Journal of Crystal Growth 81 (1987) 193–204
North-Holland, Amsterdam

POLAR-ON-NONPOLAR EPITAXY

Herbert KROEMER

Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA

One of the most fundamental problems that must be solved if device-quality GaAs is to be grown on Si substrates is that of suppressing antiphase disorder. Recent experimental evidence shows that such disorder can be suppressed not only on the (211) orientation, but also on (100), contrary to earlier theoretical expectations. A detailed discussion is given of the mechanism by which this suppression takes place, through a combination of slight misorientation and a high-temperature surface anneal, which lead to the pairing of all Si surface steps into a particular kind of double-height steps. A recent model by Aspnes and Ihm explains the energetic preference for this kind of step by postulating a drastic reconstruction of the atomic configuration at the step edge through the formation of a π-bonded chain running along the step. Another unexpected puzzle is posed by the recent observation that on a given Si(100) surface antiphase disorder-free growth with both possible Ga–As sublattice allocations can be achieved, depending on initial nucleation conditions. A new detailed nucleation model is proposed that explains these observations, by drawing heavily on earlier considerations of Harrison et al. concerning the electrostatics of a polar–nonpolar interface.

1. Introduction

When a polar (compound) semiconductor like, say, GaAs, is grown on a nonpolar (elemental) substrate of similar crystal structure, like Ge or Si, at least three new problems arise that are not present in the more conventional heteroepitaxial growth of one III/V compound upon another: (a) the problem of antiphase disorder on the compound side of the interface; (b) the problem of lack of electrical neutrality at the interface; (c) the problem of cross-doping. Although the existence of these problems has been recognized for some time, they have assumed central importance with the recent progress in the epitaxial growth of GaAs on Si substrates, a development of potentially very large practical significance.

Although the polar-on-nonpolar problems are by no means the only problems that must be solved if high-quality GaAs-on-Si epitaxy is to be achieved – the misfit dislocation problem is at least as severe – a solution of these problems is a *necessary* step without which this goal could not be achieved.

The present paper reviews the progress that has been made recently both in the understanding of the problems of polar-on-nonpolar epitaxy, and in their solution. As a result of this progress, it is probably safe to say that the principal problem in GaAs-on-Si materials quality is no longer the polar-on-nonpolar problem, but the misfit dislocation problem, a discussion of which lies outside the scope of this paper. The emphasis of the paper will be on the antiphase disorder and the interface neutrality problem, both because they are the most fundamental ones from a purely scientific point of view, and because the writer's own work has been concentrating on these. The cross-doping problem is not totally ignored; it is strongly interrelated with the the interface neutrality problem, and some comments about it will be made in that context.

2. Suppression of antiphase disorder

2.1. The problem

The diamond structure in which Si and Ge crystallize consists of two interpenetrating face-centered cubic sublattices. The two sublattices differ from each other only in the spatial orientation of the four tetrahedral bonds that connect each atom to its four nearest neighbors (which are on the other sublattice). For example, in fig. 1 the atoms with the bond orientations indicated as "A"

0022-0248/87/$03.50 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

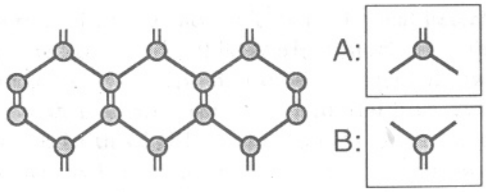

Fig. 1. Two sublattices in a Si crystal, distinguished only by bond orientation in space.

and "B" belong to different sublattices. There is no distinction between the two sublattices otherwise; both are occupied by the same atomic species.

In the zincblende structure, in which GaAs crystallizes, the two sublattices are occupied by different atoms, in the case of GaAs one by Ga atoms, the other by As atoms. In a crystal without antiphase disorder the sublattice allocation is the same throughout the crystal. But if this allocation changes somewhere inside the crystal (fig. 2), the interface between domains with opposite sublattice allocation forms a two-dimensional structural defect called an *antiphase boundary* (APB). The domains themselves are called *antiphase domains* (APDs).

Such APBs can be expected to form when GaAs is grown on Si or Ge, especially on a (100)-oriented substrate, the most widely used crystallographic orientation for MBE and MOCVD growth. Inasmuch as As forms strong bonds with Si, whereas Ga does not, the first atomic layer bonding to the Si substrate should be expected to be an As layer. Now, any *real*(100) surface will *always* exhibit steps. At any step only one atomic layer high (or an odd number of layers high) the sublattice site allocation of Ga and As on opposite sides of the step is interchanged (fig. 3), and an APB results.

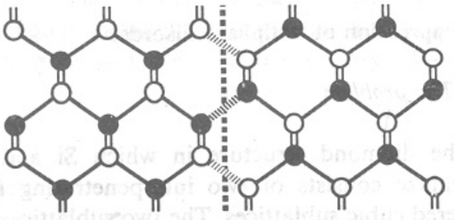

Fig. 2. Antiphase boundary (APB) formation in the zincblende structure, containing (in the case of GaAs) both Ga–Ga and As–As bonds. The configuration shown is the simplest possible one, a perfectly (110)-oriented APB, with alternating Ga–Ga and As–As bonds.

