PHYSICAL REVIEW LETTERS

Nitrogen Vacancies as Major Point Defects in Gallium Nitride

M. G. Ganchenkova and R. M. Nieminen

COMP/Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, 02015, Espoo, Finland
(Received 27 October 2005; published 15 May 2006)

We present results of *ab initio* calculations for vacancies and divacancies in GaN. Particular attention is paid to nitrogen vacancies and mixed Ga-N divacancies in negatively charged states, which in *n*-type GaN are found to be energetically comparable with gallium vacancies. We also demonstrate that the activation energy for self-diffusion over the nitrogen sublattice is lower than over the gallium one for all Fermi-level positions, which implies the nitrogen vacancies are major defects in samples annealed at high temperatures. Possibilities for direct observations of nitrogen vacancies through positron annihilation experiments are discussed.

Gallium nitride is an important wide band-gap semiconductor for micro- and optoelectronic applications, such as blue light emitting diodes and lasers. As-grown GaN samples are known to be *n*-type semiconductors without intentional doping, implying the presence of donors. These donors must be associated with defects or impurities, since perfect GaN cannot provide thermal carriers. Native defects in GaN have attracted much attention in the past decade [see the recent review [1] for an up-to-date summary]. Initially, a strong candidate responsible for the *n*-type conductivity was the positively charged vacancy in the nitrogen sublattice (V_N^+) [2]. This assumption was later discarded when calculations revealed that the formation energy, E_f, of V_N^+ in *n*-type material is close to ~3 eV [3–5] and hence the thermal equilibrium concentration of V_N^+ is too small to provide an observable number of carriers. It is currently assumed that the carriers are provided by impurity atoms, in particular, oxygen and silicon, which contaminate GaN crystals during the growth process.

Among the native defects (vacancies, interstitials, and antisites) only Ga vacancies (V_{Ga}) are assumed to appear in observable amounts in as-grown GaN. As the energies of point defects in zinc-blende and wurtzite GaN are very similar [1,5], V_{Ga} have been thought to prevail in both polytypes of as-grown GaN. This conclusion is often substantiated by results of positron annihilation studies [6] that observe only Ga vacancies in *n*-type GaN. The nitrogen vacancies are expected to play any role in the defect microstructure kinetics only in *p*-doped material [1].

It should be noticed, however, that the minor role of nitrogen vacancies in *n*-type material is inferred from calculations for positively charged vacancies only [from 3+ to 1+ [1]]. To our knowledge, other charge states (neutral or negative) have not been reported [except for Ref. [7], which indicates that V_N^0 might be not favorable]. In this Letter we demonstrate that nitrogen vacancies, including negatively charged ones, and their complexes (e.g., mixed Ga-N divacancies) can be the major participants in the defect kinetics of *n*-type GaN.

Our calculations are based on the density-functional theory (DFT) approach within the projector augmented-

wave method with the local-density approximation implemented in the VASP code [8]. In order to estimate the error due to the finite supercell (SC) size in the formation energy calculations for charged defects, we performed a thorough analysis of the total energy convergence with the supercell size (from 32 to 300 atom SC). Brillouin-zone sampling is done with the $4 \times 4 \times 4$ and $2 \times 2 \times 2$ Monkhorst-Pack *k*-point set depending on the considered SC. The atoms are allowed to relax until the residual forces fall below 0.01 eV/A [3]. We confirmed in each case that there is practically no dependence of the nitrogen vacancy configuration energy E_{conf}, that is the difference between the total energies for SC with the defect and without it, on SC size. Such behavior is found for all defect charge states from +1 to −3, with the biggest configuration energy change of ~0.04 eV for the triply negative state when the SC size changes from 64 to 300 atoms. This indicates that both the Madelung-type correction ($\sim q^2/L$, where q is the charge and L is the supercell size) and the valence band alignment correction for this defect are unimportant [the detailed description of the corrections can be found in Ref. [1]]. For the calculations of the defect configurations and energies in saddle-point positions, we use the standard "drag" method [see, e.g., Ref. [9]].

