Control of Doping by Impurity Chemical Potentials: Predictions for p-Type ZnO

Yanfa Yan and S. B. Zhang
National Renewable Energy Laboratory, Golden, Colorado 80401

S. T. Pantelides
Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235
and Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831
(Received 20 December 2000)

Theoretical work has so far focused on the role of host-element chemical potentials in determining defect formation energies that control doping levels in semiconductors. Here, we report on our analysis of the role of the dopant-impurity chemical potential, which depends on the source gas. We present first-principles total-energy calculations that demonstrate a wide variation in the possible effective chemical potential of N. We account in detail for the recent puzzling observations of doping ZnO using N_2 and N_2O and predict that the use of dilute NO or NO_2 gas would resolve the long-standing problem of achieving p-type ZnO.

DOI: 10.1103/PhysRevLett.86.5723 PACS numbers: 61.72.Bb, 61.72.Vv, 71.55.Gs

Zinc oxide (ZnO), like other transparent conducting oxides, is an important material for next-generation short-wavelength optoelectronic devices [1–3]. To realize these device applications, an important issue is to fabricate both high-quality p-type and n-type ZnO films. However, like most wide band-gap semiconductors, ZnO has the "asymmetric doping" limitation [4–8], i.e., it can be an easily doped high-quality n-type [9,10], but it is difficult to dope p-type. Nitrogen, a good p-type dopant for other II-VI semiconductors [11,12], has long been considered as a possible dopant for p-type ZnO [13], but various efforts to realize this goal have been frustrated [14,15]. The difficulty of making p-type ZnO was recently attributed to low-energy native point defects such as O vacancies (V_O) and Zn interstitials (Zn_i) [16], or to unintentional, but ever-present, hydrogen [17].

Recently, very puzzling results on ZnO films have been reported by Joseph *et al.* [18]. They basically found that they could get p-type material, but it had either high hole concentration and poor mobility or low concentration and high mobility. The key experimental results are summarized in Table I.

The realization of high-hole-concentration p-type ZnO films was attributed to the "codoping" phenomenon that had been predicted theoretically [19] and requires the simultaneous presence of two dopants, in this case Ga and N. No explanation for the other results has been reported. It is clear, however, that understanding these results may lead to a prescription that would produce p-type ZnO with both high hole concentration and high mobility.

It has long been known (see, e.g., Refs. [20] and [21]) that the solubility of dopants depends on the formation energy of the defects containing the dopants. In turn, the formation energy depends on the chemical potentials of the host elements (in this case, Zn and O) and the dopant element (N). The concentrations of native defects that may act as compensating centers also depend on the host-element chemical potentials. Finally, the concentrations of defect complexes of dopant and host atoms depend on all the chemical potentials. So far, computational results and analysis have focused on the dependence of formation energies on host-element chemical potentials. In most cases, results are presented for cation-rich or anion-rich conditions that fix the chemical potentials of the host elements [22]. The only other dependence of the formation energies is on the Fermi energy.

The dependence of doping efficiency on the *dopant* chemical potential has not attracted attention because normally there are no significant alternative dopant sources. In this paper, we show that in the case of N doping of ZnO (or other *oxides*), there is a unique and unusual opportunity. There are at least four different gases, namely, N_2, NO, NO_2, and N_2O which can be used with or without electron cyclotron resonance (ECR) plasma source. If these molecules arrive intact at the growing surface, it is their chemical potentials that determine the doping efficiency (see below). Clearly, the growth conditions need to be set up accordingly to ensure that the chosen species is available and arrives intact at the growing surface. We report first-principles total-energy calculations of the chemical potentials of the above species and the corresponding formation energies of substitutional N and possible compensating defects. With these results, we account in detail for all the observations of Joseph *et al.* [18]. In addition, we *predict* that, barring unforeseen possibilities, the use of NO or NO_2 gas should lead to p-type ZnO with high hole concentrations and high mobilities. A key feature that underlies our prediction of doping with NO or NO_2 is that these molecules supply single-N atoms, whereas one has to break N-N bonds in order to obtain desirable single-N defects when N_2 and N_2O are used. Finally, we show that the role of Ga is *not* as posited by the codoping idea; Ga serves a different

TABLE I. Key results of the ZnO doping reported by Joseph *et al.* [18].

