ELSEVIER

Physica B 302–303 (2001) 123–134

PHYSICA B

www.elsevier.com/locate/physb

Intrinsic limitations to the doping of wide-gap semiconductors

W. Walukiewicz*

Materials Sciences Division, Lawrence Berkeley National Laboratory, MS 2-200, 1 Cyclotron Rd. Berkeley, CA 94720, USA

Abstract

Doping limits in semiconductors are discussed in terms of the amphoteric defect model (ADM). It is shown that the maximum free electron or hole concentration that can be achieved by doping is an intrinsic property of a given semiconductor and is fully determined by the location of the semiconductor band edges with respect to a common energy reference, the Fermi level stabilization energy. The ADM provides a simple phenomenological rule that explains experimentally observed trends in free carrier saturation in a variety of semiconductor materials and their alloys. The predictions of a large enhancement of the maximum electron concentration in III–N–V alloys have been recently confirmed by experiment. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Compound semiconductors; Native defects; Doping

1. Introduction

It has been realized early on that many of the large variety of semiconductor materials are difficult to dope. The problem has been especially severe in wide-bandgap semiconductors where in many instances n- or p-type doping cannot be achieved at all, significantly limiting the range of applications of these materials [1–3]. These doping limitations have become even more important in new, emerging device technologies that put stringent demands on nanoscale control of electronic and structural properties of semiconductor materials. Such devices require the preparation of small size structures with very high doping levels and abrupt doping and composition profiles. The limits of the maximum doping levels and the question of the stability of dopant and compositional profiles are becoming key issues for many of the new devices.

The past several years have witnessed spectacular progress in the development of a new generation of short wavelength optoelectronic devices based on group III nitrides [4–6] and wide-gap II–VI semiconductors [7,8]. In both cases this progress was made possible through the discovery of more efficient ways to activate acceptor impurities in these material systems. Despite this progress, the high resistance of p-type layers is still a major hurdle in the development of the devices requiring high current injection levels.

There have been numerous attempts to understand the maximum doping limits in semiconductors. Most of these were aimed at explaining limitations imposed on a specific dopant in a specific semiconductor. Thus, it has been argued that in the case of amphoteric impurities in III–V

*Tel.: +1-510-486-5329; fax: +1-510-486-5530.
E-mail address: w_walukiewicz@ux5.lbl.gov
(W. Walukiewicz).

0921-4526/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0921-4526(01)00417-3

compounds, doping is limited by the impurities occupying both acceptor and donor sites, compensating each other. Redistribution of impurities can also lead to limitations of the maximum doping level in the materials with impurity diffusion strongly depending on the Fermi energy [3]. Formation of new stable solid phases involving dopant atoms can be a severe limitation in achieving high doping levels. This limitation depends on the chemical identity of the dopants and the host lattice elements and may be critical in the cases where only a limited number of potential dopants is available [9].

Passivation of donor and acceptor impurities by highly mobile impurities is another major mechanism limiting the electrical activity of dopants. Hydrogen, lithium and copper are known to passivate intentionally introduced dopants in semiconductors. Hydrogen has been an especially extensively studied impurity as it is a commonly used element in most semiconductor processing techniques and in all the growth techniques involving metalorganic precursors [10]. In some cases hydrogen can be removed during a post-growth annealing. Magnesium doped p-type GaN is frequently obtained by thermal annealing of MOCVD grown, hydrogen passivated films [11,12]. However in other instances, as in the case of N doped ZnSe, hydrogen is too tightly bound to the N acceptors and cannot be removed by a thermal annealing [13].

Over the last few years a considerable effort has been directed towards overcoming the doping limits. For example it has been proposed that one can enhance incorporation of electrically active centers by co-doping with donors and acceptors. It has been argued, based on theoretical calculations that because of the reductions of the lattice relaxation and Madelung energies formation energies of proper donor acceptor complexes can be lower than the formation energy of isolated dopant species [13]. Some preliminary experimental results indicate that indeed the co-doping method has produced p-type ZnO that cannot be achieved by any other method [14]. Further studies are needed to fully understand the issues of poor reproducibility of the results obtained by the co-doping method.

