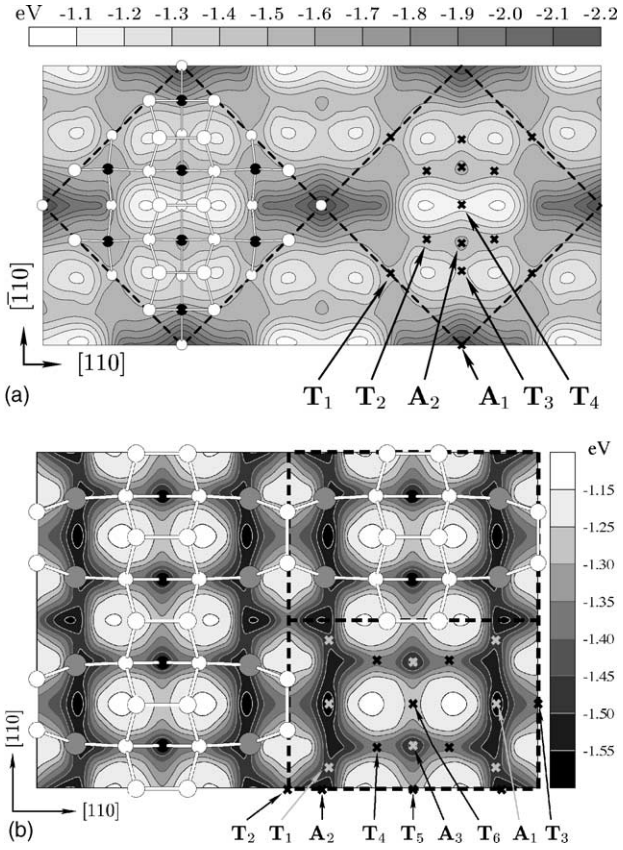

Fig. 25. (a) Potential energy surface (PES) for an In adatom on GaAs(001)–$c(4 \times 4)$. Atomic positions of the clean surface unit cell are indicated for atoms in the upper four layers, with As indicated by empty circles and Ga atoms by filled circles. (b) PES for an In adatom on $In_{2/3}Ga_{1/3}As(001)$–(2×3). Two unit cells are indicated by dashed boxes. Overlaid on the PES plot are the topmost three atomic layers, with In atoms indicated by shaded circles. Note the greater corrugation of the PES for GaAs(001)–$c(4 \times 4)$, as evidenced by the energy range compared with that for $In_{2/3}Ga_{1/3}As(001)$–(2×3) (after Penev and Kratzer [142]).

the wetting layer. First-principles calculations [142,143] reveal that the corrugation of the potential energy surface is remarkably small, showing a maximum variation of $\simeq 0.5$eV, which is appreciably less than that for GaAs(001)–$c(4 \times 4)$. Moreover, In migration on GaAs(001)–$c(4 \times 4)$ is significantly affected by strain and shows a linear dependence for small strains. For an inhomogeneously strained sample, this implies a position-dependence of the activation energy for In migration. Using a flat island on GaAs(001)–$c(4 \times 4)$, Kratzer et al. [144] have shown that the effect of strain leads to a repulsive potential with a strength of up to 0.2eV that affects both the binding energy and, to a somewhat smaller extent, the diffusion barriers for an In adatom that attempts to approach this island. This repulsive interaction can significantly impede the speed of growth of strained islands. This effect is often cited [145,146] as one of the factors responsible for the narrowing of the island-size distribution.

3.3.2. Kinetic models for 3D island nucleation

3.3.2.1. Off-lattice kinetic Monte Carlo simulations. Kinetic Monte Carlo simulations of simple lattice models have been used to model many aspects of homoepitaxial kinetics [13,147–150], but applications to thin film evolution in the presence of lattice misfit have been much slower in

development. There are two main reasons for this. The rates of atomistic processes on strained surfaces are not determined solely by the local environment of the atoms, as in the case of homoepitaxy, but may depend on nonlocal features such as the *height* of a terrace above the initial substrate, the *size* and *shape* of two- and three-dimensional islands [151,152], and their local environment [146]. This is further complicated by the competition between different strain relaxation mechanisms (e.g. alloying, misfit dislocation formation, surface profile modulations), each of which has a characteristic signature in the morphology of the substrate [128,153]. Additionally, any general theoretical approach must incorporate long-range elastic interactions, which are best treated within a continuum framework, and atomistic effects such as step-adatom interactions, alloying and possibly reconstruction changes during growth. Thus, lattice-based simulations of quantum dot kinetics [152] are appropriate for coherent (i.e. dislocation-free) island growth, but off-lattice models [154,155] are necessary to include the possibility of dislocation formation and its impact on growth morphology.

Continuous particle positions are necessary to account for any deviations from a perfect lattice structure induced by lattice misfit, including the formation of dislocations. Once an interatomic potential has been specified, a molecular dynamics simulation would be the method of choice, but this method typically proceeds by time steps of the order of 10^{-13}s, and is thereby restricted to physical times of the order of 10^{-6}s. Thus, time scales of typical epitaxial processes of seconds or minutes are not currently attainable, even with modern acceleration strategies [156].

Biehl and co-workers [157,158] have carried out off lattice KMC simulations based on a Lennard–Jones potential of the form

$$U_{ij}(U_0,\sigma) = 4U_0\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{4}$$

where the relative distance r_{ij} of particles i and j is a continuous variable. The KMC simulation proceeds by assigning to each event i an Arrhenius rate,

$$R_i = \nu_0 \exp\left(-\frac{E_i}{k_B T}\right), \tag{5}$$

where ν_0 is the attempt frequency, taken to be the same for all diffusion events, E_i is the energy barrier for the process, k_B is Boltzmann's constant, and T is the temperature. These simulations are particularly simple for one-dimensional substrates because the path between neighbouring local minima of the potential energy is uniquely determined and the transition state corresponds to the separating local maximum.

The substrate is flat prior to the onset of deposition and growth proceeds initially in a layer-by-layer manner. At a well-defined thickness h^*_{WL}, monolayer islands located on the wetting layer undergo a rapid transition to bilayer islands. The simulations reveal a significant increase of h^*_{WL} with increasing deposition flux, which has also been reported for InP/GaAs heteroepitaxy [159]. This suggests that the emergence of islands upon islands, i.e. the SK transition, is due mainly to upward particle hops onto existing monolayer islands. Alternatively, if the formation of second or third layer nuclei by freshly deposited adatoms was the dominant process, one would expect more frequent nucleation and a correspondingly earlier 2D–3D transition at higher growth rates. After the 2D–3D transition, islands grow by incorporating material from the deposition flux as well as that from the surrounding wetting layer (cf. Fig. 22). A section of a typical surface is shown in Fig. 26. Atoms near the centre of the island adopt positions close to the minimum of the nearest potential energy trough. Atoms further from the centre of the island are correspondingly further away from their nearest minima. Thus, strain relaxation occurs predominantly at the edges of islands, as expected.

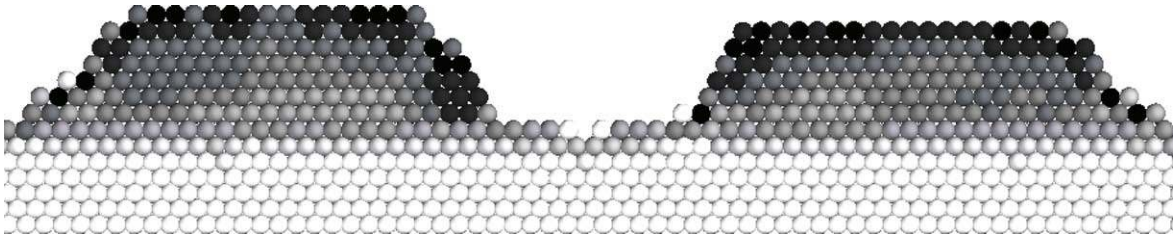

Fig. 26. A section of a simulated crystal as obtained for a flux of 7.0ML/s at a simulation temperature $T = 500$K. Islands are located on a wetting layer with $h_c \approx 1$ and the six bottom layers represent the substrate. The darker the shading of a particle, the larger the average distance from its nearest neighbours (after Biehl and Much [158]).

The main conclusions of these simulations are: (i) The strong adatom/substrate interaction favours the wetting layer formation and results in relatively slow diffusion of adatoms on the substrate. Diffusion on the wetting layer is significantly faster and the corresponding barrier decreases with the thickness of the wetting layer (cf. Fig. 25); (ii) Strain relaxation leads to a pronounced bias towards the centre on the top layer of finite mono- and multilayer islands on the wetting layer. Moreover, diffusion is slower on top of the partially relaxed islands than on the wetting layer. Whereas (i) favours the formation and persistence of the wetting layer; (ii) destabilizes layer-by-layer growth. The microscopic process that triggers the 2D–3D transition is upward adatom diffusion from the wetting layer and at island edges. The corresponding barriers decrease with the wetting layer thickness and, as a result of the competing effects, the 2D–3D transition occurs at a critical thickness which depends upon the substrate temperature and the deposition flux.

