ELSEVIER

Physica B 273–274 (1999) 976–980

PHYSICA B

www.elsevier.com/locate/physb

Overcoming doping bottlenecks in semiconductors and wide-gap materials

S.B. Zhang*, S.-H. Wei, A. Zunger

National Renewable Energy Laboratory, 1617 Cole Building, Golden, CO 80401, USA

Abstract

There often exist strong doping bottlenecks that may severely restrict potential applications of semiconductors, especially in wide-band-gap materials where bipolar doping is impossible. Recent rapid progress in semiconductor research has reached a point where these doping limitations must be overcome in order to tune semiconductors for precisely required properties. Here, we discuss how to find out what causes the doping bottlenecks. We based our discussion on a set of recent, novel developments regarding the doping limitations: the "doping limit rule" distilled from both phenomenological studies and from first-principles calculations. The thermodynamic doping bottlenecks are identified as due mainly to the formation of intrinsic defects whose formation enthalpies depend on the Fermi energy, and always act to negate the effect of doping. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Doping; Defect compensation; Wide-gap materials; Semiconductors

1. Introduction

Semiconductor-based high-tech owes its existence, in large part, to the fact that semiconductors can be doped. Materials that cannot be doped are useless for most electronic and optoelectronic applications. Indeed, failure-to-dope a class of materials is often the single most important bottleneck for advancing a semiconductor technology based on these materials. Overcoming this bottleneck can enable a whole new technology. Examples of past and present doping roadblocks include (i) p-type doping of wide-gap II–VI compounds for blue lasers, (ii) p-type doping of nitrides, (iii) the elusive n-type doping of diamond, (iv) p-type doping of (transparent conducting) oxides for displays, and (v) doping of alkali halides, wide-gap carbides and fluorides. Case (i) has been recently solved for ZnSe alone [1,2] and case (ii) for GaN alone [3,4]. All other cases remain unsolved.

The vast experimental literature on doping shows that there are three main modes of failure-to-dope:

(i) *Insoluble dopants*: The desired impurity atom cannot be introduced into the appropriate host crystal site because of limited solubility. This includes cases of formation of competing compound phases (e.g., Zn_3N_2 in ZnSe : N [5]), impurity segregation or precipitation (e.g., large impurity atoms such as Ba, Hg in III–V compounds), and substitution on the "wrong site" (e.g., the amphoteric center where the dopant may occupy different sites, thus being either a donor or an acceptor [6]).

(ii) *Deep-level dopants*: The desired dopant *is* soluble in the host, but it produces a deep, rather than shallow level, so the impurity remains un-ionized at normal temperatures. Examples of such "localized centers" include *Cd*S : Cu, *Zn*Se : Cu [7], or Diamond : P [8], in which the impurity substitutes the underlining atoms.

(iii) *Compensated dopants*: The impurity atom is both soluble and ionizable, but as it produces free carriers, a spontaneous-generated, opposite-charged native defect forms, which compensates and negates the effect of the intentional dopant. Examples include the DX center in

---

* Corresponding author. Tel.: 303-384-6622; fax: 303-384-6531.

E-mail address: szhang@nrel.gov (S.B. Zhang)

0921-4526/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0921-4526(99)00605-5

Si-doped GaAlAs [9,10], and the AX center in N-doped ZnSe [11].

Categories (i) and (ii) of failure-to-dope can sometimes be circumvented by changing the dopant or by co-doping. For example, (i) If Ba is insoluble in III–V compounds, one can attain p-doping by using instead a smaller divalent cation such as Zn. (ii) If Cu creates a deep level in ZnSe, one can use N that forms a shallower level [11]. If P in diamond is too deep, 2P + H can create a shallower center [12]. However, category (iii) of failure-to-dope is terminal, since it is not the chemical impurity that causes the failure-to-dope, but the free carriers. Thus, category (iii) represents the true "doping limit" of a material. For example, no impurity or chemical treatment has so far resulted in any p-doping of main group oxides (e.g., ZnO, CaO, MgO), or n-doping of alkali halides or stable n-doping of diamond. Also, GaAs cannot be doped p-type in excess of 10^{19} cm^{-3} [13] (while it can be doped n-type in excess of 10^{21} cm^{-3} [14]) independent of the impurity and chemical treatment. Thus, there must be a yet unknown intrinsic and fundamental doping-limiting process. We will thus focus our attention on this intrinsic limit.