When assigning defect-induced levels in the band gap, to our knowledge, there are at least two ways to do it. On the one hand, one can use the Zhang-Northrup formalism [10], referring all levels to the calculated valence band maximum (VBM) for the bulk and taking the band-gap value from the experimental data. Another possibility is to use the so-called "marker" method [11], where the bulk ionization potential and affinity can be taken as the references. The ionization potential, which corresponds to the level (+/0) for the bulk (perfect) unit cell, converges well with the unit cell size (3.52 eV for 64 atom SC and $6 \times 6 \times 6$ *k* mesh, and 3.58 for 512 atom SC and $2 \times 2 \times 2$ mesh), whereas the (0/−) affinity level is much more uncertain (varying from 6.50 eV to 5.95 for the same conditions). Therefore, we have chosen the level (+/0) (or alternatively the VBM) as the reference. All defect formation energies presented in Fig. 1 are calculated
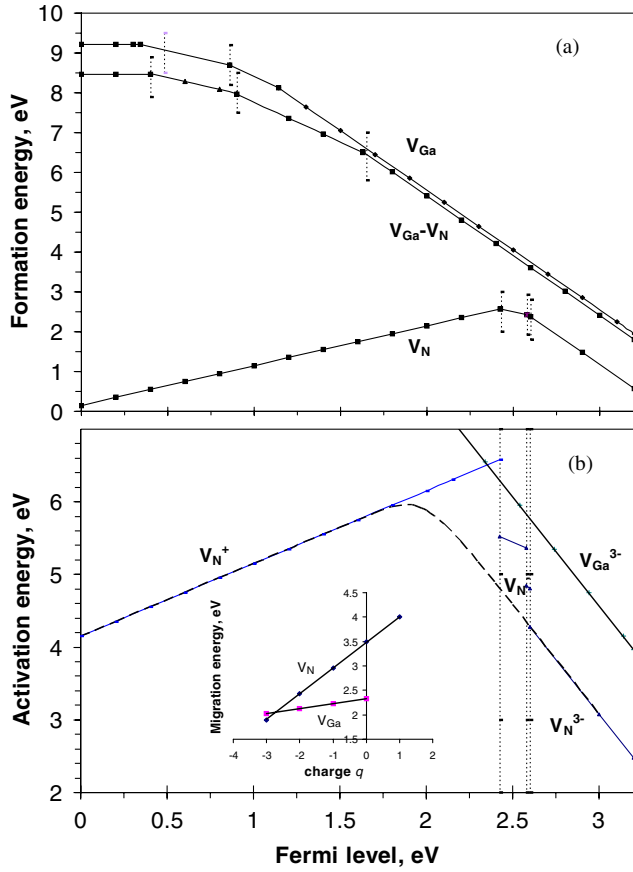
FIG. 1 (color online). Formation energies for gallium and nitrogen vacancies and mixed divacancies (a) and activation energies for self-diffusion in two sublattices (b). Solid lines represent the activation energies for the lowest energy charge state of a defect; the dashed line is the effective activation energy for vacancy-mediated self-diffusion (see detailed explanation in the text). The inset demonstrates the charge dependence of vacancy migration energies.

from the total energies of defects for a 300 atom SC and $2 \times 2 \times 2$ k-point mesh using the Zhang-Northrup formula [10] referenced to the valence band maximum.

The results for the gallium vacancy [Fig. 1(a)], considered in charge states from V_{Ga}^{+} to V_{Ga}^{3-}, almost perfectly reproduce the earlier predictions [1]. An essentially different situation is encountered in the case of the nitrogen vacancy, which was considered in charge states ranging from V_{N}^{3+} to V_{N}^{3-}. The most important conclusion from our results is that the nitrogen vacancy is a negative-U defect that exists, depending on the Fermi-level position, either in the singly positive, or in singly to triply negative charge states. Another important conclusion is that V_N is energetically more favorable than the gallium vacancy over the whole band gap and not only in p-type material, as has been assumed earlier.

