RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 50, NUMBER 11 15 SEPTEMBER 1994-I

Atomic geometry and electronic structure of native defects in GaN

Jörg Neugebauer and Chris G. Van de Walle
Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304

(Received 25 May 1994; revised manuscript received 29 June 1994)

We have studied the electronic structure, atomic geometry, and formation energies of native defects in GaN using first-principles total-energy calculations. Our results reveal the vacancies to be the dominant defects in GaN, whereas antisites and interstitials are energetically less favorable. In p-type GaN the nitrogen vacancy (a donor) has the lowest formation energy, in n-type GaN the gallium vacancy (an acceptor). Our results show that the vacancies may be important for compensation. However, isolated point defects and particularly the nitrogen vacancy can be excluded as the source for the n-type conductivity in as-grown GaN, contrary to the generally accepted picture.

The recent rapid increase in research on GaN has been stimulated even more by the successful fabrication of the first highly efficient blue-light-emitting diodes.[1] Despite their importance for future device applications little is known about intrinsic defects and their properties in GaN. As-grown GaN is usually n-type, and p-type conductivity is difficult to obtain.[2,3] For 25 years the n-type conductivity has been commonly associated with the nitrogen vacancy.[4,5] This assignment was consistent with recent tight-binding calculations by Jenkins and Dow.[6,7] In the same work defect levels of other native defects (gallium vacancy and antisites) were also calculated. Experimentally observed defect levels in the band gap of GaN have frequently been assigned to such theoretically predicted energy levels.[8] However, the tight-binding calculations have limited accuracy, and in addition give no information about atomic defect geometry (relaxation, reconstruction, etc.) and defect formation energies. The formation energy of a defect determines whether a defect can actually be formed and in which concentrations.

The aim of the present work is to calculate electronic structure *and* defect formation energies for all native defects in all relevant charge states. Atomic relaxation is fully taken into account. Based on our first-principles calculations we discuss what role native defects play in GaN and whether the nitrogen vacancy is involved in the n-type conductivity of as-grown GaN. Our analysis further reveals several defect properties strikingly different from more traditional semiconductors like Si, GaAs, or ZnSe; these properties will be explained in terms of the large mismatch between the covalent radii of Ga and N.

For calculating atomic geometry, formation energies, and electronic structure we have performed first-principles total-energy calculations based on density-functional theory using a supercell approach with 32 atoms per cell, a plane-wave basis set with 60-Ry cutoff and soft Troullier-Martins pseudopotentials.[9] Details of the method and convergency checks can be found elsewhere.[10,11] In Ref. 12 we showed that the Ga $3d$ electrons are important for the bonding properties in GaN and may change defect formation energies up to 1.5 eV. However, to reduce the computational effort we included the $3d$ electrons explicitly only for the defects with formation energies below 3 eV (i.e., for the vacancies). For all other defects the formation energies are so high that even 2-eV corrections would not affect the conclusion that the defects will not occur in any appreciable concentration. For these cases the $3d$ electrons are therefore treated using the nonlinear core correction.[13]

GaN can be grown in two phases, the cubic and wurtzite phase. We focus here on the wurtzite phase which is the stable one for bulk GaN. A comparison of defect properties in both phases was given in Ref. 12 where we found only minor differences in the electronic structure and formation energies. These differences were explained in terms of the lower symmetry of the wurtzite phase.