In this paper, the saturation of the free carrier concentration in semiconductors obtained through doping will be discussed in terms of the amphoteric defect model (ADM). In recent years, the model has been successfully applied to numerous doping related phenomena in semiconductors. It has been used to explain doping induced suppression of dislocation formation [15] as well as impurity segregation [16,17] and interdiffusion [18] in semiconductor superlattices. We will show that the ADM provides a simple phenomenological rule capable of predicting trends in the doping behavior of a large variety of semiconductor systems.

2. Amphoteric defect model

All point defects and dopants can be divided into two classes: delocalized, shallow dopants and highly localized defects and dopants. Shallow hydrogenic donors and acceptors belong to the first class. Their wave functions are delocalized and formed mostly out of the states close to the conduction band minimum or the valence band maximum. As a result the energy levels of these dopants are intimately associated with the respective band edges, conduction band for donors and the valence band for acceptors. In general the energy levels will follow the respective band edges when the locations of the edges change due to external perturbation such as hydrostatic pressure or changing alloy composition.

In contrast, wave functions of highly localized defects or dopants cannot be associated with any specific band structure extremum. They are rather formed from all the extended states in the Brillouin zone with the largest contribution coming from the regions of large density of states in the conduction and the valence band. Consequently the energy levels of such defects or dopants are insensitive to the location of the low density of states at the conduction and valence band edges. For example, it has been shown that transition metal impurities with their highly localized d shells belong to this class of dopants [19,20]. The insensitivity of the transition metal energy levels to the position of local band extrema has led to the concept of using

these levels as energy references to determine the band offsets in III–V and II–VI compounds [20] and the band edge deformation potentials in GaAs and InP [21].

Compelling evidence for the localized nature of native defects has been provided by studies of semiconductor materials heavily damaged with high gamma rays or electrons [22–28]. It has been found that for sufficiently high damage density, i.e., when the properties of the material are fully controlled by native defects, the Fermi energy stabilizes at a certain energy and becomes insensitive to further damage. The location of this Fermi level stabilization energy, E_{FS}, does not depend on the type or the doping level of the original material and therefore is considered to be an intrinsic property of a given material. As is shown in Fig. 1 the Fermi level stabilization energies for different III–V semiconductors line up across semiconductor interfaces and are located approximately at a constant energy of about 4.9 eV below the vacuum level [29]. This is a clear indication that the native defect states determining the electrical characteristics of heavily damaged materials are of highly localized nature. As can be seen in Fig. 1 the location of the stabilized Fermi energy in heavily damaged III–V semiconductors is in good agreement with the Fermi level pinning position observed at metal/semiconductor interfaces [30]. This finding strongly supports the assertion that the same defects are responsible for the stabilization of the Fermi energy in both cases.

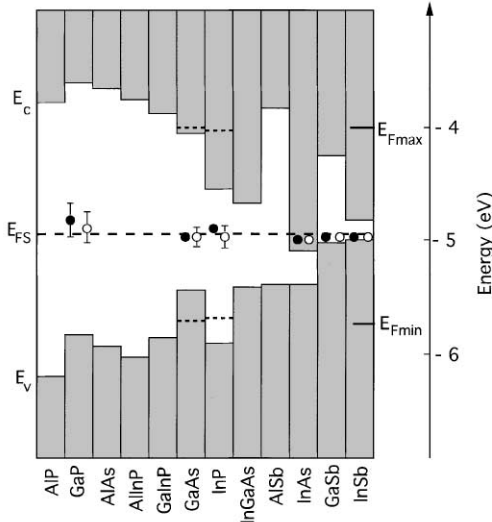

Fig. 1. Band offsets and the Fermi level stabilization energy (E_{FS}) in III–V compounds. The energy is measured relative to the vacuum level. The filled circles represent stabilized Fermi energies in heavily damaged materials, exposed to high energy radiation. The open circles correspond to the location of the Fermi energy on pinned semiconductor surfaces and at metal/semiconductor interfaces. The dashed lines show the location of the Fermi energy for a maximum equilibrium n- and p-type doping in GaAs and InP.

