Clean and As-Covered Zinc-Blende GaN (001) Surfaces: Novel Surface Structures and Surfactant Behavior

Jörg Neugebauer, Tosja Zywietz, and Matthias Scheffler

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany

John E. Northrup and Chris G. Van de Walle

Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304

We have investigated clean and As-covered zinc-blende GaN (001) surfaces, employing first-principles total-energy calculations. For clean GaN surfaces our results reveal a novel surface structure very different from the well-established dimer structures commonly observed on polar III-V (001) surfaces: The energetically most stable surface is achieved by a Peierls distortion of the truncated (1×1) surface rather than through addition or removal of atoms. This surface exhibits a (1×4) reconstruction consisting of *linear Ga tetramers*. Furthermore, we find that a submonolayer of arsenic significantly lowers the surface energy indicating that As may be a good surfactant. Analyzing surface energies and band structures we identify the mechanisms which govern these unusual structures and discuss how they might affect growth properties. [S0031-9007(98)05722-6]

Its wide direct band gap and strong chemical bonds render GaN an ideal material for optoelectronic devices in the blue/UV region of the optical spectrum. Recently, the fabrication of highly efficient blue light emitting diodes [1] and prototypes of a blue laser have been reported [2]. However, despite progress in device fabrication an understanding of the fundamental growth mechanisms is still in its infancy, and even the atomic structure of the surface is not well understood. Only recently atomically resolved scanning tunneling micrographs have been obtained for wurtzite GaN surfaces [3]. A detailed knowledge of the properties and structure of these surfaces is crucial to improve growth in a systematic way.

The stable crystal phase of GaN is the wurtzite structure. However, cubic (zinc-blende) GaN can be grown epitaxially on cubic SiC or GaAs. Cubic GaN exhibits a number of properties very appealing for device applications: It has a lower band gap than the wurtzite phase (by 0.2 eV) and can be easily cleaved. Cubic GaN has been grown, e.g., on cubic GaAs (001) substrates [4–6]. Growing on this substrate Brandt *et al.* observed in going from N-rich conditions to Ga-rich conditions a reversible sequence of reconstructions exhibiting (1×1), (2×2), and $c(2 \times 2)$ reflection high energy electron diffraction (RHEED) patterns [4]. The $c(2 \times 2)$ and (2×2) reconstructions have also been reported by other groups [6]. Scanning tunneling microscopy (STM) measurements further revealed that the (2×2) structure contains one dimer per surface cell, but the chemical nature of the dimer could not be resolved [7]. However, recently Feuillet *et al.* [8] observed a (1×4) (N-rich) and a (1×1) (Ga-rich) reconstruction for GaN (001) grown on cubic SiC. Only when exposing these surfaces to an As background pressure the two surface reconstructions commonly found for GaN on GaAs [(2×2) and $c(2 \times 2)$] were observed.

Based on these results it appears that the GaN (001) (2×2) and $c(2 \times 2)$ structures obtained in growth on GaAs substrates are stabilized by As adsorption or segregation, but that the (1×4) is an intrinsic reconstruction of the clean surface [8].

In this Letter we address these open questions, which are relevant for understanding the surface properties and growth mechanisms, by detailed density functional theory calculations. We have performed calculations for an extensive set of possible structures for clean and As-covered GaN (001) surfaces. On the basis of these results we conclude that the (1×4) structure is an intrinsic reconstruction of GaN (001) consisting of linear Ga tetramers. The (2×2) structure observed for GaN grown on GaAs (001) cannot be understood as an intrinsic reconstruction. Instead, we propose that $1/2$ monolayer of As-As dimers gives rise to this phase.

Let us first start with a few general remarks concerning cubic (001) semiconductor surfaces. At the ideal abrupt III-V (001) surface each surface atom has two dangling bonds. For cations (anions) each dangling bond is occupied with $3/4$ $(5/4)$ electrons. The high density of partially occupied surface dangling bonds renders such a surface very unstable and surface reconstruction occurs. Surface reconstructions on semiconductor surfaces are commonly driven by the following: (i) reducing the number of dangling bonds [on conventional (001) surfaces this is realized by dimer and missing dimer formation], (ii) minimizing the electronic energy (this is commonly formulated as the electron counting rule; all energetically low-lying anion dangling-bond states are doubly occupied, all cation dangling-bond states are empty), and (iii) minimizing the electrostatic energy by optimizing the arrangement of the charged surface atoms. Despite their simplicity these rules have been very successful in explaining the

structure of polar and nonpolar surfaces for a wide variety of semiconductors.