The APBs are structural defects, and we have little reason to expect that they might turn out to be the first benign defects in the history of semiconductor technology. Antiphase boundaries in GaAs contain Ga–Ga and As–As bonds. Such bonds represent electrically charged defects: A comparison of the number of bonding orbitals with the number of valence electrons available to fill them shows that Ga–Ga bonds act as acceptors, and As–As bonds as donors, with effective charges $\pm q/2$ per bond. In general, an APB will contain roughly equal numbers of both charges, thus acting as an extremely highly compensated doping sheet with very little *net* doping. The situation is least bad for an APB that follows exactly anly {110} plane, as in fig. 2: In that case Ga–Ga and As–As bonds will alternate within each crystallographic unit cell, leading to perfect *local* charge compensation. But for the deviations from this idealized arrangement, the lack of exact *local* charge balance will lead to potential fluctuations that will affect the electronic properties. Inasmuch as the initial Si surface steps are not likely to have the exact orientation within the surface plane that would lead to comparatively benign perfect-{110} APBs, the APBs actually resulting from the

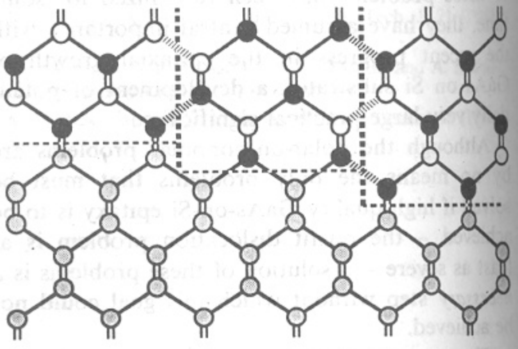

Fig. 3. Mechanism of APB formation during polar-on-nonpolar growth due to the presence of single-height steps on the substrate surface.

nucleation on a real surface must be expected to exhibit local charge fluctuations with large amplitude, and hence be harmful.

There are at least two approaches towards the essentially complete avoidance of APBs: One involves a switch to a different crystallographic orientation on which APBs do not form in principle, even in the presence of steps, like the (211) orientation employed by us [1–4] and described below. The other is to *somehow* enforce a perfect doubling of the height of *all* surface steps, an approach that has proven tenable since early 1985, contrary to all prior expectations.

2.2. The (211) solution

In our own work we have used the (211) orientation for the growth [1–4], rather than relying on perfect step doubling on the (100) surface. On a (211) surface the atomic sites of the two sublattices have a different number of back bonds to the Si substrate (fig. 4): One of the sublattices has two back bonds, the other has only one, *and this difference remains even in the presence of steps*. As a result, the two sublattice sites are no longer energetically and chemically equivalent, and elementary bonding energy considerations suggest that the As atoms, which have a much stronger tendency than Ga atoms to bond to Si, will seek out the more strongly binding doubly back-bonded sites, displacing Ga atoms to the single back-bonded sites. Using crystallographic etching techniques, we were able to demonstrate that the GaAs layers are indeed free of APBs, and that the sublattice allocation is as stated above [1–4].

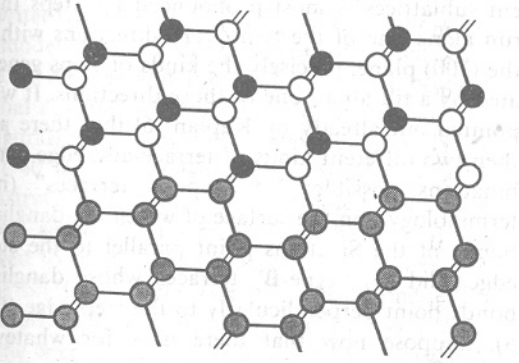

Fig. 4. A (211)-oriented polar-on-nonpolar interface. On such a surface, APBs do not form even in the presence of steps, because the two sublattices differ in the way they are back-bonded to the substrate, leading to sublattice control by chemical bonding preferences.

2.3. Step doubling on (100) surfaces

Most investigators working on GaAs-on-Si growth have preferred to continue to work with the conventional (100) orientation, or with wafers deliberately misoriented from the (100) orientation by a few degrees, relying on step doubling for the suppression of APBs.

When a step on a Si(100) surface is an even number of atomic layers high, the two sublattices on the GaAs side are in registry again, and an APB will not occur at this step. Unfortunately, it is well established experimentally that for "as-polished" exactly (100)-oriented Si surfaces the most common step height is one atomic layer [5,6] and there is in fact ample evidence [7] that not only the growth of GaAs on Si, but the growth of other III/V compounds such as GaP, on exactly (100)-oriented Si or Ge substrates *usually* exhibits copious APBs.

It was first indicated by Henzler and Clabes [5] that on misoriented Si(100) surfaces there is tendency towards step doubling with increasing annealing temperature. This was subsequently followed up in careful detail by Kaplan [6], who reported that on Si surfaces tilted by a few degrees from the (100) plane towards the (011) plane, *most* steps are two atoms high. Inasmuch as the step density on deliberately misoriented surfaces is much higher than for accurately oriented (100) surfaces, a certain amount of step doubling is to be expected, and the step doubling might be extensive if there is a simple energetic preference for double steps over single steps. However, unless the number of remaining single-height steps is drastically reduced, such tilting would not aid in the drastic suppression of APBs. In any event, it is hard to see how APBs could be avoided *completely* over the entire area of an entire wafer: In order to achieve APB-free growth, it is necessary that *all* steps be two atoms high, not just the majority of steps. At first glance, such a proposition appears hopeless. Yet it has become clear since early 1985 that such a perfect step doubling can indeed be achieved, leading to perfectly APB-free epitaxial growth of GaAs on Si(100):

(a) Recently, Fischer et al. [8,9] have reported growth on deliberately misoriented substrates, which does indeed appear to be free of APBs, judging from the anisotropic etching patterns of device structures on the epitaxial layers. Anisotropic etching is one of the simplest and most powerful techniques to test for APDs. We have had an opportunity to investigate one of the layers grown by this group, using our own etch pit technique [1–3], and we confirm the absence of APDs in that layer.

(b) Similarly convincing evidence of APD-free growth, based on an anisotropy of the RHEED patterns that was uniform over the entire wafer area, was presented by Nishi et al. [10]. The Si wafers in that work were not deliberately misoriented, but probably had a small amount of accidental misorientation. Similar results had been reported earlier by the same group for MOCVD-grown GaAs on Si [11]. RHEED evidence similar to that of Nishi et al., but less direct, had been earlier presented by Wang [12], and by hindsight it appears likely that Wang also had achieved APB-free growth.

