VOLUME 90, NUMBER 25 PHYSICAL REVIEW LETTERS week ending 27 JUNE 2003

Cluster-Doping Approach for Wide-Gap Semiconductors: The Case of p-Type ZnO

L. G. Wang and Alex Zunger
National Renewable Energy Laboratory, Golden, Colorado 80401, USA

(Received 27 March 2003; published 24 June 2003)

First-principles calculations on p-type doping of the paradigm wide-gap ZnO semiconductor reveal that successful doping depends much on engineering a stable local chemical bonding environment. We suggest a *cluster-doping* approach in which a locally stable chemical environment is realized by using few dopant species. We explain two puzzling experimental observations, i.e., that monodoping N in ZnO via N_2 fails to produce p-type behavior, whereas using an NO source produces *metastable* p-type behavior, which disappears over time.

DOI: 10.1103/PhysRevLett.90.256401 PACS numbers: 71.55.–i, 61.72.Ss, 71.15.Nc

Many novel optoelectronic applications, such as blue and UV lasers and light emitting diodes, solid state lighting, and transparent conducting contacts could become possible if one could replace the conventional 1–2 eV band gap active semiconductor materials with wide band gap (3–5 eV) materials [1–3]. However, as the material's band gap opens up, it becomes increasingly difficult to introduce free electrons (n-type doping) or free holes (p-type doping) into the system. The paradigm inorganic wide-gap ($E_g = 3.4$ eV) material used to investigate doping bottlenecks is ZnO, which has been studied very intensely recently [4–17]. N was established [11] as being the more soluble group-V impurity, having also the shallowest acceptor level relative to P and As. However, surprising behavior was uncovered: Doping with a N_2 source led to n-type, not p-type doping [15], while doping with an NO_2 or NO source led to p-type behavior which converted, over time, to n-type [15]. In this Letter we explain these experimental observations and then suggest a different approach to engineer the local bonding environment around the dopant which maximizes both the solubility and stability of the bonded dopant. This general method is based on the use of *few* atomic species for doping, designed to produce simultaneously stable bonds and low doping enthalpy, and is distinct from "codoping" [6].

The formation enthalpy for replacing "host anion" (H, a) sites by $n_{D,a}$ "dopant anions" (D, a), and replacing "host cation" (H, c) sites by $n_{D,c}$ "dopant cations" (D, c) is

$$\Delta H_f^{(D,q)}(\mu, \epsilon_F) = \Delta E_b(\epsilon_F) + \Delta E_{CRE}(\mu). \qquad (1)$$

Here

$$\Delta E_b(\epsilon_F) = [E(D, q) - E(\text{host})] + q\epsilon_F \qquad (2)$$

is the excess bond energy of the doped system (D) in charge state q, ϵ_F is the Fermi energy relative to the valence band maximum of the pure host, and E is the difference in total energy of the system and the energies of an equivalent number of free atoms. For example, in substituting oxygen by nitrogen in ZnO we replace four Zn-O bonds by four Zn-N bonds, with a bond-energy difference of ΔE_b. We will indeed see below that ΔE_b can be accurately expanded as a linear superposition of bond energies. The "chemical reservoir energy" (CRE) is

$$\begin{aligned}\Delta E_{CRE}(\mu) = {}& n_{D,a}[\mu(H, a) - \mu(D, a)] \\ &+ n_{D,c}[\mu(H, c) - \mu(D, c)],\end{aligned} \qquad (3)$$

which is the change in the energy due to exchanging *reactants* corrected by adding back the energies of the free atoms. For example, replacement of O in ZnO by N taken from an N_2 source involves removing the energy μ_N from the N_2 reservoir and adding the energy μ_O to the O_2 reservoir. High dopant solubility requires sufficiently low ΔH_f, whereas locally stable bonding requires also $\Delta E_b \ll 0$. Although we are unable to calculate ΔE_b very accurately [because of well-known local-density approximation (LDA) or generalized gradient approximation (GGA) errors in the energies of the free atoms], we will notice clear trends in ΔE_b for different bonding case.