In accordance with rule (i) common (001) surface reconstructions involve a combination of dimers and missing dimers. The only exception to this rule has been reported for the InP (001) surface where trimer units have been observed by STM [9]. The stability of dimers can be understood noting the specific topology of (001) surfaces. Each surface atom is twofold coordinated and has two back bonds. Thus, dimers can be formed simply by *rotating* the surface atoms. *Stretching* or *breaking* of back bonds (which is energetically unfavorable) is not required. In this Letter we will show that the energetically most favorable structure for GaN (001) does *not* consist of dimers. A new and hitherto never reported structure consisting of linear Ga tetramers is found to be most stable. We explain this unexpected behavior of GaN surfaces in terms of the large atomic mismatch between Ga and N and the very different cohesive and binding energies of the two species.

We have performed calculations of the total energy and atomic structure employing the local-density approximation and the first-principles pseudopotential approach. The relative stability of possible surface structures has been determined within the thermodynamically allowed range of the Ga-chemical potential: $\mu_{Ga(bulk)} - \Delta H_{GaN} < \mu_{Ga} < \mu_{Ga(bulk)}$. ΔH_{GaN} is the formation enthalpy of bulk GaN with respect to Ga bulk and N_2 molecules. The calculations have been performed with the Ga $3d$ electrons treated explicitly as valence electrons and with a plane wave cutoff of 60 Ry. Convergence checks showed that treating the Ga $3d$ electrons as valence states is crucial for calculating surface energies. Describing the $3d$ electrons in the nonlinear core correction (NLCC) which would be computationally much less demanding is not sufficient: A detailed analysis showed that the NLCC approximation for Ga systematically overestimates the strength of Ga-Ga bonding relative to Ga-N bonding. This explains, e.g., why the NLCC approximation severely underestimates the formation enthalpy ΔH_{GaN} of GaN [0.5 eV compared to 0.95 eV ($3d$ included) and 1.1 eV experiment]. For details of the method we refer to Refs. [10–13].

The surfaces were modeled by repeated slabs with two equivalent surfaces on each side. The slabs consisted of 9 layers of GaN. Tests performed with slabs containing 15 layers showed that the 9-layer slabs are adequate. Relaxation of the atoms in the top two layers of each side of the slab was found to be sufficient. A (4×4) Monkhorst-Pack mesh was used to sample the Brillouin zone [14].

We will at first focus on the so-called "ideal" surface geometry, the unreconstructed (1×1) structure. The calculated surface energies per (1×1) cell are shown in Fig. 1. Also included are the calculated surface energies for GaAs, as reference to a "traditional" semiconductor. Note that for GaAs *both* the Ga and As terminated (1×1) surfaces are energetically highly unfavorable with the Ga-terminated
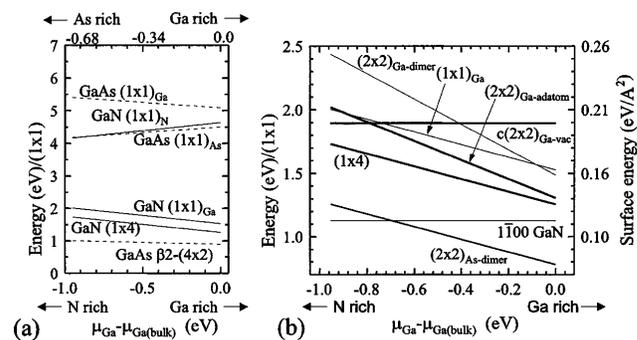
FIG. 1. Surface energies for (a) (1×1) GaN (solid lines) and GaAs (dashed lines) surfaces and (b) for clean and As-covered reconstructed GaN surfaces as a function of the Ga chemical potential μ_{Ga}. Only the thermodynamically allowed range is shown. The surface structures are described in the text. In (a) also the structures with the lowest surface energy have been included: for GaAs the $\beta2 - (4 \times 2)$ and for GaN the (1×4) tetramer structure [see Fig. 3(a)]. Also included is the surface energy of the nonpolar GaN $(1\bar{1}00)$ from Ref. [10]. Note that the thermodynamically allowed range in (a) is different for GaAs (top axis) and GaN (bottom axis).

surface slightly higher in energy than the As-terminated surface. Compared to the energetically favored $\beta2(4 \times 2)$ GaAs surface [15] the unreconstructed structures are about 5 times higher in energy [see Fig. 1(a)]. For GaN we find a strikingly different behavior: While the N-terminated surface is close in energy to the As-terminated GaAs surface the Ga-terminated surface is more than 3.5 eV per (1×1) cell lower in energy than the corresponding GaAs surface. In fact, as will be shown below the energetically most stable GaN (1×4) surface is only slightly more stable [by ≈ 0.27 eV per (1×1) cell] than the ideal Ga-terminated GaN surface.