3.3.2.2. Rate equations. Dobbs et al. [71,160,161] developed a mean-field theory for the formation of quantum dots based on rate equations [76,162]. The dynamical variables are the densities of adatoms, 2D and 3D islands (QDs). The four processes that describe how these quantities change are (i) the flux of atoms to the surface (actually to the surface of the WL, which is presumed already to have formed); (ii) the surface migration of adatoms; (iii) attachment of adatoms to form stable, immobile 2D islands containing i^* atoms. Smaller islands are thermodynamically unstable and larger islands grow by the attachment of adatoms. Provision is made for strain-induced detachment from the peripheries of 2D islands, with a barrier which tends to zero for large islands. There is no detachment from 3D islands. It is essentially these effects that control the size distribution of QDs; (iv) the transition from 2D to 3D islands, which occurs when the second (and subsequent) layers nucleate on top of the growing 2D islands, either directly from the vapour or by detachment from perimeter sites and upwards hopping. Although not based on a critical 2D island size, the nucleation rate for higher layers increases very rapidly with increasing size of the 2D islands. This approach correctly predicts the increase in number density of 3D islands with increasing In flux and decreasing temperature, and goes some way towards predicting the rapid increase in density with coverage beyond some critical value (Fig. 27). But there are limitations to this approach, most notable, the difficulty of calculating reliable island-size distributions from rate equations and, indeed, any spatial information about the islands.

3.3.2.3. Scaling theory of island-size distributions. One of the most far-reaching results to emerge from simulations and the rate equation description of homoepitaxial growth [72,165] is that, prior to coalescence, the density n_s of s-atom ($s \geq 2$) islands can be written as

$$n_s = \frac{\theta}{s_{\mathrm{av}}^2} f\left(\frac{s}{s_{\mathrm{av}}}\right) \tag{6}$$

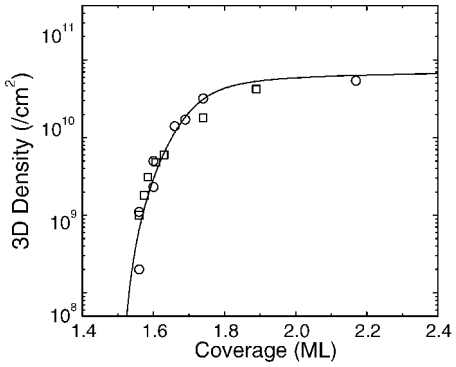

Fig. 27. 3D island density versus coverage for InAs/GaAs(001) epitaxy. The open squares represent experimental data taken from [163] at a growth temperature of 803K and a growth rate of 0.01ML/s and the open circles represent experimental data taken from [164] for a growth temperature of 773K and a growth rate of 0.22ML/s. The solid line is the result of a rate equation calculation, as described in [161]

where $s_{\rm av}$ is the average island size, θ is the coverage, and f is a scaling function. The scaling function is ''universal'' in the sense that its dependence on the coverage, the growth rate and the substrate temperature enters only through $\langle s \rangle$. Simulations [61,72,165,166] and experimental studies on metals [167,168] and semiconductors [5] are consistent with Eq. (6). This substantial body of circumstantial evidence for Eq. (6) is supplemented by an analytic justification based on solutions of rate equations [169,170].

The physical content of Eq. 6 is that the average island size $s_{\rm av}$ is the only length scale in the problem. Given that $s_{\rm av} \sim \theta^{\xi} R^{\chi}$, where θ is the coverage and $R = D/F$ is the ratio of the adatom diffusion constant D to the deposition flux F, scaling mandates that the island-size distribution for different coverages and values of R collapse onto a universal curve. As the analytic derivation of Eq. (6) makes clear [169,170], the scaling function is sensitive only to processes such as adatom attachment to and detachment from islands, evaporation, and island mobility. Structural details, as such are unimportant, so similar scaling functions are expected on different lattices (crystallographic faces) whose basic growth kinetics are similar, with possible differences for small s due to microscopic details, such as different sizes of the most stable (''magic'') islands (4 on a square lattice as opposed to 3 or 7 on a triangular lattice). Thus, the similarity of the island-size distributions for homoepitaxy on Fe(001) [61,167] and on GaAs(110) and GaAs(111)A [5] is not unexpected.

However, the applicability of such scaling concepts to heteroepitaxial systems is not altogether straightforward. Any size-dependent effects that are induced by lattice misfit, such as adatom detachment, would be expected to pre-empt the validity of Eq. (6). Nevertheless, ''scaling plots'' of heteroepitaxial island sizes, i.e. $n_s s_{\rm av}^2/\theta$ versus $s/s_{\rm av}$, which yields the profile of the scaling function f, can provide important information about nucleation and growth mechanisms even in the absence of a data collapse. The first such study that had direct relevance for the InAs/GaAs(001) system was reported by Bressler-Hill et al. [99] for 2D InAs islands on GaAs(001). Somewhat surprisingly, in light of the foregoing, is the high quality of the data collapse (Fig. 28(a)). The resulting scaling function exhibits substantial differences from that corresponding to irreversible aggregation, which is one of the most basic growth processes. There are several possible origins for these differences, including the anisotropy of the substrate and the possibility of exchange reactions between In and Ga, the latter of which is directly attributable to the lattice misfit between InAs and GaAs. But the quality of the data collapse indicates that any other effects due to strain cannot have a strong size dependence [70].

The corresponding measurements of 3D islands (Fig. 29) produced an unexpected result: the size (volume) distributions in the growth temperature range 490–540 °C exhibit a data collapse corre-

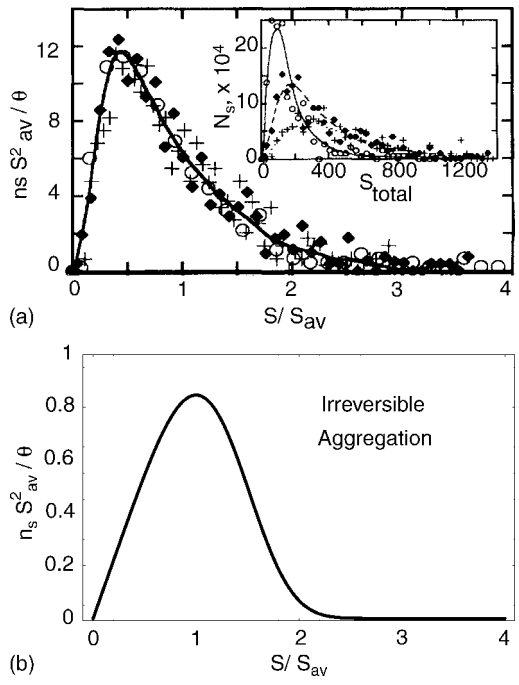

Fig. 28. (a) Scaling plot of InAs islands on nominally flat GaAs(001)–(2×4) at coverages $\theta = 0.15, 0.29, 0.35$ [99]. The inset shows the corresponding unscaled island-size distributions. Reprinted with permission from V. Bressler-Hill et al., Phys. Rev. Lett. 74 (1995) 3209. Copyright 1995 by the American Physical Society. (b) The scaling function for one of the most basic island nucleation and growth processes [73]—irreversible aggregation, whereby adatoms that attach to other adatoms or to islands do not subsequently detach. The difference in the vertical scales are due simply to normalization: based on the total number of islands in (a) and normalized to unity in (b).

sponding to irreversible aggregation (Fig. 28(b)), as does the distribution of separations. Similar results were obtained for *dislocated* InN islands on GaN(0001) [171]. This behaviour appears to indicate that the formation of QDs is determined only by the irreversible aggregation of migrating In and/or Ga adatoms, with no direct effect of strain. A possible rationalization might be obtained by parametrizing the island-size distribution by the ratio λ of the net detachment rate from an island to the net attachment rate to an island, rather than by a nominal critical island size [61,170,172]:

$$\lambda = \frac{R_{\mathrm{d}}}{(F\kappa + D\sigma n)}, \tag{7}$$

where F is the deposition flux, D the adatom diffusion constant, κ the direct capture number from the incident flux, σ the diffusive capture number with an adatom density of n and R_{d} is the escape rate of an adatom from a 3D island. Kinetic Monte Carlo simulations of 2D homoepitaxy show that λ parametrizes a continuous family of scaling functions, but for $\lambda \approx 1$, the island size distribution fits Eq. (6), even when significant detachment occurs [146]. The large adatom population prior to the 2D–3D transition can be interpreted as a increase in the effective deposition flux. These adatoms are readily incorporated into the incipient 3D islands, all of which has the effect of driving the system toward the limit of irreversible aggregation.