In olden days, it was believed that such "doping limits" are caused by the very existence of a large band gap [15]. As evidence, it was customary to cite the fact that large gap materials cannot be doped, e.g., GaN, diamond, and oxides. Today, it is clear that this is not the case. We know that some large gap materials can be doped, e.g., n-ZnO [16], n-CdS [17], p-diamond [18] and n-GaN [19]. Surprisingly, however, doping can be strongly asymmetric with respect to holes and electrons. These "doping anomalies" include [20] the facts that (i) ZnO, ZnS, and CdS can be doped only n-type, while ZnTe can be doped only p-type. (ii) $CuAlSe_2$ cannot be doped either p- or n-type, while $CuGaSe_2$ can be doped p-type only, and $CuInSe_2$ can be doped both p- and n-type, and (iii) Si, Ge can be doped both p- and n-type while diamond can be doped only p-type. The existence of such a pronounced asymmetry between n-type and p-type dopability cannot be explained simply by the existence of a large gap. We need another explanation.

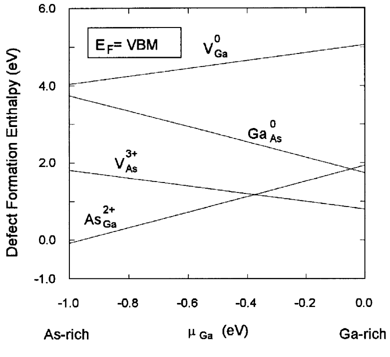

Fig. 1. Calculated point defect formation enthalpy in GaAs as a function of μ_{Ga}.

In the past, each case of failure-to-dope in semiconductors and insulators was treated surprisingly as an isolated issue. Thus, the literature on the failure to p-dope ZnSe is divorced from the literature on the failure to n-dope diamond or from the literature on doping difficulties in nitrides and carbides. It appears that there is a need to study the "science of failure-to-dope" as a generic discipline to frog-leap in this important field because as we will show, there is a common phenomenon that underlies to all of these failures.

2. Thermodynamics of doping

A key realization regarding doping is that the formation enthalpy of a charged defect A in a solid often depends on the atomic chemical potential, μ_A, and always depends on the electron Fermi level, ε_F, as described below. This holds the key to understand, and overcome, doping compensation.

2.1. Dependence of formation enthalpy on chemical potentials

The formation enthalpy of a charge-neutral intrinsic defect A^0 often depends on the atomic chemical potentials of the host atoms. For example, to form a cation vacancy in a binary compound, one cation atom is removed from the host and is placed in the atomic "reservoir" of energy, μ_A. The formation enthalpy is thus

$$\Delta H(A^0) = E_{tot}(A^0) - E_{tot}(0) + \mu_A, \qquad (1)$$

where $E_{tot}(A^0)$ is the total energy of the host crystal having a neutral defect A, $E_{tot}(0)$ is the total energy of the host without any defect. Fig. 1 shows [21] a few calculated formation enthalpies of native defects in p-type GaAs as a function of the Ga chemical potential. We see that the Ga-on-As antisite (Ga_{As}) and the As vacancy (V_{As}) are easier to form in Ga-rich conditions, while in As-rich conditions, As-on-Ga antisite (As_{Ga}) and Ga vacancy (V_{Ga}) are instead easier to form. To suppress compensation by intrinsic defect formation, it is always advantageous to prepare materials at the chemical potentials that maximize the formation enthalpy of the specific defect.

2.2. Dependence of defect formation enthalpy on the Fermi level

The formation enthalpy of a neutral defect A^0 does not depend on the Fermi energy, ε_F. However, the formation enthalpy of a charged defect does. For example, the

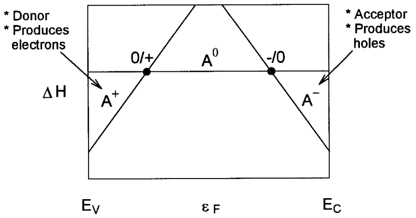

Fig. 2. Schematic change of the defect formation enthalpies (ΔH), as a function of the Fermi energy (ε_F).