(c) Perhaps the most convincing direct evidence for perfect step doubling already on the pre-growth Si(100) surface is contained in the stunning recent work by Sakamoto and Hashiguchi [13], who showed that a nominally (100)-oriented Si surface would go from a singly-stepped surface to a doubly-stepped surface during a prolonged high-temperature anneal (20 min at 1000°C), with *all* step terraces belonging to the same sublattice!

2.4. Step doubling mechanism

The empirical observation that a Si(100) surface transforms itself into a single-domain surface upon annealing raises the question: How is the possible? A little reflection shows that *perfect* long-range step doubling can *never* be explained by the elementary proposition that double steps are simply energetically preferred over single steps, *without any additional assumptions*. If that were the whole story, each single-height step on the original surface would pair at random either with the step to its right or the step to its left. In such a model there is a 50% probability that the step to the left first pairs up with the next step even further to the left, and a 50% probability that the step to the right first pairs up with the next step even further to the right, leading to a $(50\%)^2 = 25\%$ probability that *neither* of these nearest-neighbor steps are available for pairing up first with the step of interest, leaving the latter unpaired. In such a case an APB would form on the average at every fourth initial step, and a misorientation would increase the density of APBs rather than decrease it, despite the energetic preference for double steps.

In order to explain the long-range suppression of APBs, the left/right randomness of the step pairing must somehow be eliminated. In the presence of double-height steps the surface layers on all terraces belong to the same Si sublattice; hence there must clearly be a preference for one of the two Si sublattices over the other. Inasmuch as the two different kind of sublattice planes differ from each other only by a 90° rotation in space, the preference mechanism can only reside inside the atomic arrangement at the edges.

At this point it is important to realize that the dangling bond configuration at a step edge depends not only on the direction of the step but also on which of the two sublattices forms the top of the step. The difference between the two different sublattices is most pronounced for steps that run along one of the two ⟨011⟩ directions within the (100) plane, precisely the kinds of steps generated by a tilt about one of those directions. It was pointed out already by Kaplan [6] that there are then two different kinds of terrace-and-edge combinations possible: (a) "type-A terraces" (my terminology), on the surface of which the dangling bonds of the Si atoms point parallel to the step edge, and (b) "type-B" terraces whose dangling bonds point perpendicularly to the step edge (fig. 5). Suppose now that there is – for whatever reasons – a strong energetic preference for one of the two bond configurations, say, for that of a type-A step. It was recently pointed out by this writer [14] that Kaplan's LEED data strongly suggested just such a preference [15], and that under

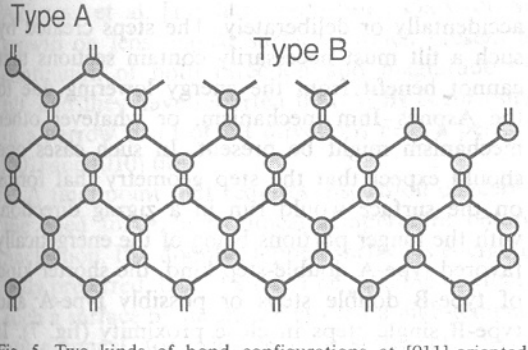

Fig. 5. Two kinds of bond configurations at [011]-oriented single-height atomic steps. For what we call a *type-A step*, the dangling bonds run parallel to the step, for the *type-B step* they run perpendicular.

such conditions, and at sufficiently high temperatures, atoms from type-B edges would diffuse towards the type-A edges, until the former had simply disappeared by forming double-height steps that are bounded by edges of type A. The result would be a perfectly doubly-stepped Si surface, with all terraces belonging to sublattice A, and leading to GaAs growth free of APBs.

The only difficulty with this hypothesis in its original form is that it is hard to see how a sufficiently large energy difference could arise between the two kinds of step edges. Single-bond arguments fail to predict any energy difference, and it is clear that a significant energetic difference requires some sort of reconstruction, but the writer was unable to give a specific reconstruction model. This difficulty has recently been overcome by Aspnes and Ihm [16], who have pointed out that at a type-A *double*-height step the atomic configuration can lower its energy significantly, by about 40 meV per edge atom, through a drastic reconstruction during which π-bonded chains are formed, similar to the π-bonded chains that are believed to be present in the 2×2 reconstruction of the Si(100) surface (fig. 6). The Aspnes–Ihm model differs from the earlier model of ref. [14] in that the energetic preference is not one for type-A single steps over type-B single steps, but one for double-height type-A steps only, over *all* other kind of steps, including single-height type-A steps. But the ultimate consequences regarding APB-free growth are of course the same under both models.

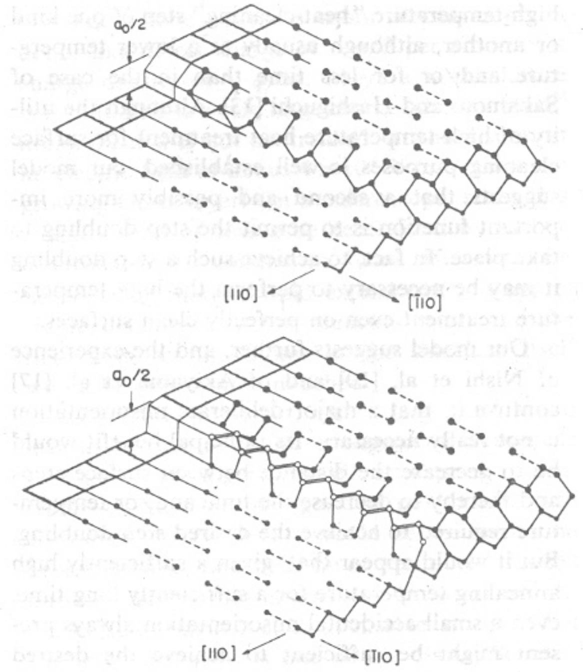

Fig. 6. Atomic reconstruction proposed by Aspnes and Ihm [16] for the type-A double step, explaining the energetic preference for double-height type-A steps over all other kinds of steps. Top: unreconstructed edge. Bottom: a π-bonded atomic chain is formed along the step edges.