The atomic geometry for the unreconstructed Ga-terminated surface is characterized by small vertical relaxations: The top layer (Ga) relaxes 0.08 Å outward; the relaxation of the second layer (N) is already negligible. Surface relaxation can thus be excluded as a possible origin for the unexpectedly large stability of this surface.

The calculated electronic band structure for this surface is shown in Fig. 2(a). For the unreconstructed surface our calculations show two surface states (S_1, S_2). The energetically lower surface state is a bonding state between neighboring surface Ga atoms. The Ga-Ga bonds are *parallel* to the $[1\bar{1}0]$ direction explaining the large dispersion along the $\overline{\Gamma J'}$ axis. The lower surface state (S_1) is partially occupied with $3/2$ electrons rendering this surface metallic. A comparison with the corresponding GaAs surface—the unreconstructed Ga-terminated (001) surface— shows a qualitatively similar band structure. The main difference is a significantly larger dispersion (more than 1 eV) of the surface bands on the GaN surface reflecting the above mentioned formation of Ga-Ga bonds. The origin of the large interaction lies in the sizable mismatch of the covalent radii of Ga and N. Because of the small
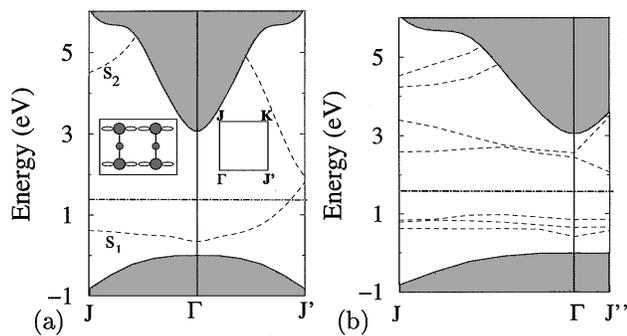
FIG. 2. Band structure calculated within the local density approximation for the relaxed Ga-terminated (a) (1×1) and (b) (1×4) surface of GaN. The shaded region corresponds to the bulk projected band structure for zinc-blende GaN. The dashed lines are surface states. In (a) the energetically lower surface state (S_1) is partially occupied with $3/2$ electrons. The upper surface state (S_2) is empty. The dot-dashed line marks the position of the Fermi energy. In (b) the lower three surface states are doubly occupied; the upper surface states are empty. $\overline{\Gamma J'}$ is parallel to the $[1\bar{1}0]$ direction.

FIG. 3. Top view of the various surfaces: (a) Ga-terminated (1×4), (b) $(2 \times 2)_{Ga-dimer}$, and (c) $(2 \times 2)_{As-dimer}$. The numbers give the distance between atoms in Å. As reference: The nearest and next nearest neighbor distances in GaN are 1.95 and 3.17 Å, respectively.

radius of the N atoms the Ga atoms in GaN have approximately the same distance as in Ga bulk [16]. The Ga atoms at the surface can form metallic bonds similar to those in bulk Ga even *without* any relaxation, thus stabilizing the Ga-terminated surface.

However, a detailed analysis of the energies showed that the formation of Ga bonds on the surface is not sufficient to explain the exceptionally low surface energy. A second mechanism stabilizing Ga-terminated surfaces with respect to N-terminated surfaces is the very different cohesive energies of the bulk phases of Ga and N: The cohesive energy of bulk Ga is 2.81 eV/atom while that of the N_2 molecule is 5.0 eV/atom. (The N-N bond in the N_2 molecule is one of the strongest bonds found in Nature.) In contrast the cohesive energy of bulk As (2.96 eV/atom) is only slightly larger than that of bulk Ga. Because of this asymmetry, more energy is required to transfer N atoms from the N reservoir to the surface than to transfer Ga atoms to the surface.