3.3.3. *Thermodynamics of the 2D–3D transition*

Korutcheva et al. [173] have extended previous thermodynamic arguments used to explain the 2D–3D transition and the ensuing narrow size distribution of QDs within a 1+1-dimensional model

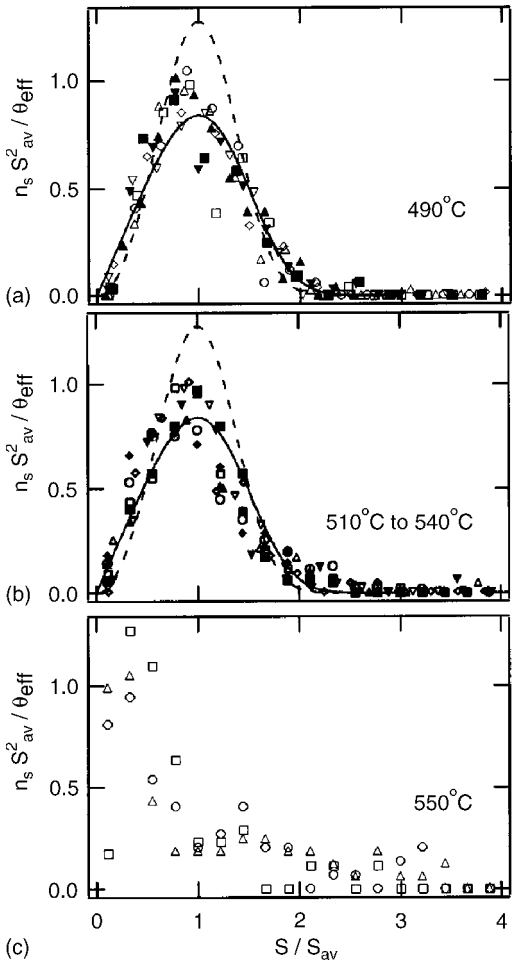

Fig. 29. Scaling plot of the volume distribution of 3D islands measured during the Stranski–Krastanov growth of InAs on GaAs(001) at temperatures of (a) 490 °C; (b) 510 °C–540 °C; and (c) 550 °C [123,124]. The quantity θ_{eff} is the coverage beyond the critical coverage (1.5ML). The solid curve is the scaling function for irreversible aggregation in Fig. 28(b). The broken curve is the scaling function [73] for a critical nucleus of $i^* = 2$, corresponding to growth with a degree of detachment. Reprinted with permission from Y. Ebiko et al., Phys. Rev. B 60 (1999) 8234. Copyright 1999 by the American Physical Society.

that takes account of the anharmonicity of the interatomic forces. The conventional assumption that the interfacial energy between the wetting layer and the QDs is small enough to be neglected, i.e. that the wetting layer completely wets the QDs would in fact eliminate 3D islanding on a thermodynamic basis, since this requires the interfacial energy to be non-zero. These authors point out that the adhesion (bonding) of atoms within the 3D islands to the wetting layer is stronger in the middle of the islands than at their free edges, since the lattice mismatch causes the atoms to be displaced by different amounts from their equilibrium positions in the potential well minima they would occupy at zero misfit (cf. Fig. 26). Consequently, the average adhesion of a finite-sized island is weaker than that of an infinite monolayer. An interface boundary therefore appears and the wetting of the island by the wetting layer is on average incomplete, which drives the formation of coherent 3D islands on the uniformly strained wetting layer.

The 1+1-dimensional model is effectively a cross-section of the real 2+1-dimensional case and the atoms in the chain are connected by bonds obeying the generalized Morse potential [174], but the

bond strength is allowed gradually to decrease with island height and there is a corresponding attenuation of strain [151]. The mechanism of QD formation is based on the premise that monolayer islands are necessary precursors to the formation of 3D islands, and that there is a critical size above which monolayer islands become unstable with respect to bilayer islands and another critical size when the bilayer islands become unstable against trilayer islands and so on. The growth of 3D islands is thus a sequential process.

The critical island size for the monolayer–bilayer transition is a strong function of lattice mismatch, decreasing rapidly as the mismatch increases. Monolayer islands are stable with respect to multilayer islands below some critical misfit (the island size goes asymptotically to infinity at some critical value), which accounts for the comparatively large misfit required for the growth of coherent 3D islands. Since this critical misfit in absolute terms is almost twice as large for layers in tension than those in compression, it also explains why coherent islands are observed much less frequently, if at all, in the former case. The narrow size distribution arises because atoms at the peripheries of islands are less tightly bound than those in the centre, so the 2D–3D transition occurs by transport from the edge to the upper surface of the islands via a sequential process. Since there is a critical size for this transition, there is correspondingly a narrow 3D island size distribution.

This is a purely thermodynamic approach, with the wetting layer already in place, so it cannot account for the orientation and reconstruction differences, nor for the rate of the 2D–3D transition, which is presumably kinetically controlled. It could, however, explain why 3D islands have been observed to form during the annealing of layers in certain systems where they do not occur during growth [175].

3.3.4. Equilibrium theories of 3D island formation

Priester and Lannoo [176,177] realized that treatments based on the premise that coherent 3D islands are more stable than a pseudomorphic layer, and that for a given amount of material the largest islands are the most stable, predict a much broader island-size distribution than that observed. Nor can direct formation of 3D islands explain the existence of a critical coverage for QD formation. They therefore set out to answer two basic questions; what is the physical origin of the critical coverage for the 2D–3D transition and what nucleation process is involved in QD formation? They based their approach on the formation of 2D platelets that are presumed to occur during the early stages of deposition of the second monolayer (of the WL). Provided surface diffusion is sufficiently rapid, the platelets reach an equilibrium size distribution which total energy calculations show to be narrow.

The experimental evidence for the existence of 2D platelets of any sort is far from convincing, however, and if this is extended to include a narrow size distribution, then nothing has been reported. Nevertheless, if we accept this concept as being valid, there is still a need to understand how the platelets suddenly transform into 3D entities with the same narrow distribution. Priester and Lannoo suggest that the platelets ''spontaneously transform into 3D islands at practically constant number of atoms in each island'' because the Schwoebel barrier at the edge of the platelets becomes almost as low as the barrier for surface migration. They do not explain, however, why this occurs for such a tiny increment of deposited material and why rapid local growth occurs in the z-direction with the incorporation of WL and substrate material.

Wang et al. [178,179] also pointed out that an equilibrium theory based only on the two energetic contributions of the island surface energy and the elastic relaxation energy cannot predict a finite equilibrium size of the islands, since at high coverage larger islands will be more stable than smaller ones because the energy gain from strain relaxation dominates. They therefore proposed a so-called constrained equilibrium theory, which enables an optimum island size to be derived at a fixed coverage

and number density from the energetic balance which drives material transport between the WL and the QDs. The number density is determined by growth kinetics.

For simplicity, they treat an idealized shape which does not occur in practice (a square-based pyramid with $\{1\,1\,0\}$ facets) and likewise the reconstruction of the WL used for surface energy calculations based on density functional theory within the local density approximation is unrealistic. They chose the $\beta 2(2 \times 4)$ InAs surface, whilst in fact it is an (In, Ga)As alloy surface, usually (1×3) reconstructed. It is nevertheless claimed that the essential physics is captured and they do indeed find good agreement between theory and experiment for QD size dependence and the predicted active involvement of the WL in QD formation is also in complete accord with measurement [129].

It is also well-established experimentally [175] that the critical thickness for the 2D–3D transition is growth-condition-dependent, which is a further prediction of the theory. Unfortunately, the functional form predicted is completely contrary to observation on two counts; first, the critical thickness increases with increasing temperature [175], whereas the theory predicts a decrease, and second, it is predicted that the critical thickness should increase with increasing number density, but experimentally it is found that while the latter increases with increasing flux, the critical thickness does not change, i.e. the critical thickness is independent of number density at constant temperature.

3.3.5. Nonlinear equation for self-organization

Continuum descriptions embody a more global view of morphological evolution than simulation methods, which is especially important for heteroepitaxial growth in the presence of misfit strain, where nonlocal elastic interactions, such as the local inter-island spacings, have hindered the widespread application of KMC simulations. Among the many advantages of continuum formulations are the extensive libraries of numerical methods for integrating deterministic and stochastic differential equations. Continuum equations also enable the consequences of elastic interactions for the self-organization of quantum dots to be identified while providing a framework, at least in principle, for the atomistic origins of this morphology to be explored.

We consider a theory in which an epitaxial thin film wets a rigid substrate. The evolution equation for the height h of this film is [180–182]

$$\frac{\partial h}{\partial t} = D\sqrt{1 + |\nabla h|^2}\nabla_S^2(E + \gamma K + \Phi), \tag{8}$$

in which D is proportional to the surface diffusivity, $E(h)$ is the elastic energy [181], γ is the surface energy and K is the curvature of the surface, and Φ is the wetting chemical potential, which is a function of the film thickness, local slope and curvature: $\Phi = \Phi(h, |\nabla h|^2, \nabla^2 h)$.

