13
Overcoming Bipolar Doping Difficulty in Wide Gap Semiconductors

Su-Huai Wei and Yanfa Yan

13.1
Introduction

Application of semiconductors as electric and optoelectronic devices depends critically on their dopability. Failure to dope a material, i.e., to produce enough free charge carriers beyond a certain limit at working temperature, is often the single most important bottleneck for advancing semiconductor-based high technology. Wide band gap (WBG) semiconductors, such as diamond, AlN, GaN, MgO, and ZnO, have unique physical properties that are suitable for applications in short-wavelength and transparent optoelectronic devices [1–9]. To realize these applications, the formation of a high-quality homo p–n junction is essential. In other words, high-quality bipolar (p- and n-type) doping are required for the same material. Unfortunately, most WBG semiconductors experience a serious doping asymmetry problem, i.e., they can easily be doped p- or n-type, but not both [10]. For example, diamond can be doped relatively easily p-type, but not n-type [1–3]. On the other hand, ZnO can easily be made high-quality n-type, but not p-type [7–9]. For materials with very large band gap such as AlN and MgO, both p- and n-type doping are difficult [11]. The doping difficulty also exists in nanostructure semiconductors where the band gaps increase due to the quantum confinement [12, 13]. These doping problems have hindered the potential applications of many WBG and nanostructure materials.

Extensive research has been done to understand the origin of the bipolar doping difficulties in WBG and nanostructure semiconductors and to find possible solutions to overcome the doping difficulty. In the past, we have proposed various approaches to overcome the bipolar doping difficulty in WBG semiconductors. These approaches have been tested by the systematical calculation of defect formation energies and transition energy levels of intrinsic and extrinsic defects in various WBG and nanostructure semiconductors using first-principles density-functional theory [14–23]. In this paper, we review the origins of the bipolar doping difficulty and describe our approaches for overcoming the doping bottleneck in WBG semiconductors.The paper is organized as follows. Section 13.2 discusses the salient features of calculating defect properties, in which the issues related to the finite size of the

Advanced Calculations for Defects in Materials: Electronic Structure Methods, First Edition.
Edited by Audrius Alkauskas, Peter Deák, Jörg Neugebauer, Alfredo Pasquarello, and Chris G. Van de Walle.

supercell and the band gap errors are discussed. Section 13.3 analyzes symmetry and occupation of defect levels. Section 13.4 describes what causes the bipolar doping difficulty in WBG semiconductors and the origin of the doping limit rules. Section 13.5 describes our proposed approaches to overcome the bipolar doping difficulties, focusing mostly on ZnO and other WBG semiconductors. Section 13.6 briefly summarizes the paper and provides an outlook for future research in this field.

13.2 Method of Calculation

We performed the band structure and total energy calculations using the first-principles density-functional theory with local density approximation (LDA) or general gradient approximation (GGA) [24, 25]. We used the supercell approach in which the defects or defect complexes are put at the center of a supercell and the periodic boundary condition is applied. For quantum dots (QDs) the surface are passivated by hydrogen or pseudohydrogen [12, 13]. In all calculations, all the atoms are allowed to relax until the Hellman–Feynman forces acting on them become negligible. For charged defects, a uniform background charge is added to keep the global charge neutrality of the supercells [19].

To determine the defect formation energy and defect transition energy levels, one needs to calculate the total energy $E(\alpha, q)$ for a supercell containing defect α in charge state q, the total energy E(host) of the same supercell without the defect, and the total energies of the involved elemental solids or gases at their stable phases. It is important to realize that the defect formation energy depends on the atomic chemical potentials μ_i and the electron Fermi energy E_F. From these quantities, the defect formation energy, $\Delta H_f(\alpha, q)$, can be obtained by:

$$\Delta H_f(\alpha, q) = \Delta E(\alpha, q) + \sum n_i \mu_i + q E_F, \tag{13.1}$$

where $\Delta E(\alpha, q) = E(\alpha, q) - E(\text{host}) + \sum n_i E(i) + q\varepsilon_{VBM}(\text{host})$. E_F is referenced to the valence band maximum (VBM) of the host. μ_i is the chemical potential of constituent i referenced to elemental solid/gas with energy $E(i)$; n_i is the number of elements and q is the number of electrons transferred from the supercell to the reservoirs in forming the defect cell. The transition energy $\varepsilon_\alpha(q/q')$ is the Fermi energy at which the formation energy of defect α at charge state q is equal to that at charge state q'. Using Eq. (13.1), the transition energy level with respect to the VBM can be obtained by

$$\varepsilon_\alpha(q/q') = \frac{\Delta E(\alpha, q) - \Delta E(\alpha, q')}{q' - q} - \varepsilon_{VBM}(\text{host}). \tag{13.2}$$

Typically, for finite supercell, the Brillouin zone integration for the charge density and total energy calculations is performed using special k-points or equivalent k-points in the superstructures. This approach gives better convergence on the calculated charged density and total energy. However, it usually gives a poor description on the symmetry and energy levels of the defect state, as well as the

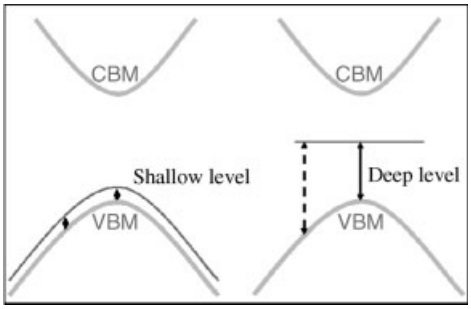

Figure 13.1 (online colour at: www.pss-b.com) Schematic plot of the defect level with respect to the band edge. It is shown that in a supercell calculation, the energy level determined at the Γ-point is correct for both shallow and deep levels.

VBM and conduction band minimum (CBM) states, and the results could be sensitive to the *k*-points sampling because the band edges are determined by the *k*-point sampling. To avoid this problem, a Γ-point-only calculation is often used to determine the total energy and transition energy level because the symmetry of the defect and the band edge states are well defined at Γ, and both the shallow and deep levels are correctly described (see Figure 13.1). However, for small supercell, Γ-point-only approach may give poor total energy convergence. Here, we propose to use a hybrid scheme to combine the advantages of both special *k*-points and Γ-point-only approaches [14, 19]. In this scheme, we first calculate the transition energy level *with respect to VBM*, which is given by

$$\varepsilon(0/q) = \varepsilon_{\mathrm{D}}^{\Gamma}(0) - \varepsilon_{\mathrm{VBM}}^{\Gamma}(\mathrm{host}) + \frac{E(\alpha, q) - (E(\alpha, 0) - q\varepsilon_{\mathrm{D}}^{k}(0))}{-q}. \tag{13.3}$$

For donor level ($q > 0$), it is usually more convenient to reference the ionization energy level to the CBM, i.e., we can rewrite Eq. (13.3) as

$$\varepsilon_{\mathrm{g}}^{\Gamma}(\mathrm{host}) - \varepsilon(0/q) = \varepsilon_{\mathrm{CBM}}^{\Gamma}(\mathrm{host}) - \varepsilon_{\mathrm{D}}^{\Gamma}(0) + \frac{E(\alpha, q) - (E(\alpha, 0) - q\varepsilon_{\mathrm{D}}^{k}(0))}{q}, \tag{13.4}$$

where a positive number calculated from Eq. (13.4) indicates the distance of the transition energy level below the CBM. In Eqs. (13.3) and (13.4) $\varepsilon_{\mathrm{D}}^{k}(0)$ and $\varepsilon_{\mathrm{D}}^{\Gamma}(0)$ are the defect levels at the special *k*-points (weight averaged) and at the Γ-point, respectively; and $\varepsilon_{\mathrm{VBM}}^{\Gamma}(\mathrm{host})$ and $\varepsilon_{\mathrm{CBM}}^{\Gamma}(\mathrm{host})$ are the VBM and CBM energies, respectively, of the host at the Γ-point. $\varepsilon_{\mathrm{g}}^{\Gamma}(\mathrm{host})$ is the band gap at the Γ-point. The first term on the right-hand side of Eqs. (13.3) or (13.4) give the single-electron defect level at the Γ-point. The second term determines the relaxation energy U (including both the Coulomb contribution and the atomic relaxation contribution) of the charged defects calculated at the special *k*-points, which is the extra cost of energy

after moving $(-q)$ charge to the neutral defect level with $E = \varepsilon_D^k(0)$. Afterwards, the formation energy of defect α at charge state q can be obtained by

$$\Delta H_f(\alpha, q) = \Delta H_f(\alpha, 0) - q\varepsilon(0/q) + qE_F, \tag{13.5}$$

where $\Delta H_f(\alpha,0)$ is the formation energy of the charge-neutral defect and E_F is the Fermi level with respect to the VBM. This new approach has been used successfully for studying defects in various semiconductors.

We want to point out that in writing down Eqs. (13.1)–(13.4), we assumed that the common reference energy level is used in the calculation of defect α in different charge states. Because in a periodic supercell calculation the zero potential energy is not well defined, therefore, we have to lineup the potential using a common reference. This is usually done by lineup core level of an atom far away from the defect center. In the case where core level is not available, average potential around the atom can also be used.

In the formula above, we also assumed that the VBM energy position is given by its eigenvalue ε_{VBM}(host). For small supercell, however, the Koopman's theorem may not hold that is

$$\delta E = \varepsilon_{VBM}(\text{host}) - [E(\text{host}, N) - E(\text{host}, N-1)] \tag{13.6}$$

is not zero, especially for VBM with localized electron states. Similar situation may also exist for the CBM energy position. In this case, the correction term δE should be added to determine the VBM or CBM energy level.