The formation energy of a defect in a charge state q is given by

$$E^f(q)=E^{\text{tot}}(q)-n_{\text{Ga}}\mu_{\text{Ga}}-n_{\text{N}}\mu_{\text{N}}-qE_F, \tag{1}$$

where n_{Ga} and n_{N} are the number of Ga and N atoms, μ_{Ga} and μ_{N} are the chemical potentials, and E_F is the Fermi energy. The total energy $E^{\text{tot}}(q)$ for a given charge state can be directly calculated. The chemical potentials, however, depend on the experimental conditions under which the material is grown. The chemical potentials for Ga and N are not independent, since both species are in equilibrium with GaN: $\mu_{\text{Ga}}+\mu_{\text{N}}=\mu_{\text{GaN}}$. Thus, only one free parameter remains, which characterizes the stoichiometry of the system. The chemical potentials are subject to certain boundary conditions: $\mu_{\text{N}}<\mu_{\text{N}_2}$ and $\mu_{\text{Ga}}<\mu_{\text{Ga(bulk)}}$. If these conditions are not satisfied, the system would be thermodynamically unstable and segregate into different components.[14]

Figure 1 summarizes all the defect formation energies as a function of the Fermi energy for the nitrogen-rich case ($\mu_{\text{N}}=\mu_{\text{N}_2}$). The slopes of the defect formation energies characterize the charge state; a change in the slope indicates a transition from one charge state to another.

The most striking feature in Fig. 1 is the low formation energy for the vacancies. Under p-type conditions the nitrogen vacancy has the lowest formation energy, while under n-type conditions the gallium vacancy is lowest. Note that these low energies are due to the fact that the vacancies can become charged, thereby significantly lowering their formation energy. Antisites and interstitials are clearly energetically less favorable. This behavior renders GaN different

0163-1829/94/50(11)/8067(4)/$06.00

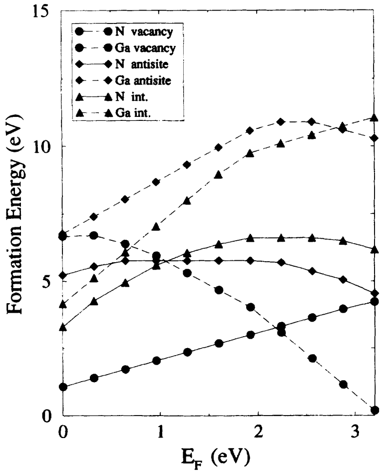

FIG. 1. Defect formation energies as a function of the Fermi level for all native defects in GaN under nitrogen-rich conditions $(\mu_N=\mu_{N_2})$. $E_F=0$ corresponds to the top of the valence band.

from more traditional semiconductors like Si, GaAs, or ZnSe where antisites and interstitials play an important role.

The key to understanding this difference lies in a specific property of GaN: a large mismatch in the covalent radii of gallium ($r_{Ga}=1.26$ Å) and nitrogen ($r_N=0.75$ Å). We investigate the consequences of this mismatch on defect formation energies for the example of the gallium antisite, where a particularly strong effect is observed. By replacing a nitrogen atom with a gallium atom the Ga-N bonds are replaced by Ga-Ga bonds. Without atomic relaxation the Ga-Ga bonds are 39% too small compared to Ga bulk. If atomic relaxation is allowed, the four neighboring Ga atoms move outward by ≈ 0.2 Å, reducing the strain in the Ga-Ga bond. The atomic relaxation is accompanied by an energy gain of 4.2 eV which is about 3 times larger than typical relaxation energies for defects in GaAs (Ref. 15) or ZnSe.[16] However, even the unusually large energy gain due to relaxation cannot avoid that the gallium antisite remains energetically unfavorable. The large energy gain is only a response to the huge internal strain which is built up by forming this antisite. Atomic relaxation can *reduce* this strain; however, it cannot completely *avoid* it. The same argumentation applies for the nitrogen antisite, where a strong tensile stress is found, and to the gallium interstitial.

This argument, however, cannot be applied to the case of the nitrogen interstitial where a *small* atom is brought additionally into the crystal. Anticipating the possibility of low-symmetry configurations for this defect, we calculated a complete total-energy surface for a N interstitial in GaN. Fixing the coordinates of the additional nitrogen atom, we allowed all the surrounding atoms to relax, for 50 positions of the nitrogen atom and for different charge states. The calculations on the fine mesh were carried out in a 16-atom supercell. The local minima by the procedure were then investigated with a better converged calculation. We identified a (100) split-interstitial configuration, in which two N atoms share the same substitutional site, as the lowest energy state [see Fig. 3(c) below]. Each nitrogen atom in the defect now has two bonds to the surrounding gallium atoms instead of four.

