

Native defects and impurities in GaN

Jörg Neugebauer and Chris G. Van de Walle

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

Summary: Applying state-of-the-art first-principles calculations we study atomic geometry, electronic structure, and energetics for all native defects and for several donor impurities (O, C, Si) in GaN. An analysis of these results gives direct insight into the defect concentrations and the solubility of impurities with respect to the growth conditions (temperature, chemical potentials) and into possible mechanisms for passivation and compensation. Particularly, we discuss in detail the role of the nitrogen vacancy, which is commonly assumed to be the source for the “auto-doping” of GaN. Our results show that GaN has distinctively different defect properties compared to more “traditional” semiconductors such as Si, GaAs or ZnSe. This is explained in terms of the large mismatch in the atomic radii of Ga and N.

1 Introduction

GaN has recently attracted widespread attention for producing blue light-emitting diodes and as a promising candidate for blue lasers and high-temperature or high-power devices [1–3]. All these applications are related to very distinct properties of GaN: a large direct bandgap of ≈ 3.5 eV, strong interatomic bonds and a high thermal conductivity. The interest in GaN has been even more stimulated by the successful fabrication of the first highly efficient blue-light emitting diodes [4]. However, despite this strong interest in GaN, little is known about intrinsic defects and impurities and their properties in GaN.

As-grown GaN is usually *n*-type conductive, and *p*-type conductivity has been difficult to achieve. Only recently *p*-type doping has been achieved by low-energy electron irradiation (LEEBI) [5] and thermal annealing [6]. For almost 25 years the *n*-type “autodoping” of GaN has been commonly associated with the nitrogen vacancy [7,8]. Only a few papers have discussed impurities (mainly oxygen) as a possible source for unintentional doping [9,10].

A first theoretical study about the electronic structure of some of the native defects in GaN has been performed by Jenkins and Dow based on tight-binding calculations [11]. According to their calculations the nitrogen vacancy is a donor. This result is often invoked in support of the nitrogen-vacancy hypothesis. However, the tight-binding calculations have limited accuracy, and in addition give no information about atomic

defect geometry (relaxation, reconstruction) and defect formation energy.

Modern first-principles calculations have had a major impact on the understanding of defects and impurities in semiconductors. With the capability to calculate total energies, it became possible to investigate the atomic structure of the defect; i.e., the stable position in the host lattice, the relaxation of the surrounding atoms, as well as the energy along a migration path [12–14]. More recently, formalisms have been developed to use the total energy of the defect to calculate its concentration, under the assumption of thermodynamic equilibrium [15,16]. The same formalism can also be applied to the calculation of impurity solubilities [17,18].

In the present paper we will give an overview about theoretical results for native defects and impurities in GaN. Particularly, we discuss the electronic structure, the atomic geometry, and the energetics for all native defects and for several donor impurities (O, C, Si) in GaN. Based on these results the dominant native defects are identified and we will discuss how the various factors may influence and limit the doping in GaN.

2 Formalism

2.1 Definitions

The energy necessary to create a defect is called the *defect formation energy*. This energy is not constant but depends on the specific growth conditions. Specifically, in GaN the relative abundance of Ga and N atoms during the crystal growth determines the formation energy. In a thermodynamic context, these abundances are described by the chemical potentials (μ_{Ga} , μ_{N}) which define the reservoir from where Ga and N atoms are taken or brought in order to create defects. If the defect is charged, the formation energy also depends on the position of the Fermi level E_F from where electrons are taken to charge the defect. Applying this concept, the formation energy of a defect is then defined as:

$$E^f(q) = E^{\text{tot}}(q) - n_{\text{Ga}}\mu_{\text{Ga}} - n_{\text{N}}\mu_{\text{N}} - qE_F \quad , \quad (2.1)$$

where n_{Ga} and n_{N} are the number of Ga and N atoms and q the charge state of the defect. $E^{\text{tot}}(q)$ is the total energy for a given charge state; it can be directly calculated. The chemical potentials for Ga and N are not independent variables, since both species are in equilibrium with GaN:

$$\mu_{\text{Ga}} + \mu_{\text{N}} = \mu_{\text{GaN(bulk)}} \quad . \quad (2.2)$$

Thus, only one chemical potential can be chosen freely. For the following discussion the gallium chemical potential is used as independent variable.

The choice of the gallium chemical potential is not completely free but has to obey certain boundary conditions. A major criterion is that the chemical potential for an element is less than the chemical potential of the corresponding bulk (or molecule for N) since otherwise this element would form the energetically more stable bulk or molecular structure. For the gallium chemical potential an upper limit is therefore given if GaN is in thermodynamic equilibrium with bulk Ga. This case will be called the Ga-rich limit. The lower limit is given for GaN in thermodynamic equilibrium with N₂ molecules; it is therefore called nitrogen-rich limit. Using these relations and Eq. (2.2) we get:

$$\mu_{\text{Ga(bulk)}} + \Delta H^{\text{GaN}} \leq \mu_{\text{Ga}} \leq \mu_{\text{Ga(bulk)}} \quad (2.3)$$

Here, ΔH^{GaN} is the *heat of formation* which is defined as:

$$\Delta H^{\text{GaN}} = \mu_{\text{GaN(bulk)}} - \mu_{\text{Ga(bulk)}} - \mu_{\text{N}_2(\text{molec})} \quad (2.4)$$

A negative heat of formation means that the reaction is exothermic. The corresponding bulk chemical potentials are calculated from the bulk forms Ga (orthorhombic), N (N₂ molecule) and GaN (wurtzite).

The above described formalism can be easily generalized to impurities by rewriting Eq. (2.1)

$$E^f(q) = E^{\text{tot}}(q) - \sum_{\text{A}} n_{\text{A}} \mu_{\text{A}} - qE_{\text{F}} \quad (2.5)$$

where A goes over Ga, N and the impurity atoms [18].

In order to compare the stability of different structures we have to employ the *free energy* instead of the *total energy*. The difference between both energies is the energy contribution $-TS$ which can be divided into vibrational and configurational entropy. For point defects the configurational entropy is simply given by the number of sites at which the defect can be created. The vibrational entropies are, at the present stage of our work, not explicitly included, which would be computationally very demanding. Such entropy contributions cancel to some extent [19], and are small enough not to affect any qualitative conclusions. However, more powerful computers and improved methods will make accurate calculations of vibrational entropies more feasible to be carried out for various systems in the near future. The effect of entropy can be included in an approximate way; experimental and theoretical results show that the entropy S is typically in the range between 0 (no entropy contributions) and $10 k_B$ (where k_B is the Boltzman constant) [20]. A simple estimate based on an Einstein model for the phonon frequencies gives values between 3 and 5 for the native defects in GaN.