	Dopant	Ga source	ECR	Pressure (mbar)	Resistivity ($\Omega \cdot$ cm)	Carrier density ($1/\text{cm}^3$)	Mobility ($\text{cm}^2/\text{V} \cdot \text{s}$)	Carrier type
1	N_2O	No	No	1×10^{-3}	4.3×10^{-2}	4.5×10^{20}	0.31	*n*
2	N_2O	No	Yes	1×10^{-3}	5×10^{5}	2.0×10^{10}	1.9×10^{3}	*p*
3	N_2O	Yes	Yes	1×10^{-3}	2.0	4×10^{19}	0.07	*p*
4	N_2	No	Yes	1×10^{-3}	0.17	1.3×10^{19}	1.3	*n*

useful role in enhancing doping efficiency, but it also leads to the formation of undesirable defects that reduce the mobility.

The calculations were carried out using density-functional theory, the generalized gradient approximation (GGA) for exchange correlation, ultrasoft pseudopotentials, and plane waves [23]. A suitable energy cutoff for the plane waves was determined to be 380 eV. A 32-atom supercell was used for defect calculations as it was found to be adequate in earlier work on similar systems (accuracy on the order of 0.2 eV) [24]. The relative energy differences that are needed for this problem are generally larger than 1 eV. In all the calculations, all atoms were allowed to relax until the Hellmann-Feynman forces acting on them became less than 0.01 eV/Å.

Under equilibrium conditions, the concentration of a point defect is mainly related to its formation energy, which depends on the chemical potentials of the host and relevant impurity atoms as defined by appropriate reservoirs [20,21]. The chemical potentials of Zn and O satisfy

$$\mu_{\text{ZnO}} < \mu_{\text{Zn}} < \mu\ (\text{Zn metal}) \quad \text{and} \quad \mu_{\text{ZnO}} < \mu_{\text{O}} < \mu\ (\text{O}_2\ \text{gas}). \tag{1}$$

Here, $\mu_{\text{ZnO}} = \mu_{\text{Zn}} + \mu_{\text{O}}$ is the formation enthalpy of ZnO. Our calculated μ_{ZnO} is -3.3 eV, close to the experimental value of -3.6 eV.

For ZnO, we considered the native defects (vacancies, interstitials, and antisite defects) and various configurations and complexes that are possible for N. A complete report on these defects will be published elsewhere [25]. For our purposes, the only relevant defects are substitutional N at an O site (N_O) and substitutional N_2 also at an O site [$(N_2)_O$] (substitutional N_2 was first studied in ZnSe by Cheong *et al.* [26]). N_O is an acceptor and $(N_2)_O$ is a double shallow donor. In the final analysis, *it is the relative concentrations of these two defects that control the doping type* (oxygen vacancies and hydrogen play a secondary role).

We start by drawing a sharp distinction between the two gases used by Joseph *et al.* [18], namely, N_2 and N_2O, and the two alternative gases, namely, NO and NO_2. The molecules of the first two gases contain *pairs* of N atoms, whereas the alternative gases contain *single*-N atoms. Let us first consider the formation of the desirable N_O defects. It is clear that N_2 and N_2O are not the best options because energy must be supplied to break their N-N bonds. In contrast, NO and NO_2 molecules can be incorporated directly in the growth surface to form N_O defects, taking advantage of the fact that O atoms also take part in the growth process as host atoms. We arrive at the same conclusion when we examine the formation of the undesirable $(N_2)_O$ defect. Now it is N_2 and N_2O that can be incorporated directly into the growth surface to form $(N_2)_O$. In contrast, NO and NO_2 molecules provide single-N atoms and can produce $(N_2)_O$ centers only by the less likely process of two N atoms arriving simultaneously at the same site on the growth surface (once a single N_O is incorporated into the bulk, it is unlikely that it can be found by a diffusing N atom because the migration energy is very high, 3 eV).