Previous efforts to lower the dopant enthalpy ΔH_f focused largely on engineering the chemical reservoir energy term [5–7]. For example, when doping the oxygen site of ZnO by nitrogen, one can lower the chemical reservoir energy contribution ΔE_{CRE} by employing O-poor (i.e., Zn-rich) chemical conditions [5,13,15] or by deliberately using a *high-energy* source of nitrogen [5,13,15], such as atomic nitrogen, not the more stable N_2; in the former case the driving force for N to leave its source and enter the lattice is greatest. But even if the solubility of the dopant is enhanced by lowering ΔE_{CRE}, doping often proved to be unstable, disappearing over time [15,16]. We propose that engineering the "local bonding term" ΔE_b holds the key to such effects, and indeed to successful doping of wide-gap materials. We explored a range of local bonding environments in ZnO, involving a nitrogen dopant plus cation additives such as Al, Ga. We will present detailed results on ZnO + (Ga, N), and illustrate also ZnO + (Al, N). Surprisingly, we find that "cluster doping" by 4N + Ga is more soluble than codoping [6] ZnO by 2N + Ga and is also locally stable. This principle of design of the local chemical bonding energy ΔE_b can be used generally to improve doping in other systems by searching for chemical additives which lower ΔE_b, while maintaining low ΔH_f and shallow ionization energies.

The allowed chemical potential ranges of Eq. (3) are decided by thermodynamic considerations. For the host ZnO, we have, at equilibrium

$$\mu_{\mathrm{Zn}} + \mu_{\mathrm{O}} = \Delta H_f(\mathrm{ZnO}), \tag{4}$$

where ΔH_f is the compound formation enthalpy relative to O_2 and solid metallic Zn. In order to assure that competing phases [i.e., elemental solids Ga and Zn, the N_2 and O_2 gases, and the compounds Ga_2O_3, GaN, $ZnGa_2O_4$] *do not* form, we require

$$\mu_i \leq 0 \qquad (i = \mathrm{Ga, N, Zn, O}), \tag{5}$$

$$2\mu_{\mathrm{Ga}} + 3\mu_{\mathrm{O}} \leq \Delta H_f(\mathrm{Ga_2O_3}), \tag{6}$$

$$\mu_{\mathrm{Ga}} + \mu_{\mathrm{N}} \leq \Delta H_f(\mathrm{GaN}), \tag{7}$$

$$\mu_{\mathrm{Zn}} + 2\mu_{\mathrm{Ga}} + 4\mu_{\mathrm{O}} \leq \Delta H_f(\mathrm{ZnGa_2O_4}). \tag{8}$$

The calculated formation enthalpies are −3.61 (− 3.6), −11.07 (− 10.9), and −16.24 eV for ZnO, Ga_2O_3, and $ZnGa_2O_4$, respectively, which compare well with available experimental values given in parentheses. The formation enthalpies of AlN and Al_2O_3 are −3.02 (− 3.3) and −17.09 (− 17.2) eV, respectively. The formation enthalpy of NO with respect to $\frac{1}{2}N_2 + \frac{1}{2}O_2$ is 0.76 (0.93) eV. Note that we define $\Delta H_f(\mathrm{GaN})$ to depend on the source of N used in experiment and is calculated to be −1.04 (expt.: −1.1) , −1.80, and −6.02 eV for N_2, NO, and free N atom sources, respectively. We consider the formation of Zn_3N_2 to be unlikely since the large unit cell is probabilistically difficult to form kinetically during doping and, actually, the phase is not observed experimentally. We also assume that the growth conditions are such that the dopant sources Ga + N are not allowed to combine outside ZnO, or precipitate macroscopic GaN within ZnO [Eq. (7)]. Based on the fact that experiments [13,15,16] are performed under N-rich conditions, we can fix the N chemical potential as $\mu_{\mathrm{N}} = \frac{1}{2}\mu_{\mathrm{N_2}}$ for the N_2 source gas, or as $\mu_{\mathrm{N}} = \mu_{\mathrm{NO}} - \mu_{\mathrm{O}}$ for the NO source gas, or as $\mu_{\mathrm{N}} = \mu(\mathrm{N\ atom})$ in the case that N atoms are used as the dopant N source. Therefore, we have actually two independent chemical potentials μ_{Ga} and μ_{O}.