From the formation energy of a defect or impurity and its entropy S we can determine the equilibrium concentration:

$$c = N_{\text{sites}} \exp^{S/k_B} \exp^{-E_f/k_B T} \quad (2.6)$$

where N_{sites} is the appropriate site concentration. For a substitutional impurity it is the number of Ga or N sites, i.e., $N_{\text{sites}} = 4.4 \times 10^{22} \text{cm}^{-3}$. Whether the assumption of thermodynamic equilibrium is satisfied depends on the mobility of the defect at the temperatures of interest. If the mobility is too low to realize thermodynamic equilibrium the bulk defect concentration would be controlled by surface kinetics.

2.2 Method

The total energies are obtained from first-principles calculations employing density-functional theory [21] and *ab initio* norm-conserving pseudopotentials [22,23]. The wave functions are expanded in a plane-wave basis set. The strong localization of the nitrogen $2s$ and $2p$ orbitals and of the gallium $3d$ electrons makes such an approach challenging. The problem becomes tractable by using soft Troulier-Martins pseudopotentials [24] and a highly efficient code [25,26]. Additionally, we implemented a method to obtain a very good guess for the initial wave functions which is based on expanding the initial wave functions in a tight-binding basis set [26]. This method gives a considerable improvement in the total-energy convergence.

We performed a great number of convergence checks to test the reliability and accuracy of the basis set, the pseudopotentials and the supercell size. These checks reveal two important aspects: (i) the Ga $3d$ electrons play an active role for the chemical bonding in GaN and cannot simply be treated as core electrons and (ii) a good k -point sampling is essential [27]. The active role of the Ga $3d$ electrons is at first glance surprising since the corresponding level in GaN is located 14 eV below the top of the valence band. In order to identify the mechanism one has to note that the Ga $3d$ electrons strongly hybridize with the N $2s$ electrons. This effect alone would have only a minor influence on the total energy since both bonding and antibonding states are occupied. However, the hybridization between the Ga $3d$ and the N $2s$ orbitals also changes the *hybridization of the nitrogen atom*, giving rise to a change in the N $2p$ levels. Since the top of the valence band is mainly characterized by the N $2p$ levels, the strong dependence of the GaN bonding properties on the deep lying Ga $3d$ electrons becomes understandable.

Taking the Ga $3d$ electrons explicitly into account is also essential for a realistic description of the bulk properties; if the d electrons are treated in a “frozen-core” approximation the calculated lattice constant is 3% too small [28]. Further, without taking the $3d$ electrons explicitly into account the formation enthalpy is too small by about 50%, indicating that the Ga $3d$ electrons enhance the bonding in GaN [26].

The second aspect – the good k -point sampling – is related to the fact that we describe not an isolated defect but a defect in a supercell, i.e., even for supercells consisting of 32 or even 72 atoms there is a strong dispersion of the defect levels due to defect-defect interaction. One possible solution to overcome this problem would be to use larger supercells; however, the computational demand would be prohibitive even for modern supercomputers. Another approach is based on the fact that an *average* of the defect band is very close to the level of the isolated defect. The averaging, which is basically

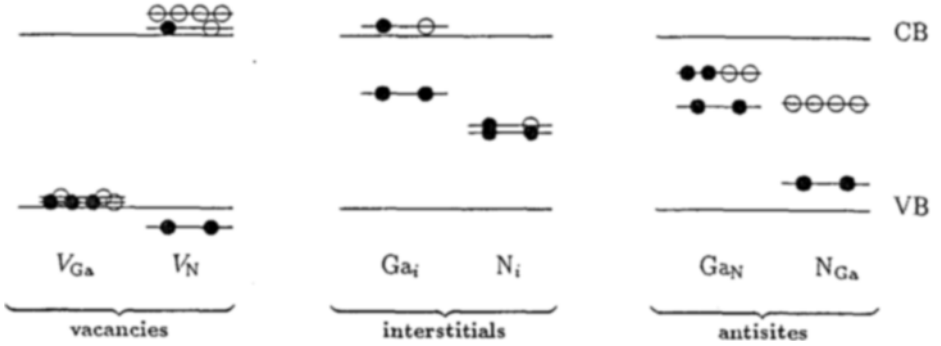


Figure 1 Schematic representation of the defect levels in GaN. The occupations are shown for the neutral charge state; filled circles indicate electrons, open circles indicate holes.

an integration over the Brillouin zone, can be performed at special k points. Based on the above argumentation it becomes immediately evident that the Γ point, which is frequently used due to its numerical simplicity, gives the worst description since there the defect-defect interaction reaches its maximum. We therefore use special k points to describe the defect levels properly. However, despite this procedure we expect that our defect levels have substantial error bars; the positions of the defect levels should be therefore used only as a semiquantitative guide.

3 Native defects in GaN

Based on the methods described above we have calculated the defect levels and the formation energies for all native defects in GaN: the vacancies (V_{Ga} , V_N), the antisites (Ga_N , N_{Ga}) and interstitials (Ga_i , N_i). All relevant charge states were taken into account. Atomic relaxation was allowed for all atoms within a sphere of 7 bohr around the defect.

3.1 Electronic structure and atomic geometry

We will first focus on the electronic structure of the native defects in GaN, particularly on the energetic position and the character of the defect-induced states in the band structure. The position of the defect levels and the occupation for the neutral charge state are shown in Fig. 1 for all native defects.