The linear stability analysis [180] of this model without the wetting potential Φ indicates that there is a critical thickness below which the film is stable and above which it is unstable. Moreover, the wavenumbers of perturbations corresponding to instability just above the critical thickness are small. Thus, the nonlinear evolution of the film can be described using a small-slope approximation. This has the virtue of transforming the free boundary elastic problem into an evolution equation for the profile H of the free surface. The small-slope approximation of Eq. (8) can be written as [182]

$$\frac{\partial h}{\partial t} = g\nabla^2 h + \nabla^4 h + \nabla^6 h + \nabla^2[h\nabla^2 h + p(\nabla h)^2 + qh^2], \tag{9}$$

where g, p, and q are adjustable constants given in terms of the elastic constants of the film and of model parameters. The linear stability of this equation can be assessed for a perturbation of the form

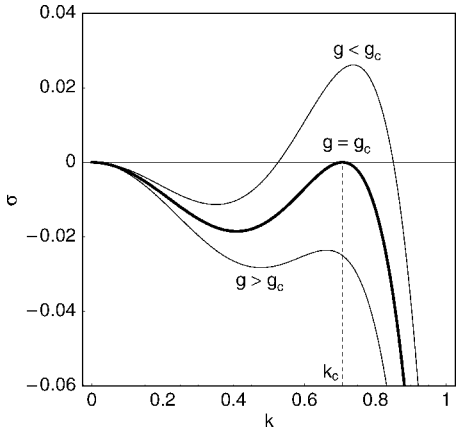

Fig. 30. Dispersion curves $\sigma(k)$ for a linear stability analysis of Eq. (9), showing the regions $g < g_c$, $g = g_c$, and $g > g_c$ and the critical wavenumber $k_c = \frac{1}{2}\sqrt{2}$. Adapted from [182].

$h \sim e^{\sigma t + i\boldsymbol{k}\cdot\boldsymbol{r}}$, which yields the dispersion relation $\sigma = -gk^2 + k^4 - k^6$. The instability is seen for $g < g_c = \frac{1}{4}$ at a wavenumber of $k_c = \frac{1}{2}\sqrt{2}$, as shown in Fig. 30. This non-zero critical value of k_c, which is a direct result of the wetting layer contribution to the elastic energy, is responsible for the emergence of solutions of Eq. (9) with stationary periodic patterns.

Different choices of g, p, and q in Eq. (9) lead to a variety of surface morphologies, including dots, pits, and wires. Fig. 31 shows two types of dots morphologies: cone-type and cap-type. The formation of these patterns is a direct result of the wetting layer interactions, which produces an instability threshold at a small but non-zero wavevector. The particular pattern and its stability are determined by the wetting potential. However, despite the presence of different morphologies in the solutions of Eq. (9), the appropriate values for any particular system are not easily determined, and the relationship to underlying atomistic processes is not at all clear, which pre-empts a direct calculation. On the other hand, the equation of motion for h in Eq. (9) is of the form one would expect to emerge from the coarse-graining of atomistic models. Once a direct connection to atomistic processes can be established, first-principles methods can be applied to determine the appropriate morphologies for particular systems as a function of their growth conditions. This remains an active research area.

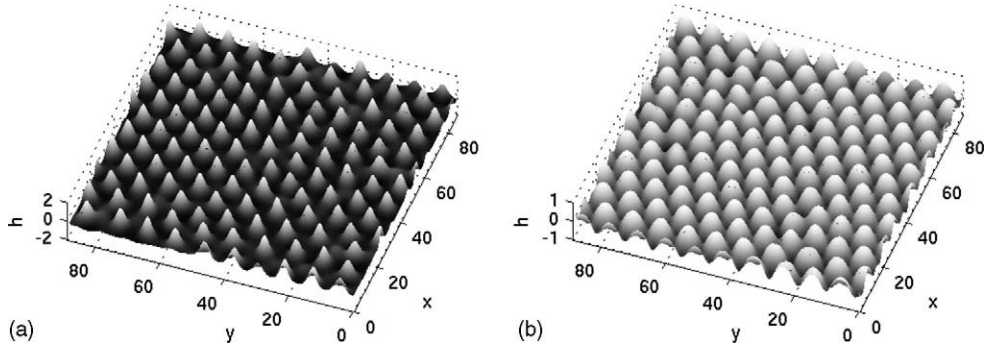

Fig. 31. Quantum dot morphologies obtained from the numerical solution of Eq. (9) for particular choices of the parameters g, p, and q: (a) hexagonal array of cone-type dots obtained for $g = 0.2$, $p = 2$, and $q = -0.5$; and (b) hexagonal array of cap-type dots, obtained for $g = 0.01$, $p = -4$, and $q = 1$. Reprinted with permission from A.A. Golovin et al., Phys. Rev. E 68 (2003) 056203. Copyright 2003 by the American Physical Society.

3.4. Quantum dot shape considerations

The precise shape of QDs is an important factor in determining their electronic structure and there have been many assertions that the basic shape is a square-based pyramid. Clearly such an idealized shape simplifies calculations of their electronic structure [183,184], but the experimental evidence shows categorically that it cannot be correct. Were it to be so, RHEED patterns from the dot structures in two orthogonal $\langle 1\,1\,0\rangle$ azimuths would have to be essentially identical, but as is evident from Fig. 21, this is very far from the case. These patterns are quite typical of those from literally hundreds that have been published in the literature, and while there may be some disagreement about their precise origin, there can be no suggestion that they could be produced by a structure whose geometry is invariant when observed in two orthogonal directions. Despite this rather obvious conclusion, there have been numerous claims, experimental as well as theoretical, that QDs are indeed essentially either square-based pyramids or truncated pyramids (which would also produce identical RHEED patterns in two orthogonal azimuths), so we need to consider the techniques that have been used for shape analysis and how the results have been interpreted. The methods used have been mainly RHEED, TEM, STM and high resolution X-ray diffraction (HRXRD).

Results using RHEED are based on the observation of chevron features (Fig. 21) extending from diffraction spots towards (occasionally away from) the shadow edge, which are attributed to faceting of the QDs [185,186,187]. The assumption is that the chevron arms correspond to reciprocal lattice rods normal to the facet plane, so the half-angle between the surface normal and the chevron direction is a measure of the angle between the (001) surface plane and the facet, whose orientation is thus determined. The facets have been identified mainly as $\{1\,1\,n\}$, where $n = 3, 4$ or 5 [186,187], although $\{1\,3\,6\}$ and $\{1\,3\,7\}$ planes have also been reported [185]. The chevron features are present in $[\bar{1}\,1\,0]$ azimuthal patterns, but not in the orthogonal [110], where only diffraction spots are observed, formed by transmission diffraction through the dots when the electron beam interacts with the surface from that direction. It is quite evident from Fig. 21 that the patterns from orthogonal $\langle 1\,1\,0\rangle$ azimuths are completely different.

There is also a problem with the interpretation of faceting based on the chevron directions. It is a simple matter to show from reciprocal lattice considerations that diffraction from facets should produce symmetrical crosses centred on the diffraction spot, not chevrons, which is obviously contrary to the results from this material system (and several others). A solution has been proposed independently by two groups using somewhat different, but related methods [80,88], and each came to the same conclusion. Pashley and co-workers [88] showed by accurate calculation of the refractive deviation of an incident beam for low grazing incidence angles that pure refraction effects could explain not only the absence of symmetrical crosses in the pattern, but also account for the length and intensity distribution of the chevrons. The effect is for the intensity to fall off rather rapidly with distance from the centre of the diffraction spot and for some in-filling between the arms of the chevron close to the spot, as seen in Fig. 21. The shape of the QDs required to produce these patterns is lenticular, as illustrated in Fig. 32; the bounding faces must be curved rather than simply facetted, and this shape also explains the occurrence of transmission spots in the [110] azimuth, since electrons leaving zones A and C in Fig. 32 will not be refracted significantly because the grazing incidence angles are too large.

Hanada et al. [116] similarly invoked refraction effects, but used kinematic theory first to calculate the points of incidence and exit for each scattering atom, and then to find the wave vectors inside and the scattering angle at the atom for given incident and out-going wave vectors in vacuum. The calculations were carried out for various possible QD shapes and compared with experimental

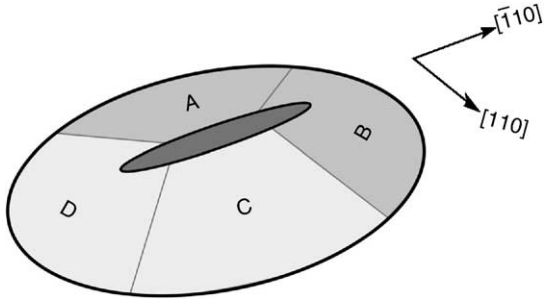

Fig. 32. Lenticular shape of QDs deduced from a refraction-based theory of the origin of chevrons observed in RHEED patterns.

RHEED patterns. They concluded that the most probable shape was an asymmetric dome and that the key requirement to produce a chevron feature was for the surface of the dot facing the $[1\,\bar{1}\,0]$ and $[\bar{1}\,1\,0]$ azimuths to be curved.