We consider N-centered clusters N-Ga_nZn_{4-n} as well as Ga-centered clusters Ga-N_nO_{4-n}, both embedded in the ZnO supercell [18]. These choices cover most forms of doping, including "monodoping" ($n = 0$), "codoping" ($n = 2$), and "cluster doping" ($n = 3, 4$). Figure 1 shows $\Delta H_f^{(D,q)}(\mu, \epsilon_F)$ as a function of ϵ_F for various dopant clusters at the chemical potentials indicated in the figure. Table I gives the formation enthalpy ΔH_f and its components ΔE_b and ΔE_{CRE} for $\epsilon_F = 0$ [i.e., at the valence-band maximum (VBM)] for the chemical potential points 1 and 2 in Fig. 1. If free N atoms were used as the N source, we find similar trends of stability as in the NO source case. Our main findings are as follows:

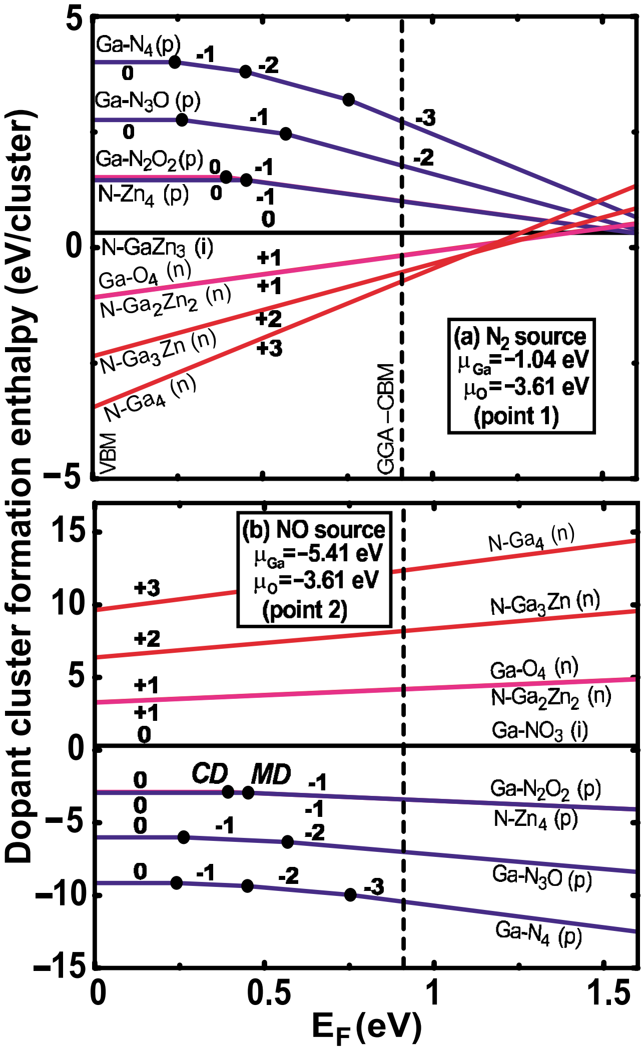

FIG. 1 (color online). The formation enthalpies of dopant clusters as a function of ϵ_F: (a) for N_2 source with the chemical potentials determined by point 1; (b) for NO source with the chemical potentials determined by point 2. The charge states q are denoted by numbers, whereas the transition energies $\epsilon(q/q')$ are denoted by solid dots. CD indicates "codoping," while MD denotes "monodoping."

(i) *Unsuccessful p doping via N_2.*—Table I shows that doping ZnO by N_2 (without Ga) leads to $\Delta H_f > 0$ and $\Delta E_b > 0$ for N-Zn_4. The positive ΔH_f implies a rather limited N solubility. Since as-grown ZnO is usually n type (because of intrinsic defects [8–10] or unintentional hydrogen impurity [4]), the low solubility of N monodoping may not create a sufficient number of holes to compensate the free electrons in order to obtain net p-type ZnO. This explains the failure of p-type doping by using N_2 sources [15].