Nitrogen vacancy

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 are split into a singlet and a doublet state

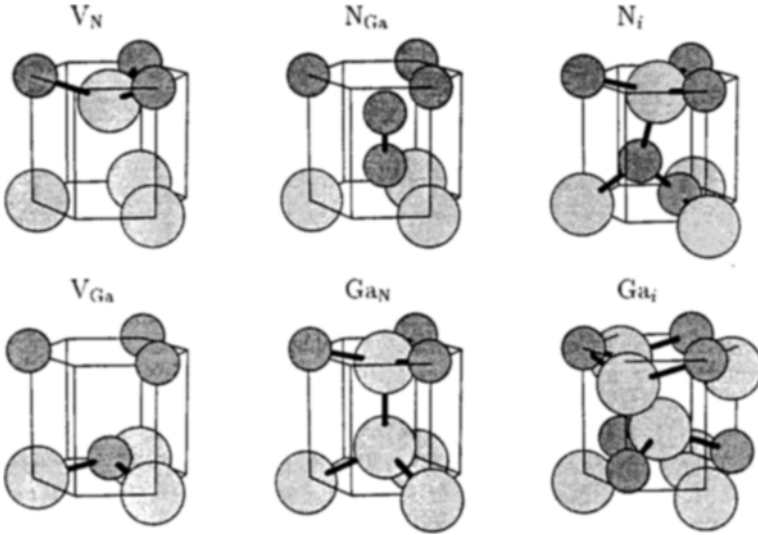


Figure 2 Ball and stick models of the atomic geometry for the native defects in wurtzite GaN. The large balls represent the Ga atoms, the smaller balls the nitrogen atoms. Bonds are drawn if the distance is smaller than 3.8 bohr. The hexagonal prisms are a guide for the eye and represent the hexagonal symmetry of the wurtzite structure. For the Ga_i a different perspective is chosen: the center of the hexagonal prism is shifted to the middle of the hexagonal rings.

which both lie as resonances in the conduction band. The splitting is a result of the lower symmetry of the wurtzite compared to the zinc-blende structure [27]. For the neutral charge state the one electron on 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 (TB) calculations [11,29], which have been used to interpret some experimental results [30]. According to the TB calculations the doubly occupied A_1 state lies close to the conduction band.

The main difference is the unusually large splitting of the A_1 and T_2 defect levels which is not correctly reproduced in the TB calculations. This large splitting originates from a peculiar property of GaN: due to the much smaller size of the nitrogen atoms compared to the Ga atoms the Ga-Ga distance in GaN is comparable to that in bulk Ga resulting in metallic-like bonds between the Ga atoms surrounding the nitrogen vacancy [31]. This strong interaction between the Ga atoms, which are second nearest-neighbors, explains the large splitting of the A_1 and T_2 defect levels. It also explains the failing of the TB-calculations where mainly first-nearest neighbor interaction is taken into account.

The atomic structure of the nitrogen vacancy is characterized by a small outward breathing relaxation. Due to the wurtzite symmetry the 4 Ga neighbors can be divided into 3 equivalent atoms lying in a plane perpendicular to the c axis and an inequivalent

atom on the c axis above the defect (see Fig. 2). The equivalent neighbors move outward by .5% of the bond length, the inequivalent neighbor by 1.5%. The relaxation energy is small (≈ 50 meV).

As pointed out by Northrup for the As vacancy in GaAs [32] and by Garcia and Northrup for the Se vacancy in ZnSe [33] a large outward relaxation of the four atoms surrounding the vacancy raises the energy of the A_1 level and eventually shifts it into the band gap. If the Fermi level lies below the defect level, the A_1 level becomes unoccupied. For the nitrogen vacancy this means that a 3+ charge state may become stable. Indeed, we find a transition level $E^{3+/+} \approx 0.16$ eV; for Fermi energies below 0.16 eV the energetically most stable charge state is 3+. This charge state is accompanied by a large outward relaxation of the surrounding Ga atoms: the three equivalent neighbors move outward by 11% and the inequivalent neighbor by 19%!

Gallium vacancy

The Ga vacancy is triply occupied with levels close to the valence band. The p -like T_2 states are split in a singlet and a doublet state. Similar as for the nitrogen vacancy we find an outward breathing relaxation. The relaxation of the three equivalent nitrogen atoms ($\approx 5\%$) is only slightly dependent on the charge state whereas the inequivalent nitrogen atom (where the singlet state is located) moves strongly outward with increasing negative charge state (3.7% ... 10%) [31].

Nitrogen interstitial

Interstitial energies were calculated not only for high symmetry positions like the tetrahedral or octahedral site but for several low symmetry configurations. Particularly for the nitrogen interstitial, where a *small* atom is brought into the crystal, we have to anticipate the possibility of low symmetry configurations and several local minima. In order to derive the *global minimum* we calculated a complete total-energy surface for a N interstitial in GaN. Fixing the coordinates of the additional nitrogen atom all surrounding atoms were allowed to relax. The calculations were performed for 50 different positions of the nitrogen atom and for different charge states in a 16 atom supercell. The local minima obtained by this procedure were then investigated in a larger supercell. We identified a (100) split-interstitial configuration, in which two nitrogen atoms share the same substitutional site (see Fig. 2), as the lowest energy configuration. In this configuration each nitrogen atom in the defect now has two bonds to the surrounding gallium atoms instead of four. The preference of this state can be understood in terms of the large binding energy nitrogen has in a nitrogen molecule: the large bond strength stabilizes a configuration where a nitrogen-nitrogen bond is formed rather than the tetrahedral or octahedral site where only less strong Ga-N bonds are formed.

In the neutral charge state this defect has two singlet states in the band gap occupied with three electrons (see Fig. 1). Depending on the position of the Fermi level the nitrogen interstitial may act both as an acceptor or a donor.

The atomic geometry of the nitrogen interstitial depends strongly on the charge state. By going from the 1− charge state (where all defect levels are completely filled) to the 3+ charge state (where all defect states are empty) the N-N bond length decreases and the Ga-N bond length increases. This can be understood by noting that the defect level are N-N ($\pi\pi$ -like) antibonding orbitals but σ GaN bonding orbitals. Thus, by systematically filling the defect levels the bond length for the bonding states (Ga-N) shortens but becomes larger for the antibonding states (N-N).

Gallium interstitial

For the gallium interstitial the octahedral interstitial site is energetically most stable. The gallium interstitial induces two defect levels, a deep donor level and a resonance in the conduction band. In the neutral charge state the deep donor level is doubly occupied, the level in the conduction band singly occupied. Similar as for the nitrogen vacancy this electron will be donated to the bottom of the conduction band where it forms a shallow donor level. The interstitial Ga atom forms nearest neighbor bonds to the three surrounding nitrogen atoms and second nearest neighbor bonds to the 6 surrounding Ga atoms. By charging the defect from the 3+ to the 1+ charge state the Ga-N bond length increases from −4% to 1%.