It therefore seems reasonable to conclude that the lenticular shape allows a self-consistent interpretation of the available RHEED data, and more recent electronic structure calculations have been based on this shape [188]. Further support can be provided since fully formed symmetric crosses are indeed present in diffraction patterns from surfaces on which facets have been formed by sequential growth and annealing, as illustrated in Fig. 33, taken from the InP(001)–InAs system. The initial pattern is indicative of the same refraction effects as those discussed above, but as the film is annealed, the features predicted for diffraction from a regularly facetted structure develop, confirming that if facets are present, RHEED can detect them.

TEM has also been widely used for the ex situ analysis of QD shapes, both in plan-view and cross-section. The problem here is ensuring the correct interpretation of diffraction contrast in the image, in particular, separating strain-related contrast from that due to atomic number (z-dependent). For example, the images of the QDs shown in Fig. 20 appear to be approximately square, bounding a Maltese cross. In this case, however, the contrast is almost entirely strain induced, so the image is not a reflection of the dot shape. Some earlier claims of a regular crystallographic shape were based on this misunderstanding, but when the contrast is evaluated correctly, the plan-view TEM results are in general agreement with the conclusions reached by RHEED and the image can be shown to be that expected from a lenticular shape [113,189].

STM imaging has also been used quite extensively, but it is not an ideal method for precise shape analysis of features between 30 and 50Å high, which is the range for typical QDs. As a consequence, there is a tendency to use electronic image enhancement to generate idealized regular facets. When height profiles are used, however, the presence of such facets appears much less convincing [190,191]. Finally, HRXRD analysis is a much more complex problem [192,193] and the results are usually in the form of strain and composition profiles, which tend to support the concept of lenticular shapes, but do not provide any definite proof.

A comparison of experimentally determined shapes with those calculated theoretically shows there to be no significant level of agreement. Moll et al. [194] used a combination of density functional theory calculations with elasticity theory to deduce the equilibrium shape of the 3D islands and found them to be bounded by $\{1\,1\,0\}$ and $\{1\,1\,1\}$ planes, with a (001) top. This is clearly not the experimentally measured shape, so either the calculation has shortcomings or growth is not an equilibrium process. There is no doubt that there is a significant degree of kinetic control, so the calculation may simply be inapplicable to the experimental situation. Spencer and Tersoff [195]

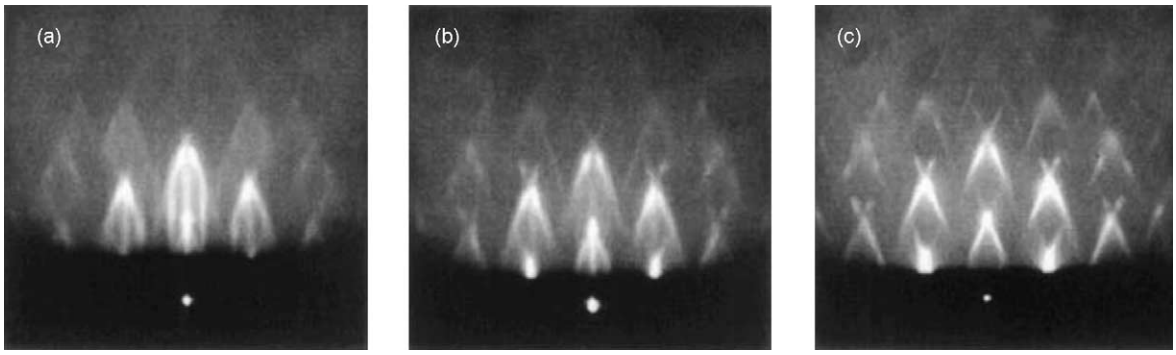

Fig. 33. RHEED patterns in the [110] azimuth from the growth of 2.5ML InAs on InP(001): (a) followed by annealing; (b, c) showing the final development of symmetrical crosses from the initial chevrons.

studied the equilibrium of a coherent strained layer that first wets the substrate and found that with further growth discrete islands formed with a contact angle of zero to this film. Small islands had a minimum width and hence an arbitrarily small aspect ratio, while very large islands had a shape that approached that of a sphere on the WL, but there was no faceting. Again, however, only equilibrium growth could be considered, so although the results were rather closer to those actually observed, it would be unwise to conclude that accurate modelling of the shape development has yet been achieved.

3.5. Summary and outlook

3.5.1. Self-assembly processes in GaAs homoepitaxy

The most important feature for homoepitaxy of GaAs is the complexity of the surface almost universally used as the starting template for growth on the (001)-oriented surface, the $\beta 2(2 \times 4)$ reconstruction. We have discussed in some detail the basis of both the ideal unit cell and the additional effects introduced by various types of disorder observed in the long-range structure. On (110)- and (111)A-oriented substrates the surfaces are much less complicated and we have shown how this affects the island kinetics, resulting in an essentially ''conventional'' nucleation and growth scenario, as reflected in the distributions of island sizes on these surfaces.

The range of surface stoichiometry which occurs on (001) surfaces means that Ga and As_2 have to be treated with equal importance in growth models. One consequence of this is the need to include a weakly bound mobile precursor state for As_2 prior to possible incorporation in the growing lattice, which is borne out by experiment, first principles calculations and KMC simulation. It also leads to a growing surface with a higher As population than that present on the starting $\beta 2(2 \times 4)$ template. This facilitates the insertion of Ga into the As–As dimer bond, which is an essential step in the growth process.

3.5.2. Quantum dot formation

In this part of the review we have first discussed the orientation and reconstruction specificity of the formation of coherent self-assembled quantum dots by a modified Stranski–Krastanov mechanism in the InAs–GaAs system. Several theoretical approaches have been developed to explain various features of QD formation, including first principles calculations, KMC simulations, rate equations,

equilibrium theories, and continuum approaches. So far, however, none is capable of explaining all of the experimental observations, due in large part to the essentially multiscale nature of the phenomenon. The basic phenomenology and the theoretical support, where available, as well as open questions, are:

- QDs only form on the GaAs(001)–$c(4 \times 4)$ surface, at least as far as the low-index oriented substrates are concerned. On the other low-index surfaces and other reconstructions, growth follows a layer-by-layer 2D mode and strain is relaxed by the formation of misfit dislocations. We have not considered higher index surfaces, since they are made up from low index components. Although there have been discussions about the relative efficacies of various strain relaxation mechanisms, there is no theory that explains the heteroepitaxial growth modes on the low-index GaAs surfaces. But (deterministic) continuum theories reveal that the wetting layer exerts a strong influence on the nature of morphological evolution and facilitates the appearance of ordered arrays of quantum dots.
- When QDs are formed, a 2D ''wetting'' layer develops first. Several properties of this wetting layer are revealed by experiments, first-principles calculations, and simulations: it is an alloy phase whose migration barriers for In adatoms are far lower than on the GaAs substrate. Energy barriers to attachment to large islands and steps, as well as the possibility of Ga–In atom exchange in the outermost surface region, leads to a large (and very mobile) adatom population prior to the 2D-to-3D transition. Support for this concept can be found in the final state composition of the QDs, which are not pure InAs, but an (In, Ga)As alloy, where the In concentration increases with decreasing In flux to the surface.
- The 2D–3D transition is very rapid, occurring with an incremental deposition of <0.1ML. The atomistic nucleation mechanism has not been conclusively established, but simulations suggest that upward adatom diffusion from the wetting layer at island edges. The sharpness of the 2D–3D transition can be attributed to the large mobile adatom population that is poised to respond quickly to any changes to the surface chemical potential. Indeed, this may explain why the measured 3D island-size distributions exhibit the profile of irreversible homoepitaxial systems: the large mass current toward 3D island edges corresponds to a large effective flux, which pushes the system towards the irreversible limit. However, the existence of 2D precursors for the 3D islands must be regarded with some caution and may, in fact, be an artefact of the quenching process rather than an essential kinetic intermediate.
- Experiments and simulations indicate that growing 3D islands consume mass both from the adatom population and from the nearby wetting layer. The barriers to attachment found in first-principles calculations would suggest a size regularization of the 3D islands, resulting in a narrow island size distribution. However, the large initial growth rates of the 3D islands noted above may pre-empt this.

Several theoretical approaches have been developed to explain various features of QD formation, including first principles calculations, KMC simulations, rate equations, equilibrium theories, and continuum approaches. So far, however, none is capable of explaining all of the experimental observations, due in large part to the essentially multiscale nature of the phenomenon.

Finally, we have considered QD shape and concluded that both TEM and RHEED observations can only be accounted for consistently if the dots are effectively lenticular. More regular, facetted structures which have been proposed are not in accord with the totality of experimental results.

Acknowledgements

The support of the (UK) Engineering and Physical Sciences Research Counciil is gratefully acknowledged.