These results lead to conclusions that differ drastically from the current views [cf. e.g., Fig. 1 in Ref. [1]]. In order to clarify possible reasons for such a discrepancy, we have performed a set of calculations for V_N in the same range of

charge states using smaller SCs (from 32 up to 216 atoms) in both zinc blende and wurtzite. The earlier results for positively charged nitrogen vacancies are well reproduced qualitatively in smaller supercells, no matter how the d electrons are accounted for and which polytype is considered. However, even in a small SC the negative charge states of the V_N are definitely more stable in the upper part of the band gap than the positive charge states, a fact that has been overlooked before. The calculated ionization levels are well converged for the 300 atom SC calculations. The values are similar when calculated by the Zhang-Northrup formalism [$(+/-)$ at 2.43 eV, $(-/2-)$ at 2.58 eV, and $(2-/3-)$ at 2.60 eV] and the marker method [$(+/-)$ at 2.46 eV, $(-/2-)$ at 2.61 eV, and $(2-/3-)$ at 2.63 eV], when E_ν and the $(+/0)$ bulk level are taken as references.

The calculated ionization levels, based on total energy differences, lie above the DFT single-particle gap calculated for the bulk. In order to confirm that the defect-induced single-particle states are localized in the defect region and not resonant with the conduction band (E_c), we have carefully analyzed their distribution. In the case of V_N the a_1 (s-like) state goes below the valence band edge, whereas the t_2 (p-like) states lie inside the calculated Kohn-Sham (KS) gap. The appearance of these levels is accompanied with the upwards shift of the conduction band edge with the gap opening up to 2.7 eV. When V_N is negatively charged, extra electrons progressively fill the t_2 states, which remain below E_c. The filling of the defect levels is accompanied with changes in the relaxation and bonding patterns. When the charge state changes from V_N^{+} to V_N^{3-}, the relaxation changes from outward (for V_N^{+}) to inward (for all other charge states). With each electron the inward relaxation increases, while the distance between the nearest neighbors falls down (from 3.17 Å for V_N^{+} to 2.83 Å for V_N^{3-}, while the equilibrium distance is 3.15 Å). Also, the T_d symmetry of V_N^{+} changes to the D_{2d} symmetry for all negatively charged vacancies. This symmetry decrease is the result of a Jahn-Teller distortion, which evidences a bonding trend between the nearest neighbors of the vacancy. In turn, the bonding can be only due to the localization of the added electrons. Indeed, the electron density distribution pattern of the highest occupied electron states clearly shows the contribution of localized electrons to the newly formed bonds. Examining the radial distribution of the partial charge densities (with respect to the vacancy center) we observe very pronounced localization of the electrons occupying the highest KS levels within the sphere of the radius of approximately second nearest neighbor distance.

If a rigid "scissor" correction is applied to the band gap and defect-derived (t_2) level states corresponding to V_N, the ionization levels $(+/-)$, $(-/2-)$, and $(2-/3-)$ move upwards by 15% to lie between 2.8 and 3.0 eV, but remain below the band gap, whereas the shift of levels associated with V_{Ga} is nearly 30%. This result comes from the differ-

ences in calculated conduction band bottom positions in the supercells containing V_N or V_{Ga}. Thus the scissor corrections change the absolute values of the ionization levels, but leave V_N, more favorable than V_{Ga}.