It thus appears that the Aspnes–Ihm model is the best model proposed so far to offer an explanation of why single-domain growth of GaAs on Si can be achieved.

2.5. The role of temperature and misorientation

An essential ingredient of the step-doubling hypothesis is a surface temperature sufficiently high for a sufficiently long time to permit diffusion of Si atoms from the energetically unfavorable steps to the favorable ones. In Kaplan's LEED work, these conditions were clearly met: the Si surfaces had been heated to 1100°C, and because of the deliberate large misorientation the distance between steps was small. A study of the recent papers reporting reasonably convincing evidence of a single-domain surface indicates that in all cases the pre-growth Si surface was subjected to a high-temperature "heat-cleaning" step of one kind or another, although usually at a lower temperature and/or for less time than in the case of Sakamoto and Hashiguchi [13]. Although the utility of high-temperature heat treatment for surface cleaning purposes is well established, our model suggests that a second and possibly more important function is to permit the step doubling to take place. In fact, to achieve such a step doubling it may be necessary to perform the high temperature treatment even on perfectly clean surfaces.

Our model suggests further, and the experience of Nishi et al. [10] and of Akiyama et al. [17] confirm it, that a major deliberate misorientation is not really necessary. Its principal benefit would be to decrease the distance between surface steps and thereby to decrease the time and/or temperature required to achieve the desired step doubling. But it would appear that, given a sufficiently high annealing temperature for a sufficiently long time, even a small accidental misorientation always present might be sufficient to achieve the desired goal.

Another important question concerns the *direction* rather than the magnitude of the surface tilt. In our above discussion we had assumed, for simplicity, that the misorientation of the surface away from the exact [100] orientation corresponds to a rotation about one of the two ⟨011⟩ directions, thus leading to steps that can line up parallel to that direction. It is this assumption that led to two different kinds of steps with different kinds of dangling bond configurations. The situation would be quite different for a rotation about either the [010] or the [001] direction. In that case both kinds of steps have dangling bonds whose projections upon the [100] plane runs at the same angle relative to the edge direction (differing only in the sign of that angle), and which are energetically equivalent by symmetry. For such a tilt there is no mechanism enforcing a coherent step doubling with no remaining single steps, and it has in fact been reported [17,18] that GaAs growth on such surfaces leads to copious APBs.

In practice, the exact direction of tilt will rarely agree with an ideal exact ⟨011⟩ rotation, but will be somewhere in between a favorable ⟨011⟩ and the unfavorable [010] or [001] orientation, either accidentally or deliberately. The steps created by such a tilt must necessarily contain sections that cannot benefit from the energy lowering due to the Aspnes–Ihm mechanism, or whatever other mechanism might be present. In such cases one should expect that the step geometry that forms on the surface would run in a zigzag direction, with the longer portions being of the energetically favored type-A double-step kind, the shorter kind of type-B double steps or possibly type-A and type-B single steps in close proximity (fig. 7). If the stabilization energy of the type-A double steps is sufficiently large, APBs would still be suppressed even for very large deviation of the tilt axis from a ⟨011⟩ axis, except possibly for small regions between any pair of single steps. Empirically, this is apparently what happens:

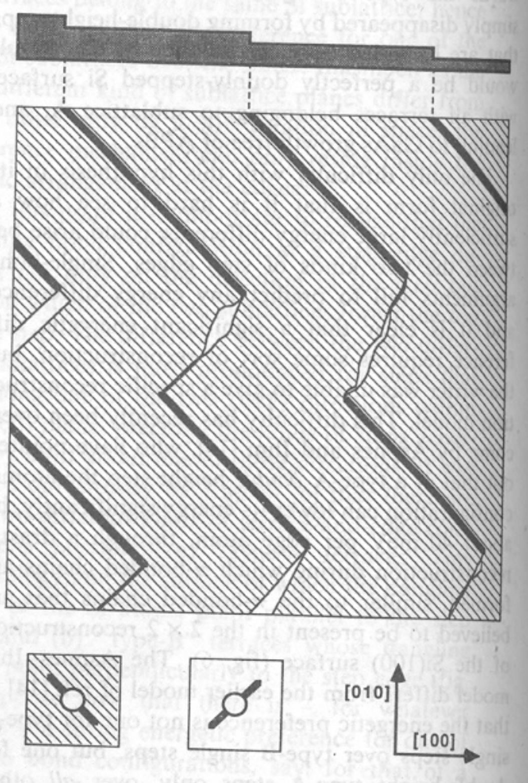

Fig. 7. Zigzag edge model for surfaces with a tilt axis deviating from ⟨011⟩. Sections of type-A double steps alternate with shorter sections that might be type-B double steps or regularly or irregularly shaped pairs of single steps.

Akiyama et al. [17] have performed GaAs-on-Si growth on lens-shaped surfaces, which present a continuum of both direction and magnitude of tilt, and they have reported that APBs occur only in a narrow band of tilt directions near a perfect [001] and [010] tilt.

A final point that requires discussion concerns the need for a high anneal temperature before nucleation. If a doubly-stepped surface is energetically preferred, then it should be possible to create such a surface by means other than a high-temperature anneal, for example by a sufficiently weak chemical etch, which might attack the Si surface only at the weak steps, but leaves the stronger steps alone, until the entire surface consists only of the more stable steps. In fact, the literature on GaAs-on-Si growth contains many reports of seemingly APB-free growth following heat treatments much gentler than those reported by Sakamoto et al. [13], suggesting that a certain amount of step doubling might already take place during the chemical polish and/or etch treatment currently employed. This is clearly a fertile field for future experimentation.

