Defects, Impurities and Doping Levels in Wide-Band-Gap Semiconductors

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

Received July 21, 1995

A fundamental understanding of the mechanisms governing the behavior of defects and impurities is essential to control doping in semiconductors. Wide-band-gap semiconductors, in particular, often exhibit doping-related problems. We discuss how first-principles theoretical techniques can be applied to the calculation of formation energies and concentrations of native defects and dopant impurities. The approach is illustrated with results for GaN.

I. Introduction

Wide-band-gap semiconductors are becoming increasingly widely used. Two main areas exist for technological applications: high-temperature devices and optoelectronics. For high-temperature applications, advantage is taken of the fact that the large band gap leads to a small intrinsic carrier concentration, even at elevated temperatures. The various doped layers in a semiconductor device will therefore retain their electronic characteristics up to much higher temperatures than would be the case in Si. SiC is extensively being used for this class of applications, and diamond is being pursued as an even better candidate material.

The other class of applications centers around light emission and detection in the wavelength range corresponding to the band gap of the wide-band-gap materials, covering the entire visible spectrum into the UV. The II-VI compounds (mostly ZnSe and ZnSe-based alloys) play a prominent role here; blue and green light-emitting diodes (LEDs), as well as lasers have been demonstrated. Another class of materials which has shown great promise is the III-nitrides (mostly GaN and GaN-based alloys). Blue and green GaN LEDs are already commercially available. The recently achieved ability to cover the entire visible spectrum with semiconductor-based light emitters offers unprecedented possibilities for development of full-color displays, white light sources, etc.

The recent progress in the development of wide-band-gap semiconductor devices can be attributed to important advances in growth techniques, as well as in the ability to control doping. Doping of the semiconductor is an essential requirement for all devices; it requires the ability to produce undoped material, as well as material which is n type (doped with donor impurities) and p type (doped with acceptors). Wide-band-gap semiconductors have traditionally posed severe problems in this respect. In ZnSe, a major breakthrough occurred when p-type doping was demonstrated using lithium[1] and subsequently nitrogen[2]. Similarly, p-type doping of GaN used to present major difficulties, until a process was developed to activate Mg acceptors[3,4]. In addition, achieving low-conductivity, undoped GaN still poses a challenge.

In spite of the progress, serious issues remain to be addressed regarding doping of the wide-band-gap semiconductors, requiring a basic understanding of the fundamental mechanisms that govern impurity incorporation and defect formation. First-principles theory has established itself as an important tool for achieving such basic knowledge. In this paper, we will give a brief description of the methods involved in first-principles calculations of defects and impurities, and summarize some of the results that we have obtained for wide-band-gap semiconductors. The aim here is not to be comprehensive, but rather to present a broad overview as well as pointers to other references which contain more details. We will also limit ourselves to discussing results for GaN; results for ZnSe have been discussed

in Refs. [5-9].

II. Theoretical approach

The first-principles calculations are based on density-functional theory[10] and *ab initio* pseudopotentials[11,12], using a plane-wave basis set (essentially expanding wave functions, etc., in Fourier series). The strong potential of the first-row element, nitrogen, and the localization of the $3d$ electrons in Ga make such an approach challenging: high Fourier components have to be included in the expansions, significantly increasing the size of the matrices in the eigenvalue problem. This problem has been addressed in a pure plane-wave method by utilizing more tractable pseudopotentials[12] and a highly optimized code[13,14]. The role of the Ga $3d$ electrons in GaN was discussed in Ref. [15]. All our calculations are carried out in a supercell geometry; two shells of host atoms around the defect or impurity are relaxed. More details about convergence and other aspects of the calculations can be found in Refs. [15] and [16].

