RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 51, NUMBER 23 15 JUNE 1995-I

Native defects in gallium nitride

P. Bogusławski
North Carolina State University, Raleigh, North Carolina 27695-8202
and Institute of Physics, Polskiej Akademii Nauk, 02-668 Warsaw, Poland

E. L. Briggs and J. Bernholc
North Carolina State University, Raleigh, North Carolina 27695-8202
(Received 31 January 1993; revised manuscript received 15 March 1995)

The results of an extensive theoretical study of native defects in hexagonal GaN are presented. We have considered cation and anion vacancies, antisites, and interstitials. The computations were carried out using *ab initio* molecular dynamics in supercells containing 72 atoms. N vacancy introduces a shallow donor level, and may be responsible for the n-type character of as-grown GaN. Due to the wide gap of nitrides, self-compensation effects strongly reduce both n-type and p-type doping efficiencies due to the formation of gallium vacancy and interstitial Ga, respectively.

Gallium nitride, together with other wide gap nitrides, holds substantial promise for electronic applications.[1] As-grown undoped samples of GaN are almost always n type, with the concentration of conduction electrons ranging typically from 10^{17} to 10^{20} cm^{-3}. These values are much higher than concentrations of detected impurities.[2,3] This strongly suggests that the doping is due to native defects. The residual donor was tentatively identified with the nitrogen vacancy.[2–6] The present calculations support this possibility. We point out, however, that among the point defects another candidate is the interstitial Ga. Like V_N, Ga(I) is an effective-mass donor, and its calculated abundance is comparable to that of V_N under certain conditions. It is also well known that GaN is difficult to dope p type. According to the present results, this behavior may be due to an efficient intrinsic compensation mechanism involving the formation of native defects that introduce shallow donor levels. This is a much more serious problem in the wide band-gap GaN (and even more in AlN) than in, say, GaAs, although a reduction of the n-type doping efficiency by the formation of Ga vacancies was observed in GaAs.[7] In this paper we summarize the results of an extensive study of native defects in wurtzite GaN. We describe their electronic structure and use the calculated formation energies to identify the dominant native defects and to evaluate doping efficiencies and stoichiometry effects. The results compare well with the experimental data and lead to a tentative identification of the dominant native defects in as-grown GaN.

The calculations were carried out using *ab initio* molecular dynamics,[8] using an efficient relaxation procedure for optimization of the atomic geometries.[9] For Ga, a standard nonlocal Ga pseudopotential was used,[10] while a soft nonlocal pseudopotential was generated for N.[11] The kinetic energy cutoff for the plane wave basis set was 30 Ry. Tests for pure GaN show that these potentials reproduce well previous theoretical and experimental results.[12] Calculations for defects were carried out in a large supercell that would contain 72 atoms in the case of the perfect crystal. Due to the size of the cell, only the Γ point was used for Brillouin zone summations. Some results for the Ga interstitials were corroborated by a newly developed multigrid method[13] that allows for the use of a much higher cutoff and the explicit inclusion of Ga $3d$ electrons.

Assuming that the GaN crystal is in equilibrium with an atomic reservoir of either Ga or N, the formation energy of a defect in a charge state q is given by

$$E_{\rm form}(q)=E_{\rm tot}(q)-n_{\rm Ga}\mu_{\rm Ga}-n_{\rm N}\mu_{\rm N}+qE_F\,, \tag{1}$$

where $E_{\rm tot}$ is the total energy of the supercell with the defect, $n_{\rm Ga}$ and $n_{\rm N}$ are the numbers of Ga and N atoms, $\mu_{\rm Ga}$ and $\mu_{\rm N}$ are the chemical potentials, and E_F is the Fermi energy. The chemical potentials must satisfy the relation $\mu_{\rm Ga}+\mu_{\rm N}=\mu_{\rm GaN}$.[14] Assuming elemental bulk sources for either Ga or N, the computed range of the allowed values for $\Delta\mu$, the difference between the chemical potentials of Ga and N, is from -1.8 eV (N-rich limit) to $+1.8$ eV (Ga-rich limit), since it is bound by the theoretical value of the heat of formation of GaN.