The experimental observations show evidence both in favor and against the existence of V_N, in n-type GaN. The results of x-ray absorption fine structure measurements [12] on both cubic and hexagonal GaN samples grown at 600 °C by electron-cyclotron-resonance molecular-beam epitaxy without intentional doping (and hence, most probably—n-type) are explicitly interpreted by their authors in terms of the presence of V_N. On the other hand, positron annihilation spectroscopy (PAS) easily observes V_{Ga} [5–7], but bears no evidence of V_N in n-type GaN. This sounds contradictory to our expectation that in a heavily n-type doped GaN crystal, where the difference between the formation energies of gallium and nitrogen vacancies is the smallest (~0.98 eV and ~0.30 eV before and after the scissor correction, respectively), the ratio of thermal equilibrium concentrations is $C_{V_N}^{th}/C_{V_{Ga}}^{th} > 10$–$10^4$ at typical annealing temperatures of 800–1000 °C. One can imagine several possible explanations for this apparent discrepancy.

First, it is plausible that the negatively charged V_N cannot be observed in PAS experiments because their signal is indistinguishable from that from the bulk. In case of negatively charged gallium vacancy one observes a positron lifetime of 235 ps, much longer than the typical value of 160–163 ps for positron lifetimes in the bulk [6]. The positron lifetimes estimated for our electron density distribution in the bulk and V_{Ga}^{3-} are 130 ps and 200 ps, respectively, in full agreement with the value published in Ref. [13]. Such a systematic difference between the calculated and experimental positron lifetimes is associated with the use of the LDA enhancement factor and neglect of positron-induced relaxation [see Refs. [14,15] and references therein for more details]. Our estimates of positron lifetimes for the nitrogen vacancy are ~131 ps for V_N^-, and ~132 ps for V_N^{2-} and V_N^{3-}, respectively, essentially the same as in the bulk [13]. Thus, indeed, the difficulty of extracting information of negative nitrogen vacancies from the background camouflages the presence of V_N in the bulk, even if their abundant formation at high temperatures is prompted by low formation energies.

Second, in order to achieve an experimentally observable concentration of vacancies one has to heat up the samples and wait until the vacancy concentration reaches thermal equilibrium values. The rate of vacancy accumulation is proportional to vacancy currents, which depend on the sums of formation and migration energies. While the migration energy of V_{Ga}^{3-} is predicted to be 1.9 eV [16], the nitrogen vacancy migration energies can be quite high [e.g., for V_N^{3+} and V_N^+ they are predicted to be 2.6 and 4.6 eV, respectively [16]]. These predictions are in good agreement with our calculations that give 2.0 and 4.0 eV for V_{Ga}^{3-} and V_N^+, respectively. The difference in migration energies could mean that the time required for nitrogen

vacancies to reach the thermal equilibrium concentration is too long as compared to the experiment duration time.

However, it is important to consider migration energies for V_N in proper charge states [shown in the inset in Fig. 1(b)]. The vacancy migration energies fall down linearly with the increase of the number of electrons in both V_{Ga} and V_N. The migration energies of V_N^{2-} and V_N^{3-} are not much different from the value of 2.0 eV obtained for V_{Ga}^{3-} [see Fig. 1(b)]. In fact, for V_N^{3-} the migration energy of 1.9 eV is even lower than for V_{Ga}^{3-}. The resulting activation energies for vacancy diffusion are summarized in Fig. 1(b), where the activation energy is a sum of formation and migration energies calculated without scissor correction. The solid lines in Fig. 1(b) represent activation energies for the lowest-energy charge state of V_N at each Fermi-level position. However, the sharp transitions between "lowest-energy defect" regions are not physically reasonable, because the defect concentration close to a transition point involves comparable contributions from both relevant charge states. In order to take this into account, it is convenient to introduce an effective activation energy for the vacancy diffusion $Q_{ef}(\mu_e)$:

$$Q_{ef}(\mu_e) = -k_B T \ln\left(\sum_q e^{-Q_q(\mu_e)/k_B T}\right), \qquad (1)$$

where Q_q is the activation energy of a vacancy of charge q, μ_e is the electron chemical potential, and $k_B T$ has its usual meanings. The dependence of the effective activation energy for the vacancy-mediated self-diffusion on the Fermi-level position is shown in Fig. 1(b) by dashed line. The calculated activation energy for V_N^{3-} and V_N^{2-} [see Fig. 1(b)] changes from 4.28 eV to 2.48 eV, and from 4.85 eV to 3.61 eV, respectively, when the Fermi level shifts from 2.60 to 3.2 eV. This nicely correlates with the experimental data for self-diffusion that gives (4.1 ± 0.4) eV for the samples with the Fermi level 30 meV below E_c at room temperature [17].