We now present a detailed analysis and numerical results for the production of the desirable N_O defects. The process is controlled by thermodynamics as there are no apparent kinetic constraints. In the case of N_2 and N_2O molecules, which entail bond breaking, the formation energy of a defect containing a single-N atom and any number of excess or missing host atoms is given by

$$E_f = E(n_{\text{Zn}}, n_{\text{O}}, \text{N}) - n_{\text{Zn}}\mu_{\text{Zn}} - n_{\text{O}}\mu_{\text{O}} - \mu_{\text{N}}. \tag{2}$$

Here, $E(n_{\text{Zn}}, n_{\text{O}}, \text{N})$ is the total energy of a supercell containing the defect, n_{Zn} and n_{O} are the numbers of host atoms, and μ_{N} is the chemical potential of N relative to the designated source. When the source gas is N_2, we have $\mu_{\text{N}} = \frac{1}{2}\mu_{\text{N}_2}$. When the source gas is N_2O, we have $\mu_{\text{N}} = \frac{1}{2}\mu_{\text{N}_2\text{O}} - \mu_{\text{O}}$.

In the case of NO and NO_2 molecules, N can be incorporated as part of a molecule. The key point here is that these molecules arrive *intact* at the growing surface. For defects that contain a single N, the formation energy is

$$E_f = E(n_{\text{Zn}}, n_{\text{O}}, \text{N}) - n_{\text{Zn}}\mu_{\text{Zn}} - (n_{\text{O}} - 1)\mu_{\text{O}} - \mu_{\text{NO}} \tag{3}$$

or

$$E_f = E(n_{\text{Zn}}, n_{\text{O}}, \text{N}) - n_{\text{Zn}}\mu_{\text{Zn}} - (n_{\text{O}} - 2)\mu_{\text{O}} - \mu_{\text{NO}_2}. \tag{4}$$

In these cases, the molecules may simply be supplied as such or they may be produced by a reaction in the gas phase. For example, NO molecules can be created by $N_2O \Leftrightarrow NO + N$. In this case, $\mu_{\text{NO}} = \mu_{\text{N}_2\text{O}} - \mu_{\text{N}}$. Of course, other reactions occur (e.g., $2N \Leftrightarrow N_2$) so that, in general, one has a set of coupled equations for the chemical potentials. In this conditions, one must simultaneously consider doping by N_2O and N.

In Fig. 1(a) we show the results of calculations for the formation energy of N_O for all the four gases using the above equations. The Fermi energy is assumed to be at the valence-band edge, with corresponds to optimal *p*-type doping. The difference between N_2/N_2O and NO/NO_2 is very clear, i.e., the use of NO/NO_2 leads to significantly reduced formation energies for N_O because it does not entail an energy supplement to break N-N bonds. The negative formation energies of N_O at Zn-rich conditions indicate that NO or NO_2 molecules can be incorporated spontaneously to form N_O defects.

We now consider the formation of the undesirable defect $(N_2)_O$. In this case, when N_2 and N_2O are used, the process is controlled by thermodynamics since, once more, there are no apparent kinetic constraints. The formation energy of a defect containing a pair of N atoms and any number of excess or missing host atoms is given by

$$E_f = E(n_{\mathrm{Zn}}, n_{\mathrm{O}}, \mathrm{N}) - n_{\mathrm{Zn}}\mu_{\mathrm{Zn}} - n_{\mathrm{O}}\mu_{\mathrm{O}} - \mu_{\mathrm{N_2}} \quad (5)$$

and

$$E_f = E(n_{\mathrm{Zn}}, n_{\mathrm{O}}, \mathrm{N}) - n_{\mathrm{Zn}}\mu_{\mathrm{Zn}} - (n_{\mathrm{O}} - 1)\mu_{\mathrm{O}} - \mu_{\mathrm{N_2O}}, \quad (6)$$

respectively.