formation enthalpy of a positively charged defect A^+ is equal to the energy of a neutral defect A^0, minus the energy $\varepsilon(0/+)$ needed to ionize A^0 to form A^+, plus the energy of the ionized electron. Since this electron resides in the Fermi reservoir, its energy is ε_F:

$$\Delta H(A^+) = \Delta H(A^0) - \varepsilon(0/+) + \varepsilon_F. \tag{2}$$

For a double donor we will have $+2\varepsilon_F$, etc. Thus, as shown in Fig. 2, the higher the Fermi energy, the larger the energy needed to form A^+. So donors (that produce electrons in the reaction $A^0 \rightarrow A^+ + e^-$) are more difficult to form in electron-rich (n-type) materials. Similarly, for acceptors, the formation energy decreases as ε_F increases:

$$\Delta H(A^-) = \Delta H(A^0) + \varepsilon(-/0) - \varepsilon_F. \tag{3}$$

So acceptors (that produce holes in the reaction $A^- + h^+ \rightarrow A^0$) are more difficult to form in hole-rich (p-type) materials.

These simple considerations show that

(a) If we dope a material intentionally n-type via some donor impurity, as ε_F moves up in the gap, the formation energy of native-acceptors $\Delta H(A^-)$ decreases. At some points, the formation energy is so low that such native acceptors (e.g., cation vacancy or DX centers) could form spontaneously, thus negating the effect of the intentionally introduced donors.

(b) If we dope a material intentionally p-type via some acceptor impurity, as ε_F moves down in the gap, the formation energy of native donors $\Delta H(A^+)$ decreases. At some point the formation energy is so low that such native donors (e.g., the AX center) could form spontaneously, thus negating the effects of the intentionally introduced acceptors.

Thus, the spontaneous formation of native defects determines the maximum and minimum pinning energy positions over which the Fermi energy can vary.

3. The phenomenological "doping limit rule"

Based on an earlier "vacuum pinning rule" [22] (that established the universality of the energetic positions of deep, transition-metal-impurity levels in semiconductors), Walukiewicz [23–25] studied the trends in doping limits in many semiconductors. This study and the follow-on studies by Tokumitsu [26], Ferreira et al. [27], and more recently by Zhang et al. [20] established a remarkable "phenomenological doping limit rule". It showed that there are common and surprisingly simple principles that cut across failure to dope in different material classes. Failure-to-dope is not related to the size of the band gap per-se, but rather to the position of the valence band maximum (VBM) with respect to the p-like pinning energy $\varepsilon_{\text{pin}}^{(p)}$, and the position of the conduction band minimum (CBM) with respect to the n-like pinning energy $\varepsilon_{\text{pin}}^{(n)}$.

From the discussion in Section 2, doping stops when there are too many spontaneously generated defects that compensate the intentional dopants. The net concentration $N^{(n/p)}(T, \varepsilon_F)$ of free carriers (electrons or holes) in a semiconductor is determined, in the single, parabolic band approximation, by the position of the pinning Fermi energy,

$$N^{(n/p)}[T, \varepsilon_F^{(n/p)}] = \frac{1}{2\pi^2}[2m^{*,(n/p)}]^{3/2} \times \int_0^4 \frac{\varepsilon^{1/2} d\varepsilon}{\exp[\beta(\varepsilon - \varepsilon_F^{(n/p)})] + 1}, \tag{4}$$

where $\beta = 1/kT$ is the temperature factor, and m^* is the appropriate effective mass. If we know the measured maximum electron or hole concentration, $N_{\max}^{(n/p)}$, we may obtain $\varepsilon_{\text{pin}}^{(n)}$ and $\varepsilon_{\text{pin}}^{(p)}$ simply by inverting Eq. (4).

Fig. 3 shows the values of $\varepsilon_{\text{pin}}^{(n)}$ and $\varepsilon_{\text{pin}}^{(p)}$ obtained from measured maximum doping in all the III–V compounds. In this figure, the VBMs were aligned. While the data for $\varepsilon_{\text{pin}}^{(p)}$ are scattered within a relatively small range of 0.5 eV, the data for $\varepsilon_{\text{pin}}^{(n)}$ are scattered over a wide range of 1.2 eV, showing no emerging trend.