3. Interface atomic structure and neutrality

In our original discussion of the initial nucleation of GaAs on a Si(100) surface we made the simplifying assumption that the last Si plane was an unbroken plane, in which case chemical bonding arguments led to the conclusion that the first atomic plane on the GaAs side should be an unbroken As plane, as inside bulk GaAs. However, it was pointed out by Harrison, Kraut, Waldrop, and Grand (HKWG) [19] already in 1978 in the context of a GaAs/Ge interface that an atomic configuration composed of unbroken bulk planes at a polar–nonpolar (100) interface is energetically highly unfavorable [4]. This can easily be understood by arguments more adapted to our needs than those given by HKWG, as follows. Recall that Ga–Ga and As–As bonds are changed defects, each carrying a charge $\pm q/2$. The same argument applies to Ga–Si and As–Si bonds, except that the defect change is only half as large, $\pm q/4$ per bond. If the first atomic plane adjacent to the Si substrate were a perfect As plane, the two back bonds per As atom would imply a donor-like defect charge of $+q/2$ per atom, or a charge density of $+q/a^2$, where a is the lattice constant. This is a very large charge (about 3×10^{14} donors/cm^2). If not neutralized, it would support an electric field of about 4×10^7 V/cm inside the growing GaAs layer! On a macroscopic scale, this charge would of course be neutralized by mobile electrons in the conduction band. However, the neutralizing charge would extend over an appreciable distance into the semiconductor, and the field seen by the atoms at the interface itself would be almost undiminished.

As HKWG point out, such a large field would lead to massive atomic re-arrangements during the high-temperature growth itself, attempting to neutralize the interface charge. One possibility – not considered by HKWG – would be the formation of a very large concentration of negatively charged antisite defects (Ga atoms on As sites) on the GaAs side. Harrison et al. themselves propose that the re-arrangement is one of the GaAs/Si interface itself, in such a way that a significant fraction of the Si atoms is removed from the top Si layer and replaced by Ga atoms, whose back bonds to the Si substrate have the opposite charge imbalance and hence neutralize the As–Si bond charge [20]. Electrical neutrality would be reached when the number of Ga–Si bonds created in this break-up of the last Si plane equals the number of As–Si bonds, and the authors propose that the reconstruction proceeds close to this stage.

Harrison et al. consider two limiting cases of idealized atomic arrangements, both of which would restore a perfectly neutral interface, shown in fig. 8. In the first of these, all Si atoms broken out from the top Si plane (referred to as plane No. 0 in what follows) are removed from the vicinity of the interface, to the surface of the growing epilayer. Interface neutrality is then achieved when one-half of the Si atoms are replaced by either Ga or As atoms. In the second arrangement, all broken-out Si atoms remain in plane 1, the plane directly atop the original Si surface. In this case, neutrality is achieved already when one-quarter of the Si atoms are broken out. The authors point out that the first of these arrangements, while free

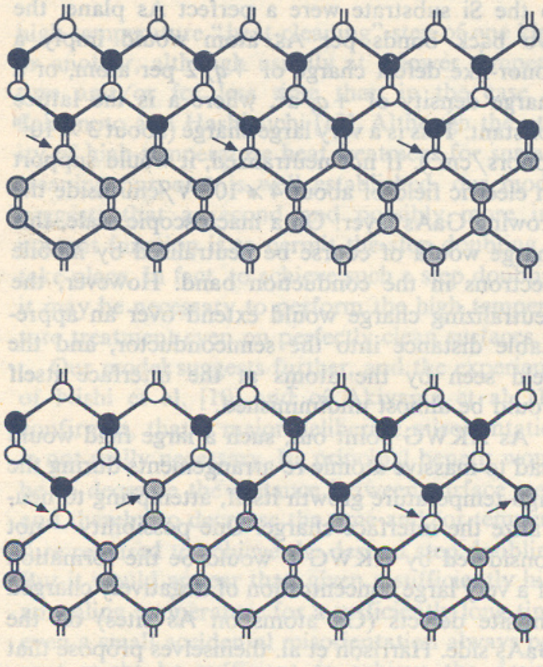

Fig. 8. Two models of atomic re-arrangement proposed by Harrison et al. [19] to achieve electrical neutrality at the interface, both involving the removal of Si atoms from the top Si layer. Top: single transition layer model, requiring removal of one-half the Si atoms, without their re-incorporation into the GaAs near the interface. Bottom: double transition layer model, requiring the removal of one-quarter of the Si atoms, and their re-incorporation into the first atomic layer on the GaAs side; this configuration has the lower energy.

of a net electric charge, still carries a residual electric dipole, whereas the second arrangement is free of both, and hence represents a state of lower overall electrostatic energy.

At this point an interesting question arises. If the HKWG re-arrangement towards an essentially neutral interface should indeed go to completion, or at least near-completion, then there would be no longer any energetic preference for this first plane after the original Si surface to be an As plane, and hence not for a specific sublattice allocation for Ga and As on the GaAs side! Which of the two sublattices would be which would then be decided by the kinetics of the nucleation process.

One would certainly expect that the As-first hypothesis remains valid under nucleation conditions under which the Si is first exposed to an As flux, permitting the formation of copious As–Si bonds, before turning on the Ga flux. But the outcome is far less clear if the initial exposure of the Si surface is to Ga rather than As, as advocated by us for the (211) orientation. The question clearly calls for an experimental answer.

The question is readily tested by the etch pattern geometry generated by anisotropic etched on deliberately misoriented (100) surfaces. The orientation of the etch pits relative to the rotation axis depends on the sublattice allocation. Such tests have been performed by Fischer et al. [8,9], using two different nucleation conditions, by depositing either an As prelayer or a Ga prelayer before the actual growth. They found that APB-free growth could be achieved in both cases, but the sublattice ordering depended on the nature of the prelayer. Evidently a sublattice switch does take place! In their most recent work [22], Fischer et al. show that the sublattice allocation also changes with nucleation temperature, under what are implied to be otherwise unchanged conditions: For nucleation at low temperatures of 450–500°C they find APB-free growth with one particular sublattice ordering. There follows a temperature range (of unstated width) inside which copious APBs are observed. Above a certain (unstated) temperature, APB-free growth is again achieved, but with a sublattice ordering opposite to that at low temperatures.