In contrast, when NO and NO_2 are used, the process of forming $(N_2)_O$ is not controlled by thermodynamics, but by kinetics, because two N atoms from different source points must arrive simultaneously at the same site on the growth surface. The rate of the process is proportional to $(C_N/C_O)^2$, where C_N is the concentration of NO or NO_2 molecules and C_O is the concentration of available nominal O sites. By ensuring that $C_N \ll C_O$, the formation of $(N_2)_O$ can be suppressed altogether. For comparison purposes, we show in Fig. 1(b) the formation energies of $(N_2)_O$ for each of the four gases. The solid lines indeed control the formation of $(N_2)_O$ when N_2 or N_2O is used. It is clear that if N_2O gas is used and the molecules arrive intact at the growing surface, high concentration of $(N_2)_O$ centers will be observed at Zn-rich conditions. The dotted lines show that the formation energies of $(N_2)_O$ are higher than N_O at O-rich conditions, but lower at Zn-rich conditions when NO or NO_2 molecules are used. The lower formation energies of $(N_2)_O$ at zinc-rich conditions would control the formation of $(N_2)_O$ *only* if the process is allowed to achieve equilibrium. The negative formation energies for N_O and $(N_2)_O$ at Zn-rich conditions indicate that the formation of N_O and $(N_2)_O$ is controlled completely by kinetics. As we just saw, the formation of $(N_2)_O$ can be suppressed by working in the dilute gases where $C_N \ll C_O$.

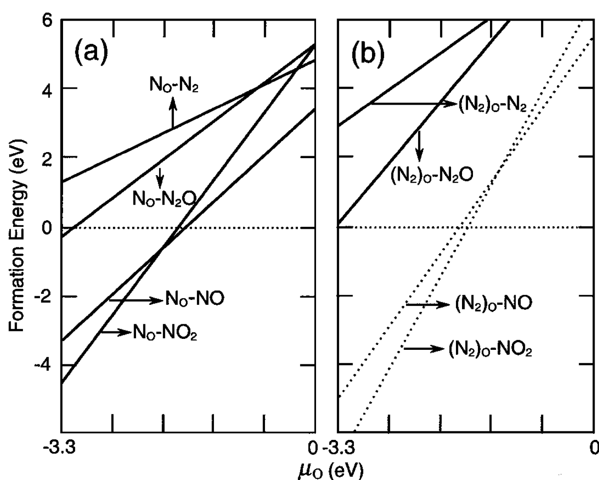

FIG. 1. (a) Calculated formation energies of N_O as functions of the O chemical potential formed by atomic N (stripped off an N_2 or an N_2O molecule), NO, or NO_2 molecule. $\mu_O = -3.3$ is the Zn-rich limit condition and $\mu_O = 0$ is the O-rich limit condition. (b) Calculated formation energies of a $(N_2)_O$ as functions of the O chemical potential for the defects formed by N_2, N_2O, NO, and NO_2 molecules.

We are now ready to discuss the experimental data of Joseph *et al.* [18]. We note that the growth conditions are Zn rich because when Zn and O atoms are ablated from the target, O atoms may form O_2 molecules through simple collisions, whereas Zn atoms would have to nucleate Zn metal somewhere in the chamber and have other Zn atoms find that nucleus for continued precipitation (we calculated the binding energy of Zn_2 and O_2 molecules to be 0.2 and 6.9 eV, respectively, showing that Zn atoms do not form molecules at the temperature of interest). We now discuss line-by-line the key results of the experiments shown in Table I.