A similar problem of absence of apparent trends in energy levels existed in another field, namely that of

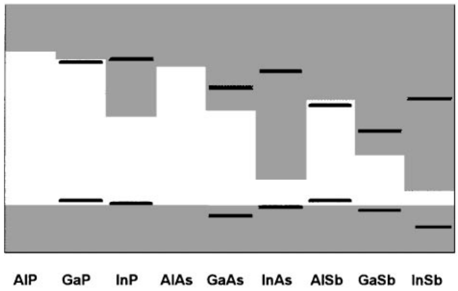

Fig. 3. Calculated pinning energies, by inverting Eq. (4), for conventional III–V compounds. The valence band maxima are lined up in the plot, without any absolute energy scale. Not large scatter near the conduction band.

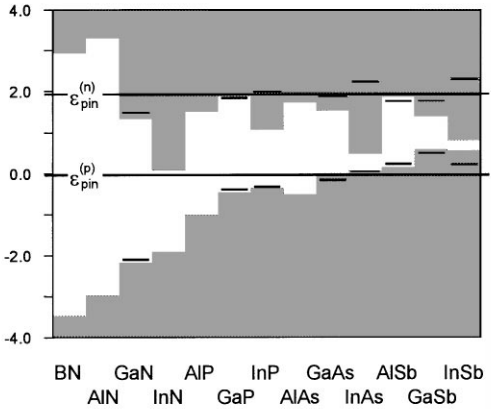

Fig. 4. Phenomenological pinning energies in III–V compounds (solid bars) obtained by inverting Eq. (4). The bands were aligned with calculated valence band offsets on an absolute energy scale. The heavy lines denote the average pinning energies. The energy zero is the VBM of GaAs.

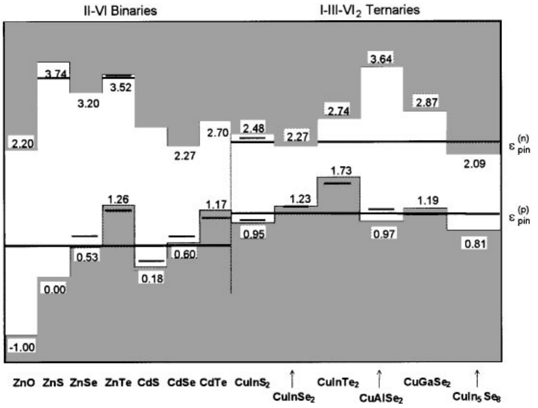

Fig. 5. Phenomenological pinning energies in II–VI and I–III–VI$_2$ compounds. The legend is the same as in Fig. 4. The energy zero is the VBM of ZnS.

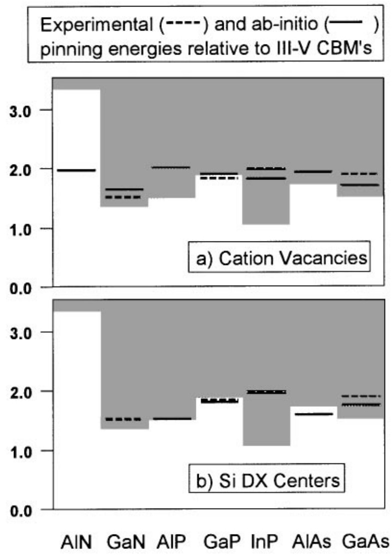

Fig. 6. LDA (solid lines) vs. the experimental (dashed lines) n-type pinning energies in seven III–V compounds. The energy zero is the VBM of GaAs.

transition metal impurities in semiconductors. However, Caldas et al. [22] showed that if one refers the impurity levels to the vacuum level, the states of a given impurity in different hosts all line up.

Thus, following Walukiewicz [23–25] we will do the same for our calculated p- and n-type Fermi energy pinning levels. However, now we will use modern, calculated values of unstrained band offsets [28] that are believed to have correct chemical trends. This is shown in Fig. 4 for III–V compounds. The scatter in $\varepsilon_{\rm pin}^{(n)}$ is approximately 0.5 eV while the scatter in $\varepsilon_{\rm pin}^{(p)}$ is somewhat larger, most noticeably for p-type GaN. Results for II–VI and I–III–VI$_2$ ternary compounds are shown in Fig. 5. This newly discovered doping limit rule tells us that when referred to an absolute energy scale, all $\varepsilon_{\rm pin}^{(n)}$ and separately all $\varepsilon_{\rm pin}^{(p)}$ tend to align for each class of materials. The emerging "doping limit rule" is:

(a) A material for which $\varepsilon_{\rm pin}^{(n)} \ll \varepsilon_{\rm CBM}$ cannot be doped n-type.