The equilibrium concentration of a defect is given by

$$C=N_{\rm sites}\exp(S_{\rm form}/k_B-E_{\rm form}/k_BT), \tag{2}$$

where $N_{\rm sites}$ is the concentration of sites, and $S_{\rm form}$ is its formation entropy. Here, we assume a typical value of $6k_B$ for all defects. Since the formation energies of defects in GaN are much greater than the usual entropic contributions $(4-10k_BT)$, this choice does not significantly affect the results. The concentrations of defects are thus determined by three independent parameters: temperature, $\Delta\mu$, and the number of carriers supplied by external dopants. The position of the Fermi level and the formation energies of charged defects [cf. Eq. (1)] are self-consistently determined as a function of $\Delta\mu$ from the charge neutrality condition, given the temperature and the number of external dopants.

Density functional theory (DFT) is well known in underestimating semiconductor band gaps. In the present case, the calculated gap of 2.5 eV is smaller by 1 eV than the experimental value. This error does not influence the formation energies of neutral defects, but may strongly affect those of charged defects. For example, $E_{\rm form}$ of a shallow donor

 © 1995 The American Physical Society

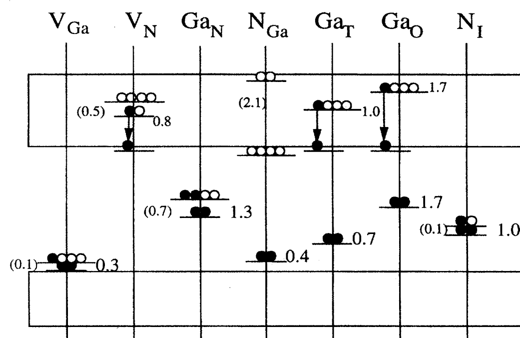

FIG. 1. Energy levels of neutral native point defects in GaN.

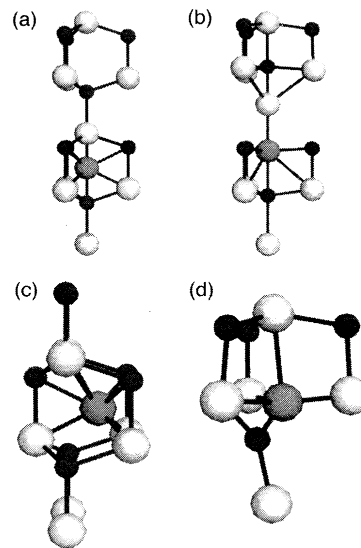

FIG. 2. Atomic configurations of interstitials: (a) the ideal and (b) the relaxed Ga(T) interstitial, (c) the ideal Ga(O) interstitial, and (d) the relaxed N interstitial. Big and small spheres represent Ga and N atoms, respectively.

should be reduced by essentially the energy of the band gap in p-type samples. Thus, an underestimate of E_{gap} by 1 eV overestimates E_{form} by the same amount. For higher charge states the error is even greater. We have attempted to partially eliminate this error by correcting both the band-gap value and the positions of the gap states of charged defects by the following procedure: For the perfect GaN crystal, GW calculations[15] obtain the correct band-gap and show that the major difference with DFT is a rigid upward shift of the conduction bands. Thus, the relevant level(s) is first projected onto valence and conduction states of perfect GaN, which form a complete basis. We then correct the formation energy by assuming that the level(s) has shifted upwards by the fraction of the band-gap correction given by its conduction band content.

It is convenient to analyze the results obtained for the wurtzite structure in terms of zinc-blende symmetry with a superimposed hexagonal perturbation. In the zinc-blende structure, a substitutional defect, e.g., a vacancy, has four equivalent nearest neighbors. In the wurtzite structure, the atom along the c axis relative to the defect (called here type-1 neighbor) becomes inequivalent to the three remaining neighbors (called here type-2 neighbors). The lowering of the point symmetry is also reflected in the electronic structure: the defect states that are threefold degenerate in the zinc-blende structure split into singlets and doublets in the wurtzite structure. In the following, we refer to such singlet and doublet pairs as quasitriplets. The resulting energy levels are schematically shown in Fig. 1 with the hexagonal splittings given in parentheses.