The mechanism explaining the defect induced stabilization of the Fermi energy is based on the concept of amphoteric native defects. The stabilization of the Fermi energy can be understood if we assume that the type of defects formed during high energy particle irradiation or metal deposition on the semiconductor surface depends on the location of the Fermi energy with respect to E_{FS}. For Fermi energy $E_F > E_{FS}$ ($E_F < E_{FS}$) acceptor-like (donor-like) defects are predominantly formed resulting in a shift of the Fermi energy towards E_{FS}. Consequently, the condition $E_F = E_{FS}$ is defined as the situation where the donor and acceptor like defects are incorporated at such rates that they perfectly compensate each other leaving the Fermi energy unchanged.

Such an amphoteric behavior of simple native defects is supported by theoretical calculations that show that depending on the location of the Fermi energy vacancy like defects can acquire either negative or positive charge acting as acceptors or donors, respectively. In the case of GaAs it was shown that both gallium and arsenic vacancies can undergo amphoteric transformations [31]. For example, as shown in Fig. 2 V_{Ga} is a triple acceptor for $E_F > E_v + 0.6\,\mathrm{eV}$. However for lower Fermi energies this configuration is unstable and the vacancy undergoes a relaxation in which one of the first neighbor As atoms moves towards the vacant Ga site. The transformation is schematically represented by the reaction:

$$V_{Ga} \Leftrightarrow (V_{As} + As_{Ga}). \quad (1)$$

In arsenic rich GaAs the calculated formation energy of V_{Ga} is below 1 eV for E_F at the conduction band edge [32].

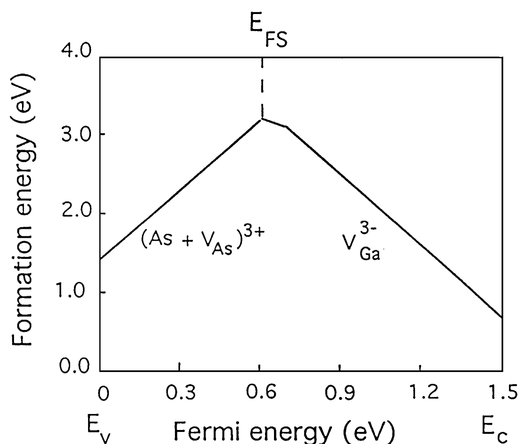

Fig. 2. Formation energy of a gallium vacancy and the related donor defect as function of the Fermi energy in the GaAs band gap [31,32].

A similar amphoteric behavior is also predicted for V_{As} where the transformation is given by the reaction [31]:

$$V_{As} \Leftrightarrow (Ga_{As} + V_{Ga}) \quad (2)$$

In this case the V_{As} donor that is stable in GaAs with E_F larger than about $E_v + 0.8\,eV$ and transforms to an acceptor-like $(V_{Ga} + Ga_{As})$ configuration for $E_F < E_v + 0.8\,eV$ [27]. It is worth noting that these theoretical values of E_{FS} are very close to experimentally determined ones, ranging from $E_v + 0.5\,eV$ to $E_v + 0.7\,eV$ [23].

Most recent theoretical calculations have shown that the amphoteric behavior of native defects is a feature common to many different compound semiconductor systems, including II–VI and III–V semiconductors and the group III-Nitrides [33]. The calculations have confirmed that the reaction (1) is responsible for the amphoteric behavior of V_{Ga}. However it has been found that in the case of V_{As} a transformation from a donor like V_{As} to an acceptor like configuration occurs through a dimerization of the three-fold coordinated Ga atoms surrounding the As vacancy rather than reaction (2). Although, a different type of a structural relaxation is predicted in this case, it does not change the overall conclusion that both cation and anion site vacancies are amphoteric defects and, when introduced in large concentrations, will lead to a stabilization of the Fermi energy.

Since E_{FS} is associated with highly localized defects, its location is not correlated with the positions of the conduction or valence band edges. Thus, as can be seen in Fig. 1, E_{FS} can be located anywhere in the gap or even in the conduction band. In the case of GaAs, E_{FS} is located close to the midgap energy. Therefore high energy radiation damage always leads to a high resistivity GaAs [23]. On the other hand, in the unusual case of InAs, E_{FS} is located deep in the conduction band. Consequently, any high energy radiation damage leads to high n-type conductivity in this material [26]. It has been shown that the location of E_{FS} relative to the band edges is the single most important factor affecting defect-related phenomena in semiconductors.