The first-principles calculations produce a wealth of information about defects and impurities, including details of the electronic structure (wave functions, charge densities, defect levels in the band gap, hyperfine parameters) as well as the atomic structure (position in the lattice, relaxation of the host atoms, migration paths). The ability to calculate the total energy also enables the derivation of defect and impurity concentrations. Indeed, the equilibrium concentration, C, of a defect or impurity in the semiconductor is determined by its formation energy, E_{form}:

$$C = N_{\mathrm{sites}} \exp\left[-\frac{E_{\mathrm{form}}}{kT}\right] \quad (1)$$

where N_{sites} is the appropriate site concentration; e.g., for a substitutional impurity on the N site in GaN, N_{sites} is the number of nitrogen lattice sites (per cm^3). The energy appearing in Eq. (1) is in principle a Gibbs free energy; at the present stage of our work, we are not explicitly calculating vibrational entropies (a computationally demanding task). Such entropy contributions cancel to some extent[17], and are small enough not to affect our conclusions.

Let us illustrate the definition of the formation energy with the example of an oxygen impurity on a substitutional N site (O_N) in GaN, which behaves as a donor. The formation energy is determined by the relative abundance of Ga, N, and O atoms in the environment in which the crystal is grown. In a thermodynamic context, these abundances are described by the chemical potentials (energies of reservoirs) for the N and O atoms, μ_N and μ_0. The Ga chemical potential is not an independent variable, since the sum of μ_{Ga}, and μ_N has to be constant and equal to the energy of a two-atom unit of GaN. Forming a substitutional O_N impurity requires the addition of one O atom to the crystal and the removal of one N atom; the formation energy of the oxygen donor in the positive charge state (which the donor assumes after donating an electron to the conduction band) is therefore

$$E_{\mathrm{form}}(GaN:O_N^+) = \varepsilon(GaN:O_N^+) - \mu_O|\mu_N + E_F \quad (2)$$

where $\varepsilon(GaN : O_N^+)$ is the total energy derived from a first-principles calculation for substitutional O, and E_F is the position of the Fermi level, which acts as the reservoir for electrons. Equation (2) illustrates the basic form of the expression for the formation energy; the expression is easily generalized to other defects and impurities[5,7,16].

The N and O chemical potentials are variable parameters; they are, however, subject to the following boundary conditions. The upper limit of the N chemical potential is given by GaN in equilibrium with N_2 molecules, while the lower limit is for GaN in equilibrium with bulk Ga; μ_N can therefore vary over a range given by the heat of formation of GaN, ΔH. To establish an upper bound on the chemical potential of the impurity (oxygen), one needs to consider which other phases the element can form, either by itself, or in its interactions with the constituents of the host material. In the case of oxygen, the most stringent bound is due to the formation of the compound Ga_2O_3.

III. Results for native defects in GaN

Fig. 1 displays the main results of our first-principles calculations for native defects in GaN. In previous work, we have found that the formation energies are largely insensitive to the structure (zincblende vs. wurtzite) of the material[15]. The formation energies for these defects are defined similarly to our example in Eq. 2; they are therefore a function of the Fermi level and of the atomic chemical potentials. The results shown in Fig. 1 are for a choice of chemical potentials corresponding to Ga-rich conditions (which appear to be most common in experimental growth situations) as

well as N-rich conditions. The formation energies displayed in Fig. 1 are slightly different from those given in Ref. [16], resulting from inclusion of additional nearest neighbor shells in the atomic relaxation.

We note that the self-interstitials (Ga_i and N_i) and the antisite defects (N_{Ga} and Ga_N) have high formation energies, and are therefore very unlikely to occur in any significant concentration. The vacancies, on the other hand, have lower formation energies. Under conditions of thermodynamic equilibrium, a low formation energy is required for the defect to occur in large concentrations. We will focus here on the vacancy defects; more information about the atomic and electronic structure of all the native defects can be found in Ref. [16].