The authors interpret their observations in terms of a model in which the first atomic plane following an unbroken Si surface is either an unbroken As plane or an unbroken Ga plane, depending on nucleation conditions. They suggest that the switch from As to Ga for the first layer is simply a consequence of the loss of As by evaporation at higher temperatures.

We would not wish to rule out an unbroken As plane model for the As-dominated case, despite the electrostatic argument that speak against it. But the formation of a simple unbroken Ga plane bonded to an unbroken Si plane as a result of As loss by evaporation is extremely unlikely, on the purely chemical grounds of the very different strength of Ga–Si and As–Si bonds. In our own work on the growth of GaP on Si we found that during the thermal decomposition and desorption of a GaP film the last Ga would evaporate long before the last phosphorous [23]. One would expect the same to be true for GaAs on Si, and recent work by Bringans et al. [24] strongly supports this expectation. Hence the formation of an unbroken Ga plane bonded to an unbroken Si surface is extremely unlikely so long as there is any As present at all.

The observations of Fischer et al. [8,9,22] clearly call for a different explanation. In the next section we propose a mechanism for the Ga-dominated nucleation mode that leads to the observed final result, but from diametrically opposite initial assumptions.

4. Proposed nucleation model

4.1. Ga-dominated nucleation: the As–Si site exchange postulate

We make the following two initial postulates:

(a) Arriving Ga atoms bond to Si atoms *only* when at least one Ga–As bond can be formed along with every Ga–Si bond. The idea behind this postulate is that the 1/4 electron excess of the Ga–As bond is transferred to the Ga–Si bond, where it helps forming the latter bond.

(b) Even the formation of As–Si bonds is facilitated if at the same time Ga–Si bonds are formed, to take up the electron excess of the As–Si bond.

These postulates lead to the idea that the initial nucleation of GaAs on Si does not simply take the form of either As or – much less likely – Ga first bonding to the original Si surface, especially not in the presence of a sufficiently large non-bonded and hence mobile Ga concentration on the Si surface. Instead, we therefore make the *central postulate* that:

(c) the initially arriving As atoms will undergo an exchange reaction during which a Si atom from the last Si plane (plane 0) exchanges sites with an arriving As atom, with two Ga atoms simultaneously bonding to both the As atom and two adjacent Si atoms of the original surface, as shown in fig. 9.

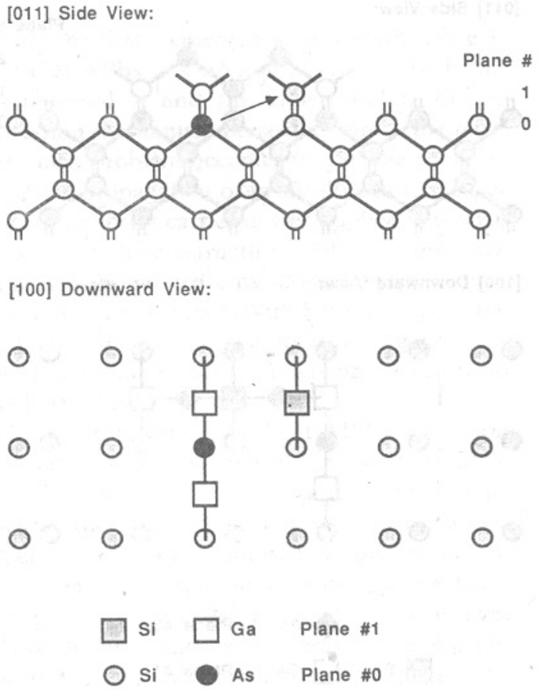

Fig. 9. Proposed first stage of the nucleation of GaAs on Si under Ga-rich conditions: an incoming As atom interchanges sites with a Si atom in plane 0 (the original top Si plane), simultaneously bonding two "waiting" Ga atoms. The ejected Si atom is placed on an adjacent site in plane 1. Top: [011] view; bottom: [100] (i.e. downward) view.

Note that in this initial nucleation step two As–Si bonds and two Ga–Si bonds are formed; hence the nucleus is an electrically neutral object in the sense of the work of HKWG. The number of Si–Si bonds does not change during the site exchange.

If we make the plausible assumption that the Si atom ejected from plane 0 is simply placed on one of the adjacent sites in plane 1, the site in plane 2 that has back bonds to both this Si atom and to the adjacent Ga atom, will form a natural bonding site for another As atom, and an As–Ga pair will connect between the remaining bond of the Si atom in plane 1 and a Si atom in plane 0, as shown in fig. 10. During this second nucleation step again two As–Si and two Ga–Si bonds are formed.

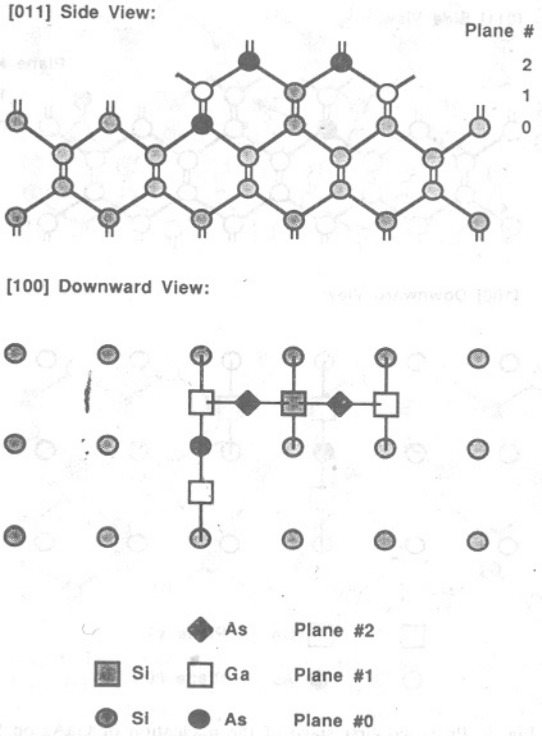

Fig. 10. Proposed second stage of the nucleation of GaAs on Si under Ga-rich conditions: two As atoms bond to the Si atom in plane 1, and a Ga atom closes the bonding loop of the outer of the two As atoms to the Si surface.