For the vacancies, we expect an A_1 singlet and a quasitriplet, corresponding to the T_2 state in zinc-blende symmetry. As in the zinc-blende structure, the A_1 state is in the valence bands for both vacancies. Since the electronic states of V_N are mainly composed of Ga dangling bonds, the energy of the quasitriplet is quite high. The singlet state of the quasitriplet is about 0.8 eV above the bottom of the conduction band. The hexagonal splitting is 0.5 eV and the quasitriplet level contains one electron. However, since the quasitriplet is a resonance, the electron autoionizes to the bottom of the conduction band, where it forms an effective-mass state bound by the Coulomb tail of the vacancy potential.

Since the formation of the Ga vacancy creates N dangling bonds, its levels should be close to the top of the valence bands. Indeed, the quasitriplet is located about 0.3 eV above the valence bands edge, and the hexagonal splitting is only 0.1 eV. Since the quasitriplet is populated by three electrons in the neutral charge state, V_{Ga} can trap both electrons and holes.

The gallium antisite introduces a quasitriplet close to the middle of the band gap. In the neutral charge state, the singlet at $E_v+1.4$ eV and the doublet at $E_v+2.1$ eV contain two electrons each. The outward relaxation around Ga_N is large. The bond lengths with type-1 and type-2 neighbors increase from the ideal value of 3.70 a.u. to 4.12 and 4.27 a.u., respectively.

In the neutral charge state, N_{Ga} introduces a doubly occupied singlet at $E_v+0.4$ eV, and an empty doublet at $E_c-0.2$ eV. The N_{Ga} strongly distorts along the c axis. The bond distance to the type-1 neighbor is reduced by 29% and becomes comparable with the bond length in the N_2 molecule, 2.07 a.u. Due to the distortion, the remaining bond lengths increase by 11%. The empty doublet at $E_c-0.2$ eV cannot be occupied even by one electron, due to the large value of the electron-electron repulsion parameter for this level (0.95 eV).

We now turn to interstitials. In the wurtzite structure, there are two high-symmetry interstitial positions, T and O, shown in Fig. 2. The point symmetry is C_{3v} in both cases. The T site is located in the middle of the line connecting nonbonded Ga and N atoms. An atom at the T site has two nearest neighbors and six next nearest neighbors. The O site has six nearest neighbors, with distances greater by 28% than at the T site, and 7% greater than the equilibrium bond length. Due to the lack of a reflection plane perpendicular to the c axis, neither site can be the equilibrium position of an interstitial defect, except by accident. In particular, a relaxation along the c axis is expected. Indeed, we find that both the T and O positions are highly unstable for native interstitials.

The equilibrium position of the interstitial Ga is strongly charge dependent. Figure 2 shows both the initial and the

relaxed configurations of the neutral Ga(T). The relaxation leads to an upward displacement of three atoms: Ga(T), the host Ga, and the host N. The Ga(T) moves by 1.2 a.u., gaining 11 eV. After the relaxation, the Ga-Ga distance increases to 4.17 a.u., which is close to the average Ga-Ga distance for Ga_N antisite. This configuration is similar to the bridge-bond geometry of interstitial Ga in GaAs.[16] With an increasing charge the complex moves down, assuming an off-axis position for $Ga^{3+}(T)$. $Ga_T{}^0$ introduces a deep level at $E_c-1.8$ eV, occupied by two electrons, and a resonance at $E_c+1.0$ eV. The one electron that should occupy the resonance autoionizes and becomes trapped in an effective-mass level. Finally, Ga(T) exhibits a negative U behavior, since the $+2$ charge state is unstable for all positions of the Fermi level. The energy gain for the reaction 2 $Ga^{2+}(T) \rightarrow Ga^{3+}(T) + Ga^{+}(T)$ is 1.8 eV. The (+/3+) level is located at 1.7 eV above the top of the valence band.

In the case of Ga(O) shown in Fig. 2(c), the relaxation is much weaker. The interstitial is displaced upwards by 0.55 a.u. and the first three Ga neighbors move down by 0.5 a.u. The change of the charge state has only a minor effect on the relaxation pattern. The electronic structures of Ga(T) and Ga (O) are quite similar, despite the very different surroundings. Ga(O) also introduces a resonance in the conduction bands that leads to an effective-mass state, and a deep level at E_c − 0.8 eV. The (+/3+) level is located at about $E_v+1.7$ eV and the U parameter is nearly zero. Finally, the computed formation energy of the neutral Ga(O) is only 0.2 eV greater than that of Ga(T), which falls within the margin of error of our calculations.