Boguslawski *et al.* find, using a similar method, the tetrahedral site as another local minimum [34]. According to their calculations both the tetrahedral and the octahedral site are energetically nearly equivalent (with the tetrahedral site more stable by 0.2 eV). Our calculations, in contrast, show that this site is energetically less favorable by nearly 1 eV. The transition levels are $E^{3+/2+} \approx 1.9$ eV, $E^{2+/+} \approx 2.6$ eV. The fact that $E^{3+/2+} < E^{2+/+}$ is characteristic of a “negative- U ” effect and indicates that the 2+ charge state is thermodynamically unstable for this configuration.

Gallium antisite

The gallium antisite Ga_N has a singlet and a doublet state in the band gap. The substitutional Ga atom forms four covalent-like bonds to the surrounding nearest neighbor Ga atoms (see Fig. 3). The bond length between the Ga atoms is ≈ 4.0 bohr (neutral charge state) and much shorter than the bond length in bulk Ga (4.61 bohr) [35]. The shorter Ga-Ga bond length for the defect compared to the metallic bond length is characteristic for a covalent bond (see also Fig. 3). Further, the Ga-Ga bond length is 12% larger than the bulk Ga-N bond length indicating a large strain around this defect. Indeed, atomic relaxation lowers the formation energy by nearly 5 eV. The strong outward relaxation of the neighboring Ga atoms decreases by occupying the defect levels, from 23% (+4 charge state) to 12% (−2 charge state).

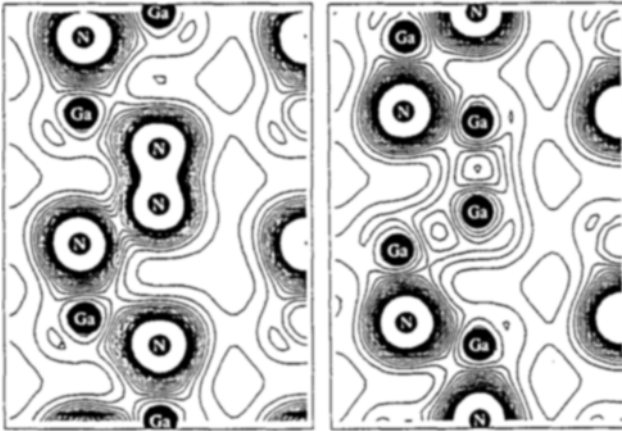


Figure 3
Contour plot of the valence charge density for the N antisite (left) and Ga antisite (right). The contour spacing is $0.01 e \text{ bohr}^{-3}$.

Nitrogen antisite

The nitrogen antisite has a singlet and a doublet state in the band gap. Contrary to the gallium antisite, where the substitutional Ga atom remains nearly on the lattice site, for the nitrogen antisite a strong distortion is found. The substitutional nitrogen atom moves toward the inequivalent N neighbor and forms a strong N–N bond (see Fig. 2). For the $2+$ charge state the bond length is 2.2 bohr, very close to the bond length of a N_2 molecule (2.07 bohr). The N–N bond length is very sensitive to the charge state: filling up the defect levels increases the bond length up to 2.8 bohr ($4-$ charge state). The distance to the 3 equivalent neighbors is much larger (3.9 bohr) indicating that no bond is formed (see also Fig. 3). Thus, the nitrogen antisite is characterized only by one N–N bond. This can be explained, again, in terms of the large bond strength of a nitrogen atom in a N_2 molecule.

Discussion

From the electronic structure calculations we can immediately classify the defects into donors, acceptors and amphoteric defects. The nitrogen vacancy and the gallium interstitial are donors, the gallium vacancy an acceptor and the nitrogen interstitial, and both antisites are amphoteric. First-principles calculations performed by Boguslawski *et al.*, which were mainly restricted to the neutral charge state, are in qualitative agreement with this conclusion [34]. We note, however, that their calculated formation energies deviate significantly for some of the defects.

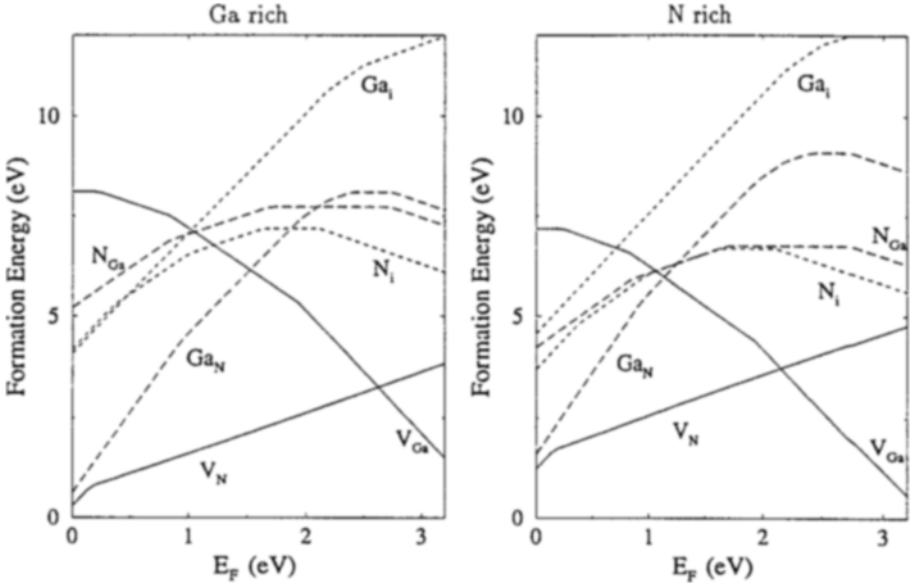


Figure 4 Defect formation energy as a function of the Fermi energy for vacancies (solid lines), antisites (long-dashed lines), and interstitials (dashed lines). Ga-rich conditions (left) and N-rich conditions (right) are assumed.

3.2 Energetics

We have calculated the formation energy, as a function of the Fermi level and of the chemical potentials, for all native point defects in GaN. The results are displayed in Fig. 4 for both Ga-rich conditions (which appear to be common in experimental growth conditions) and N-rich conditions. The formation energies are slightly different from those given in Ref. [31]; the atomic relaxation includes now additional nearest neighbor shells. The most striking feature is the high formation energy of the self-interstitials (Ga_i , N_i) and antisite defects (Ga_N , N_Ga). The only exception is Ga_N under extreme p -type conditions. However, under these conditions the nitrogen vacancy is lower in energy rendering it the dominant defect. Antisites and interstitials are therefore unlikely to occur in any significant concentrations. This behavior is different from that found for more “traditional” semiconductors like Si, GaAs, or ZnSe where antisites and interstitials are well known to play an important role. In the following discussion we will point out two mechanisms responsible for this behavior: the large mismatch in the atomic radii between Ga and N and the wide band gap.