(ii) *n-type doping with Ga and N_2 source.*—Figure 1(a) shows that using the N_2 source and Ga favors ($\Delta H_f < 0$) the clusters N-Ga_4 and N-Ga_3Zn which promote *n*-type doping, whereas the clusters Ga-N_4 and Ga-N_3O which promote *p*-type doping are less favorable ($\Delta H_f > 0$) under these conditions. Table I reveals that the stability of N-Ga_4 results from a highly favorable bonding energy ΔE_b comprising 4 Ga-N and 12 Ga-O bonds. This explains the surprising observation [13] that using Ga with an N_2 source produces *n*-type doping of ZnO, even though nitrogen is naively expected to dope II-VI's *p* type.

(iii) *Metastable p-type monodoping with the NO source.*—Figure 1(b) and Table I show that using the NO source (without Ga) leads to ΔH_f(N-Zn_4) < 0, i.e., soluble isolated nitrogen. This is so because $\Delta E_{CRE} \ll 0$ due to the use of a high-chemical energy N source. However, the excess bond energy for the four Zn-N bonds is $\Delta E_b > 0$, so these bonds might not be locally stable and could change over time. Note that even though $\Delta E_b > 0$ suggests a *driving* force for the local bonding to change, the *mechanism* of such changes depends on detailed kinetic factors such as lattice locations and the activation energy for N displacement, studied, e.g., in Refs. [7,19]. The local instability of a single N-on-O substitution means that this center could attract a second nitrogen, leading to N_2-on-O, which is a *donor* [5], or N-on-O can diffuse away, leaving an oxygen vacancy, which is also a donor [9,10]. This is consistent with the observation [15] that despite the introduction into ZnO of large quantities of N via a high-energy NO source, the *p*-type conductivity disappeared over time.

(iv) *Stable p-type cluster doping with Ga and NO source.*—Figure 1(b) shows that using the NO source with Ga stabilizes the Ga-centered dopant clusters Ga-N_4 and Ga-N_3O which promote *p*-type doping. Table I shows that these clusters have $\Delta H_f < 0$ *and* $\Delta E_b < 0$, so they are both soluble in ZnO *and* locally stable. In contrast, Table I shows that ΔH_f for codoping [6] (Ga-N_2O_2) is not better than monodoping via N-Zn_4. Thus, adding N and Ga with a 4:1 or 3:1 ratio promises to create both good solubility and local stability for *p*-type doping of ZnO. This is our most important result and awaits experimental testing. Although our present work is a thermodynamic study in nature and does not consider kinetic effects, the highly negative formation enthalpies that we find for these dopant clusters should provide a strong driving force for their formation. We estimate that the configurational entropy contribution at growth temperatures is about 0.5 eV in favor of stabilizing the monodoping (the vibrational entropy has even a much smaller contribution). Therefore, ignoring entropy contributions in the present study does not change the trends in Fig. 1.

(v) *Shallower acceptor levels for cluster doping.*—Solubility and local stability are not sufficient for successful doping; ionizability of the impurity is also needed. The solid dots in Fig. 1 indicate the ionization transition energies, i.e., the value of the Fermi energy at which there is a transition between different charge states [these are independent on μ, thus being the same in Figs. 1(a) and 1(b)]. We see that N monodoping (MD) via N-Zn_4 creates a rather deep acceptor at $E(0/-) = E_v + 0.45$ eV, codoping (CD) via Ga-N_2O_2 has $E(0/-) = E_v + 0.39$ eV, whereas cluster doping via Ga-N_4 has the shallowest acceptor level at $E(0/-) = E_v + 0.24$ eV. Thus, not only is cluster doping stable, but it also produces shallower defect levels.

(vi) *A simple bond-additive model explains trends in ΔE_b.*—The trends in ΔE_b can be understood by counting the number of bonds formed or destroyed in each doping scenario, as shown in the last three columns of Table I. We

TABLE I. Total formation enthalpy ΔH_f for various dopant clusters and its components ΔE_{CRE} [Eq. (3)] due to "chemical reservoir energy" and ΔE_b [Eq. (2)] due to "local bonding configuration." The chemical potentials μ_N, μ_O, μ_{Ga}, and μ_{Zn} are the same as in Fig. 1. The last three columns gives the number of bonds related to forming dopant clusters. In the column ΔE_b we give in the parentheses the values obtained by the approximate expression Eq. (9) as a sum of excess bond energies.