For the N interstitial, starting from either the T or O site, leads to the same final configuration, shown in Fig. 2(d). A similar split-interstitial geometry has been obtained for the As interstitial in GaAs.[16] The displacement from the ideal T site is over 2 a.u., and the relaxation energy is 7.6 eV. The final N–N bond length is 2.64 a.u., which is very close to the N-N distance in the case of N_{Ga}. Considering the electronic structure, N(T) introduces a nearly degenerate pair of deep levels at about $E_v+1.0$ eV, separated by 0.1 eV and occupied by three electrons.

The calculated formation energies were used to study stoichiometry effects and to identify the dominant defects in GaN. Bulk samples are grown at $T_g=1300-1500\,^{\circ}$C and under Ga-rich conditions,[6] while molecular beam epitaxy and chemical vapor deposition proceed typically at 600–950 and 900–1100 °C, respectively. We have considered several temperatures in the 600–1300 °C range, and computed defect concentrations assuming equilibrium thermodynamics [cf. Eq. (2)]. However, GaN is currently grown under conditions that deviate significantly from equilibrium (see below). Nevertheless, the results still provide an indication of the identity of the dominant defects and of general effects governing their formation. We describe below the results for $T_g=1300\,^{\circ}$C, and examine three scenarios: (i) no doping, i.e., intrinsic sample, (ii) n-type, and (iii) p-type samples with concentrations of external carriers of $10^{18}-10^{21}$ cm^{-3}. Lowering of temperature to 900 °C results in an overall decrease of defect concentrations by a few orders of magnitude, but trends remain the same.

Our results reveal that the stoichiometry and the doping efficiency of GaN are strongly influenced by the large value of the band gap, and have properties that should be universal for all wide band-gap semiconductors. The formation energies of neutral N and Ga vacancies are 3.2 and 8.1 eV, respectively, at Ga-rich conditions. We thus expect a negligible concentration of V_{Ga}. However, according to Eq. (1), formation energies of defects in highly charged states may be reduced by an energy of up to 2–3 times the band gap, i.e., of the order of 10 eV. In n-type samples, the energy gain associated with transferring three electrons from the Fermi level to the low-lying acceptor states of the vacancy makes V_{Ga}^{-3} the prevalent native defect. In p-type samples, this effect occurs for Ga^{3+}(I), where three donor electrons are transferred to acceptor levels. Therefore, *doping of wide band-gap semiconductors under conditions of thermal equilibrium leads to very strong self-compensation effects*. An attempt to dope the crystal will result in an increasing concentration of the appropriate compensating defect rather than an increase of free carriers. Consequently, one should expect low doping efficiencies and high doping-induced deviations from the ideal stoichiometry. Growth under conditions far from thermal equilibrium or introduction of dopants by ion implantation may lead to greater doping efficiency.

We illustrate the above ideas for GaN at T_g = 1300 °C. In both intrinsic and doped material, the concentration of N(I) decreases from about 10^{17} cm^{-3} at N-rich conditions to about 10^{11} cm^{-3} at Ga-rich conditions. In an intrinsic crystal, V_N^+ is a dominant defect for $-1.0 \leq \Delta\mu \leq 1.8$ eV, with concentration of about 10^{18} cm^{-3} in the Ga-rich limit. Accordingly, the concentration of V_{Ga}^{-3} is about three times smaller. The concentration of conduction electrons increases with $\Delta\mu$ from 10^{14} to 10^{16} cm^{-3}, and intrinsic GaN is slightly n type, in qualitative agreement with experiment.

Doping of GaN substantially modifies its stoichiometry. The n-type crystal is N rich over almost the whole range of $\Delta\mu$, and the dominant defect is V_{Ga}. Assuming the concentration of ionized external donors of 10^{18} cm^{-3} we find that the concentration of V_{Ga}^{-3} is nearly three times smaller, and is practically constant over the entire range of $\Delta\mu$. Only in the Ga-rich limit V_N starts to dominate and reaches a concentration of about 10^{18} cm^{-3}. The p-type crystal is Ga rich, due primarily to the formation of Ga(I), although V_N again starts to dominate at the Ga-rich limit. As in the case of n doping, the concentration of Ga(I) is determined by the concentration of external acceptors.