The kinks in the formation energies (see Fig. 4) are caused by a change in the defect charge state; increasing the Fermi level makes charge states with a higher filling more favorable. Both the sign and the value of the slope identify the charge state: a positive slope indicates a positive charge state, a negative slope a negative charge state. The

position of the Fermi level at which the slope changes, i.e., the position of the kink, identifies the transition level.

The high formation energy of antisites and interstitials can be explained in terms of the large mismatch in the covalent radii of Ga and N. When creating a native defect, two cases have to be distinguished. If a *large atom* (Ga) is brought into the crystal (either replacing a small N atom, or as an interstitial) the atoms around the defect have to be moved away from the defect which causes a lot of strain energy. Although atomic relaxation significantly reduces the formation energy [by up to several eV (!)] it cannot completely compensate the high formation energy.

If a *small atom* (N) is brought into the crystal (N_i , N_{Ga}) the initial bond length is too *long* to form nearest neighbor bonds. The system has two possibilities to shorten the bond length: reducing symmetrically all bond lengths of the surrounding neighbors or breaking the symmetry and forming low-symmetry configurations. For both the N interstitial and the N antisite we find that low-symmetry configurations are preferred. The reason is the tendency of nitrogen to rather form *one, strong* N-N bond (like in a nitrogen molecule) than assume a configuration where a nitrogen atom has a higher coordination with *several, weak* N-N bonds.

It is further interesting to note that the formation energy of all native defects is highest for the neutral charge state; other charge states can occur with much lower energies. Thus, to compare different defects it is extremely important to take *all* charge states into account.

The defects with the lowest formation energies are the vacancies: under *p*-type conditions the nitrogen vacancy has a very low formation energy, under *n*-type conditions the Ga vacancy. Under conditions of thermodynamic equilibrium, a low formation energy is required for the defect to occur in large concentrations. Based on Eq. (2.6) the defect concentration as function of the temperature can be calculated. It is important to note that the defects are not independent from each other: they are coupled via the condition of charge neutrality. The total sum of electrons in the conduction band, holes in the valence band and charges on the defects must be zero [20].

Using this formalism we can directly calculate the concentration of the native defects as a function of growth temperature (Fig. 5). For this calculation undoped GaN is assumed; no impurity atoms are included. As expected from the low formation energies only the vacancies occur in significant concentrations. For all growth temperatures the nitrogen vacancy is the defect with the highest concentration. However, even at growth temperatures as high as 1300 K (typical for metalorganic chemical vapor deposition) the nitrogen vacancy concentration is orders of magnitude too small to explain the observed high carrier concentration ($10^{18} \dots 10^{20} \text{ cm}^{-3}$) in "auto-doped" GaN. In *thermodynamic equilibrium*, the nitrogen vacancy can therefore be excluded as a source for the *n*-type doping of as grown GaN.

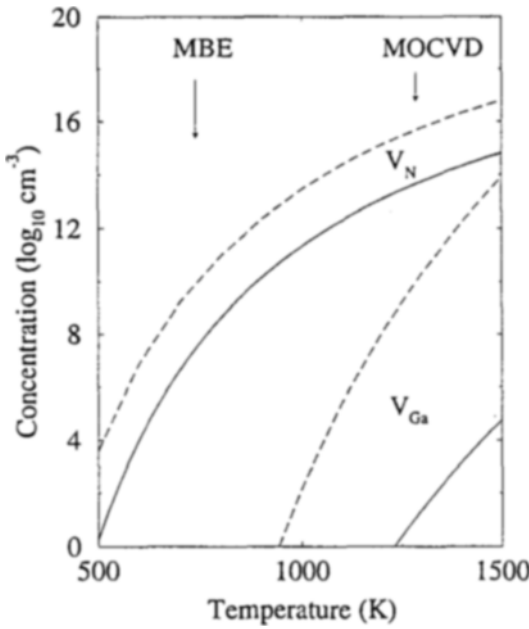


Figure 5

Equilibrium concentration of native defects in undoped GaN versus growth temperature. Ga-rich conditions are assumed. Lower (solid line; $S = 1k_B$) and upper limit (dashed line; $S = 10k_B$) are shown.

3.3 Kinetics

What about the possibility that the nitrogen vacancy is incorporated via a *non-equilibrium process*? Non-equilibrium implies that the concentration of nitrogen vacancies is prevented from reaching its equilibrium value, i.e., the nitrogen vacancies would be frozen in. This would only be possible if the nitrogen vacancies are sufficiently *immobile* to prevent any significant flux of nitrogen vacancies between bulk-like regions and possible sinks or sources of nitrogen atoms (such as surfaces or dislocations). It is therefore important to study the diffusion properties of nitrogen vacancies or equivalently of nitrogen atoms.

The migration of the nitrogen atoms can be realized by several diffusion channels. Nitrogen atoms could diffuse interstitially (and eventually annihilate nitrogen vacancies); however, the high formation energy of the nitrogen interstitial (see Fig. 4) implies a low probability for this channel. The other possible mechanism is the migration of the nitrogen vacancy: one of the neighboring nitrogen atoms around the nitrogen vacancy jumps into the nitrogen vacancy annihilating the vacancy and creating a new vacancy at the original position of the nitrogen atom. We have studied the migration path of the nitrogen vacancy in detail [26]. The saddle-point configuration which determines the diffusion barrier is a so called split-vacancy: the migrating nitrogen atom is shared by two nitrogen vacancies. Preliminary results give a value of ≈ 4 eV for the diffusion barrier indicating that at low growth temperatures the nitrogen vacancy is immobile; at higher temperatures, it becomes sufficiently mobile to realize thermodynamic equilibrium.

This conclusion is also consistent with the experimentally observed fact that the carrier concentration of as-grown GaN *decreases* with *increasing* growth rate [9]. Indeed, if the nitrogen vacancy concentration (and hence the carrier concentration) would be controlled via a non-equilibrium process the opposite behavior would be expected: the deviation from thermodynamic equilibrium should be larger at higher growth rate. The fact that a *higher* growth rate leads to a *lower* carrier concentration is indicative of another doping mechanism – for instance, unintentional incorporation of donor impurities, to be discussed later.