References

[1] B.A. Joyce, D.D. Vvedensky, in: S. Mahajan (Ed.), Handbook of Semiconductors, vol. 3, Elsevier, The Netherlands, 1994, pp. 275–368.
[2] J.R. Arthur, Surf. Sci. 500 (2002) 189.
[3] J.H. Neave, P.J. Dobson, B.A. Joyce, J. Zhang, Appl. Phys. Lett. 47 (1985) 100.
[4] T. Shitara, D.D. Vvedensky, M.R. Wilby, J. Zhang, J.H. Neave, B.A. Joyce, Phys. Rev. B 46 (1992) 6815.
[5] A.R. Avery, H.T. Dobbs, D.M. Holmes, B.A. Joyce, D.D. Vvedensky, Phys. Rev. Lett. 79 (1997) 3938.
[6] T. Shitara, D.D. Vvedensky, M.R. Wilby, J. Zhang, J.H. Neave, B.A. Joyce, Phys. Rev. B 46 (1992) 6825.
[7] J.H. Neave, B.A. Joyce, P.J. Dobson, Appl Phys. A 34 (1984) 179.
[8] M. Hata, A. Watanabe, T. Isu, J. Cryst. Growth 111 (1991) 83.
[9] J.E. Northrup, S. Froyen, Phys. Rev. Lett. 71 (1993) 227.
[10] J.E. Northrup, S. Froyen, Phys. Rev. B 50 (1994) 2015.
[11] H.H. Farrell, J.P. Harbison, L.D. Peterson, J. Vac. Sci. Technol. B 5 (1987) 1482.
[12] M.D. Pashley, Phys. Rev. B 40 (1989) 10481.
[13] A. Madhukar, S.V. Ghaisas, CRC Crit. Rev. Solid State Mater. Sci. 14 (1988) 1.
[14] M. Itoh, G.R. Bell, A.R. Avery, T.S. Jones, B.A. Joyce, D.D. Vvedensky, Phys. Rev. Lett. 81 (1998) 633.
[15] M. Itoh, G.R. Bell, B.A. Joyce, D.D. Vvedensky, Surf. Sci. 464 (2000) 200.
[16] K. Shiraishi, Thin Solid Films 272 (1996) 345.
[17] A. Kley, P. Ruggerone, M. Scheffler, Phys. Rev. Lett. 79 (1997) 5278.
[18] K. Shiraishi, T. Itoh, Phys. Rev. B 57 (1998) 6301.
[19] C.G. Morgan, P. Kratzer, M. Scheffler, Phys. Rev. Lett. 82 (1999) 4886.
[20] K. Shiraishi, T. Ito, Surf. Sci. 357–358 (1996) 451.
[21] P. Kratzer, C.G. Morgan, M. Scheffler, Phys. Rev. B 59 (1999) 15246.
[22] T. Ito, K. Shiraishi, Surf. Sci. 357–358 (1996) 486.
[23] T. Ito, K. Shiraishi, Jpn. J. Appl. Phys. Part 2, 35 (1996) 949.
[24] I. Kamiya, D.E. Aspnes, L.T. Florez, J.P. Harbison, Phys. Rev. B 46 (1992) 15894.
[25] I. Kamiya, D.E. Aspnes, H. Tanaka, L.T. Florez, E. Colas, J.P. Harbison, Appl. Surf. Sci. 60–61 (1992) 543.
[26] I. Kamiya, D.E. Aspnes, H. Tanaka, L.T. Florez, J.P. Harbison, R. Bhat, Phys. Rev. Lett. 68 (1992) 627.
[27] J.R. Creighton, K.C. Banton, Surf. Sci. 409 (1998) 372.
[28] M. Wassermeier, J. Bahrend, J.-T. Zettler, K. Stahrenberg, K. Ploog, Appl. Surf. Sci. 107 (1996) 48.
[29] J.N. Eckstein, C. Webb, S.-L. Weng, K.A. Bertness, Appl. Phys. Lett. 51 (1987) 1833.
[30] J.N. Eckstein, C. Webb, S.-L. Weng, K.A. Bertness, J. Vac. Sci. Technol. B 6 (1988) 736.
[31] H. Tsuda, J. Appl. Phys. 73 (1993) 1534.
[32] J.J. Zinck, D.H. Chow, Appl. Phys. Lett. 66 (1995) 3524.
[33] C.T. Foxon, B.A. Joyce, Surf. Sci. 64 (1977) 293.
[34] L. Däweritz, R. Hey, Surf. Sci. 236 (1990) 15.
[35] D.J. Chadi, J. Vac. Sci. Technol. A 5 (1987) 834.
[36] T. Hashizume, Q.-K. Xue, A. Ichimiya, T. Sakurai, Phys. Rev. B 51 (1995) 4200.
[37] Y. Garreau, M. Sauvage-Simkin, N. Jedrecy, R. Pinchaux, M.B. Veron, Phys. Rev. B 54 (1996) 2989.
[38] V.P. LaBella, H. Yang, D.W. Bullock, P.M. Thibado, P. Kratzer, M. Scheffler, Phys. Rev. Lett. 83 (1999) 2989.
[39] A. Ohtake, M. Ozaki, T. Yasuda, T. Hanada, Phys. Rev. B 65 (2002) 165315.
[40] P.J. Dobson, J.H. Neave, B.A. Joyce, Surf. Sci. 119 (1982) L339.
[41] M.D. Pashley, K.W. Harberen, W. Friday, J.M. Woodall, P.D. Kirchner, Phys. Rev. Lett. 60 (1988) 2176.
[42] P.K. Larsen, D.J. Chadi, Phys. Rev. B 37 (1988) 8282.
[43] A.R. Avery, C.M. Gorringe, D.M. Holmes, J.L. Sudijono, T.S. Jones, Phys. Rev. Lett. 76 (1996) 3344.
[44] A. Ichimiya, Surf. Sci. 23 (1990) 75.
[45] A. Ichimiya, Y. Nishikawa, M. Uchiyama, Surf. Sci. 493 (2001) 232.