As can be seen in Fig. 1(b), the activation energy for self-diffusion in the nitrogen sublattice is noticeably lower than in the gallium one over the whole band gap. Correspondingly, V_N accumulation rate in the bulk due to the diffusion from the sample surface is higher than for V_{Ga}. Even during transients involved in the establishment of thermal equilibrium vacancy concentrations the total amount of nitrogen vacancies in the sample should exceed that of gallium vacancies. Moreover, for the obtained vacancy migration energies the typical time of thermal vacancy equilibration in a 5 μm thick GaN layer at 900 °C is about few minutes for V_{Ga}^{3-} and V_N^{3-}, being much shorter than, e.g., typical times of many hours required for self-diffusion experiments [17].

Further strong evidence of the low formation energy of V_N comes from the fact that GaN samples prepared by conventional techniques (e.g., metalorganic chemical vapor deposition) are subject to thermal decomposition at temperatures above ~900 °C. The thermal decomposition

is unequivocally due to the evaporation of nitrogen from the sample surface, which implies a high concentration of nitrogen vacancies at the sample surface, and, consequently, their low formation energy. It is hard to explain this fact in the framework of earlier assumption that V_N^+ has the lowest energy over the whole band gap, because in the upper part of the band gap the formation energy of V_N^+ is close to 3 eV (corresponding to concentration $\sim 10^8$–10^9 cm^{-3} at 900 °C) and the diffusion activation energy is as high as 7 eV V_N. On the contrary, the existence of negatively charged V_N implies an activation energy of (3.53 ± 0.75) eV and a concentration above 10^{19} cm^{-3} at the same temperature and in the Fermi-level range of 2.6–3.1 eV above the valence band. Such concentration is sufficiently high to promote thermal decomposition and, in conjunction with a reasonably low migration energy, can result in enhanced self-diffusion. Indeed, Ambacher *et al.* [17,18] found that within the accuracy of their measurements, the nitrogen diffusion process has the same activation energy as the decomposition process, that is (3.9 ± 0.2) eV [18] for samples with the Fermi-level 30 meV below E_c. These results imply that at high temperatures a high concentration of defects is created close to the sample surface, which significantly enhances the diffusion process [17,18].

Correspondingly, the improvement of the quality and stability of GaN samples synthesized at high nitrogen pressures (>1 GPa[6]) and temperatures as high as 1500 °C, can be explained by suppression of thermal V_N production at sample surface by the increase of free nitrogen concentration in the ambient. However, the high pressures should suppress V_{Ga} as well, though for a different reason. According to our calculations, the relaxation of nitrogen atoms around a V_{Ga} is outwards (reaching $\sim 8\%$ of interatomic distance for V_{Ga}^{3-}). This normally suggests the increase of the defect formation energy under compressive elastic loads and, indeed, hydrostatic compression of 1.5 GPa was found to increase the energy of V_{Ga}^{3-} by 0.5 eV.

The high concentrations of both V_{Ga} and V_N reached in n-type GaN at high temperatures prompt the question whether other vacancy defects, in particular, mixed Ga-N divacancies (V_{Ga}-V_N), can appear in n-type crystals. Earlier calculations [5] indicated that the binding of two vacancies into a mixed divacancy is energetically favorable over the whole band gap. We find that for Fermi levels above 1.5 eV from the top of the valence band the most stable charge state becomes $(V_{Ga}$-$V_N)^{3-}$. This defect not only has a substantial binding energy (~ 2.34 eV for dissociation into $V_N^0 + V_{Ga}^{3-}$), but its formation energy turns out to be lower than that of gallium vacancies [Fig. 1(a)].