The strong dispersion along the $\overline{\Gamma J'}$ direction and the metallic character suggest that the Ga-terminated (1×1) surface might be Peierls unstable along the $[1\bar{1}0]$ direction. Electron counting considerations indicate that the smallest unit cell which allows an energy gap is a (1×4) structure. In fact, our calculations show that the (1×1) surface is unstable against formation of a (1×4) reconstruction. The reconstructed surface is semi-insulating with a band gap of 1.2 eV [see Fig. 2(b)]. The atomic geometry is characterized by *linear Ga tetramers* [see Fig. 3(a)]. The energy gain is 1.1 eV per (1×4) cell compared to the unreconstructed surface. The three bonds in the tetramer give rise to three almost dispersionless surface states close to the valence band which are each doubly occupied. The two dangling-bond orbitals [see Fig. 3(a)] are unoccupied and give rise to the two upper surface states [Fig. 2(b)].
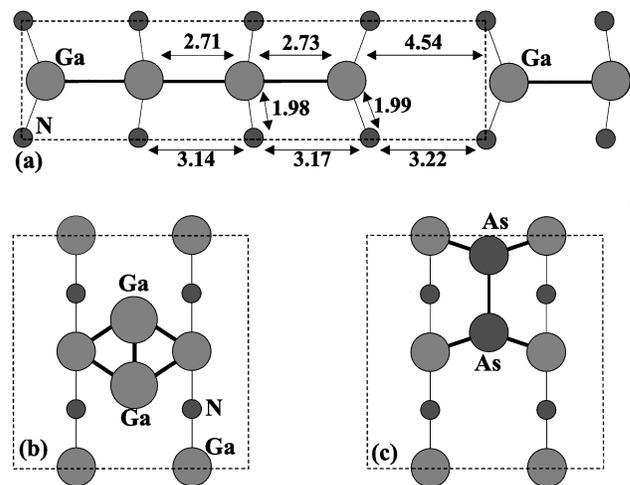
A surface consisting of linear tetramers is unique and has neither been observed experimentally nor studied theoretically for other III-V (001) surfaces. Our calculations show indeed that on the corresponding GaAs surface the tetramers are unstable and spontaneously form dimers. The very different stability of tetramers on the two materials originates again from the sizable mismatch of the covalent radii of Ga and N. In order to form a tetramer on GaN breaking of back bonds is not required and the Ga-N bond length remains almost unchanged [<0.1 Å; see Fig. 3(a)]. On GaAs, however, forming a tetramer is not possible without breaking back bonds.

In order to identify the experimentally observed reconstructions we have studied a large number of possible configurations with different surface stoichiometries and starting from very different initial geometries and symmetries. These calculations will be described elsewhere in more detail [17]. Initially we focused on structures which exhibit low surface energies on other III-V (001) surfaces and which obey the electron counting rule. Examples are the $c(2 \times 2)$ Ga and N vacancy structure and the $\beta2(4 \times 2)$ structure. However, all these surfaces are energetically less favorable than the Ga-terminated (1×4) structure.

We then focused on the experimentally observed (2×2) periodicity. STM measurements by Wassermeier *et al.* indicate that the surface consists of *one* dimer per (2×2) cell. During the initial stages of our investigations we focused on clean GaN surfaces, but discovered that all possible dimer structures are energetically higher in energy than the tetramer structure. Starting with N dimers resulted always in the formation of N_2 molecules which are bound by less than 50 meV, i.e., in a physisorbed state. Ga dimers are energetically most stable on a Ga-terminated surface [the geometry is shown in

Fig. 3(b)]; however, the energy gain is too small to make them favorable [see Fig. 1(b)]. We performed these calculations not only for the standard dimer geometry but also for rotated and translated dimers. On the basis of these results we conclude that there is no stable (2×2) dimer structure on *clean* GaN (001) surfaces. We emphasize, however, that the tetramer structure found here to be stable is a strong candidate to explain the (1×4) RHEED pattern observed by Feuillet *et al.* [8].