Dopant	Doping	ΔH_f (eV)		ΔE_{CRE} (eV)			Number of bonds		
cluster	type	N_2	NO	N_2	NO	ΔE_b (eV)	Ga-N	Ga-O	Zn-N
N-Ga_4[a]	n	−3.46	9.64	11.93	25.03	−15.39 (−15.44)	4	12	0
N-Ga_3Zn[a]	n	−2.35	6.38	9.02	17.75	−11.37 (−11.29)	3	9	1
N-Ga_2Zn_2[b]	n	−1.08	3.27	6.12	10.48	−7.20 (−7.14)	2	6	2
N-$GaZn_3$[c]	i	0.33	0.31	3.21	3.20	−2.89 (−2.99)	1	3	3
N-Zn_4[d]	p	1.46	−2.93	0.31	−4.07	1.15 (1.16)	0	0	4
Ga-O_4[d]	n	−1.07	3.29	2.90	7.27	−3.98 (−3.96)	0	4	0
Ga-NO_3[c]	i	0.33	0.31	3.21	3.20	−2.89 (−2.99)	1	3	3
Ga-N_2O_2[b]	p	1.53	−2.87	3.52	−0.87	−2.00 (−2.02)	2	2	6
Ga-N_3O[a]	p	2.77	−6.01	3.83	−4.94	−1.07 (−1.05)	3	1	9
Ga-N_4[a]	p	4.01	−9.14	4.14	−9.01	−0.13 (−0.08)	4	0	12

[a]Cluster doping. [b]Codoping. [c]Compensated pair. [d]Monodoping.

see, for example, that forming the N-Ga_4 cluster in ZnO involves creation of 4 Ga-N and 12 Ga-O bonds and destroying 16 Zn-O bonds. On the other hand, creating the Ga-N_4 cluster inside ZnO involves making 4 Ga-N and 12 Zn-N bonds, destroying 16 Zn-O bonds. Since we have three unknown excess bond energies ($E_{\text{Ga-O}}$, $E_{\text{Zn-N}}$, and$E_{\text{Ga-N}}$) and many more calculated ΔE_b values (Table I), it is straightforward to determine the excess bond energies for any doping configuration σ from a simple cluster expansion [20]

$$\Delta E_b(\sigma, \epsilon_F) = N^{(\sigma)}_{\text{Ga-N}} E_{\text{Ga-N}} + N^{(\sigma)}_{\text{Ga-O}} E_{\text{Ga-O}} + N^{(\sigma)}_{\text{Zn-N}} E_{\text{Zn-N}}, \tag{9}$$

where $N^{(\sigma)}_{\alpha\text{-}\beta}$ are the number of α-β bonds in configuration σ (last three columns in Table I). We obtain the excess bond energies $E_{\alpha\text{-}\beta}$ by the least-squares fit. The values of ΔE_b predicted by the model are given in parentheses in Table I and agree well with the directly calculated GGA values. We repeated the ten cluster calculations for Al + N doping of ZnO, the excess bond energies are the following:

$$\begin{aligned} E_{\text{Ga-N}} &= -0.89\text{ eV}, & E_{\text{Al-N}} &= -1.57\text{ eV}, \\ E_{\text{Ga-O}} &= -0.99\text{ eV}, & E_{\text{Al-O}} &= -1.90\text{ eV}, \\ E_{\text{Zn-N}} &= 0.29\text{ eV}. \end{aligned} \tag{10}$$

We now see that monodoping (N-Zn_4) is locally unfavorable since it replaces four Zn-O bonds by the tetrahedrally less stable four Zn-N bonds, whereas N-Ga_4 cluster doping is very favorable since it replaces 16 Zn-O bonds by 4 Ga-N and 12 Ga-O bonds that are more stable. We also see that doping with Al + N is better than doping with Ga + N since the corresponding III-N and III-O bonds are stronger for III = Al [Eq. (10)]. Inspection of a table of bond energies [21,22] can then be used to design optimal local bonding environments in various wide-gap solids. For example, in doping p-type ZnO Al + N and Ga + N are better than In + N, Al + P, or Ga + As.