It the supercell calculation, the periodic boundary conditions introduces a spurious Coulomb interaction between the charged defects in different cells. To estimate the magnitude of this interaction, point charges immersed in neutralizing jellium are usually assumed. Attempts to including higher-order terms are difficult, because the higher-order multipoles of the defect charge are not uniquely defined in the supercell approach [26]. However, in reality, the charged defect does not have a delta-function-like distribution, especially for shallow defects, which have a relatively uniform charge distribution. Therefore, direct application of the Makov and Payne [26] correction using only point charge often overestimates the effect [19]. Thus, for shallow low charge state defects we usually assume the Makov and Payne correction are not important. However, this correction term could be large if the defect level is localized or the defect is in a small QD.

Another issue that leads to uncertainty in defect calculation is caused by the fact that LDA or GGA calculations underestimate the band gap of a semiconductor. One way to correct this error is to project the defect level to the CBM and VBM states, and shift the defect level accordingly when the band edges are shifted to correct the band gap error [27]. Another way to correct this error is using high level DFT calculation such as GW approach [28]. However, currently, this type of full scale calculation is still formidable. Recently, more empirical hybrid density-functional method [29, 30] which mixes certain amount of Hartree–Fock potential with the GGA potential is used for defect calculation. Although this kind of approach can correct the band gap in some empirical way, the symmetry breaking caused by the Hartree–Fock potential

and the accuracy of the calculated defect levels, in our opinion, still needs experimental verification.

13.3
Symmetry and Occupation of Defect Levels

It is often very useful to know the symmetry and character of the single-particle defect level before we start the calculation, because it can help us identify the defect level and because defects with different symmetry and character will behave differently. Moreover, to modify defect states or correct the band gap error, it is also important to know the symmetry of the state. For example, although both anion vacancy and cation interstitial have the same a_1 defect levels in II–VI semiconductors [23], anion vacancy has the a_1^v character derived from the valence band or cation dangling bond states, whereas cation interstitial has instead the a_1^c character derived from the conduction band. Thus, the energy level of the cation interstitial is expected to follow closely with the CBM, whereas the energy level of anion vacancy state will not.

For simple extrinsic impurities, one can predict in principle whether a dopant is a donor with a single-particle energy level close to the CBM or an acceptor with a single-particle energy level close to the VBM by simply counting the number of the valence electrons of the dopants and the host elements. For example, in CdTe, one can expect that group-I elements substituting on the Cd site, X_{Cd}^{I} create acceptors, whereas group-VII elements substituting on the Te site, Y_{Te}^{VII} creates donors. Generally speaking, to produce a shallow acceptor, it is advantageous to use a more electronegative dopant, whereas to produce a shallow donor, it is advantageous to use a less electronegative dopant.

For intrinsic defects, the situation is more complicated. Figure 13.2 shows the single-particle energy levels of tetrahedrally coordinated charge-neutral defects in CdTe [23]. Generally speaking, when a high-valence atom is replaced by a low-valence

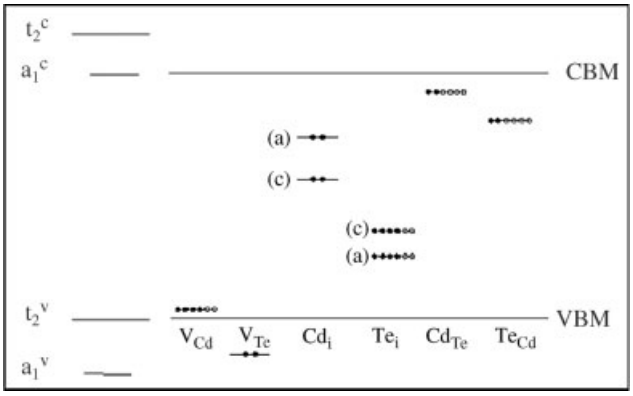

Figure 13.2 (online colour at: www.pss-b.com) Single-particle defect levels for the tetrahedrally coordinated neutral intrinsic defects in CdTe. The solid (open) dots indicate the state is occupied (unoccupied).

atom (e.g., Cd_{Te}) or by a vacancy V_{Cd} and V_{Te}, defect states are created from the host valence (v) band states that move upward in energy. The defect states consist of a low-lying singlet a_1^v state and a high-lying threefold-degenerate t_2^v state. Depending on the potential, both a_1^v and t_2^v can be above the VBM. These states are occupied by the nominal valence electrons of the defect plus the valence electrons contributed from the neighboring atoms (e.g., in CdTe, six electrons if the defect is surrounded by four Te atoms or two electrons if it is surrounded by four Cd atoms). For example, for charge-neutral V_{Cd}, the defect center has a total of $0 + 6 = 6$ electrons. Two of them will occupy the a_1^v state and the remaining four will occupy the t_2^v states just above the VBM, so V_{Cd} is an acceptor. On the other hand, if a low-valence atom is replaced by a high-valence atom (e.g., Te_{Cd}), or if a dopant goes to an interstitial site (e.g., Cd_i and Te_i), the a_1^v and t_2^v are pulled down and will remain inside the valence band. Instead, the defect states a_1^c and t_2^c are created from the host conduction band states that move down in energy. Depending on the potential, both the a_1^c and t_2^c states can be in the gap. For example, for charge-neutral Te_{Cd}, $6 + 6 = 12$ electrons are associated with this defect center. Eight of them will occupy the bonding a_1^v and t_2^v states, two will occupy the a_1^c state, and the remaining two will occupy the t_2^c state. Since the partially occupied t_2^c state is close to the CBM, Te_{Cd} is also a donor. For the interstitial defect, Cd_i has two electrons that will fully occupy the a_1^c state and is thus expected to be a donor. The Te_i defect center has six electrons. Two will occupy the a_1^c state and the remaining four will occupy the t_2^c states. Since the partially occupied t_2^c states are closer to the VBM, Te_i is expected to be a deep acceptor.

13.4 Origins of Doping Difficulty and the Doping Limit Rule

In general, there are three main factors that could cause the doping limit in a semiconductor material [10, 14, 19, 22, 23]: (i) the desirable dopants have limited solubility; (ii) the desirable dopants have sufficient solubility, but they produce deep levels, which are not ionized at working temperatures; and (iii) there is spontaneous formation of compensating defects. The first factor depends highly on the selected dopants and growth conditions. The second factor only depends on the selected dopants. Thus, these two factors can sometimes be suppressed by carefully selecting appropriate dopants and controlling the growth conditions. The third factor is an intrinsic problem for semiconductors; thus, it is the most difficult problem to overcome, especially for WBG semiconductors. This is because the formation energy of charged compensating defects depends linearly on the position of the Fermi level, E_F [see Eq. (13.1)]. When a semiconductor is doped, the Fermi level shifts, which can lead to spontaneous formation of the compensating charged defects. For example, when a semiconductor is doped p-type, E_F moves close to the VBM. In this case, the formation energy of the charged donor defects decreases because they will donate their electrons into the Fermi reservoir (Figure 13.3). In WBG semiconductors with low VBM, the formation energy decrease of donor defects can be so large that at some Fermi energy $E_F = \varepsilon_{pin}^{(p)}$ the formation energy of certain donor defect becomes zero,

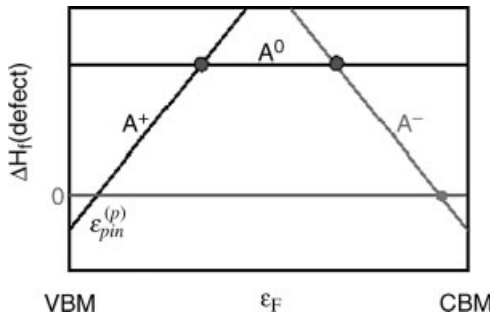

Figure 13.3 (online colour at: www.pss-b.com) Schematic plot of the dependence of the formation energy of charged defects on the Fermi energy position. The p-type pinning energy $\varepsilon_{pin}^{(p)}$ is the Fermi energy E_F at which the formation energy of donor A has zero formation energy.

i.e., it can form spontaneously, so further shift of the Fermi energy is not possible. Moreover, low VBM also leads to high ionization energy. Therefore, *a semiconductor with low VBM is difficult to be doped p-type*. The trend for n-type doping is similar, i.e., *a semiconductor with high CBM is difficult to be doped n-type*. This doping limit rule explains why a semiconductor with large band gap usually cannot be doped one type or even both types under equilibrium thermodynamic growth conditions. It also provides a general guideline about whether a material can be doped p- or n-type if we know the band alignment between different compounds. For example, Figure 13.4

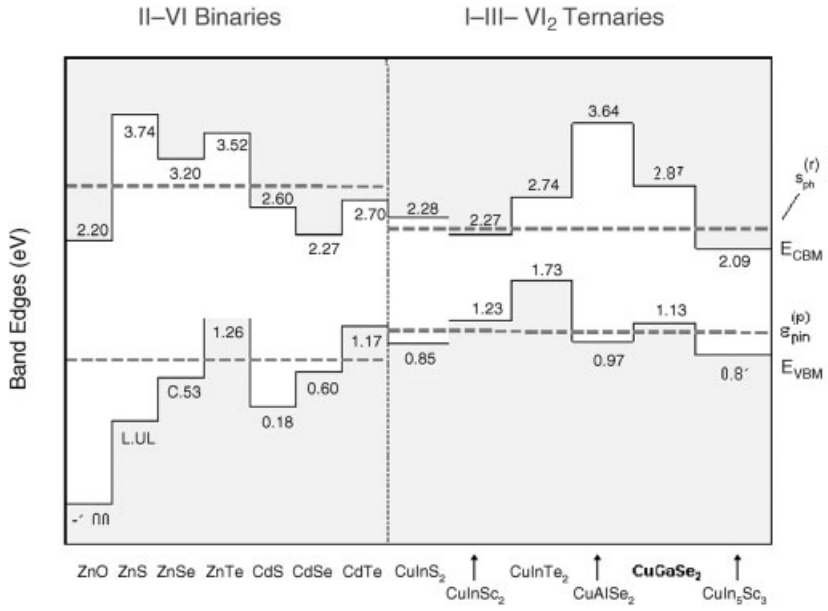

Figure 13.4 (online colour at: www.pss-b.com) Band alignment and n- and p-type pinning energy of II–VI and I–III–VI semiconductors (Ref. [10]).

shows the calculated band alignment for II–VI and I–III–VI semiconductors [10]. We see that ZnO has very low CBM and VBM, so it can be easily doped n-type, but not p-type. On the other hand, ZnTe with high VBM energy can be easily doped p-type, but not n-type. For I–III–VI compounds, $CuInSe_2$ can be doped both p- and n-type, but for $CuGaSe_2$, n-type doping will be difficult.