Compensation effects due to intrinsic defects are very large for both types of doping, leading to a very small doping efficiency, see Fig. 3. This is particularly true for p-type doping in the Ga-rich limit, where, e.g., 10^{21} cm^{-3} acceptors result in 10^{14} cm^{-3} holes. The computed very low p-type doping efficiency is in agreement with the experimental data for Mg doping, even in cases where no post-growth electron beam activation is required.[17] The efficiency of n-type doping is significantly larger, see Fig. Fig. 3(b).

Comparing to previous theoretical results, a recent *ab initio* study of native defects in GaN by Neugebauer and Van De Walle[18] obtained results similar to ours for both vacancies and antisites, but there are differences for interstitials. Vacancies and antisites were also studied by Jenkins and Dow[19] using a model tight-binding Hamiltonian. Their results differ significantly from those based on *ab initio* methods. Turning

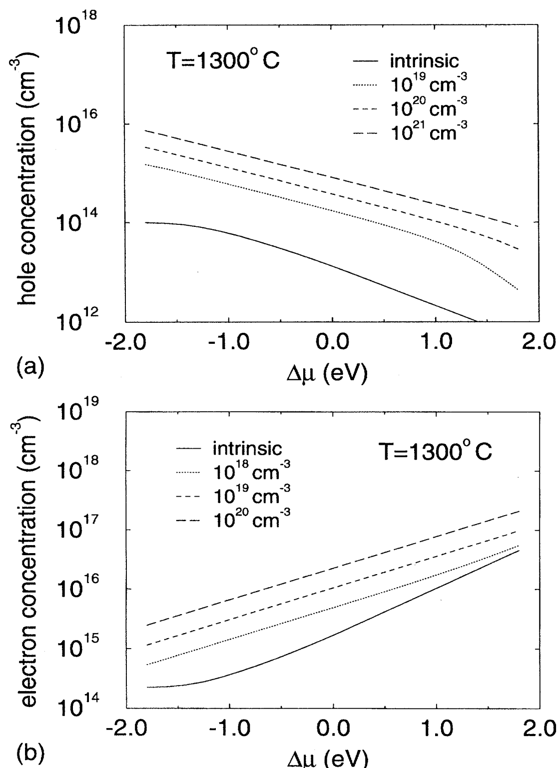

FIG. 3. Calculated concentration of (a) free electrons and (b) free holes for various doping rates.

to the experiment, as-grown undoped samples of GaN are almost always n-type, which was suggested to be due to the formation of nitrogen vacancies,[2–6] but the residual donor(s) were not positively identified. The experimental donor energies are strongly scattered, falling in three ranges:[4,20] 10–40, about 100, and 200–400 meV. Possible explanations include the presence of more than one type of point defect, and/or interaction between point defects and extended defects, which could lead to energy shifts. According to our calculations, under the usual Ga-rich conditions there are two prevalent shallow donors, V_N and Ga(I), but the concentration of V_N is in general higher. The presence of either defect can explain the recently observed pressure-induced freeze-out of free electrons.[6,21] However, the computed formation energies are too large to account for the observed high electron concentrations. This discrepancy may be due to far-from-equilibrium growth conditions that are also suggested by the large number of extended defects present in even state-of-the-art samples. Further experimental work is needed to positively identify the residual donor in GaN. Most likely the donor is a native one, but the n-type character of high-resistance samples could be due to residual impurities. The optically detected magnetic resonance data of Glaser *et al.*[5] indicate the presence of a deep donor level as well as a shallow one for a wide variety of samples. The level structure of Ga(I), unlike that of V_N, could account for this observation. Turning to acceptors, luminescence measurements[22,23,4] systematically find a residual acceptor at ~ 0.2 eV. Lagerstedt and Monemar[22] suggested that it is a native defect, which was later assigned to the gallium vacancy.[4] Our calculations indeed show that V_{Ga} induces a shallow acceptor level.