In conclusion, we propose new design rules for doping with few dopant species, thus promoting a locally stable chemical environment, which has overall good solubility and shallow ionization energy.

The authors thank X. Li and T. Coutts for helpful discussions and for providing us a copy of Ref. [15] prior to its publication. This work was supported by U.S. DOE under Contract No. DE-AC36-98-GO10337.

---

[1] H. L. Hartnagel, A. L. Dewar, A. K. Jain, and C. Jagadish, *Semiconducting Transparent Thin Films* (IOP Publishing, Bristol, 1995).

[2] E. F. Schubert, *Doping in III-V Semiconductors* (Cambridge University Press, Cambridge, U.K., 1993).

[3] R. Bhargava, *Properties of Wide-Gap II-VI Semiconductors* (Inspec Publication, London, 1997).

[4] C. G. Van de Walle, Phys. Rev. Lett. **85**, 1012 (2000).

[5] Y. Yan, S. B. Zhang, and S. T. Pantelides, Phys. Rev. Lett. **86**, 5723 (2001).

[6] T. Yamamoto and H. Katayama-Yoshida, Jpn. J. Appl. Phys. **38**, L166 (1999); J. Cryst. Growth **214–215**, 552 (2000); Physica (Amsterdam) **302B–303B**, 155 (2001).

[7] E. C. Lee *et al.*, Phys. Rev. B **64**, 85120 (2001).

[8] F. Oba *et al.*, J. Appl. Phys. **90**, 824 (2001).

[9] A. F. Kohan *et al.*, Phys. Rev. B **61**, 15 019 (2000).

[10] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. B **63**, 75205 (2001).

[11] C. H. Park, S. B. Zhang, and S.-H. Wei, Phys. Rev. B **66**, 73202 (2002).

[12] K. Shimomura, K. Nishyama, and R. Kadono, Phys. Rev. Lett. **89**, 255505 (2002).

[13] M. Joseph *et al.*, Jpn. J. Appl. Phys. **38**, L1205 (1999); M. Joseph *et al.*, Physica (Amsterdam) **302B–303B**, 140 (2001).

[14] Y. R. Ryu *et al.*, J. Cryst. Growth **216**, 330 (2000).

[15] X. Li *et al.*, Electrochem. Solid State **6**, c56 (2003); X. Li *et al.*, J. Vac. Sci. Technol. (to be published).

[16] D. C. Look *et al.*, Appl. Phys. Lett. **81**, 1830 (2002); in Proceedings of the Second International Workshop on Zinc Oxide, Dayton, Ohio, USA, 2002.

[17] G. F. Neumark, Mater. Sci. Eng. **R21**, 1 (1997).

[18] We use the pseudopotential plane-wave total-energy method with the VASP code [G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996)], the generalized gradient approximation (GGA) for the exchange-correlation potential, and the Vanderbilt ultrasoft pseudopotentials. A 2 × 2 × 2 Monkhorst-Pack grid (including Γ) of wave vectors in the Brillouin zone of a 72-atom ZnO supercell and the plane-wave basis cutoff energy of 380 eV were used in our calculations. Both host and dopant atoms in the supercell were allowed to relax until the force on atoms are smaller than 0.025 eV/Å. Band gap error effects on ΔH_f were not corrected.

[19] S. Gundel and W. Faschinger, Phys. Rev. B **65**, 035208 (2002).

[20] A. Zunger *et al.*, Modeling Simul. Mater. Sci. Eng. **10**, 685 (2002); A. Zunger, *Statics and Dynamics of Alloy Phase Transformations*, edited by P. E. A. Turchi and A. Gonis (Plenum Press, New York, 1994), p. 361.

[21] J. E. Huheey, E. A. Keiter, and R. L. Keiter, *Inorganic Chemistry: Principles of Structure and Reactivity* (Harper Collins College Publishers, New York, 1993), 4th ed., pp. A25–A33.

[22] W. A. Harrison, *Electronic Structure and the Properties of Solids: the Physics of the Chemical Bond* (Dover Publications, Inc., New York, 1989), p. 176.