Based on the above understanding, we will search for corresponding solutions to overcome the doping limit. We will focus on the following approaches: (i) increase defect solubility by "defeating" bulk defect thermodynamics using non-equilibrium growth methods such as extending the achievable chemical potential through molecular doping or raising the host energy using surfactant; (ii) reduce defect ionization energy level by designing shallow dopants or dopant complexes; and (iii) reduce defect compensation and ionization level by modifying the host band structure near the band edges. As examples, we will discuss doping in some representative WBG semiconductors such as n-type doping in ZnTe and diamond and p-type doping in ZnO. The principles discussed here are general and are applicable to other WBG semiconductors.

13.5 Approaches to Overcome the Doping Limit

13.5.1 Optimization of Chemical Potentials

13.5.1.1 Chemical Potential of Host Elements

As Eq. (13.1) indicates, the formation energy of a defect, which determines the solubility of dopants, depends sensitively on the atomic chemical potentials of both the host elements and the dopants [14, 19]. Thus, optimization of the growth conditions and dopant source is critical to enhance the doping ability. So far, computational results and analysis have focused on the dependence of formation energies on host-element chemical potentials [27, 31]. For example, N substituting O (N_O) is expected to be a p-type dopant for ZnO. The formation energy of N_O depends on the chemical potentials of Zn, O, and N. Under thermal equilibrium growth conditions, there are some thermodynamic limits on the achievable values of the chemical potentials. First, to avoid precipitation of the elemental dopant and host elements, the chemical potentials are limited by

$$\mu_{Zn} \leq \mu(\text{Zn metal}) = 0, \tag{13.7}$$

$$\mu_O \leq \mu(O_2\ \text{gas}) = 0, \tag{13.8}$$

$$\mu_N \leq \mu(N_2\ \text{gas}) = 0. \tag{13.9}$$

Second, μ_{Zn} and μ_O are limited to the value of maintaining ZnO. Therefore,

$$\mu_{Zn} + \mu_O = \Delta H_f(\text{ZnO}). \tag{13.10}$$

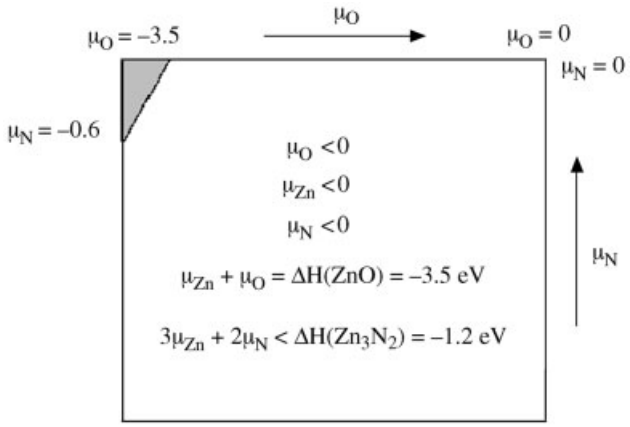

Figure 13.5 The achievable chemical potential region for N doped ZnO under equilibrium growth condition.

Here, ΔH_f(ZnO) is the formation energy of bulk ZnO. The calculated value is about −3.5 eV. Finally, to avoid the formation of the Zn_3N_2 secondary phase, μ_N is also limited by

$$3\mu_{ZN} + 2\mu_N \leq \Delta H_f(Zn_3N_2). \tag{13.11}$$

The calculated formation energy of Zn_3N_2 is about −1.2 eV. Using the equations above, the achievable chemical potential region is shown in Figure 13.5.

Figure 13.6 shows the calculated formation energies of charge neutral defects as a function of the O chemical potential. Here, μ_N is derived from N_2 gas. It is seen that the formation energy of N_O is lower at the O-poor condition, but higher at the O-rich

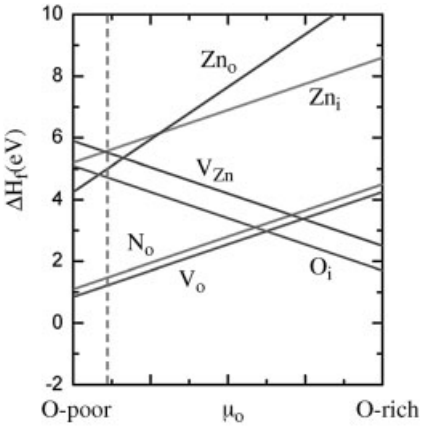

Figure 13.6 (online colour at: www.pss-b.com) Calculated formation energies of charge-neutral defects as a function of O chemical potential. The dashed line indicates the growth condition at which (left region) Zn_3N_2 will precipitate.

condition. Thus, to enhance the solubility of N, ZnO films should be synthesized at O-poor conditions. It should be noted that the formation energies of other intrinsic defects also depend on the growth conditions. At O-poor conditions, the formation energies for "acceptor-killer" defects, such as Zn interstitials (Zn_i) and O vacancies (V_O), are decreased. Thus, there is an intrinsic problem for enhancing the p-type doping using N as the dopant. In Section 13.5.4, we will discuss how this dilemma may be eliminated by selecting appropriate dopants.

The left region of the dashed line indicates the growth condition at which Zn_3N_2 will precipitate under equilibrium growth. This indicates that the O chemical potential should not go into this region so as to avoid the precipitation of Zn_3N_2. Thus, the precipitation of a secondary phase can limit the solubility of dopants. To overcome this, it has been shown that the precipitation of a secondary phase can be suppressed through epitaxial growth [32]. For example, the calculated thermodynamic solubility of N in bulk GaAs is only $[N] < 10^{14}\,\mathrm{cm}^{-3}$ at $T = 650\,^{\circ}\mathrm{C}$ due to the formation of a fully relaxed, secondary GaN phase. However, single-phase epitaxial films grown at $T = 400$–$650\,^{\circ}\mathrm{C}$ with $[N]$ as high as ~10% have been reported. Zhang and Wei [32] found that if coherent surface strain is considered, the formation of the secondary GaN phase could be suppressed during epitaxial growth. As a result, the solubility of N can be enhanced significantly. A similar approach could be used to avoid forming Zn_3N_2 in ZnO:N.

Avoiding the formation of secondary phase can also lead to some unexpected consequences. For example, substituting Zn by Al for n-type doping in ZnO, the formation energy of Al_{Zn} depends on $(\mu_{Zn} - \mu_{Al})$, thus one may expect that the formation energy of Al_{Zn} should reach minimum under Zn-poor condition. However, to avoid the formation of Al_2O_3, we need to satisfy the following condition:

$$2\mu_{Al} + 3\mu_O \leq \Delta H_f(Al_2O_3). \tag{13.12}$$

Combine Eqs. (13.10) and (13.12), we have

$$(\mu_{Zn} - \mu_{Al}) \geq \frac{\mu_O + 2\Delta H_f(ZnO) - \Delta H_f(Al_2O_3)}{2}. \tag{13.13}$$

That is, in the achievable chemical potential region, Al_{Zn} has the lowest formation under O-poor or Zn-rich condition.

13.5.1.2 Chemical Potential of Dopant Sources

Although the dependence of doping efficiency on the host element's chemical potential has been studied extensively, the dependence of doping efficiency on *dopant* chemical potential has not attracted much attention because normally there are no significant alternative dopant sources. However, for 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 that can be used as the dopant source. If these molecules arrive intact at the growing surface, their chemical potentials will determine the doping efficiency. We found that the N solubility can be enhanced significantly when metastable NO or NO_2 gases are used as the dopant sources [15]. A key feature that underlies our idea of doping with NO or NO_2 is that

these molecules can supply single N atoms by breaking only the weak N–O bonds, whereas one has to break the strong N–N bonds to obtain desirable single-N defects when N_2 and N_2O are used. For example, when the NO, N_2O, or NO_2 molecules arrive *intact* at the growing surface, the formation energy of a charge-neutral N_O defect is given by

$$\Delta H_f(N_O, 0) = E(N_O, 0) - E(\text{host}) + 2\mu_O - \mu_{NO} \tag{13.14}$$

or

$$\Delta H_f(N_O, 0) = E(N_O, 0) - E(\text{host}) + 1.5\mu_O - \frac{\mu_{N_2O}}{2}, \tag{13.15}$$

$$\Delta H_f(N_O, 0) = E(N_O, 0) - E(\text{host}) + 3\mu_O - \mu_{NO_2}. \tag{13.16}$$

Our calculated formation energy is about 0.9, 0.2, and −0.2 eV for NO, N_2O, and NO_2 molecules, respectively. The NO and NO_2 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, $\mu_{NO} = \mu N_2O - \mu_N$. Figure 13.7 shows the calculated formation energy of charge-neutral N_O for four different gases. 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 any energy to break the 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, if these molecules are intact before they are incorporated into ZnO.