(b) A material for which $\varepsilon_{\rm pin}^{(p)} \gg \varepsilon_{\rm VBM}$ cannot be doped p-type.

This remarkable rule permits one to guess rather accurately if a material can be doped or not merely by positioning its band-edge energies on a diagram such as Figs. 4 and 5. This phenomenological rule explains all the "doping anomalies" noted in Section 1. However, the microscopic origin of this generally successful rule, i.e., the identity of the compensating defects, remains unclear.

4. First-principles study of the n-doping limit in III–V compounds

Our recent first-principles pseudopotential total energy calculations reveal that for n-doping of III–V compounds, the spontaneously-formed acceptor "killer defects" are indeed the microscopic origin of failure-to-dope. These defects are the cation vacancies and the DX centers depending on host materials. For the cation vacancies, the pinning energy $\varepsilon_{\rm pin}^{(n)}$ is defined as the Fermi energy at which the vacancy formation enthalpy is zero,

$$\Delta H(\varepsilon_{\rm F} = \varepsilon_{\rm pin}^{(n)}) = 0. \quad (5)$$

For the DX centers, $\varepsilon_{\mathrm{pin}}^{(\mathrm{n})}$ is defined as the defect transition energy $\varepsilon(+/-)$ at which the donor to acceptor transition takes place,

$$\varepsilon_{\mathrm{F}} = \varepsilon_{\mathrm{pin}}^{(\mathrm{n})} = \varepsilon(+/-). \tag{6}$$

Fig. 6 compares the ab initio $\varepsilon_{\mathrm{pin}}^{(\mathrm{n})}$ values in seven III–V materials whose band diagram is aligned with ab initio band offsets [28]. The dashed lines give the experimental pinning energies. Fig. 6 shows a clear tendency of line-up of $\varepsilon_{\mathrm{pin}}^{(\mathrm{n})}$ with respect to the vacuum level. In the case of the cation vacancy, the scatter among 7 calculated $\varepsilon_{\mathrm{pin}}^{(\mathrm{n})}$ is less than 0.4 eV. This scatter would, however, be an order of magnitude larger (3.3 eV), should one line up, as in Fig. 3 the valence band edges without the band offsets.

5. Strategies to suppress the "killer defects"

The understanding in (a) the identity of the "killer defects" and in (b) the microscopic meaning of the phenomenological doping limit rule in a series of key materials enables new ways of overcoming the doping limit. For example, a long-standing problem in the field of oxides is that while they can be made n-type, they cannot be made p-type. It will be a great success to be able to make a "p-type transparent conductor". Doping of carbides and fluorides poses another outstanding milestone challenge — so far, most experimental attempts at doping had failed (except, in part, SiC [29] and CdF_2 [30]) for reasons that remain a rather mysterious puzzle. Success here will open a hitherto unexplored future class of electronic materials and possibly pave the way for new technologies. How could this study lead to new design principles to overcome such doping roadblocks?

(i) Work within the bulk defect thermodynamics. One may design new dopable materials by adjusting the band edges with respect to the pinning energies. Suppose that the pinning energy $\varepsilon_{\mathrm{pin}}^{(\mathrm{p})}$ is too high relative to the VBM so the material cannot be made p-type. Since $\varepsilon_{\mathrm{pin}}^{(\mathrm{p})}$ is fixed, one may consider increasing $\varepsilon_{\mathrm{VBM}}$, as an alternative. Thus, one strategy to suppress the killer defect would be to modify the host material so that its VBM will be higher with respect to the vacuum level (i.e., reduce the work function).

(ii) "Defeat" bulk defect thermodynamics. An example is [31–33] the co-doping of GaN by Mg and H. Mg is a substitutional acceptor while H is an interstitial donor. The two forms a charge neutral defect pair [33], thus the Fermi energy is not moved away from near the midgap during the growth and the formation enthalpy of the killer defect (according to Eq. (1)) remains to be high. After growth, the H atoms can be removed at low-enough temperatures at which no killer defect can form, thus freeing the Mg as acceptors. This may explain the apparent discrepancy of the p-type GaN from the "doping limit rule" in Fig. 4.

Acknowledgements

Supported by the US Department of Energy under contract No. DE-AC36-98-GO10337.