Based on the above discussion we can therefore assume that the concentration of nitrogen vacancies is allowed to reach its equilibrium value – which, however, is too low to explain the observed *n*-type conductivity in as-grown GaN. If the nitrogen vacancy can be excluded as a source, what causes the *n*-type conductivity? First, we note that our results apply only to *isolated point defects*. It is possible that defect related *complexes* or *extended defects* are involved. Further, the calculations were performed for *bulk* GaN. Defect formation energies may be very different near extended defects such as grain boundaries and dislocations which are known to occur in unusually large concentrations in most epitaxial GaN grown to date [36]. As an alternative mechanism we discuss unintentional incorporation of donor impurities during growth in Sec. 4.

4 Donor impurities in GaN

In the previous Section we pointed out that unintentional incorporation of donor impurities during growth may be responsible for the *n*-type doping of as-grown GaN. Silicon, oxygen and carbon are common contaminants in the systems typically used for growth of GaN. In this Section we will therefore focus on the electronic properties of these impurities and, based on thermodynamic models, estimate their solubility.

The first question which has to be addressed for each impurity is its location, i.e., which site is preferred. Possible sites are the Ga and N substitutional sites, as well as interstitial configurations. The site preferred may depend on the chemical potentials (growth conditions) or the position of the Fermi level [18].

4.1 Silicon

For a Si impurity in GaN we find that the nitrogen substitutional site and the interstitial configurations are energetically unfavorable. Si_{Ga} is an energetically very stable configuration. This result can be understood by noting that silicon has an atomic radius very similar to gallium. Thus, while easily fitting in on a Ga site, it causes a large strain if it replaces a small N atom or goes on an interstitial site.

For the following discussion we focus therefore on Si on the Ga site. For this position we find silicon to be a single donor with transition levels about 0.1 eV below the bottom

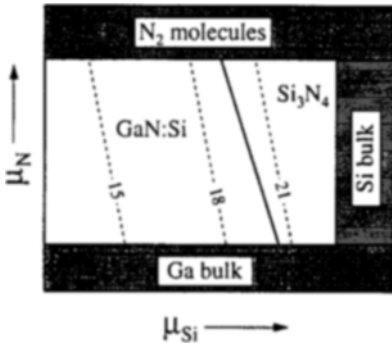


Figure 6

Equilibrium concentration of Si donors in GaN as a function of the nitrogen and silicon chemical potentials. The thick solid lines are the phase boundaries for the thermodynamically most stable phases. The numbers at the contour lines give the \log_{10} Si concentration.

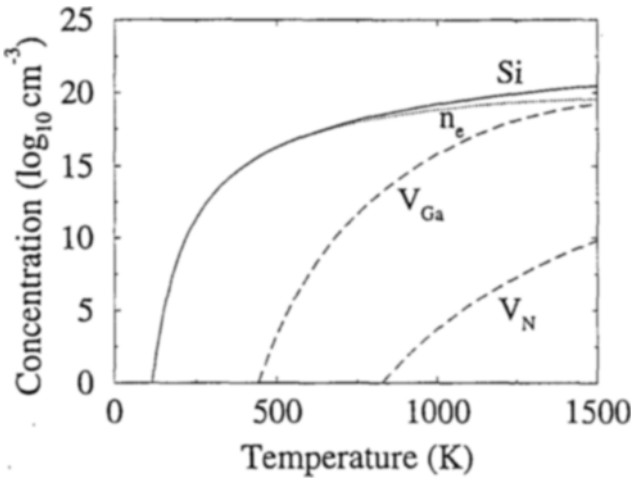
of the conduction band indicating a shallow donor. In order to decide whether this level is really a shallow donor or whether it lies only accidentally close to the conduction band we calculated the pressure dependence of the defect level. With increasing pressure the donor level closely follows the conduction band; we can therefore conclude that Si_{Ga} is a shallow donor.

The formation energy for this site can be separated into two independent processes: (i) the removal of a Ga atom (creation of a Ga vacancy) and (ii) the chemisorption of a Si atom on this site. For both processes the energy depends on the specific growth conditions which are characterized by the chemical potentials. The removed Ga atom has to be brought to a reservoir where it has the energy μ_{Ga} . If this energy is high (gallium-rich environment) the formation energy will be high as well. The same argumentation can be applied to the chemisorption of Si. The formation energy thus depends on two external parameters (μ_{Ga} and μ_{Si}), and will be minimum if μ_{Ga} reaches its lowest value (N-rich conditions) and μ_{Si} reaches its maximum value (Si-rich environment). In order to find an upper limit for the Si chemical potential one has to consider not only the formation of bulk Si but also chemical compounds of Si with Ga and N. Indeed, we find that the thermodynamically most stable configuration is not bulk Si but Si_3N_4 . The chemical potentials then have to obey the relation:

$$3\mu_{\text{Si}} + 4\mu_{\text{N}} < \mu_{\text{Si}_3\text{N}_4} \quad ; \quad (4.7)$$

otherwise Si_3N_4 precipitates would be formed.

Another value which enters the formation energy is the position of the Fermi level. However, the Fermi level is not a free parameter but unambiguously determined by the condition of charge neutrality. The computation of the impurity concentration therefore has to be done self-consistently; in this way compensation by native defects is taken inherently into account [18]. The formation energy and via Eq. (2.6) the equilibrium concentration of the impurities can be expressed as a function of the Si and Ga chemical potentials and the temperature. Entropy effects are neglected in the following discussion. Figure 6 shows the concentration of Si atoms in GaN at a temperature of 1300 K which

**Figure 7**

Solubility of Si donors in GaN as function of temperature assuming thermodynamic equilibrium. Also included are the calculated concentration of native defects (dashed lines) and the electron carrier concentration (dotted line).

is a typical growth temperature for MOCVD. As expected, the concentration increases going from gallium-rich conditions to nitrogen-rich conditions if the silicon chemical potential is kept constant. The solubility is limited by the formation of Si_3N_4 . It is interesting to note that the maximum solubility is not reached under N-rich conditions (where the removal of the Ga atom is easiest) but under Ga-rich conditions. The reason is that nitrogen-rich conditions promote the formation of the Si_3N_4 compound.