However, in practical growth conditions, to avoid the precipitation of the secondary phases such as Zn_3N_2, besides satisfying

$$\mu_O + \mu_N = \mu_{NO} \quad \text{for NO gas as dopant,} \tag{13.17}$$

$$2\mu_O + \mu_N = \mu_{NO_2} \quad \text{for } NO_2 \text{ gas as dopant,} \tag{13.18}$$

$$\mu_O + 2\mu_N = \mu_{N_2O} \quad \text{for } N_2O \text{ gas as dopant,} \tag{13.19}$$

we also need to satisfy Eqs. (13.10) and (13.11), which set a low limit for the achievable O chemical potential. In Figure 13.7, the left, middle, and right dashed vertical lines indicate the low limits for O chemical potentials for N_2O, NO, and NO_2 molecules, respectively.

13.5.2
H-Assisted Doping

As we discussed above, the solubility of both the dopants and compensating defects depend sensitively on the position of the Fermi level. If we can control the Fermi level at a desirable position, then we may enhance the solubility of dopants and suppress the formation of compensating defects. Electron or hole injection could be a method to control the position of the Fermi level during film growth. Another popular

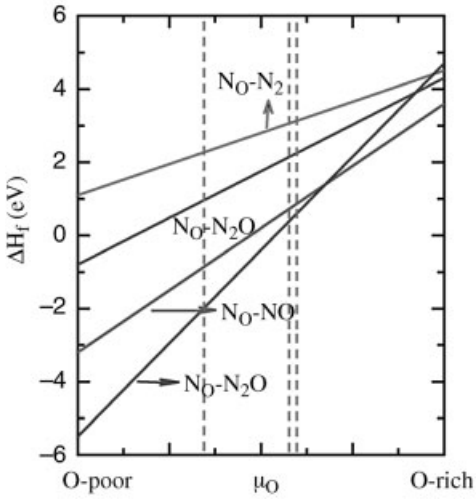

Figure 13.7 (online colour at: www.pss-b.com) Calculated formation energy of charge-neutral N_O with four different dopant sources: N_2, N_2O, NO, and NO_2. The left, middle, and right dashed lines indicate the low limits for achievable O chemical potentials for N_2O, NO, and NO_2 molecule doping.

approach is passivating the dopants by H atoms. For example, in Mg-doped GaN, the introduction of H can prevent such a shift. As a result, the concentration of Mg can be enhanced [34]. After film growth, H can be annealed out to achieve p-type conductivity.

For N-doping in ZnO (also other oxides), the co-existence of H can, in addition to preventing the Fermi level shift, directly passivate N dopants, forming a molecular NH complex on O site [$(NH)_O$]. The binding energy for $(NH)_O$ is 2.9 eV. The $(NH)_O$ complexes electronically mimic O atoms and cause smaller lattice distortion than N_O. Thus, the concentration of $(NH)_O$ in ZnO can be much higher than N_O [14]. Figure 13.8 shows the calculated formation energy for $(NH)_O$ as a function of O chemical potential. For comparison, the formation energies of N_O and some hole-killer defects are also shown. It is seen that the formation energy of $(NH)_O$ is lower than any other defects in the O-poor condition. In addition, the existence of H also pins the Fermi energy level, so the formation of compensating defects enhanced by a shifted Fermi level is also suppressed. Therefore, p-type doping could be achieved after subsequently driving out the hydrogen atoms from the sample by thermal annealing.

13.5.3
Surfactant Enhanced Doping

To lower the defect formation energy, which is the total energy difference between the final doped state and the initial state, we can either increase the initial dopant energy, as discussed in the previous section, or increase the energy of the host. Recently, we have shown that this can be done by introducing an appropriate surfactant during

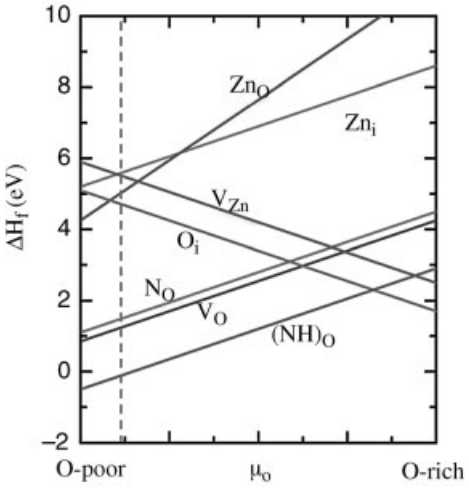

Figure 13.8 (online colour at: www.pss-b.com) Calculated formation energy of $(NH)_O$ in ZnO.

epitaxial growth [35]. The general concept for enhancing dopant solubility via epitaxial surfactant growth is schematically described in Figure 13.9. It is known experimentally that the surfactants in epitaxial growth float on the top surface of the growth front. The enhancement of dopant solubility initiates in the sublayers below the surface. For p-type doping, dopants introduce acceptor levels with holes near the VBM of the host system. On the other hand the surfactants on the growth surface will introduce surfactant levels. If the surfactant levels are higher in energy than the acceptor levels and have electrons available, the surfactants will donate the electrons to the acceptors and consequently leads to a Coulomb binding between the surfactant and the dopant. Such charge transfer reduces the energy of the system and consequently leads to effective reduction on the formation energy of dopant incorporation in the host. The formation energy reduction is large if the energy level

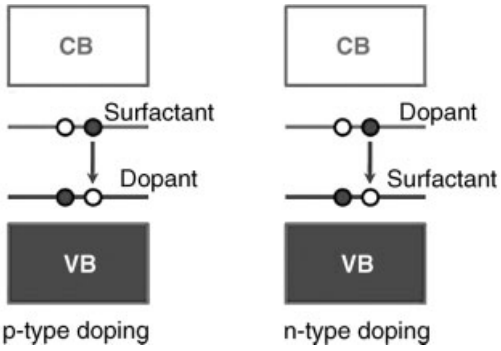

Figure 13.9 (online colour at: www.pss-b.com) Schematic plot of the mechanism of surfactant enhanced doping during epitaxial growth.

difference between the surfactant and acceptor levels is large. The same principle holds for n-type doping, except that the charge transfer is from dopant level to empty surfactant level. In this case, the positions of dopant levels are close to the CBM and the surfactant levels must be lower in energy than the dopant levels. Thus, the key for this concept is how to ensure that the surfactant levels are higher (lower) in energy than the dopant levels and there are indeed electrons (holes) in the surfactant levels in p-type (n-type) doping. We have calculated the formation energy for substitutional Ag_{Zn} in the sublayer of ZnO $(000\bar{1})$ surface and found that the formation energy is lowered by 2.3 eV with S as surfactant as compared to that without surfactant.

13.5.4 Appropriate Selection of Dopants

There are two general rules for choosing an appropriate dopant to produce shallow defect levels. First, an appropriate dopant should favor the growth conditions that will suppress the formation of compensating defects. As we discussed in Section 13.2, the solubility of dopants and concentration of intrinsic defects depend sensitively on the growth conditions. It is highly desirable to have a growth condition that enhances the dopant solubility and suppresses the formation of intrinsic compensating defects. This may be achieved by identifying the compensating defects and choosing suitable dopants. For example, the major defects that compensate acceptors in ZnO are Zn_i and V_O. Figure 13.6 shows that to suppress the formation of these hole-killer defects, an O-rich growth condition is preferred. Of course, this condition is not preferred for N incorporation. However, such a growth condition is preferred for doping at cation sites.

Second, dopants at cation sites in compound semiconductors generally produce shallower acceptor levels than dopants at anion sites. This is because for most cation–anion compound semiconductors, the valance bands are derived mainly from the anion atoms. Dopant substituting at cation site would, in general, cause smaller perturbation than dopants at anion sites on the anion-derived valance band. Thus, theoretical studies have found that Group-I elements such as Li and Na have low acceptor levels, whereas Group-V elements such as N, P, As, and Sb have deep acceptor levels in ZnO [33]. The large ionization energy for Group-V acceptors in ZnO can be understood as follows: The acceptor level, especially the shallow acceptor level, has a wave-function character similar to that of the VBM state, which consists mostly of anion p, and small amounts of cation p and cation d orbitals. Therefore, to have a shallow acceptor level, the dopant should be as electronegative as possible, that is, it should have low p orbital energy. For example, because the atomic p orbital energy level of N is the lowest (Figure 13.10), i.e., most electronegative, among the Group-V elements, N has been the preferred acceptor dopant for II–VI semiconductors because it produces the lowest acceptor levels compared to the other Group-V dopants. However, due to the low VBM of the oxides, the level of N_O in ZnO is still relatively deep [19, 33] at about 0.4 eV above the VBM, making acceptor ionization difficult. The other Group-V elements are less electronegative than the N atom; therefore, they have much larger ionization energy than N_O [33]. This explains why it

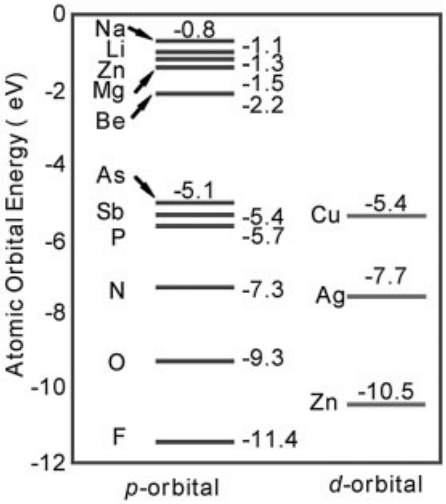

Figure 13.10 (online colour at: www.pss-b.com) LDA-calculated valence p and d energy levels of neutral atom to show the general chemical trends.

is difficult to achieve anion site shallow acceptors for the oxides. Recent experiments, however, have demonstrated good p-type conductivity for As, as well as p-doped ZnO. However, theoretical studies revealed that As impurities actually occupy Zn antisites, forming $As_{Zn} + 2V_{Zn}$ complexes [36]. The real contribution to the p-type conductivity is from V_{Zn}. The ionization energy is reduced due to the interaction between V_{Zn} and As_{Zn}.