References

[1] R.M. Park, M.B. Troffer, C.M. Rouleau, J.M. DePuydt, M.A. Haase, Appl. Phys. Lett. 57 (1990) 2127.
[2] K. Ohkawa, T. karasawa, T. Mitsuyu, Jpn. J. Appl. Phys. 30 (1991) L152.
[3] S. Nakamura, J. Vac. Sci. Technol. A 13 (1995) 705.
[4] I. Akasaki, S. Sota, H. Sakai, T. Tanaka, M. Koike, H. Amano, Electron Lett. 32 (1996) 1105.
[5] D.B. Laks, C.G. Van de Walle, G.F. Neumark, P.E. Blochl, S.T. Pantelides, Phys. Rev. B 45 (1992) 10965.
[6] J.E. Northrup, S.B. Zhang, Phys. Rev. B 47 (1993) 6791.
[7] A. Zunger, in: H. Ehrenreich, D. Turnbull (Eds.), Solid State Physics, 39, Academic Press, New York, 1986, p. 275.
[8] T. Nishimori, K. Nakano, H. Sakamoto, Y. Takakuwa, S. Kono, Appl. Phys. Lett. 71 (1997) 945.
[9] D.J. Chadi, K.J. Chang, Phys. Rev. Lett. 61 (1988) 873.
[10] S.B. Zhang, D.J. Chadi, Phys. Rev. B 42 (1990) 7174.
[11] C.H. Park, D.J. Chadi, Pyhs. Rev. Lett. 75 (1995) 1134.
[12] H. Katayama-Yoshida et al., Centennial Meeting, Bulletin Am. Phys. Society (1999) 1194.
[13] M.C. Wu, Y.K. Su, K.Y. Cheng, C.Y. Chang, Solid State Electron. 31 (1988) 251.
[14] T. Yamada, E. Tokumitsu, K. Saito, T. Akatsuka, M. Miyauchi, M. Konagai, K. Takahashi, J. Crystal Growth 95 (1989) 145.
[15] J. Van Vechten, in: T.S. Moss (Ed.), Handbook of Semiconductors, Vol. 3, Amsterdam, North-Holland, 1980, p. 3.
[16] H.L. Hartnagel et al., Semiconducting Transparent Thin Films, Inst. Of Physics, Bristol, 1995.
[17] J.L. Boone, G. Cantwell, J. Appl. Phys. 57 (1985) 1171.
[18] T.H. Borst, O. Weis, Phys. Stat. Sol. A 154 (1996) 423.
[19] J. Neugbauer, C.G. Van de Walle, Phys. Rev. B 50 (1994) 8067 and references therein.
[20] S.B. Zhang, S.-H. Wei, A. Zunger, J. Appl. Phys. 83 (1998) 3192.
[21] S.B. Zhang, J.E. Northrup, Phys. Rev. Lett. 67 (1991) 2339.
[22] M. Caldas, A. Fazzio, A. Zunger, Appl. Phys. Lett. 45 (1984) 671.
[23] W. Walukiewicz, J. Vac. Sci. Technol. B 5 (1987) 1062.
[24] W. Walukiewicz, J. Vac. Sci. Tech. B 6 (1988) 1257.
[25] W. Walukiewicz, in: K. Wada, S. Pang (Eds.), Defects in Optoelectronic Materials. C&B Science Publishers, Newark, 1999, and references therein.
[26] E. Tokumitsu, Jpn. J. Appl. Phys. 29 (1990) L698.
[27] S.O. Ferreira, H. Sitter, W. Faschinger, R. Krump, G. Brunthaler, J. Crystal Growth 146 (1995) 418.
[28] S.-H. Wei, A. Zunger, Appl. Phys. Lett. 72 (1998) 2011.
[29] H. Wirth et al., Appl. Phys. Lett. 74 (1999) 979.
[30] R.P. Khosla, Phys. Rev. 183 (1969) 695.
[31] S. Nakamura, N. Iwasa, M. Senoh, T. Mukai, Jpn. J. Appl. Phys. 31 (1992) 1258.
[32] J.A. Van Vechten, J.D. Zook, R.D. Hornig, B. Goldenberg, Jpn. J. Appl. Phys. 31 (1992) 3662.
[33] J. Neugebauer, C.G. Van de Walle, Appl. Phys. Lett. 68 (1996) 1829.