In summary, we have reported the results of large scale *ab initio* calculations for native point defects in GaN. The results suggest that the residual donors responsible for the n-type character of as-grown GaN are nitrogen vacancies. However, the concentration of Ga interstitials under equilibrium conditions in the usual Ga-rich material can become comparable to that of the vacancy. Both n-type and p-type doping efficiencies are substantially reduced by the formation of V_{Ga}, Ga(I), and V_N, even in the absence of passivating impurities, such as hydrogen.

This work was supported in part by Grants No. ONR N00014-92-J-1477, No. NSF DMR 9408437, and No. KBN 2-P302-124-07. The calculations were carried out at the Pittsburgh Supercomputing Center and ICM, Warsaw.

[1] R. F. Davis, Physica B **185**, 1 (1993); S. Strite and H. Morkoc, J. Vac. Sci. Technol. B **10**, 1237 (1992); H. Morkoc *et al.*, J. Appl. Phys. **76**, 1363 (1994).
[2] H. P. Maruska and J. J. Tietjen, Appl. Phys. Lett. **15**, 327 (1969); B. Monemar and O. Lagerstedt, J. Appl. Phys. **50**, 6480 (1979).
[3] M. Ilegems and M. C. Montgomery, J. Phys. Chem. Solids **34**, 885 (1973).
[4] T. L. Tansley and R. J. Egan, Phys. Rev. B **45**, 10 942 (1992), and references therein.
[5] E. R. Glaser *et al.*, Appl. Phys. Lett. **63**, 2673 (1993).
[6] P. Perlin *et al.*, in *22nd International Conference on the Physics of Semiconductors*, edited by D. J. Lockwood (World Scientific, Singapore, 1995), p. 2383.
[7] W. Walukiewicz, Appl. Phys. Lett. **54**, 2094 (1989).
[8] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[9] C. Wang, Q.-M. Zhang, and J. Bernholc, Phys. Rev. Lett. **69**, 3789 (1992).
[10] G. B. Bachelet, D. R. Hamann, and M. Schluter, Phys. Rev. B **26**, 4199 (1982); X. Gonze, R. Stumpf, and M. Scheffler, *ibid.* **44**, 8503 (1991).
[11] G. Li and S. Rabii (unpublished).
[12] See P. Bogusławski, E. L. Briggs, T. A. White, M. G. Wensell, and J. Bernholc, 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, 1994), p. 693.
[13] E. L. Briggs, D. J. Sullivan, and J. Bernholc (unpublished).
[14] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991); D. B. Laks *et al., ibid.* **66**, 648 (1991).
[15] A. Rubio *et al.*, Phys. Rev. B **48**, 11 810 (1993).
[16] J. D. Chadi, Phys. Rev. B **46**, 9400 (1992).
[17] T. W. Weeks, Jr., M. Bremser, and R. F. Davis (unpublished).
[18] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).
[19] D. W. Jenkins and J. D. Dow, Phys. Rev. B **39**, 3317 (1989).
[20] R. J. Molnar, T. Lei, and T. D. Moustakes, Appl. Phys. Lett. **62**, 72 (1993).
[21] P. Bogusławski and J. Bernholc (unpublished).
[22] O. Lagerstedt and B. Monemar, J. Appl. Phys. **45**, 2266 (1974).
[23] R. Dingle *et al.*, Phys. Rev. B **4**, 1211 (1971); M. Illegems, R. Dingle, and R. A. Logan, J. Appl. Phys. **43**, 3797 (1972).

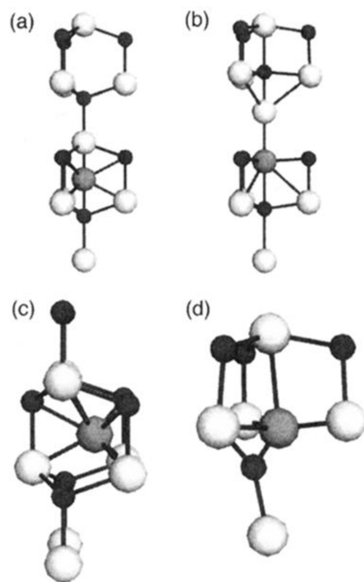

FIG. 2. Atomic configurations of interstitials: (a) the ideal and (b) the relaxed Ga(T) interstitial, (c) the ideal Ga(O) interstitial, and (d) the relaxed N interstitial. Big and small spheres represent Ga and N atoms, respectively.