As discussed above the maximum solubility is achieved under Ga-rich conditions and if μ_{Si} reaches its maximum value consistent with Eq. (4.7). Figure 7 shows the maximum solubility as a function of temperature. This Figure also includes the concentrations of the native defects. One should keep in mind that these concentrations are not independent but coupled via the condition of charge neutrality. The dominant native defect is the Ga vacancy; the nitrogen vacancy occurs in significantly smaller concentrations. All other native defects occur only in negligible concentrations.

The maximum silicon concentration which can be incorporated into GaN increases with temperature. However, with increasing temperature the Ga vacancy concentration increases, and approaches the Si concentration. Since the Ga vacancy is an acceptor, it will partly compensate the Si donors. This mechanism becomes more significant with increasing temperature. Eventually, the carrier concentration remains constant although the Si concentration increases (see Fig. 7). The reason is that with increasing Si (donor) concentration the Fermi level increases which makes the Ga vacancy energetically more favorable (see Fig. 4).

We note that the concentrations shown in Fig. 7 should be used only as a semiquantitative guide; the expected error bars for the formation energies of a few tenths of an eV may give rise to changes in the concentrations of more than an order of magnitude. Further, all the concentrations were computed assuming that the system is in thermodynamic equilibrium. This assumption works well for many impurities in semiconductors

such as Si, GaAs, or ZnSe; however, little is known about the diffusivity of dopants in GaN.

4.2 Oxygen

Oxygen is another impurity which is likely to occur during growth. Whereas Si is generally accepted to be a shallow donor which can be incorporated in high concentrations in GaN, little is known about oxygen. The few experimental results which are available indicate that oxygen is a donor and can occur in high concentrations [9,10].

We have studied different sites for O in GaN: the substitutional sites, O_{Ga} and O_N , as well as several interstitial configurations. Among the interstitial sites a split-interstitial configuration was found to be most stable indicating the tendency of nitrogen and oxygen to form stable molecules. For oxygen on the Ga site, the atomic geometry is similar to the nitrogen antisite: the oxygen impurity moves toward the inequivalent nitrogen neighbor and forms a strong N-O bond. The resulting bond length is 2.9 bohr, i.e. significantly shorter than the Ga-N bulk bond length (3.57 bohr) but also clearly longer than the bond length of a NO molecule (2.17 bohr). The distance to the inequivalent neighbors is 4.1 bohr, indicating that no N-O bonds to the inequivalent neighbors are formed. O_{Ga} is mainly realized by *one* strong N-O bond.

It is tempting to conclude that the formation of molecular-like structures is a general property of light elements on the Ga site in GaN. However, we will see that, e.g., carbon prefers a symmetric configuration.

Within the physically allowed range of chemical potentials and Fermi energies oxygen on a nitrogen site is the energetically preferred configuration. This can be understood by the fact that the atomic radius of oxygen is very similar to that of nitrogen. Therefore, substituting nitrogen with oxygen will cause only a small elastic strain around the defect. Indeed, we find that the O-Ga bond length is only 0.5% larger than the bulk Ga-N bond.

We can therefore conclude that oxygen sits on a nitrogen site. On this site oxygen has one level (close to the conduction band) in the band gap which is singly occupied in the neutral charge state. Oxygen therefore acts as a single donor, as expected for a group VI element on a group V site. Using the same formalism as applied for the Si impurity we can estimate the solubility of O in GaN. The maximum oxygen concentration which can be achieved is limited by the formation of gallium oxides (GaO , Ga_2O_3). The most stringent condition arises from the formation of Ga_2O_3 . At a growth temperature of 1300 K we find a solubility limit of $\approx 10^{18} \text{ cm}^{-3}$.

To summarize, oxygen can be built in in high concentrations in GaN where it acts as a donor. If the oxygen incorporation exceeds a certain limit, and assuming thermodynamic equilibrium, Ga_2O_3 precipitates may be formed which could affect the electrical and optical properties of GaN in an undesirable way.

4.3 Carbon

The role of carbon in GaN is controversial. In 1980 Ogino and Aoki [37] presented strong evidence that carbon is related to the parasitic yellow luminescence which hampers optoelectronic applications. These authors proposed that the yellow luminescence is realized by a transition from a shallow donor level to a deep acceptor level at 0.8 eV. The deep acceptor level was tentatively assigned to a complex between carbon and a Ga-vacancy (C_N-V_{Ga}), where C substitutes for one of the nitrogen atoms around the Ga vacancy [37]. Very recently Glaser *et al.* [38], based on ODMR measurements, suggested an alternative model: the yellow luminescence was attributed to a transition between a deep donor level (0.8 eV below the conduction band) and a shallow acceptor which was assumed to be C on a N site. The deep donor level was proposed to arise from a native defect, possibly the Ga interstitial [38].

We have studied several possible geometries for C in GaN: the substitutional sites (C_{Ga} and C_N), interstitial geometries, and additionally the C_N-V_{Ga} complex as suggested by Ogino and Aoki [37].

C_{Ga} exhibits a very different atomic structure than calculated for the chemically similar defect structures such as the N_{Ga} antisite or the O_{Ga} impurity. Whereas N and O on the Ga site prefer an asymmetric configuration with only *one* bond to a nitrogen atom C remains on the substitutional site and forms *four* nearly symmetric bonds to its neighbors. The C-N bond length is ≈ 3.0 bohr (16% smaller than the Ga-N bulk bond length).

C_N is a shallow acceptor and has a very low formation energy varying from -0.2 eV (Ga rich) to 0.9 eV (N rich) under extreme *n*-type conditions. Thus, if carbon is present during growth it can be easily incorporated in *n*-type GaN and will partly compensate part of the *n*-type conductivity.

For the C_N-V_{Ga} complex we find a doublet state at ≈ 0.7 eV and a singlet state at ≈ 1.2 eV. In the neutral charge state the doublet state is doubly occupied and the singlet state is empty. The formation energy for the neutral defect is very high (8.1 eV); however, under extreme *n*-type conditions its formation energy is only slightly larger than for C_N .