According to the above discussion, Group-Ia (Li, Na) and Group-Ib (Cu, Ag) elements may be better choices for producing p-type ZnO. So far, only doping with Group-Ia elements has been studied extensively. There are very few studies on doping of ZnO with Group-Ib elements. However, very few p-type ZnO films have been achieved using Group-Ia elements as dopant. Theoretical studies have revealed the possible reasons for the difficulty. Substitutional Group-Ia elements (Li and Na) at T_d sites are indeed shallow acceptors [33]. However, when the Fermi energy is close to the VBM, Group-Ia elements prefer to occupy the interstitial sites in ZnO, which are electron donors. As a result, Group-Ia elements fail to dope ZnO p-type. The reason why Li and Na prefer the interstitial sites rather than substitutional sites is largely due to the low ionization energies of the valence *s* electron and large size mismatch of ions of the Group-Ia elements. Such mismatches are much less for Group-Ib elements. Thus, Group-Ib elements may be better candidates than Group-Ia elements for p-type ZnO doping. Therefore, we have studied the doping effect with Group-Ib elements in ZnO.

Our electronic structure calculations have revealed that Cu, Ag, or Au occupying a Zn site creates a single-acceptor state above the VBM of ZnO. Our calculated GGA transition energies $\varepsilon(0/-)$ are at about 0.7, 0.4, and 0.5 eV above the VBM for Cu_{Zn}, Ag_{Zn}, and Au_{Zn}, respectively [20]. These results indicate that (i) the acceptor level

created by Ag_{Zn} is shallower than the acceptor levels created by Cu and Au and (ii) the transition energies for the substitutional Group-Ib elements are much deeper than that of the substitutional Group-Ia elements. The reason for (ii) can be understood as the following: The substitutional elements induced acceptor level is derived mostly from the VBM state, which has the anion p and cation d characters. For Group-Ib elements, their occupied d orbital energies are near the oxygen p level. Because both the O, p and the Group-Ib d orbitals have the same t_2 symmetry in the tetrahedral environment, there is strong p–d repulsion between the two levels, pushing the acceptor levels higher. On the other hand, Group-Ia elements have no active valence d orbitals, so their defect levels are shallower than the Group-Ib substitutional defects. Among the three Group-Ib elements, Ag has the largest size and lowest atomic d orbital energy, so the p–d repulsion is the weakest. This explains why Ag_{Zn} has the lowest transition energy level among the three Group-Ib elements.

As we discussed above, although the Group-Ia substitutional acceptor levels are shallower, the Group-Ia elements prefer to occupy interstitial sites in p-type ZnO samples, forming shallow donors. In this case, p-type ZnO cannot be realized due to strong self-compensation. Thus, we have also calculated the formation energy of the Group-Ib dopants at interstitial sites. Figure 13.11 shows the defect formation energies as a function of the Fermi level calculated under the oxygen-rich condition for Group-Ib elements at different sites. The solid dots indicate the transition energy level positions for substitutional Cu, Ag, and Au. We find that the self-compensation is very small with Group-Ib elements in ZnO. This is because Group-Ib elements do not prefer to occupy the interstitial sites, even when the Fermi level is close to the

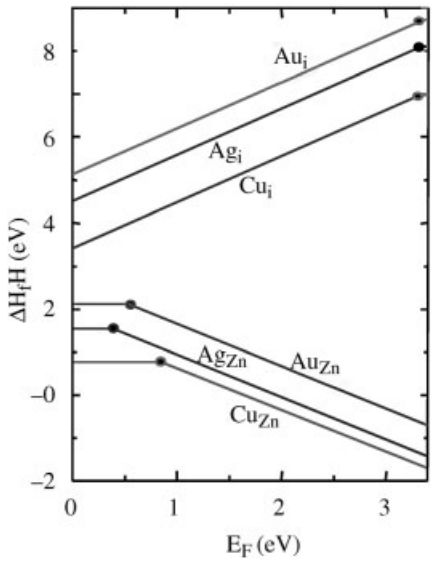

Figure 13.11 (online colour at: www.pss-b.com) Calculated formation energies as a function of Fermi level for Group-Ib elements in ZnO.

VBM. Our calculations reveal that Group-Ib elements may be better candidates than the Group-Ia elements for p-type doping of ZnO.

Our calculations revealed that the acceptor levels created by Group-Ib elements are not very shallow. However, because the formation energies of the substitutional Group-Ib elements in ZnO are very low at the O-rich conditions, a high concentration of dopants can be easily achieved. At this growth condition, the compensation by intrinsic donor defects can be effectively suppressed. Therefore, p-type doping in ZnO could still be achieved with these elements, especially for Ag doping. It is important to point out that the calculated $(0/-)$ transition energy level for Ag_{Zn} is comparable to the calculated $(0/-)$ transition energy for N_O, which is currently one of the most favorable dopants for p-type doping in ZnO. For N doping, ZnO thin films should be grown under an O-poor condition to incorporate N efficiently at O sites. This growth condition also promotes the formation of hole-killer defects, such as O vacancies and Zn and N interstitials. On the other hand, for incorporating Ag at Zn sites, the growth should be done at an O-rich condition, which suppresses the formation of major hole-killer defects. In addition, self-compensation can also be avoided for Ag doping. Therefore, our results suggest that Ag may be a better dopant than N for p-type doping in ZnO, especially when it is combined with passivating donors to form defect complexes (see Section 13.5.5 below). Our conclusion is supported by recent experiment results on p-type ZnO thin films with Ag and Cu dopants [37].

13.5.5
Reduction of Transition Energy Levels

To reduce the acceptor transition energy level in ZnO, co-doping or cluster doping has been proposed [38]. In conventional co-doping, two single acceptors (e.g., N_O) are combined with a single donor (e.g., Ga_{Zn}) to form an acceptor defect complex. It is expected that through donor–acceptor level repulsion, shallow acceptor levels can be created. However, detailed theoretical analyses show that for direct-band gap semiconductors such as ZnO, the reduction of this type of conventional co-doping on the ionization energy is rather small. This is because the donor and acceptor levels usually have different symmetries and wave-function characters: the donor state has the *s*-like a_1 character, whereas the acceptor has the p-like t_2 character. Furthermore, in the case of two acceptors plus one donor (e.g., $2N_O + Ga_{Zn}$), because the two acceptors are forced to be fcc (or hcp in the wurtzite structure) nearest neighbors, the acceptor–acceptor level repulsion can even raise the ionization energy [23].

To avoid the problem discussed above, we have explored a different and novel idea in which a fully occupied deep donor is used to attract a second partially occupied donor to lower its ionization energy [16]. In particular, we studied a double donor (either Si, Ge, or Sn on the Zn site) paired with a single donor (either F, Cl, Br, or I on the Te site) in ZnTe. Different from the Coulomb binding that exists in charged donor–acceptor complexes in the co-doping approach the binding between the two donors results from the level repulsion between the two donor states (Figure 13.12). The level repulsion significantly reduces the energy of the fully occupied lower level,

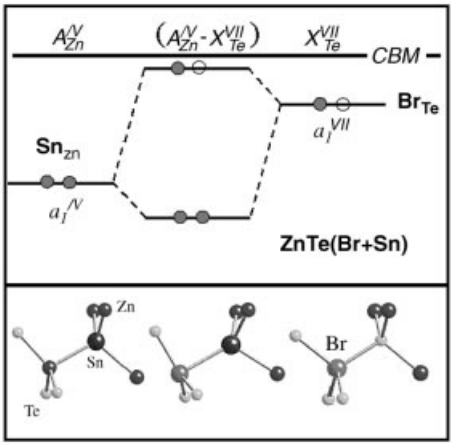

Figure 13.12 (online colour at: www.pss-b.com) Illustration of the interaction between the single donor and the double donor states associated with Br_{Te} and Sn_{Zn} in the formation of complexes Br_{Te}–Sn_{Zn} in ZnTe.

stabilizing the donor–donor pair, while it increases the energy of the partially occupied upper level, thus reducing the ionization energy. Notice that because the doubly occupied a_1^{IV}-derived state is *charge neutral*, there is no Coulomb repulsion between the two nominally donor impurities. Furthermore, because the two donor states have the same symmetry and atomic character the level repulsion is very efficient. For example, we find that the formation of a Br_{Te}–Sn_{Zn} pair in ZnTe is exothermic with a binding energy of 0.9 eV. It lowers the electron ionization energy of Br_{Te} by a factor of more than three from 240 to 70 meV, resulting in an effective shallow donor. Similar idea has also been proposed to enhance n-type doping in diamond [17]. Recently Kim and Park [18] have also suggested that the same idea can be applied to explain oxygen vacancy assisted n-type doping in ZnO by forming Zn_i–V_O paire to lower the formation energy and transition energy levels of Zn_i in ZnO.