Nitrogen vacancies have commonly been invoked to explain the n-type conductivity of as- grown GaN. Fig. 1 shows that the formation energy of the nitrogen vacancy (V_N) is actually quite high in n-type material (Fermi level high in the gap, near the conduction band). In thermodynamic equilibrium, the concentration of nitrogen vacancies should therefore be quite low, and they should not be responsible for the observed n-type conductivity. What about the possibility that incorporation of nitrogen vacancies occurs 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 is only possible if the nitrogen species is sufficiently *immobile* to prevent any significant flux of nitrogen between the bulk of the material and a source or sink of nitrogen atoms.

Several channels can in principle contribute to the diffusion of nitrogen. Nitrogen could diffuse interstitially (and eventually annihilate vacancies); however, the high formation energy of the nitrogen interstitial (see Fig. 1) renders this diffusion mechanism unlikely. The other mechanism involves diffusion of the nitrogen vacancy. Our calculated migration barrier is ~4 eV, which should allow for vacancy diffusion at the high temperatures at which MOCVD (metal-organic chemical vapor deposition) or bulk growth of GaN is carried out. If one can assume efficient transport between the vacancy defects in the bulk and the reservoir of nitrogen atoms, the concentration of N vacancies will be allowed to reach its equilibrium value. We calculate that, at a typical growth temperature of 1300 K, and in the absence of dopant impurities, this vacancy concentration would be less than 10^{16}cm^{-3}.

While we conclude that isolated nitrogen vacancies are unlikely to occur in n-type material, Fig. 1 shows that their formation energy is actually very low in p-type material. They may, therefore, play a role in compensating acceptors. Similarly, the Ga vacancy may have low enough energy to occur as a compensating center in n-type material.

IV. Results for dopant impurities in GaN

In the previous section, we argued that nitrogen vacancies are not responsible for the observed n-type conductivity in as-grown GaN. We therefore have to look for another source for the conductivity. First, we stress that our conclusion about nitrogen vacancies applies to *isolated point defects*. It is still possible that a vacancy-related *complex* or *extended defect* is involved, if it acts as a donor. We also propose an alternative explanation for the n-type conductivity, namely that it is caused by unintentional incorporation of donor impurities.

Silicon, oxygen, and carbon are common contaminants in the systems typically used for growth of GaN. We have performed first-principles calculations for these impurities[18], and found that silicon and oxygen indeed behave as donors, and can potentially be incorporated in high concentrations. Carbon, if it is located on the Ga site (C_{Ga}) also behaves as a donor - however, this site is less favorable than the C_N site, on which C behaves as an acceptor.

V. Conclusions

We have given a brief overview, illustrated with examples for GaN, of the power of first-principles calculations for addressing problems related to doping of semiconductors. It should be emphasized that the methods are general, and can be applied to *any* defect in *any* semiconductor. Indeed, doping problems still occur in many semiconductors, including Si. The applications of wide-band-gap semiconductors are rapidly on the rise, however, and a better understanding of the basic mechanisms is especially important in these materials. First-principles theory is an important tool that can be used for achieving that goal.

Acknowledgments

Thanks are due to N. M. Johnson and W. Götz for useful discussions. This work is supported in part by ARPA and by the DAAD (German Academic Exchange Service).