It is evident that this nucleation process automatically leads to a placement of all Ga atoms into plane 1 and of all As atoms into planes 0 and 2, the opposite choice from what would be present if the As–Si site exchange did not take place, but exactly the order observed by Fischer et al. under nucleation conditions of sufficiently high temperature and Ga flux. There can be little doubt that any subsequent lateral spreading of each nucleus thus formed will retain this sublattice order. This spreading can proceed either by adding As–Ga pairs, or by adding two As atoms and performing an As–Si site exchange. Because the two processes lead to opposite bond charges, a combination of both may be expected to take place, to aid in the establishment of electrical neutrality, but it is doubtful that perfect neutralization will take place.

We note however that, if perfect neutralization *did* take place, the resulting composition of the different planes adjacent to the interface would be exactly what is demanded by the HKWG 2-layer reconstruction model.

4.2. The As-dominated case

The experiments of Fischer et al. [8,9,22] make it clear that under lower-temperature As-dominated conditions the As–Si site exchange postulated for Ga-dominated nucleation does not occur. The reason for this is probably one or both of the following: either the presence of an excess of Ga may be necessary to drive the exchange, or the temperature is too low for the exchange reaction to overcome an energetic reaction barrier that is very likely present, for example the As_2 dissociation barrier. Quite possibly both may play a role.

Nor is it clear whether or not the HKWG interface reconstruction process takes place at the lower temperatures and, if not, exactly how electrical neutrality is subsequently established – if indeed it is. Quite possibly the HKWG mechanism occurs even then, but with the Si atoms ejected from plane 0 now replaced by Ga rather than As atoms, because of the prior formation of a tightly-bonded partial As coverage in plane 1. An alternate possibility would be that the top Si plane remains intact, but is neutralized by the formation of a very high concentration of Ga-on-As site antisite defects, as mentioned earlier. Or maybe the interface remains highly charged, being neutralized only by mobile electrons in the conduction band. This is evidently another fertile field for future research.

4.3. Residual defects

The high-temperature Ga-dominated nucleation model presented above is an idealization. In practice, one must expect numerous defects to occur, especially the following three:

(a) Occasionally, As atoms may end up in plane 1, by bonding to Si plane 0 without undergoing the site exchange reaction. So long as these atoms remain in the minority, they would simply appear as local antisite point defects (double donors) on what is otherwise a Ga sublattice, without causing actual finite-size antiphase domains.

(b) As mentioned already, in those areas of the interface where the growth proceeds by lateral spreading from the initial nuclei, the neutralization of As–Si bonds by Ga–Si bonds formed may be incomplete, leading to a net doping of the interface, which may be either donor-like or acceptor-like, depending on the exact details of the process.

(c) Finally, some of the Si atoms removed from plane 0 might not be incorporated into plane 1, but are taken up by the growing GaAs bulk instead. On purely thermodynamic grounds one should expect at least the GaAs layers closest to the interface to be Si doped to the thermodynamic solubility limit.

Additional Si atoms may accumulate on the GaAs surface, from where they are gradually incorporated into the growing GaAs as bulk dopant.

Because of these various kinds of charged defects, the GaAs-on-Si(100) interface will almost certainly be one with a residual interface charge sufficiently large that it cannot be ignored for device purposes, combined with a heavy Si doping of at least the near-interface region of the crystal. The extent to which these two effects will take place will depend strongly on details of the exact growth procedure.

Up to a point, these defects are largely inconsequential, so long as they remain confined to within a few atomic monolayers of the original interface, and do not have a deleterious effect on the quality of subsequent layers. The defect structure near the interface is likely to be dominated by the very large density of misfit dislocations there, compared to which the other defects are a comparatively minor disturbance. Largely because of the misfit dislocations, the GaAs/Si interface itself is not likely to be usable as a part of the "intrinsic" device for most devices under current consideration, and its short-range properties are therefore not of primary concern. Apart from the dislocations, the defect making itself felt farthest from the interface is probably Si uptake by the growing GaAs, and its suppression probably deserves the highest priority after the suppression of the propagation of misfit dislocations.

5. Conclusions

With the 1985 emergence of convincing evidence that APBs in GaAs-on-Si (100) growth can be suppressed – and the understanding of the suppression mechanism presented here – the most urgent next problem becomes that of the suppression of the propagation of misfit dislocation, especially for minority carrier device applications such as lasers. For heterostructure FETs, the interface charge and the related cross-doping problem may be of similar importance. With both problems, the long-term issue is not just to make "good" devices, but to do so without having to resort to thick buffer layers.

The achievement of excellent APB suppression on the (100) orientation makes the switch to (211), long advocated by us [1–4], a less urgent one. However, the (211) orientation not only remains a perfectly viable option, but we continue, in fact, to believe that the *long-term* potential of the (211) orientation is still the better one for many devices. Most of the nucleation and interface charge problems discussed in sections 2–4 of this paper pertained almost exclusively to the problems of the (100) orientation, and are absent on the (211) orientation. Not only is there no problem about the sublattice allocations during nucleation, there also is no natural interface charge: on a perfect (211) surface, the numbers of As–Si and Ga–Si bonds are exactly the same (see fig. 4), leading naturally to an electrically neutral interface. The lack of a charge implies the lack of an electric field driving any atomic re-arrangement; hence any takeup of Si by the growing GaAs should also be much weaker. This has indeed been observed: In their work on the MBE growth of GaP on Si, Wright et al. [2] found that the uptake of Si by a (211) layer was much less than that by a (100) layer grown side by side. All of these could be sizeable advantages of the (211) orientation over the (100) orientation for applications in which the near-interface quality, and especially a low near interface impurity uptake of the GaAs are important.