We have also proposed two approaches to reduce the ionization energy in p-type doping of ZnO [21]. The proposals are based on the following considerations: (i) as discussed in the previous section, to lower the ionization level, one should find a dopant with low valence p orbital energy (more electronegative), preferably at the anion site. Because the wave function of the V_{Zn} has a large distribution on the neighboring O atomic sites (Figure 13.13a), replacing one of the neighboring O atoms by the more electronegative F (the F 2p level is 2.1 eV lower in energy than the O 2p level, see Figure 13.10) is expected to reduce the energy level of V_{Zn}. The binding energy between the F_O single donor and the V_{Zn} double acceptor is also expected to be large. Furthermore, this defect complex pair $V_{Zn} + F_O$ contains only one acceptor, so there will be no acceptor–acceptor repulsion to raise the ionization level; and (ii) we notice that one of the reasons that the N_O defect level is deep in ZnO is because the N 2p level strongly couples to the nearest-neighbor Zn 3d orbitals (Figure 13.13c),

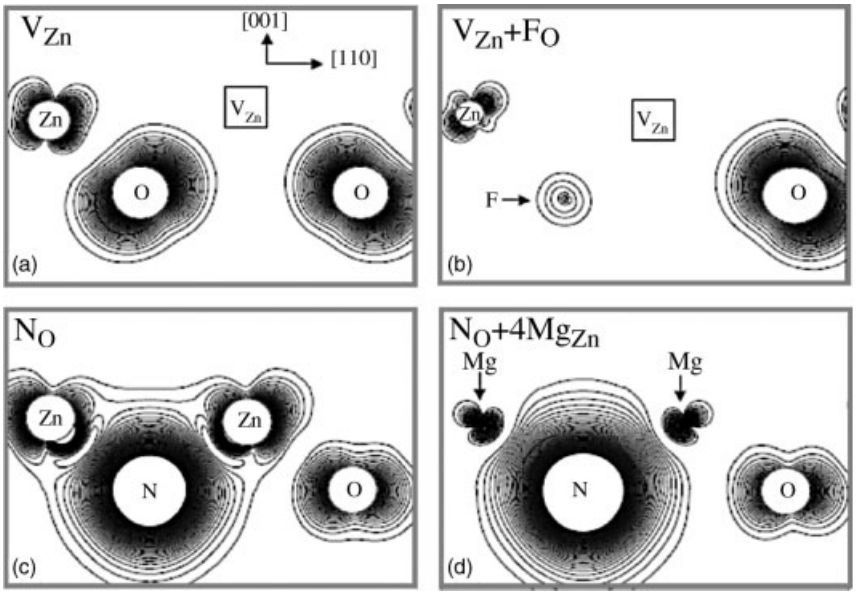

Figure 13.13 (online colour at: www.pss-b.com) Charge density plot of defect levels in ZnO: (a) V_{Zn}, (b) $V_{Zn} + F_O$, (c) N_O, and (d) $N_O + 4Mg_{Zn}$.

which both have t_2 symmetry in this tetrahedron environment. If we can replace the Zn atom by an isovalent Mg atom that has a similar atomic size as Zn but no occupied d orbital, the defect transition energy level of $N_O + nMg_{Zn}$ should be lower than that of N_O in ZnO. The effect should be most efficient for $n = 4$, when the tetrahedral environment around N_O is preserved and no level splitting occurs.

Figure 13.13b shows the charge density plot of the $V_{Zn} + F_O$ defect level. When F is introduced, it creates defect levels inside the valence band, removing one of the oxygen dangling-bond contributions to the acceptor level and making the transition energy lower. The calculated $(0/-)$ transition energy level of $V_{Zn} + F_O$ is 0.16 eV, which is much smaller than the corresponding $(-/2-)$ transition energy level of V_{Zn} at 0.34 eV. It is also lower in energy than the $(0/-)$ transition energy level of V_{Zn} at 0.18 eV. The calculated $V_{Zn} + F_O$ binding energy is -2.3 eV, indicating that the defect pair is very stable with respect to the isolated defects. This large binding energy can be understood by noticing that to form the defect complex, F_O donates one of its electrons to V_{Zn}, which results in a large Coulomb interaction between V_{Zn}^- and F_O^+. Based on this study, we believe that adding a small amount of F in ZnO to form a $V_{Zn} + F_O$ defect pair is beneficial to p-type doping in ZnO. However, we also want to point out that F_O itself is a donor, so too much F_O (more than the amount of V_{Zn}) in the sample can over compensate the acceptors.

Figure 13.13d shows the defect level charge density of $N_O + 4Mg_{Zn}$. Compared to $N_O + 4Zn_{Zn}$, we see that the cation d character is removed and the defect level is more localized on the N atomic site. The calculated $(0/-)$ transition energies are 0.29 eV for $N_O + Mg_{Zn}$ and 0.23 eV for $N_O + 4Mg_{Zn}$ eV, shallower than that for N_O.

However, the calculated binding energy for $N_O + Mg_{Zn}$ is positive at 0.3 eV, indicating that N does not like to bind with Mg in ZnO. This is because the N−Zn bond is stronger than the N−Mg bond. Our calculations show that both N−Zn and Mg−O bonds are shorter than the Zn−O bond, but the N−Mg bond length is longer than the Zn−O bond length. However, for ZnMgO alloys with relatively high Mg concentrations, the opportunity to form $N_O + nMg_{Zn}$ is reasonably high due to the entropy contribution. Furthermore, the VBM of the ZnMgO alloys is similar to that of ZnO, because the wave function is more localized on the ZnO region. This may explain why some ZnMgO alloys can be doped p-type [39]. Further lowering of the acceptor transition energy level is expected if we replace Mg by Be, because the Be 2p orbital energy is much lower than the 3p orbital of Mg (Figure 13.10). Indeed, we find that the (0/−) transition energy levels of $N_O + Be_{Zn}$ and $N_O + 4Be_{Zn}$ are at 0.22 and 0.12 eV, respectively.

Other successful co-doping schemes include the one demonstrated by Limpijumnong *et al.* [36] who show that As_{Zn}–$2V_{Zn}$ in ZnO creates relatively shallow acceptor levels. In this complex, the shallow acceptor level is realized because the two V_{Zn} acceptors are connected by the As_{Zn} (or P_{Zn}) antisite donor through a cation sublattice; so the separation between the two V_{Zn} is large and the level repulsion between them is weak.

13.5.6
Universal Approaches Through Impurity-Band Doping

We recently proposed a universal approach to overcome the long-standing doping polarity problem for WBG semiconductors [22]. The approach is to reduce the ionization energies of dopants and the spontaneous compensation from intrinsic defects by creating a passivated impurity band, which can be achieved by introducing passivated donor–acceptor complexes or isovalent impurities. In this case, the ionization energy is reduced by shifting the band edge through the impurity band, which is higher than the VBM or lower than the CBM, rather than through the shifting of defect energy levels. When the same element is used to create the impurity band and as dopant, the ionization energy is always small. Furthermore, due to a smaller Fermi level shift, charge compensation is also reduced. Our density-functional theory calculations demonstrate that this approach provides excellent explanations for the available experimental data of n-type doping of diamond and p-type doping of ZnO, which could not be understood by previous theories. In principle, this universal approach can be applied to any WBG semiconductors, and therefore, it will open a broad vista for the use of these materials. Our concept agrees well with the observation by Kalish *et al.* [40], who suggested that impurity bands could play a role in co-doped diamond.

We first demonstrate our approach for n-type doping in diamond. It is known that n-type doping of diamond is extremely difficult because the donor levels are usually 0.6 eV or deeper below the CBM for most dopants such as N and P [41, 42]. Some n-type diamonds have been reported by using N and P as dopants and the mechanism has been studied theoretically. However, the most exciting n-type doping of diamond

in the last few years is the co-doping of B with deuterium. It is reported that through this co-doping, n-type diamond has been realized with an activation energy of about 0.2–0.3 eV [3].

We now explain how and why our new concept can explain the experimental results of n-type doping by deuteration of B-doped diamonds. It is reported that the deuteration of B-doped diamond undergoes two clear steps: (i) the passivation of B acceptors by deuterium and (ii) the excess deuterium doping that leads to the formation of shallow donors. The experiments suggest strongly that (B, D) complexes are responsible for the shallow donors; here, D indicates deuterium. In our calculation, we use H for deuterium. Our calculation shows that the ionization energy level for an isolated H in diamond is about 2.8 eV below the CBM, which is consistent with the calculated results reported by others [43]. Isolated B + 2H complexes in diamond have also been found theoretically to be deep donors [44]. Our calculations reveal that the passivated (B + H) complexes generate fully unoccupied impurity bands, which lie about 1.0 eV below the host CBM. An isolated H atom in diamond has two low-energy sites: bond center (C−H−C) or anti-bond (C−C−H) sites. When B atoms are available in diamond, H atoms preferentially bond to B atoms, because in their mutual presence, B atoms are negatively charged and H atoms are positively charged. The energy of the bond-center configuration is lower than the anti-bond configuration because an H^+ ion prefers to sit at a high electron-density site. Figure 13.14 shows the calculated total density of states (DOS) for pure diamond host (green curve) and a supercell containing a (B + H) complex (red curve), with the B–H–C configuration. It reveals clearly that the formation of a passivated (B + H) complex does not change the basic electronic structure, but only generates an unoccupied impurity band below the CBM. Our results, therefore,

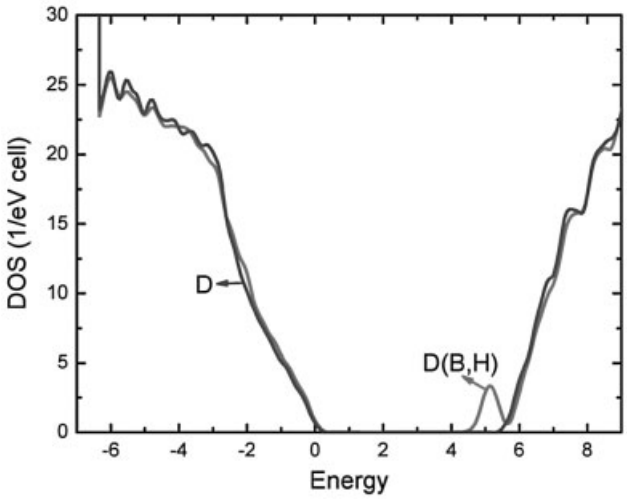

Figure 13.14 (online colour at: www.pss-b.com) Calculated DOS for pure diamond host (green curve) and a supercell containing a (B + H) complex (red curve), with the B–H–C configuration.

suggest that the first step of the deuteration of B-doped diamonds is to passivate the B acceptors, and create the fully unoccupied impurity bands below the CBM.