Summing up, the following conclusion can be made from this study: contrary to the earlier expectations, the nitrogen vacancy should be the dominant defect in GaN in the whole range of Fermi-level positions in the band gap.

[1] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[2] H. P. Maruska and J. J. Tietjen, Appl. Phys. Lett. **15**, 327 (1969); P. Perlin, T. Suski, H. Teisseyre, M. Leszczynski, I. Grzegory, J. Jun, S. Porowski, P. Boguslawski, J. Bernholc, J. C. Chervin, A. Polian, and T. D. Moustakas, Phys. Rev. Lett. **75**, 296 (1995).

[3] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, R8067 (1994).

[4] T. Mattila, A. P. Seitsonen, and R. M. Nieminen, Phys. Rev. B **54**, 1474 (1996).

[5] T. Mattila and R. M. Nieminen, Phys. Rev. B **55**, 9571 (1997).

[6] K. Saarinen, T. Laine, S. Kuisma, J. Nissilä, P. Hautojärvi, L. Dobrzynski, J. M. Baranowski, K. Pakula, R. Stepniewski, M. Wojdak, A. Wysmolek, T. Suski, M. Leszczynski, I. Grzegory, and S. Porowski, Phys. Rev. Lett. **79**, 3030 (1997); J. Oila, V. Ranki, J. Kivioja, K. Saarinen, P. Hautojärvi, J. Likonen, J. M. Baranowski, K. Pakula, T. Suski, M. Leszczynski, and I. Grzegory, Phys. Rev. B **63**, 045205 (2001); K. Saarinen, T. Suski, I. Grzegory, and D. C. Look, Phys. Rev. B **64**, 233201 (2001).

[7] P. Boguslawski, E. L. Briggs, and J. Bernholc, Phys. Rev. B **51**, R17255 (1995).

[8] G. Kresse and J. Hafner, Phys. Rev. B **47**, R558 (1993); G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11 169 (1996); P. E. Blöchl, Phys. Rev. B **50**, 17 953 (1994); G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[9] H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998), p. 385.

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

[11] C. D. Latham, M. Alatalo, R. M. Nieminen, R. Jones, S. Öberg, and P. R. Briddon, Phys. Rev. B **72**, 235205 (2005).

[12] M. Katsikini, H. Rossner, M. Fieber-Erdmann, E. Holub-Krappe, T. D. Moustakas, and E. C. Paloura, J. Synchrotron Radiat. **6**, 561 (1999).

[13] S. Hautakangas, I. Makkonen, V. Ranki, M. J. Puska, K. Saarinen, X. Xu, and D. C. Look, Phys. Rev. B (to be published).

[14] T. Torsti, V. Lindberg, I. Makkonenl, E. Ogando, E. Räsänen, H. Saarikoski, M. J. Puska, and R. M. Nieminen, Psi-k Newsletter **65**, 105 (2004).

[15] I. Makkonen, M. Hakala, and M. J. Puska, Phys. Rev. B **73**, 035103 (2006).

[16] S. Limpijumnong and C. G. Van de Walle, Phys. Rev. B **69**, 035207 (2004).

[17] O. Ambacher, F. Freudenberg, R. Dimitrov, H. Angerer, and M. Stutzmann, Jpn. J. Appl. Phys. **37**, 2416 (1998).

[18] O. Ambacher, M. S. Brandt, R. Dimitrov, T. Metzger, M. Stutzmann, R. A. Fischer, A. Miehr, A. Bergmaier, and G. Dollinger, J. Vac. Sci. Technol. B **14**, 3532 (1996).