Probably the biggest unknown in the (100) versus (211) competition is the behavior of misfit dislocations on the two orientations. Nothing is known yet about differences between the two orientations in this regard, but the differences could be severe, and they could decide the issue.

Acknowledgements

The writer gratefully acknowledges contributions from many individuals. Dr. B.A. Joyce (Philips) was a discussion partner during the early phases of this work. Dr. W.I. Wang (IBM) was the first to try to persuade me that APB-free growth could indeed be achieved on mis-oriented substrates, followed by Professor H. Morkoc and Dr. R.J. Fischer (University of Illinois). Drs. M. Akiyama and S. Nishi (Oki Electric Co.) and especially Dr. T. Sakamoto (ETL-Tsukuba) completed that persuasion. To Professor Morkoc and Dr. Fischer additional thanks are due for intense discussions, for providing a wafer to test for freedom from APBs, and for making valuable information available prior to publication. Dr. D.E. Aspnes (Bellcore) resolved many puzzles with a preprint of his work. My former co-worker, Dr. P.N. Uppal, now at Martin Marietta, participated in many discussions; he also performed the tests for APBs on the Illinois wafer. Dr. E.A. Kraut (Rockwell) served as a patient sounding board throughout this work. Last but not least, much gratitude is due to the US Army Research Office for supporting this work.

References

[1] S.L. Wright, M. Inada and H. Kroemer, J. Vacuum Sci. Technol. 21 (1982) 534.

[2] S.L. Wright, H. Kroemer and M. Inada, J. Appl. Phys. 55 (1984) 2916.

[3] P.N. Uppal and H. Kroemer, J. Appl. Phys. 58 (1985) 2195.

[4] For a review, see also: H. Kroemer, Surface Sci. 123 (1983) 543.

[5] M. Henzler and J. Clabes, in: Proc. 2nd Intern. Conf. on Solid Surfaces, Kyoto, 1974 [Japan. J. Appl. Phys. Suppl. 2, Part 2 (1974) 389].

[6] R. Kaplan, Surface Sci. 93 (1980) 145.

[7] For extensive references see Uppal and Kroemer [3].

[8] R.J. Fischer, N.C. Chand, W.F. Kopp, H. Morkoc, L.P. Erickson and R. Youngman, Appl. Phys. Letters 47 (1985) 397.

[9] R.J. Fischer, N. Chand, W.F. Kopp, C.-K. Peng, H. Morkoc, K.R. Gleason and D. Scheitlin, IEEE Trans. Electron Devices ED-33 (1986) 206.

[10] S. Nishi, H. Inomata, M. Akiyama and K. Kaminishi, Japan. J. Appl. Phys. 24 (1985) L391.

[11] M. Akiyama, Y. Kawarada and K. Kaminishi, Japan. J. Appl. Phys. 23 (1984) L843.

[12] W.I. Wang, Appl. Phys. Letters 44 (1984) 1149.

[13] T. Sakamoto and G. Hashiguchi, Japan. J. Appl. Phys. 25 (1986) L57.

[14] H. Kroemer, in: Heteroepitaxy on Si Technology, Materials Research Society Proc. 1986 Spring Meeting, in the press.

[15] In ref. [6], Kaplan himself writes: "On relatively low step density, i.e. accurately cut (100) crystals, regions terminating on the different sublattices occur with equal probability. This should be true also of high step density vicinal surfaces, *unless the surface energy is highly sensitive to the dangling bond configuration relative to the steps*" [emphasis mine]. Kaplan did not follow up this remark any further; his data show that type-A terraces *must* be present, but he fits his LEED data to a model that assumes both kinds of terraces to be present, without stating whether the data could be fitted just as well or better by an "A-only" model. In ref. [14] the present writer argues that one could have concluded already from Kaplan's data that there must be a strong energetic preference for type-A steps present.

[16] D.E. Aspnes and J. Ihm, to be published.

[17] M. Akiyama, K. Kawarada, S. Nishi and K. Kaminishi, 1986 Spring Meeting, Materials Research Society.

[18] S. Sakai, T. Soga, M. Takeyasu and M. Umeno, 1986 Spring Meeting, Materials Research Society.

[19] W.A. Harrison, E.A. Kraut, J.R. Waldrop and R.W. Grant, Phys. Rev. B18 (1978) 4402.

[20] Our discussion here and in the rest of this section goes beyond HKWG in specifically assuming that the first atomic plane above the original Si surface is, at least initially, an As plane. The arguments in Harrison et al., apart from discussing GaAs-on-Ge rather than GaAs-on-Si, apply independently of which of the two sublattices on the GaAs side is which, so long as the two different kinds of atoms occur in alternating planes.

[21] R.J. Fischer, W.T. Masselink, J. Klem, T. Henderson, T.C. McGlinn, M.V. Klein, H. Morkoc, J. Mazur and J. Washburn, J. Appl. Phys. 45 (1985) 374.

[22] R. Fischer, H. Morkoc, C. Choi, N. Otsuka, M. Longerbone and L.P. Erickson, to be published.

[23] S.L. Wright, PhD Dissertation, University of California, Santa Barbara, CA (1982), unpublished.

[24] R.D. Bringans, R.I.G. Uhrberg, M.A. Olmstead, R.Z. Bachrach and J.E. Northrup, 1986 Spring Meeting, Materials Research Society.