Line 1 of Table I: N_2O gas without ECR plasma.— Under these conditions, *n*-type ZnO films with very high carrier concentrations (4.5×10^{20} cm^{-3}) were obtained. Our results for direct incorporation of N_2O molecules [Fig. 1(b)] show that the formation energy of $(N_2)_O$ is nearly zero at the Zn-rich limit, accounting for the very high level of *n*-type doping. Without ECR, the N_2O molecules do not crack in the gas phase, and therefore do not form N_O defects. Thus, the theoretical result constitutes a clear "prediction" (postdiction), not simply an explanation.

Line 2 of Table I: N_2O gas with ECR plasma on.— Under this condition, *p*-type ZnO films with very low carrier concentration (2.0×10^{10} cm^{-3}), but high mobility (1.9×10^3 cm^2/V $\cdot$ s), were obtained. In this case, the ECR plasma provides energy that drives the reaction [27]

$$N_2O \Leftrightarrow NO + N \quad (7)$$

in the *forward* direction. Because the energy for breaking NO molecules, 6.6 eV, is 2.2 eV higher than that for N_2O molecules (4.4 eV) [28], the further dissociation of NO can generally be neglected [27]. Once the reaction products, NO and N, enter the chamber, reaction (7) would go in reverse. In addition, N atoms would also form N_2 molecules. The net result is a mix of NO, N_2, and N_2O. These results show that NO will introduce low-formation energy

N_O, whereas N_2O molecules will introduce low-formation energy $(N_2)_O$. A calculation of the branching ratios at the temperature of interest can, in principle, tell us which concentration would be higher, but that is beyond the scope of this paper. The data show high resistivity and slightly p-type doping, which suggests that the branching ratios are such that NO wins slightly.

Line 3 of Table I: N_2O gas with ECR plasma source and additional Ga source.— Under these conditions, p-type material with very high carrier concentration (4×10^{19} cm^{-3}), but very low mobility (0.07 cm^2/V · s), was obtained. This is similar to the previous case. In this case, however, when NO and N enter the chamber, they also find Ga. N and Ga atoms can form GaN molecules (binding energy 3.4 eV), reducing the concentration of N that could reconvert NO to N_2O. Consequently, more NO molecules survive and reach the growing surface, leading to p-type ZnO with high carrier concentration. Clearly, the wt % of Ga_2O_3 in the ZnO target is important because it controls the branching ratios. The poor mobility is probably the result of GaN precipitates in the film.

Line 4 of Table I: N_2 gas with ECR plasma on.— Under these conditions, n-type material with high carrier concentration (1.3×10^{19} cm^{-3}), but low mobility (1.3 cm^2/V · s), was obtained. Because the dissociation energy for N_2 molecules, 9.9 eV, is 5.5 eV higher than that for N_2O molecules (4.4 eV) [28], the ECR that cracks N_2O molecules cannot crack N_2 molecules. However, the ECR source can transfer energy to N_2 molecules and change the chemical potential of the molecules. In this case, the effective chemical potential of N_2 is $\mu'_{N_2} = \mu_{N_2} + \Delta E$, where μ_{N_2} is the chemical potential of N_2 molecules without ECR source and ΔE is the energy transferred from the ECR source. From Eq. (5), it is seen that the ECR plasma decreases the formation energy of $(N_2)_O$ by ΔE. In the experiment of Joseph *et al.* [18], the ECR can crack the N_2O molecules. Assuming that ΔE is close to the dissociation energy of N_2O molecules, i.e., 4.4 eV, the formation energy of $(N_2)_O$ formed by N_2 molecules [shown in Fig. 1(b)] should be shifted downward about 4.4 eV, giving a negative formation energy at Zn-rich conditions. This result explains the growth of n-type ZnO with high carrier concentration.