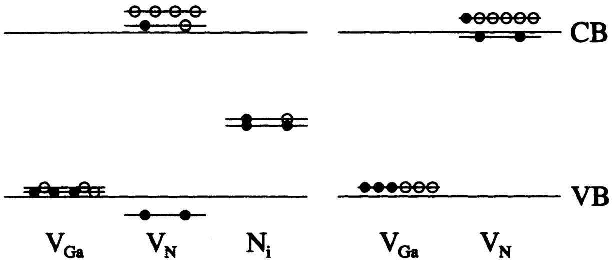

FIG. 2. Schematic representation of the location of defect levels for the gallium and nitrogen vacancies, and the nitrogen split interstitial. The results of a tight-binding calculation by Jenkins and Dow (Ref. 6) are shown on the right. The occupations are shown for the neutral charge state; filled circles indicate electrons, open circles indicate holes.

We now focus on the electronic structure of the native defects in GaN, particularly on the energetic position and character of the defect-induced states in the band structure (see Fig. 2). The *gallium vacancy* in the neutral charge state is triply occupied, with levels close to the valence band which can be filled with three more electrons, thus acting as a triple acceptor. The T_2 states are split into a doublet and a singlet, where the singlet state can be described as a p-like dangling bond mainly located on the nitrogen atom along the c axis through the Ga vacancy [Fig. 3(a)]; the doublet is a linear combination of the dangling bonds located at the three equivalent neighbors [Fig. 3(d)]. The splitting of the T_2 state into a doublet and a singlet state is a result of the lower symmetry of the wurtzite compared to the zinc-blende structure.[12] Filling the defect states changes the charge state of the Ga vacancy from $+3$ to -3. With increasing occupation of the defect levels the surrounding nitrogen atoms move inward, have their largest inward relaxation for the neutral charge state, and move outward again for negative charge states (Table I).

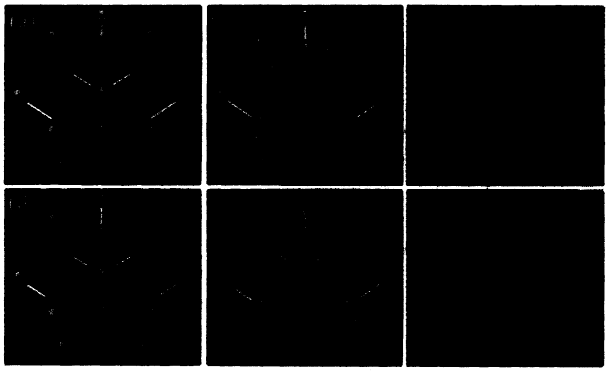

FIG. 3. Atomic geometry and charge density contour surfaces for the gallium vacancy [(a) and (d)] and the nitrogen vacancy [(b) and (e)]. (a) and (b) show the singlet state and (d) and (e) the doublet state of the T_2 vacancy state. (c) shows the geometry of the nitrogen split interstitial together with one of the defect states. For comparison the T_2 state of the As vacancy in GaAs is shown in (f). Green balls mark the position of the cations (Ga), blue balls the position of the anions (N, As). The contour surface shown as a transparent pink-colored object was taken at $\rho=0.05e/\text{bohr}^3$.

TABLE I. Change of defect geometry with charge state for the gallium and nitrogen vacancies and the nitrogen split interstitial. Units are in % with respect to the values for an unrelaxed defect. $r_{\text{N-N}}$ is given in % with respect to the theoretical bond length of a nitrogen molecule (1.1 Å). $r_{\parallel}$ is the distance from the vacancy center to the three equivalent neighbors, $r_{\perp}$ is the distance to the remaining neighbor.