We can therefore derive two conclusions: (i) under *n*-type conditions, C_N (either as a point defect or a C_N-V_{Ga} complex) can be incorporated into GaN in high concentrations whereas (ii) under *p*-type conditions C_N has a high formation energy and therefore a low solubility. Let us now come back to the problem of the yellow luminescence. The main difference between the two models (Ogino and Aoki & Glaser *et al.*) is whether C is responsible for a deep acceptor level (at ≈ 0.86 eV) [37] or for a shallow acceptor level [38]. From the electronic structure calculations we find that C_N acts as a shallow acceptor, and the $V_{Ga}-C_N$ complex as a deep acceptor. Since both defects have under *n*-type conditions a low formation energy our results are consistent with both models.

If the model proposed by Glaser *et al.* [38] is correct one question still remains: What is the origin of the deep donor level? Native defects are one possibility. The native

defect with the lowest formation energy under *n*-type conditions is the Ga vacancy; it is therefore tempting to associate the deep level with the $E^{2-}/3-$ transition. However, deep-level spectroscopy by Götz *et al.* shows not *one* but *four* deep levels which would also account for the unusual broadness of the yellow luminescence [39]. It is possible that the native defect in question occurs not as an isolated point defect, but near an extended defect such as a dislocation or grain boundary, which would affect its transition levels. Very recent experiments by Ponce *et al.* using spatially resolved cathodoluminescence spectroscopy indicate that the yellow luminescence is sharply localized around grain boundaries [40]. This result may indicate that the deep donor level is not a bulk property but related to the grain boundaries. Further experimental and theoretical studies are necessary to answer this question conclusively.

Acknowledgements

This work was supported in part by the DAAD (German Academic Exchange Service) and by ARPA under agreement no. MDA972-95-3-0008. We gratefully acknowledge stimulating discussions with W. Götz, N. Johnson and F. Ponce.

References

- [1] R. F. Davis, Proceedings IEEE **79**, 702 (1991).
- [2] S. Strite and H. Morkoc, J. Vac. Sci. Technol. **10**, 1237 (1992).
- [3] H. Morkoc, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov, and M. Burns, J. Appl. Phys **76**, 1363 (1994).
- [4] S. Nakamura, T. Mukai, and M. Senoh, Appl. Phys. Lett. **64**, 1687 (1994).
- [5] H. Amano, M. Kito, K. Hiramatsu, and I. Akasaki, Jpn. J. Appl. Phys **28**, L2112 (1989).
- [6] S. Nakamura, N. Iwasa, M. Senoh, and T. Mukai, Jpn. J. Appl. Phys **31**, 1258 (1992).
- [7] H. P. Maruska and J. J. Tietjen, Appl. Phys. Lett. **15**, 327 (1969).
- [8] M. Ilegems and H. C. Montgomery, J. Phys. Chem. Solids **34**, 885 (1973).
- [9] W. Seifert, R. Franzheld, E. Butter, H. Sobotta, and V. Riede, Crystal Res. & Technol. **18**, 383 (1983).
- [10] B.-C. Chung and M. Gershenson, Appl. Phys. Lett. **72**, 651 (1992).
- [11] D. W. Jenkins and J. D. Dow, Phys. Rev. B **39**, 3317 (1989).
- [12] G. A. Baraff and M. Schlüter, Phys. Rev. B **28**, 2296 (1983).
- [13] Y. Bar-Yam and J. D. Joannopoulos, Phys. Rev. Lett. **52**, 1128 (1984).
- [14] R. Car, P. J. Kelly, A. Oshiyama, and S. T. Pantelides, Phys. Rev. Lett. **54**, 360 (1985).

- [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, and S. T. Pantelides, Phys. Rev. Lett. **66**, 648 (1991).
- [17] J. E. Northrup and S. B. Zhang, Phys. Rev. B **47**, 6791 (1993).
- [18] C. G. Van de Walle, D. B. Laks, G. F. Neumark, and S. T. Pantelides, Phys. Rev. B **47**, 9425 (1993).
- [19] G.-X. Qian, R. M. Martin, and D. J. Chadi, Phys. Rev. B **38**, 7649 (1992).
- [20] D. B. Laks, C. G. Van de Walle, G. F. Neumark, P. E. Blöchl, and S. T. Pantelides, Phys. Rev. B **45**, 10965 (1992).
- [21] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
- [22] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
- [23] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
- [24] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
- [25] R. Stumpf and M. Scheffler, Comp. Phys. Commun. **79**, 447 (1994).
- [26] J. Neugebauer and C. G. Van de Walle, to be published.
- [27] J. Neugebauer and C. G. Van de Walle, in *Materials Research Society Symposia Proceedings*, edited by C. H. Carter Jr., G. Gildenblat, S. Nakamura, and R. J. Nemanich (Materials Research Society, Pittsburgh, Pennsylvania, 1994), Vol. 339, p. 687.
- [28] V. Fiorentini, M. Methfessel, and M. Scheffler, Phys. Rev. B **48**, 1739 (1992).
- [29] D. W. Jenkins, J. D. Dow, and M. H. Tsai, Appl. Phys. Lett. **72**, 4130 (1992).
- [30] T. L. Tansley and R. J. Egan, Phys. Rev. B **45**, 10942 (1992).
- [31] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).
- [32] J. E. Northrup and S. B. Zhang, Phys. Rev. B **50**, 4962 (1994).
- [33] A. Garcia and J. E. Northrup, Phys. Rev. Lett. **74**, 1131 (1995).
- [34] P. Boguslawski, E. Briggs, and J. Bernholc, in *The Physics of Semiconductors*, edited by D. J. Lockwood (Materials Research Society, World Scientific, Singapore, 1994), Vol. 3.
- [35] R. W. G. Wyckoff, *Crystal structures* (Interscience Publishers, New York, 1963), Vol. 1.
- [36] S. D. Lester, F. A. Ponce, M. G. Craford, and D. A. Steigerwald, Appl. Phys. Lett. **66**, 1249 (1994).
- [37] T. Ogino and M. Aoki, Jpn. J. Appl. Phys **19**, 2395 (1980).
- [38] E. R. Glaser, T. A. Kennedy, K. Doverspike, L. B. Rowland, D. K. Gaskill, J. A. Freitas, Jr., M. Asif Khan, D. T. Olson, J. N. Kuznia, and D. K. Wickenden, submitted to Phys. Rev. B.
- [39] W. Götz, N. Johnson, R. A. Street, H. Amano, and I. Akasaki, Appl. Phys. Lett. **66**, 1340 (1995).
- [40] F. A. Ponce, D. P. Bour, W. Götz, and P. J. Wright, Appl. Phys. Lett. **68**, 57 (1996).