It is clearly seen that our results account for all the puzzling observations of ZnO doping reported by Joseph *et al.* [18]. We found that NO or NO_2 is the most efficient way to dope p-type. In contrast, N_2 and N_2O incorporate substitutional N_2, making the material n-type. The role of Ga is *not* as posited by the codoping idea. Instead, Ga serves a different useful role in enhancing doping efficiency; namely, it helps deplete atomic N produced by reaction (7) so that NO molecules survive and reach the growing surface, instead of reconverting to N_2O. However, Ga also leads to the formation of undesirable GaN precipitates that reduce the mobility. We thus predict that the use of NO or NO_2 without an ECR plasma source and without Ga would achieve better p-type ZnO. The principle applies to other transparent conducting oxides, such as CdO, SnO_2, Ga_2O_3, and In_2O_3.

We thank Dr. S.-H. Wei and Dr. C. H. Park for useful discussions. The work at NREL was supported by the U.S. DOE under Contract No. DE-AC36-99GO10377. The work at Vanderbilt University was supported by NSF Grant No. DMR-9803768 and the William A. and Nancy F. McMinn Endowment.

[1] J. R. Tuttle *et al.,* Prog. Photovolt. Res. Appl. **3**, 235 (1995).
[2] A. E. Delahoy and M. Cherny, Mater. Res. Soc. Symp. Proc. **426**, 467 (1996).
[3] P. Yu *et al.,* in *Proceedings of the 23rd International Conference on the Physics of Semiconductors, Berlin, 1996* (World Scientific, Singapore), Vol. 2, p. 1453.
[4] G. F. Neumark, Phys. Rev. Lett. **62**, 1800 (1989).
[5] T. Minami *et al.,* Jpn. J. Appl. Phys. **24**, L781 (1985).
[6] S. B. Zhang, S.-H. Wei, and A. Zunger, J. Appl. Phys. **83**, 3192 (1998).
[7] D. J. Chadi, Phys. Rev. Lett. **72**, 534 (1994).
[8] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. Lett. **84**, 1232 (2000).
[9] T. Minami *et al.,* Jpn. J. Appl. Phys. **24**, L781 (1985); *ibid.* **25**, L776 (1986).
[10] J. Hu and R. G. Gordon, Sol. Cells **30**, 437 (1991).
[11] R. M. Park *et al.,* Appl. Phys. Lett. **57**, 2127 (1990).
[12] K. Ohkawa, T. Karasawa, and T. Mitsuyu, Jpn. J. Appl. Phys. **30**, L152 (1991).
[13] A. Kobayashi, O. F. Sankey, and J. D. Dow, Phys. Rev. B **28**, 946 (1983).
[14] Y. Sato and S. Sato, Thin Solid Films **281–282**, 445 (1996).
[15] K. Minegishi *et al.,* Jpn. J. Appl. Phys. **36**, L1453 (1997).
[16] A. F. Kohan *et al.,* Phys. Rev. B **61**, 15 019 (2000); S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. B **63**, 75 205 (2001).
[17] C. G. Van de Walle, Phys. Rev. Lett. **85**, 1012 (2000).
[18] M. Joseph, H. Tabata, and T. Kawai, Jpn. J. Appl. Phys. **38**, L1205 (1999).
[19] T. Yamamoto and H. Katayama-Yoshida, Jpn. J. Appl. Phys. **38**, L166 (1999).
[20] D. B. Laks *et al.,* Phys. Rev. Lett. **66**, 648 (1991).
[21] S. B. Zhang and J. E. Northup, Phys. Rev. Lett. **67**, 2339 (1991).
[22] J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75**, 4452 (1995).
[23] M. C. Payne *et al.,* Rev. Mod. Phys. **64**, 1045 (1992).
[24] F. A. Reboredo and S. T. Pantelides, Phys. Rev. Lett. **82**, 1887 (1999).
[25] Y. Yan, S. J. Pennycook, and S. T. Pantelides (unpublished).
[26] B.-H. Cheong, C. H. Park, and K. J. Chang, Phys. Rev. B **51**, 10 610 (1995).
[27] L. Date *et al.,* Surf. Coat. Technol. **116–119**, 1042 (1999).
[28] T. A. Cleland and D. W. Hess, J. Electrochem. Soc. **136**, 3103 (1989).