	V_{Ga}		V_{N}		N_i	
Charge	$r_{\perp}$	$r_{\parallel}$	$r_{\perp}$	$r_{\parallel}$	$r_{\text{N-N}}$	$r_{\text{N-Ga}}$
−3	10	4.7				
−2	8.3	5.2				
−1	6.3	5.0			30	−7.7
0	3.7	3.5			22	−5.4
1	4.1	4.1	1.5	0.5	14	−2.8
2	5.9	6.1			8.7	−0.2
3	7.1	7.3			6.0	1.4

This complex relaxation pattern is surprising at first glance since according to simple tight-binding arguments an inward relaxation would be expected with increasing occupation [as observed for the anions on III-V (110) semiconductor surfaces]. With increasing occupation the dangling bond will change from sp^3-like towards more s-like character, in order to decrease the dangling bond one-particle energy. This transition changes the character of the backbonds from sp^3-like towards more p^3-like character; thus a reduction of the angles between the backbonds and subsequently an inward relaxation is expected. The main difference between a charged dangling bond in a vacancy and at a surface is that there is no effective screening of the electrostatic interaction inside the vacancy. Thus, to reduce the electrostatic interaction in a charged defect an outward displacement occurs to increase the distance between the charged dangling bonds. The dominance of the electrostatic interaction is also consistent with the fact that the outward relaxation is independent of the *sign* of the charge state. Charging the defect in either direction results in an outward relaxation of the neighboring atoms to decrease the repulsive electrostatic interaction.

For the *nitrogen vacancy* the s-like A_1 state lies as a resonance below the band edge in the valence band. The p-like T_2 states form a resonance in the conduction band (see Fig. 2). Similar to the gallium vacancy the p-like state is split into a singlet and a doublet. For the neutral charge state the one electron of the singlet state is transferred to the lower-lying bottom of the conduction band; the nitrogen vacancy acts as a single donor. This result is *qualitatively* different from previous tight-binding calculations[6,7] which are widely used to interpret experimentally observed defect levels. According to those calculations the A_1 state lies close to the bottom of the conduction band.

The main difference between the two results is that we find a large splitting between the A_1 and T_2 defect levels, shifting the A_1 level into the valence band. The T_2 level is split into a singlet and a doublet state, for which the charge densities are shown in Figs. 3(b) and 3(e). Again, the singlet state corresponds to a dangling bond located at the neighboring nitrogen atom along the c axis, the doublet is a linear combination of the dangling bonds located on the three equivalent neighbors. A comparison between Figs. 3(d) and 3(e) reveals a remarkable difference in the character of the defect states; whereas for the gallium vacancy the nitrogen dangling bonds are strongly localized, indicating a weak interaction, for the nitrogen vacancy the gallium dangling bonds are strongly overlapping. This state can rather be considered as a metallic bond between the gallium atoms, than as a weak interaction of dangling bonds located at second-nearest neighbors. This strong interaction explains the large splitting between the A_1 and T_2 defect levels. It also explains the failing of the tight-binding calculations,[6,7] where only nearest-neighbor interaction was taken into account.

The origin of the large Ga-Ga interaction again lies in the sizable mismatch of the covalent radii of gallium and nitrogen. Due to the small radius of the nitrogen atoms the gallium atoms in GaN have approximately the same distance as in Ga bulk.[12] Thus, the Ga atoms around the nitrogen vacancy form strong metalliclike bonds similar to those in bulk Ga. In order to discuss this argument in more detail and also to learn more about chemical trends we calculated the arsenic vacancy in GaAs. The main difference between GaN and GaAs is that GaAs has a 20% larger lattice constant, making the Ga-Ga distance 20% larger. The T_2 defect state for the As vacancy in GaAs is shown in Fig. 3(f). Contrary to the nitrogen vacancy in GaN the dangling bonds are strongly localized and only weakly interacting.