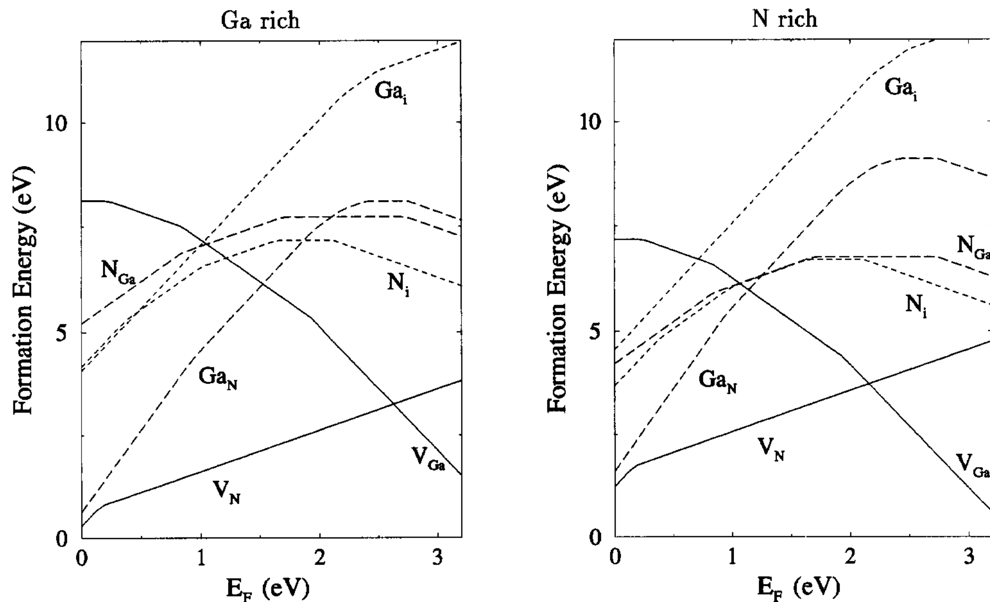

Figure 1. Defect formation energy as a function of the Fermi energy for vacancies (solid lines), antisites (long-dashed lines), and interstitials (dashed lines). Ga-rich conditions (left) and N-rich conditions (right) are assumed. $E_F = 0$ corresponds to the top of the valence band.

References

1. M. A. Haase, H. Cheng, J. M. DePuydt and J. E. Potts, J. Appl. Phys. **67**, 448 (1990).
2. M. A. Haase, J. Qiu, J. M. DePuydt and H. Cheng, Appl. Phys. Lett. **59**, 1272 (1991).
3. H. Amano, M. Kito, K. Hiramatsu and I. Akasaki, Jpn. J. Appl. Phys. **28**, L2112 (1989).
4. S. Nakamura, N. Iwasa, M. Senoh and T. Mukai, Jpn. J. Appl. Phys. **31**, 1258 (1992).
5. D. B. Laks, C. G. Van de Walle, G. F. Neumark and S. T. Pantelides, Phys. Rev. Lett. **66**, 648 (1991).
6. D. B. Laks, C. G. Van de Walle, G. F. Neumark, P. E. Blochl and S. T. Pantelides, Phys. Rev. B**45**, 10965 (1992).
7. C. G. Van de Walle, D. B. Laks, G. F. Neumark and S. T. Pantelides, Phys. Rev. B **47**, 9425 (1993)
8. D. B. Laks, C. G. Van de Walle, G. F. Neumark and S. T. Pantelides, Appl. Phys. Lett. **63**, 1375 (1993).
9. C. G. Van de Walle and D. B. Laks, Solid State Communications **93**, 447 (1995).
10. P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, *ibid.* **140**, A1133 (1965).
11. D. R. Hamann, M. Schlüter and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
12. N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
13. R. Stumpf and M. Scheffler, Computer Phys. Commun. **79**, 447 (1994).
14. J. Neugebauer and C. G. Van de Walle, to be published.
15. J. Neugebauer and C. G. Van de Walle, in *Diamond, SiC and Nitride Wide Bandgap Semiconductors*, edited by C. H. Carter Jr., G. Gildenblat, S. Nakamura, and R. J. Nemanich, Materials Research Society Symposia Proceedings, Vol. 339 (Materials Research Society, Pittsburgh, Pennsylvania). p. 687.
16. J. Neugebauer and C. G. Van de Walle, Phys. Rev. B**50**, 8067 (1994).
17. G.-X. Qian, R. M. Martin and D. J. Chadi, Phys. Rev. B **38**, 7649 (1988); N. Chetty and R. M. Martin, *ibid.* **45**, 6089 (1992).
18. J. Neugebauer and C. G. Van de Walle, in *Proceedings of the 22th International Conference on the Physics of Semiconductors,* Vancouver, 1994 (World Scientific Publishing Co Pte Ltd., Singapore), p. 2327.