[46] P.K. Larsen, P.J. Dobson, J.H. Neave, B.A. Joyce, B. Bölger, J. Zhang, Surf. Sci. 169 (1985) 176.
[47] J.M. McCoy, U. Korte, P.A. Maksym, Surf. Sci. 418 (1998) 273.
[48] D.W. Pashley, J.H. Neave, B.A. Joyce, to be published.
[49] Q.-K. Xue, T. Hashizume, T. Sakurai, Prog. Surf. Sci. 56 (1997) 1, and references therein.
[50] S.B. Zhang, A. Zunger, Phys. Rev. B 53 (1996) 1343.
[51] M.A. Salmi, N. Alatalo, T. Ala-Nissila, R.M. Nieminen, Surf. Sci. 425 (1999) 31.
[52] T. Ohno, Thin Solid Films 272 (1996) 331.
[53] W.G. Schmidt, F. Bechstedt, Phys. Rev. B 54 (1996) 16742.
[54] B.A. Banse, J.R. Creighton, Appl. Phys. Lett. 60 (1992) 856.
[55] C.T. Foxon, J.A. Harvey, B.A. Joyce, J. Phys. Chem. Solids 34 (1973) 1693.
[56] C. Sasaoka, K. Yato, A. Usui, Surf. Sci. 265 (1992) L239.
[57] M.J. Begarney, L. Li, C.H. Li, D.C. Law, Q. Fu, R.F. Hicks, Phys. Rev. B 62 (2000) 8092.
[58] A.I. Shkrebtii, N. Esser, W. Richter, W.G. Schmidt, F. Bechstedt, B.O. Fimland, A. Kley, R. Del Sole, Phys. Rev. Lett. 81 (1998) 721.
[59] K. Kamisawa, H. Yamaguchi, Phys. Rev. B 56 (1997) 12080.
[60] G.R. Bell, J.G. Belk, C.F. McConville, T.S. Jones, Phys. Rev. B 59 (1999) 2947.
[61] C. Ratsch, P. Šmilauer, A. Zangwill, D.D. Vvedensky, Surf. Sci. 329 (1995) L599.
[62] C. Ratsch, A. Zangwill, P. Šmilauer, D.D. Vvedensky, Phys. Rev. Lett. 72 (1994) 3194.
[63] D. Kandel, Phys. Rev. Lett. 78 (1997) 499.
[64] P. Jensen, H. Larralde, A. Pimpinelli, Phys. Rev. B 55 (1997) 2556.
[65] P.-M. Lam, R. Tashakkori, K. Yu, Phys. Rev. B 56 (1997) 4893.
[66] J.E. Vasek, Z. Zhang, C.T. Salling, M.G. Lagally, Phys. Rev. B 51 (1995) 17207.
[67] P. Šmilauer, K. Mizushima, D.D. Vvedensky, Phys. Rev. Lett. 81 (1998) 5600.
[68] I. Markov, Phys. Rev. B 50 (1994) 11271.
[69] D. Kandel, E. Kaxiras, Phys. Rev. Lett. 75 (1995) 2742.
[70] C. Ratsch, A. Zangwill, P. Šmilauer, Surf. Sci. Lett. 314 (1994) L937.
[71] H.T. Dobbs, D.D. Vvedensky, A. Zangwill, J. Johansson, N. Carlsson, W. Seifert, Phys. Rev. Lett. 79 (1997) 897.
[72] J.W. Evans, M.C. Bartelt, J. Vac. Sci. Technol. A 12 (1994) 1800.
[73] J.G. Amar, F. Family, Phys. Rev. Lett. 74 (1995) 2066.
[74] E. Bauer, Z. Krist. 110 (1958) 372.
[75] G. LeLay, R. Kern, J. Cryst. Growth 44 (1978) 372.
[76] J.A. Venables, G.D.T. Spiller, M. Hanbücken, Rep. Prog. Phys. 47 (1984) 399.
[77] X.M. Zhang, D.W. Pashley, J.H. Neave, P.N. Fawcett, J. Zhang, B.A. Joyce, J. Cryst. Growth 132 (1993) 331.
[78] J.G. Belk, J.D. Sudijono, X.M. Zhang, J.H. Neave, T.S. Jones, B.A. Joyce, Phys. Rev. Lett. 78 (1997) 475.
[79] D.M. Holmes, J.G. Belk, J.L. Sudijono, J.H. Neave, T.S. Jones, B.A. Joyce, J. Vac. Sci. Technol. A 14 (1996) 849.
[80] J.G. Belk, D.W. Pasley, C.F. McConville, J.L. Sudijono, B.A. Joyce, T.S. Jones, Phys. Rev. B 56 (1997) 10289.
[81] J.G. Belk, D.W. Pashley, C.F. McConville, B.A. Joyce, T.S. Jones, Surf. Sci. 410 (1998) 82.
[82] J.G. Belk, D.W. Pashley, B.A. Joyce, T.S. Jones, Phys. Rev. B 58 (1998) 16194.
[83] D. Maroudas, L.A. Zepeda-Ruiz, W.H. Weinberg, Appl. Phys. Lett. 73 (1998) 753.
[84] L.A. Zepeda-Ruiz, D. Maroudas, W.H. Weinberg, J. Appl. Phys. 85 (1999) 3677.
[85] R. Bonnet, Philos. Mag. A 43 (1981) 1165.
[86] R. Bonnet, Phys. Rev. B 53 (1996) 10978.
[87] N. Oyama, E. Ohta, K. Takeda, K. Shiraishi, H. Yamaguchi, Surf. Sci. 433 (1999) 900.
[88] X.M. Zhang, D.W. Pashley, L. Hart, J.H. Neave, P.N. Fawcett, B.A. Joyce, J. Cryst. Growth 131 (1993) 300.
[89] L. Hart, P.F. Fewster, Inst. Phys. Conf. Ser. No. 134, Presented at Microsc. Semicond. Mater. Conf., Oxford, 1993, p. 569.
[90] H. Yamaguchi, M.R. Fahy, B.A. Joyce, Appl. Phys. Lett. 69 (1996) 776.
[91] H. Yamaguchi, J.G. Belk, X.M. Zhang, J.L. Sudijono, M.R. Fahy, T.S. Jones, D.W. Pashley, B.A. Joyce, Phys. Rev. B 55 (1997) 1337.
[92] E. Tournié, O. Brandt, C. Gianni, K.H. Ploog, J. Cryst. Growth 127 (1993) 765.
[93] C.W. Snyder, B.G. Orr, Phys. Rev. Lett. 70 (1993) 1030.
[94] N. Grandjean, J. Massies, Phys. Rev. Lett. 70 (1993) 1031.
[95] Q.K. Xue, Y. Hasegawa, T. Ogino, H. Kiyama, T. Sakurai, J. Vac. Sci. Technol. B 15 (1997) 1270.
[96] T.L. Lee, C. Kumpf, A. Kazimirov, P.F. Lyman, G. Scherb, M.J. Bedzyk, N. Nielsen, R. Feidenhans'l, R.L. Johnson, B.O. Fimland, J. Zegenhagen, Phys. Rev. B 66 (2002) 235301.

[97] C. Kumpf, L.D. Marks, D. Ellis, D. Smilgies, E. Landemark, M. Nielsen, R. Feidenhans'l, J. Zegenhagen, O. Bunk, J.H. Zeysing, Y. Su, R.L. Johnson, Phys. Rev. Lett. 86 (2001) 3586.
[98] A. Trampert, E. Tournié, K.H. Ploog, Appl. Phys. Lett. 66 (1995) 2265.
[99] V. Bressler-Hill, S. Varma, A. Lorke, B.Z. Nosho, P.M. Petroff, W.H. Weinberg, Phys. Rev. Lett. 74 (1995) 3209.
[100] T. Ito, K. Shiraishi, Jpn. J. Appl. Phys. 36 (1997) L1532.
[101] G.R. Bell, T.J. Krzyzewski, P.B. Joyce, T.S. Jones, Phys. Rev. B 61 (2000) R10551.
[102] J.G. Belk, C.F. McConville, J.L. Sudijono, T.S. Jones, B.A. Joyce, Surf. Sci. 387 (1997) 213.
[103] T.J. Krzyzewski, P.B. Joyce, G.R. Bell, T.S. Jones, Surf. Sci. 517 (2002) 8.
[104] J.G. Belk, J.L. Sudijono, D.M. Holmes, C.F. McConville, T.S. Jones, B.A. Joyce, Surf. Sci. 365 (1996) 735.
[105] M. Sauvage-Simkin, Y. Garreau, R. Pinchaux, M.B. Veron, J.P. Landesman, J. Nagle, Phys. Rev. Lett. 75 (1995) 3485.
[106] L.G. Wang, P. Kratzer, N. Moll, M. Scheffler, Phys. Rev. B 62 (2000) 1897.
[107] L.G. Wang, P. Kratzer, M. Scheffler, Jpn. J. Appl. Phys. 39 (7B) (2000) 4298.
[108] A.G. deOliveira, S.D. Parker, R. Droopad, B.A. Joyce, Surf. Sci. 227 (1990) 150.
[109] M.R. Fahy, M.J. Ashwin, J.J. Harris, R.C. Newman, B.A. Joyce, Appl. Phys. Lett. 61 (1992) 1805.
[110] L.H. Germer, J.W. May, R.J. Szostak, Surf. Sci. 8 (1967) 430.
[111] J.E. Houston, R.L. Park, Surf. Sci. 21 (1970) 209.
[112] J.G. Belk, Ph.D. thesis, University of London, 1999, unpublished, p. 66, et. seq.
[113] J. Zou, X.Z. Liao, D.J.H. Cockayne, R. Leon, Phys. Rev. B 59 (1999) 12279.
[114] A. Rosenauer, U. Fischer, D. Gerthsen, A. Förster, Appl. Phys. Lett. 71 (1997) 3868.
[115] D.W. Pashley, J.H. Neave, B.A. Joyce, Surf. Sci. 476 (2001) 35.
[116] T. Hanada, B.H. Koo, H. Totsuka, T. Yao, Phys. Rev. B 64 (2001) 165307.
[117] R. Heitz, T.R. Ramachandran, A. Kalburge, Q. Xie, I. Mukhametzhanov, P. Chen, A. Madhukar, Phys. Rev. Lett. 78 (1997) 4071.
[118] T.R. Ramachandran, A. Madhukar, I. Mukhametzhanov, R. Heitz, A. Kalburge, Q. Xie, P. Chen, J. Vac. Sci. Technol. B 16 (1998) 1330.
[119] T.J. Krzyzewski, P.B. Joyce, G.R. Bell, T.S. Jones, Phys. Rev. B 66 (2002) R121307.
[120] T.J. Krzyzewski, P.B. Joyce, G.R. Bell, T.S. Jones, Surf. Sci. 532–545 (2003) 822.
[121] F. Patella, S. Nufis, F. Arciprete, M. Fanfoni, E. Placidi, A. Sgorlata, A. Balzarotti, Phys. Rev. B 67 (2003) 205308.
[122] P.B. Joyce, T.J. Krzyzewski, G.R. Bell, T.S. Jones, S. Malik, D. Childs, R. Murray, Phys. Rev. B 62 (2000) 10891.
[123] Y. Ebiko, S. Muto, D. Suzuki, S. Itoh, K. Shiramine, T. Haga, N. Yokoyama, Phys. Rev. Lett. 80 (1999) 2650.
[124] Y. Ebiko, S. Muto, D. Suzuki, S. Itoh, H. Yamakoshi, K. Shiramine, T. Haga, K. Unno, M. Ikeda, Phys. Rev. B 60 (1999) 8234.
[125] M. Krishnamurthy, J.S. Drucker, J.A. Venables, J. Appl. Phys. 69 (1991) 6461.
[126] B.A. Joyce, D.D. Vvedensky, W.K. Liu, M.B. Santos, Thin Films: Heteroepitaxial Systems, World Scientific, Singapore, 1999, pp. 368–400.
[127] A. Madhukar, Q. Xie, P. Chen, A. Konkar, Appl. Phys. Lett. 64 (1994) 2727.
[128] B.A. Joyce, J.L. Sudijono, J.G. Belk, H. Yamaguchi, X.M. Zhang, H.T. Dobbs, A. Zangwill, D.D. Vvedensky, T.S. Jones, Jpn. J. Appl. Phys. 36 (6B) (1997) 4111.
[129] P.B. Joyce, T.J. Krzyzewski, G.R. Bell, B.A. Joyce, T.S. Jones, Phys. Rev. B 58 (1998) R15981.
[130] N. Grandjean, J. Massie, G. Tottereau, Phys. Rev. B 55 (1997) R10189.
[131] X.Z. Liao, J. Zou, D.J.H. Cockayne, R. Leon, C. Lobo, Phys., Rev. Lett. 82 (1999) 5148.
[132] A. Rosenauer, D. Gerthsen, D. Van Dyck, M. Arzberger, G. Böhm, G. Abstreiter, Phys. Rev. B 64 (2001) 245334.
[133] I. Kegel, T.H. Metzger, A. Lorke, J. Peisl, J. Stangl, G. Bauer, J.M. Garca, P.M. Petroff, Phys. Rev. Lett. 85 (2000) 1694.
[134] N. Liu, J. Tersoff, O. Baklenov, A.L. Holmes, C.K. Shih Jr., Phys. Rev. Lett. 84 (2000) 334.
[135] A.G. Cullis, D.J. Norris, T. Walther, M.A. Migliorato, M. Hopkinson, Phys. Rev. B 66 (2002) R081305.
[136] T. Walther, A.G. Cullis, D.G. Norris, M. Hopkinson, Phys. Rev. Lett. 86 (2001) 2381.
[137] O. Dehaese, X. Wallart, F. Mollot, Appl. Phys. Lett. 66 (1995) 52.
[138] S.-H. Lee, W. Moritz, M. Scheffler, Phys. Rev. Lett. 85 (2000) 3890.
[139] C. Ratsch, W. Barvosa-Carter, F. Grosse, J.H.G. Owen, J.J. Zinck, Phys. Rev. B 62 (2000) R7719.
[140] W.G. Schmidt, S. Mirbt, F. Bechstedt, Phys. Rev. B 62 (2000) 8087.
[141] C. Ratsch, Phys. Rev. B 63 (2001) 161306(R).
[142] E. Penev, P. Kratzer, in: B.A. Joyce, P.C. Kelires, A.G. Naumovets, D.D. Vvedensky (Eds.), Quantum Dots: Fundamentals, Applications, Frontiers, Kluwer, Dordrecht, 2004, pp. 27–42.
[143] P. Kratzer, E. Penev, M. Scheffler, Appl. Surf. Sci. 216 (2003) 446.