Not finding a stable (2×2) surface and in view of recent experiments [8] we studied the influence of arsenic on the surface composition and structure. We find that the energetically most stable structure is a (2×2) structure with an arsenic dimer on a Ga-terminated surface. Each As atom has two back bonds to the underlying Ga atoms [see Fig. 3(c)]. The As-As dimer length is 2.46 Å. Other structures with As in the second or third surface layer, with mixed Ga-As bonds, and with higher coverages of As atoms have been found to be energetically less stable. The surface energy of the (2×2) As dimer structure is plotted in Fig. 1(b). Though arsenic-poor conditions $(\mu_{As} = \mu_{As(bulk)} - \Delta H_{GaAs})$ have been assumed, the As-covered surface is energetically clearly more favorable than the clean GaN (001) surfaces; i.e., removing As from GaAs and putting it on GaN is an *exothermic* reaction. The only assumption made here is that the system is in thermodynamic equilibrium with GaAs. Considering the comparatively large As vapor pressure of GaAs at typical GaN growth temperatures this should be the case if growing on GaAs substrate or in an As-contaminated growth chamber. We thus conclude that the chemisorption of As atoms can significantly reduce the surface energy.

Our results further show that structures with As in the second or deeper surface layer are energetically less favorable than structures where As stays in the top surface layer. This result is consistent with previous experimental [18] and theoretical studies [19] according to which As is virtually immiscible in GaN. Both results, the reduction in surface energy and the low miscibility, suggest that As might be a good surfactant for the growth of zinc-blende GaN. Generally, we expect also for other group V elements a similar behavior. In particular, if As incorporation should turn out to be a problem using elements with an atomic radius *larger* than As (e.g., Sb, Bi) would be an interesting alternative.

Our results for clean and As-covered surfaces allow some interesting conclusions concerning the growth of GaN (001). First, both for clean and As-covered GaN the surface energy drops if going from N-rich conditions toward Ga-rich conditions. The reason is that on both surfaces the number of Ga atoms exceeds the number of N atoms: The stable surfaces are always Ga rich. This behavior is very different from what has been found for the nonpolar surfaces which are stoichiometric except for extremely Ga-rich conditions [11]. Second, GaN (001)

has a significantly higher surface energy than the nonpolar GaN surfaces implying that the polar (001) surface is less stable. The difference in surface energies *increases* and the stability of the (001) surface *decreases* when going towards more N-rich conditions. This behavior might explain why the growth morphology of cubic GaN severely degrades when growing under N-rich conditions. It also casts severe doubts on the old paradigm of pushing growth towards extreme N-rich conditions in order to avoid a N deficiency. Instead, from a thermodynamic point of view we expect best growth conditions under more Ga-rich conditions where the surface energy is low. Finally it should be mentioned that the large difference in atomic radii between Ga and N atoms and the extreme binding energy of N_2, which drive the formation of Ga-rich surface structures, is a general feature of the III-nitride semiconductors. For AlN, InN, and its alloys with GaN we expect therefore a similar behavior.

[1] S. Nakamura, T. Mukai, and M. Senoh, Appl. Phys. Lett. **64**, 1687 (1994).
[2] S. Nakamura *et al.,* Appl. Phys. Lett. **70**, 868 (1997).
[3] A. R. Smith *et al.,* Phys. Rev. Lett. **79**, 3934 (1997).
[4] O. Brandt *et al.,* Phys. Rev. B **52**, R2253 (1995).
[5] H. Yang *et al.,* Appl. Phys. Lett. **68**, 244 (1996).
[6] D. Schikora *et al.,* Phys. Rev. B **54**, R8381 (1996).
[7] M. Wassermeier and O. Brandt (private communication).
[8] G. Feuillet *et al.,* Appl. Phys. Lett. **70**, 1025 (1997).
[9] C. D. MacPherson, R. A. Wolkow, C. E. J. Mitchell, and A. B. McLean, Phys. Rev. Lett. **77**, 691 (1996).
[10] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).
[11] J. E. Northrup and J. Neugebauer, Phys. Rev. B **53**, R10477 (1996).
[12] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997).
[13] J. Neugebauer and C. G. Van de Walle, Proc. Mater. Res. Soc. Symp. **408**, 43 (1995).
[14] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[15] J. E. Northrup and S. Froyen, Phys. Rev. Lett. **71**, 227 (1993).
[16] J. Neugebauer and C. G. Van de Walle, Proc. Mater. Res. Soc. Symp. **339**, 687 (1994).
[17] J. Neugebauer *et al.* (to be published).
[18] M. Weyers, M. Sato, and H. Ando, Jpn. J. Appl. Phys. **31**, L853 (1992).
[19] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **51**, 10 568 (1995).