The defect states of the *nitrogen split interstitial* are shown in Fig. 2. Figure 3(c) reveals that the defect states are antibonding πp-like states for the N-N bond but bonding states for the Ga-N interaction. The splitting of the two states is again caused by the reduced symmetry of the wurtzite structure. Because this is an antibonding state, leaving these levels unoccupied reduces the N-N distance (the bond becomes stronger), and the N-Ga distance is increased (see Table I).

Summarizing, we have found that the dominant intrinsic point defects in GaN are the vacancies. Under p-type conditions the nitrogen vacancy (a single donor) dominates, under n-type conditions the gallium vacancy (a triple acceptor). These results also shed light on the issue of whether the nitrogen vacancy is the source of the n-type conductivity in as-grown GaN. Under n-type conditions the formation energy of the nitrogen vacancy is about 4 eV.[17] For a system in thermodynamic equilibrium, such a high formation energy implies that nitrogen vacancies cannot occur in appreciable concentrations. Even if deviations from thermodynamic equilibrium occur, they would be unlikely to stabilize a defect with such a high formation energy. Instead of nitrogen vacancies, under n-type conditions gallium vacancies are actually formed which may partly compensate the origin of the n-type conductivity. The isolated nitrogen vacancy can therefore not be a source for n-type conductivity. However, the formation of clusters of nitrogen vacancies cannot be excluded. Indeed, preliminary results indicate a large binding energy of nitrogen vacancies which could result in clustering. Another possibility is that impurities are the origin of the n-type conductivity. Further theoretical and experimental studies are necessary to investigate these possibilities.

This work was supported in part by the DAAD (German Academic Exchange Service) and by the AFOSR.

[1] S. Nakamura, T. Mukai, and M. Senoh, Appl. Phys. Lett. **64**, 1687 (1994).

[2] H. Amano, M. Kito, K. Hiramutsu, and I. Akasaki, Jpn. J. Appl. Phys. **28**, L2112 (1989).

[3] I. Akasaki, H. Amano, Y. Koide, K. Hiramatsu, and N. Sawaki, J. Cryst. Growth **98**, 209 (1989).

[4] H. P. Maruska and J. J. Tietjen, Appl. Phys. Lett. **15**, 327 (1969).

[5] M. Ilegems and H. C. Montgomery, J. Phys. Chem. Solids **34**, 885 (1973).

[6] D. W. Jenkins and J. D. Dow, Phys. Rev. B **39**, 3317 (1989).

[7] D. W. Jenkins, J. D. Dow, and M. H. Tsai, J. Appl. Phys. **72**, 4130 (1992).

[8] T. L. Tansley and R. J. Egan, Phys. Rev. B **45**, 10 942 (1992); Physica B **185**, 190 (1993).

[9] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[10] J. Neugebauer and C. G. Van de Walle (unpublished).

[11] R. Stumpf and M. Scheffler, Comput. Phys. Commun. (to be published).

[12] J. Neugebauer and C. G. Van de Walle, in Diamond, SiC and Nitride Wide-Bandgap Semiconductors, edited by C. H. Carter, Jr., G. Gildenblat, S. Nakamura, and R. J. Nemanich, MRS Symposia Proceedings No. 339 (Materials Research Society, Pittsburgh, in press).

[13] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[14] C. G. Van de Walle, D. B. Laks, G. F. Neumark, and S. T. Pantelides, Phys. Rev. B **47**, 9425 (1993).

[15] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).

[16] D. B. Laks, C. G. Van de Walle, G. F. Neumark, P. E. Blöchl, and S. T. Pantelides, Phys. Rev. B **45**, 10 965 (1992).

[17] Note that this is the value for nitrogen-rich conditions. For nitrogen-poor conditions the formation enthalpy of GaN (1.1 eV) has to be subtracted.