[144] P. Kratzer, E. Penev, M. Scheffler, Appl. Phys. A 75 (2002) 79.
[145] A. Madhukar, J. Cryst. Growth 163 (1996) 149.
[146] H.M. Koduvely, A. Zangwill, Phys. Rev. B 60 (1999) R2204.
[147] J.D. Weeks, G.H. Gilmer, Adv. Chem. Phys. 40 (1979) 157.
[148] S. Clarke, D.D. Vvedensky, Phys. Rev. Lett. 58 (1987) 2235.
[149] H. Metiu, Y.-T. Lu, Z.Y. Zhang, Science 255 (1992) 1088.
[150] T. Shitara, D.D. Vvedensky, M.R. Wilby, J. Zhang, J.H. Neave, B.A. Joyce, Phys. Rev. B 46 (1992) 6815.
[151] C. Ratsch, A. Zangwill, Surf. Sci. 293 (1993) 123.
[152] C. Ratsch, P. Šmilauer, D.D. Vvedensky, A. Zangwill, J. Phys. I (Fr.) 6 (1996) 575.
[153] F.K. LeGoues, MRS Bull. 21 (4) (1996) 38.
[154] B.W. Dodson, P.A. Taylor, Phys. Rev. B 34 (1996) 2115.
[155] D.A. Faux, G. Gaynor, C.L. Carson, C.K. Hall, J. Bernholc, Phys. Rev. B 42 (1990) 2922.
[156] A.F. Voter, F. Montalenti, T.C. Germann, Ann. Rev. Mater. Res. 32 (2002) 321.
[157] F. Much, M. Ahr, M. Biehl, W. Kinzel, Europhys. Lett. 56 (2001) 791.
[158] M. Biehl, F. Much, Europhys. Lett. 63 (2003) 14.
[159] J. Johansson, W. Seifert, J. Crystal Growth 234 (2002) 132, 134 (2002) 239.
[160] H.T. Dobbs, D.D. Vvedensky, A. Zangwill, Appl. Surf. Sci. 123–124 (1998) 646.
[161] H.T. Dobbs, A. Zangwill, D.D. Vvedensky, M. Tringides, Surface Diffusion, Plenum, New York, 1998, pp. 263–275.
[162] J.A. Venables, Philos. Mag. 27 (1973) 697.
[163] D. Leonard, K. Pond, P.M. Petroff, Phys. Rev. B 50 (1994) 11687.
[164] N.P. Kobayashi, T.R. Ramachandran, P. Chen, A. Madhukar, Appl. Phys. Lett. 68 (1996) 3299.
[165] M.C. Bartelt, J.W. Evans, Phys. Rev. B 46 (1992) 12675.
[166] C. Ratsch, A. Zangwill, P. Šmilauer, D.D. Vvedensky, Phys. Rev. Lett. 72 (1994) 3194.
[167] J.A. Stroscio, D.T. Pierce, Phys. Rev. B 49 (1994) 8522.
[168] B. Müller, L. Nedelmann, B. Fischer, H. Brune, K. Kern, Phys. Rev. B 54 (1996) 17858.
[169] M.C. Bartelt, J.W. Evans, Phys. Rev. B 54 (1996) R17359.
[170] D.D. Vvedensky, Phys. Rev. B 62 (2000) 15435.
[171] Y.G. Cao, M.H. Xie, Y. Liu, S.X. Hu, Y.F. Ng, H.S. Wu, S.Y. Tong, Phys. Rev. B 68 (2003) 161304.
[172] M.C. Bartelt, L.S. Perkins, J.W. Evans, Surf. Sci. 344 (1995) L1193.
[173] E. Korutcheva, A.M. Turiel, I. Markov, Phys. Rev. B 61 (2000) 16890.
[174] I. Markov, Phys. Rev. B 48 (1993) 14016.
[175] B. Lambert, A. Le Corre, V. Drouot, H. L'Haridon, S. Loualiche, Semicon. Sci. Technol. 13 (1998) 143.
[176] C. Priester, M. Lannoo, Phys. Rev. Lett. 75 (1995) 93.
[177] C. Priester, M. Lannoo, Appl. Surf. Sci. 104–105 (1996) 495.
[178] L.G. Wang, P. Kratzer, M. Scheffler, N. Moll, Phys. Rev. Lett. 82 (1999) 4042.
[179] L.G. Wang, P. Kratzer, N. Moll, M. Scheffler, Phys. Rev. B 62 (2000) 1897.
[180] B.J. Spencer, S.H. Davis, P.W. Voorhees, J. Appl. Phys. 73 (1993) 4955.
[181] B.J. Spencer, S.H. Davis, P.W. Voorhees, Phys. Rev. B 47 (1993) 9760.
[182] A.A. Golovin, S.H. Davis, P.W. Voorhees, Phys. Rev. E 68 (2003) 056203.
[183] C. Pryor, Phys. Rev. B 60 (1999) 2869.
[184] L.-W. Wang, J. Kim, A. Zunger, Phys. Rev. B 59 (1999) 5678.
[185] H.L. Lee, R. Lowe-Webb, W. Yang, P. Cercel, Appl. Phys. Lett. 72 (1998) 812.
[186] Y. Nabetini, T. Ishikawa, S. Noda, A. Sasaki, Appl. Phys. Lett. 66 (1994) 3347.
[187] R.P. Mirin, J.P. Ibbetson, K. Nishi, A.C. Gossard, J.E. Bowers, Appl. Phys. Lett. 67 (1995) 3785.
[188] A.J. Williamson, L.W. Wang, A. Zunger, Phys. Rev. B 62 (2000) 12963.
[189] Y. Androussi, T. Benabbas, A. Lefevre, Philos. Mag. Lett. 79 (1999) 201.
[190] Y. Hasegawa, H. Kigama, Q.K. Xue, T. Sakurai, Appl. Phys. Lett. 72 (1998) 2265.
[191] J. Márquez, L. Gulhar, K. Jacobi, Appl. Phys. Lett. 78 (2001) 2309.
[192] I. Kegel, T.H. Metzger, A. Lorke, J. Peisl, J. Stangl, G. Bauer, N. Nordland, W.V. Schoenfeld, P.M. Petroff, Phys. Rev. B 63 (2001) 035318.
[193] K. Zhang, C. Heyn, W. Hansen, T. Schmidt, J. Falta, Appl. Phys. Lett. 76 (2000) 2229.
[194] N. Moll, M. Scheffler, E. Pehlke, Phys. Rev. B 58 (1998) 4566.
[195] B.J. Spencer, J. Tersoff, Phys. Rev. Lett. 79 (1997) 4858.