When excess deuterium/H atoms are available after the first step, they will start to dope the passivated system, i.e., they effectively dope the new host with the unoccupied impurity band, rather than the original conduction band. Thus, in calculating the ionization energy, the term $\varepsilon^{\Gamma}_{CBM}$(host) in Eq. (13.4) should now be replaced by the impurity-band minimum (IBM), $\varepsilon^{\Gamma}_{IBM}$. In other words, the transition now occurs between the H defect levels and the unoccupied impurity bands, rather than the original conduction bands. As a result, the transition energy can be reduced dramatically.

For H doping in the (B + H)-passivated diamonds, the excess H atoms bind to the (B + H) complexes, forming (H–B–H) triplets. For charge-neutral H atoms, the lowest energy configuration is shown in Figure 13.15a, where the excess H is at the B antibonding site. We call this configuration (H–B–H)–AB. When the excess H atom is positively charged ($q = +1$), the fully relaxed structure is shown in Figure 13.15b. We see in Figure 13.15b that the H^+ ion at the antibonding site becomes energetically unstable, and it moves to a bond-center site with high electron density to lower the Coulomb energy. This atomic displacement results in significant bond rearrangements and a large energy lowering of the charged defect (−1.8 eV), which leads to significant reduction of the ionization energy [see Eq. (13.4)]. The calculated $\varepsilon(0/+)$ transition energy level is 0.3 eV below the unoccupied impurity-band edge. We also studied a metastable (H–B–H)–BC triplet defect, where both H atoms are at the puckered B–C bond-center sites. The atomic configurations for neutral and charged defect complexes are shown in Figure 13.15c and d, respectively. This configuration is about 0.6 eV higher in energy than the (H–B–H)–AB complex due to strong H^+–H^+ Coulomb repulsion; but the calculated transition energy level is 0.2 eV, which is 0.1 eV lower than that for the (H–B–H)–AB complex due to less crystal-field splitting.

The calculated transition energies agree very well with the experimentally measured ionization energies, suggesting that the second step of deuteration of B-doped diamond is to effectively dope the (B + H) impurity bands. This new concept, therefore, explains why (B, H) co-doping can create shallow donors in diamonds. It should be noted that to form the impurity bands and have reasonable transport properties, a critical concentration threshold is needed. Furthermore, the edge of the impurity band depends on the concentration of B atoms. The higher B concentration results in a more-broadened (B + H) impurity band. Consequently, the ionization energy will be reduced. This explains another experimental observation, i.e., diamonds with a higher B concentration exhibit shallower donor levels.

Our approach can also be applied to explain p-type doping of ZnO. As discussed above, p-type doping of ZnO is difficult. However, Ga and N co-doping has produced good p-type ZnO [8, 9]. The doping mechanism is not well understood. Most reliable theoretical calculations predicted that the ionization energy for N acceptors in ZnO is about 0.4 ± 0.1 eV above the VBM [21, 33, 42]. But the experimentally measured N acceptor ionization energy in p-type ZnO is much shallower, only 0.1–0.2 eV above the VBM [6, 7]. The conventional co-doping concept cannot explain the discrepancy

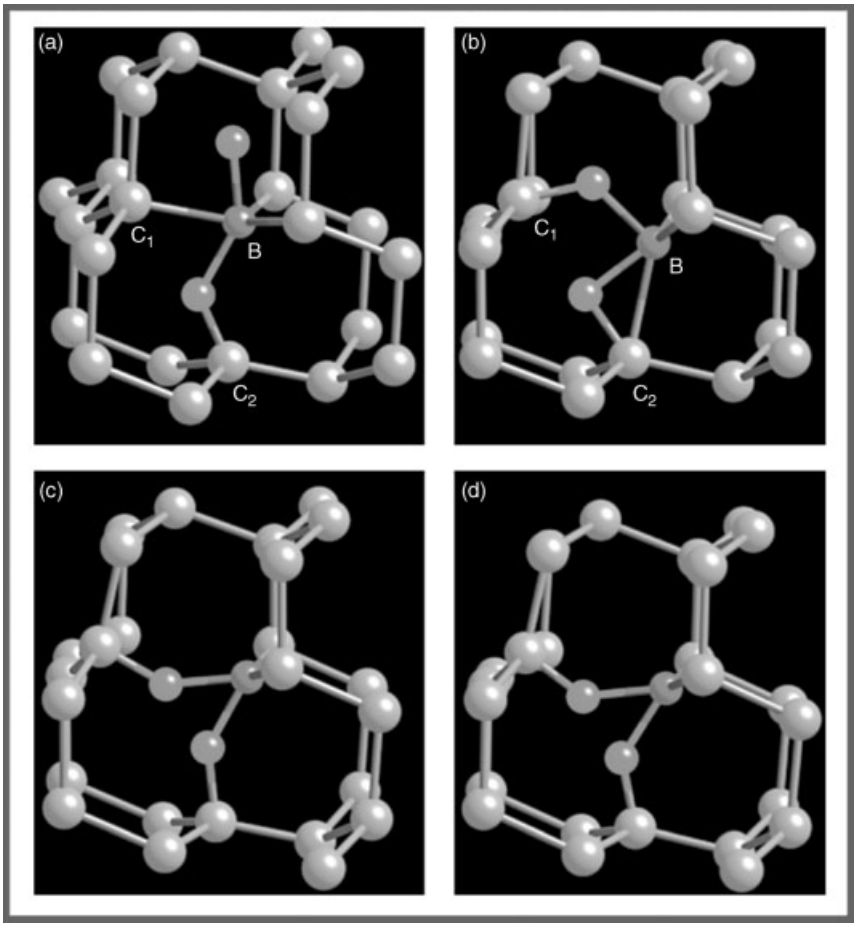

Figure 13.15 (online colour at: www.pss-b.com) Relaxed structures for B + 2H complexes in diamond with charge-neutral and +1 charged states. The blue balls are C atoms. The red balls are B atoms. The green balls are H atoms. (a) Neutral state for complex (H–B–H)–AB, (b) +1 charged state for complex (H–B–H)–AB, (c) neutral state for complex (H–B–H)–BC, and (d) +1 charged state for complex (H–B–H)–BC.

because the calculated ionization level of an isolated Ga + 2N impurity is still deep, at about 0.4 eV.

Here, we show that to successfully use Ga and N co-doping to obtain p-type ZnO, the first step is to form passivated stoichiometric (Ga + N) complexes, and create a fully occupied impurity band above the VBM of ZnO. Ga and N bind together strongly in ZnO because they passivate each other. Figure 13.16 shows the calculated total DOS for pure ZnO host (blue curve) and a system containing a (Ga + N) complex (red curve). It reveals clearly that the formation of a passivated (Ga + N) complex does not change the basic electronic structure, but only generates an additional fully occupied band above the VBM. When excess N atoms are available,

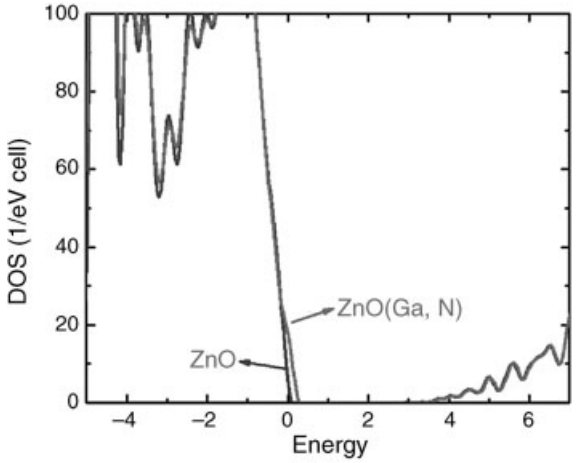

Figure 13.16 (online colour at: www.pss-b.com) Calculated DOS for pure ZnO (green curve) and a supercell containing a (Ga, N) complex.

they will dope the passivated system. The transition will occur between the N defect levels and the fully occupied impurity bands, rather than the original valence bands. Thus, the term $\varepsilon^{\Gamma}_{\mathrm{VBM}}$(host) in Eq. (13.3) should now be replaced by the impurity-band maximum, $\varepsilon^{\Gamma}_{\mathrm{IBM}}$.

Previous calculations suggested that for the Ga + 2N complexes, the first N occupies the first nearest-neighboring O site of the Ga, which occupies a Zn site [34]. The second N occupies the second nearest-neighboring O site. This N atom does not bind directly to the Ga atom. We call this configuration (N–Ga–N)–A. However, our calculations reveal that the excess N atoms bind to the (Ga + N) sites, forming

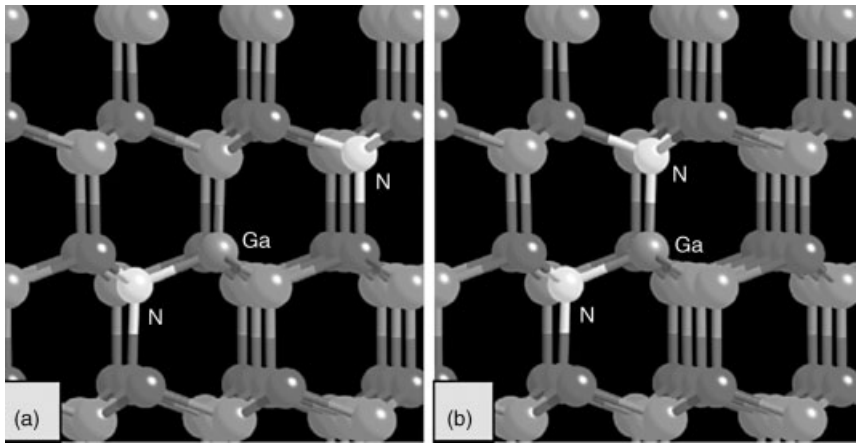

Figure 13.17 (online colour at: www.pss-b.com) Relaxed structures for (a) configuration A and (b) configuration B.

a (N–Ga–N)–B complex with both N atoms occupying the first nearest-neighboring O sites of the Ga atom. The relaxed structures for A and B configurations are shown in Figure 13.17a and b, respectively. The B configuration is about 0.5 eV lower in energy than the A configuration. We have calculated the acceptor ionization energies for both configurations, considering effective doping of the passivated (Ga + N) impurity bands. The calculated ionization energies are 0.2 and 0.1 eV for configurations A and B, respectively. Our results, therefore, are able to explain the puzzling experimentally measured ionization energies for N acceptors. Again, we want to point out that to form (Ga + N) impurity bands and have reasonable transport properties, critical Ga and N concentrations are needed. The transition energy is also expected to be reduced in the ZnO with higher Ga concentration.

With our approach, we are able to explain experimentally observed B and H co-doped n-type diamonds and Ga and N co-doped p-type ZnO, which could not be understood by previous theories. The physical principle behind this new concept is clear; that is, we can first create a fully passivated impurity band and then dope the impurity band. This approach can be applied, in principle, to any WBG semiconductors to overcome the doping polarity problems found in these materials. It should be pointed out that to be successful, the concentration of the defects inducing the impurity band must exceed a certain percolation limit, so that reasonable transport properties can be achieved. The small band gap reduction caused by forming an impurity band can also be easily adjusted by alloying with other elements. For example, adding a small amount of Mg or Be in ZnO can easily open the band gap without changing the doping property [45, 46].

13.6 Summary

We have reviewed three main origins for the doping limit in WBG semiconductors, i.e., (i) low dopant solubility; (ii) deep ionization energy levels; and (iii) spontaneous formation of compensating defects. We have also proposed solutions to overcome the doping bottlenecks, which include (i) increase defect solubility by defeating bulk defect thermodynamics using non-equilibrium growth methods such as extending the achievable chemical potential through molecular doping or rasing the host energy using surfactant; (ii) reduce defect ionization energy level by designing shallow dopants or dopant complexes; and (iii) reduce defect compensation and ionization level by modifying the host band structure near the band edges. The issues related to the defect calculations are discussed. We believed that the band gap correction is an important issue in computational defect physics and more studies are needed.

Acknowledgement

We would like to thank A. Janotti, M.M. Al-Jassim, J. Li, S.-S. Li, S. Limpijumnong, C.-H. Park, D. Segev, J.-B. Xia, L. Zhang, and S.B. Zhang for their contributions and

helpful discussions in this work. This work is supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308.

References

1 Koizumi, S., Watanabe, K., Hasegawa, M., and Kanda, H. (2001) *Science*, **292**, 1899.
2 Isberg, J., Hammersberg, J., Johansson, E., Wikstrom T., Twitchen D.J., Whitehead, A.J., Coe S.E., and Scarsbrook, G.A. (2002) *Science*, **297**, 1670.
3 Teukam, Z., Chevallier, J., Saguy C., Kalish, R., Ballutaud, D., Barbé, M., Jomard, F., Tromson-Carli, A., Cytermann, C., Butler, J.E., Bernard, M., Baron, C., and Deneuville, A. (2003) *Nature Mater.*, **2**, 482.
4 Taniyasu, Y., Kasu, M., and Makimoto, T. (2006) *Nature*, **441**, 325.
5 Huang, M.H., Mao, S., Feick, H., Yan, H., Wu, Y., Kind, H., Weber, E., Russo, R., and Yang, P. (2001) *Science*, **292**, 1897.
6 Tsukazaki, A., Ohtomo, A., Onuma, T., Ohtani, M., Makino, T., Sumiya, M., Ohtani, K., Chichibu, S.F., Fuke, S., Segawa, Y., Ohno, H., Koinuma, H., and Kawasaki, M. (2005) *Nature Mater.*, **4**, 42.
7 Tsukazaki, B., Kubota, M., Ohtomo, A., Onuma, T., Ohtani, K., Ohno, H., Chichibu, S.F., and Kawasaki, M. (2005) *Jpn. J. Appl. Phys.*, **44**, L643.
8 Joseph, M., Tabata, H., and Kawai, T. (1999) *Jpn. J. Appl. Phys.*, **38**, L1205.
9 Look, D.C., Claflin, B., Alivov, Y.I., and Park, S.J. (2004) *Phys. Status Solidi A*, **201**, 2203.
10 Zhang, S.B., Wei, S-H., and Zunger, A. (1998) *J. Appl. Phys.*, **83**, 3192.
11 Neumark, G.F. (1997) *Mater. Sci. Eng.*, **R21**, 1.
12 Li, J., Wei, S.-H., and Wang, L.-W. (2005) *Phys. Rev. Lett.*, **94**, 185501.
13 Li, J., Wei, S.-H., Li, S.-S., and Xia, J.-B. (2008) *Phys. Rev. B*, **77**, 113304.
14 Yan, Y. and Wei, S.-H. (2008) *Phys. Status Solidi B*, **245**, 641.
15 Yan, Y., Zhang, S.B., and Pantelides, S.T. (2001) *Phys. Rev. Lett.*, **86**, 5723.
16 Janotti, A., Wei, S.-H., and Zhang, S.B. (2003) *Appl. Phys. Lett.*, **83**, 3522.
17 Segev, D. and Wei, S.-H. (2003) *Phys. Rev. Lett.*, **91**, 126406.
18 Kim, Y.S. and Park, C.H. (2009) *Phys. Rev. Lett.*, **102**, 086403.
19 Wei, S.-H. (2004) *Comput. Mater. Sci.*, **30**, 337.
20 Yan, Y., Al-Jassim, M., and Wei, S.-H. (2006) *Appl. Phys. Lett.*, **89**, 181912.
21 Li, J., Wei, S.-H., Li, S.-S., and Xia, J.-B. (2006) *Phys. Rev. B*, **74**, 081201.
22 Yan, Y., Li, J., Wei, S.-H., and Al-Jassim, M. (2007) *Phys. Rev. Lett.*, **98**, 135506.
23 Wei, S.-H. and Zhang, S.B. (2002) *Phys. Rev. B*, **66**, 155211.
24 Wei, S.-H. and Krakauer, H. (1985) *Phys. Rev. Lett.*, **55**, 1200.
25 Kresse, G. and Hafner, J. (1993) *Phys. Rev. B*, **47**, 558.
26 Makov, G. and Payne, M.C. (1995) *Phys. Rev. B*, **51**, 4014.
27 Zhang, S.B., Wei, S.-H., and Zunger, A. (2001) *Phys. Rev. B*, **63**, 75205.
28 Rinke, P., Janotti, A., Scheffler, M., and Van de Walle, C.G. (2009) *Phys. Rev. Lett.*, **102**, 026402.
29 Oba, F., Togo, A., Tanaka, I., Paier, J., and Kresse, G. (2008) *Phys. Rev. B*, **77**, 245202.
30 Janotti, A., Varley, J.B., Rinke, P., Umezawa, N., Kresse, G., and Van de Walle, C.G. (2010) *Phys. Rev. B*, **81**, 085212.
31 Lee, W., Kang, J., and Chang, K.J. (2006) *Phys. Rev. B*, **73**, 024117.
32 Zhang, S.B. and Wei, S.-H. (2001) *Phys. Rev. Lett.*, **86**, 1789.
33 Park, C.H., Zhang, S.B., and Wei, S.-H. (2002) *Phys. Rev. B*, **66**, 073202.
34 Neugebauer, J. and van de Walle, C.G. (1995) *Phys. Rev. Lett.*, **75**, 4452.
35 Zhang, L., Yan, Y., and Wei, S.-H. (2009) *Phys. Rev. B*, **80**, 073305.
36 Limpijumnong, S., Zhang, S.B., Wei, S.-H., and Park, C.H. (2004) *Phys. Rev. Lett.*, **92**, 155504.
37 Kang, H.S., Ahn, B.D., Kim, J.H., Kim, G.H., Lim, S.H., Chang, H.W., and Lee, S.Y. (2006) *Appl. Phys. Lett.*, **88**, 202108.

38 Katayama-Yoshida, H. and Yamamoto, T. (1997) *Phys. Status Solidi B*, **202**, 763.

39 Heo, Y.W., Kwon, Y.W., Li, Y., Pearton, S.J., and Norton, D.P. (2004) *Appl. Phys. Lett.*, **84**, 3474.

40 Kalish, R., Saguy, C., Cytermann, C., Chevallier, J., Teukam, Z., Jomard, F., Kociniewski, T., Ballutaud, D., Butler, J.E., Baron, C., and Deneuville, A. (2004) *J. Appl. Phys.*, **96**, 7060.

41 Kajihara, S.A., Antonelli, A., Bernhole, J., and Car, R. (1991) *Phys. Rev. Lett.*, **66**, 2010.

42 Wang, L.G. and Zunger, A. (2002) *Phy. Rev. B*, **66**, 161202(R).

43 Goss, J.P., Jones, R., Heggie, M.I., Ewels, C.P., Briddon, P.R., and Oberg, S. (2002) *Phys. Rev. B*, **65**, 115207.

44 Goss, J.P., Briddon, P.R., Sque, S., and Jones, R. (2004) *Phys. Rev. B*, **69**, 165215.

45 Ohtomo, A., Kawasaki, M., Koida, T., Masubuchi, K., Koinuma, H., Sakurai, Y., Yoshida, Y., Yasuda, T., and Segawa, Y. (1998) *Appl. Phys. Lett.*, **72**, 2466.

46 Ryu, Y.R., Lee, T.S., Lubguban, J.A., Corman, A.B., White, H.W., Leem, J.H., Han, M.S., Youn, C.J., and Kim, W.J. (2006) *Appl. Phys. Lett.*, **88**, 052103.