3. Maximum doping limits in GaAs

Numerous electronic and optoelectronic applications have made GaAs one of the most extensively studied compound semiconductor. It has been realized very early that it is rather easy to dope GaAs with acceptors. Very high concentrations, in excess of $10^{20}\,cm^{-3}$, can be readily obtained by doping with group II atoms [34]. Even higher concentrations close to $10^{21}\,cm^{-3}$ were obtained by doping with carbon [35]. On the other hand n-type doping is much more difficult to achieve. The doping becomes less efficient for donor concentrations larger than about $3 \times 10^{18}\,cm^{-3}$ and the maximum electron concentration saturates at a level slightly above $10^{19}\,cm^{-3}$ [36–39]. The maximum concentration does not depend on the dopant species or the method by which the dopants are introduced into the crystal. Therefore, this limitation appears to be an intrinsic property of the material rather than a feature attributable to the chemical or electronic characteristics of the dopants.

Over the years numerous attempts were made to understand the nature of this limitation. For example it has been proposed that at high concentrations Se donors form electrically inactive complexes [36]. In the case of group IV dopants an

obvious explanation was based on the amphoteric nature of these impurities. It was argued that at high doping levels the dopants begin to occupy both sites forming donors and acceptors that compensate each other [39]. It would be rather surprising if these dopant specific explanations could account for the universal nature of the electron concentration limits.

These results point at the intrinsic nature of the mechanism limiting the free electron concentration in GaAs. Calculations of the electron concentration as function of the doping levels were performed assuming that triple negatively charged V_{Ga} are responsible for the compensation [18]. The results are shown in Fig. 3. A good fit to experimental data was obtained assuming that the formation energy of V_{Ga} $E_f = 2.4\,eV$ for the Fermi energy located at the intrinsic level. As is seen in Fig. 3, the results of the calculations quite well reflect the overall dependence of the electron concentration on the doping level, N_d. At low N_d the Fermi energy is located well below the conduction band, E_f is large and the concentration of V_{Ga} small. Under these conditions all donors are electrically active and $n = N_d$. With increasing doping the Fermi energy shifts upwards towards the conduction band, resulting in a lower E_f and a higher $[V_{Ga}]$. Gallium vacancies compensate the donors and n becomes a sublinear function of N_d. In fact, it can be shown that in a limited concentration range, n is proportional to $(N_d)^{1/3}$. The $\frac{1}{3}$ power dependence reflects the fact that V_{Ga} is a triply charged acceptor. Such a dependence is expected when electrons can still be described by nondegenerate statistics. At even higher doping levels the Fermi energy enters the conduction band and becomes strongly dependent on electron concentration [40]. This leads to a rapid reduction of E_f, an increase of V_{Ga} and as a consequence saturation of n.

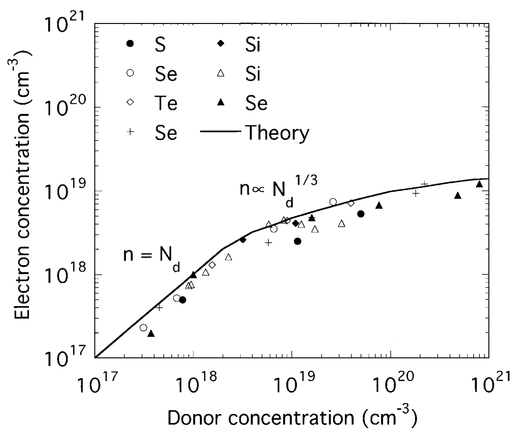

Fig. 3. Electron concentration as a function of donor doping in GaAs. The data points represent experimental results for several different donor species (●—S [37], ○—Se [37], ◆—Si [36] △—Si [39], ◇—Te [37], +—Se [36] and ▲—Se [38]).

It is important to note that the value of $E_f = 2.4\,eV$ appears to be consistent with other determinations of the formation energy of V_{Ga} in intrinsic GaAs. Detailed studies of Ga self-diffusion in undoped GaAs provided the value of the diffusion activation energy, that is a sum of the formation and migration energies of V_{Ga}, $E_{f+m} = E_f + E_m = 3.7\,eV$ [41]. The entropy of $S = 3.5\,k$ has also been determined in this study. In addition, extensive investigations of V_{Ga} facilitated diffusion of As_{Ga} defects in non-stoichiometric, low temperature grown GaAs have provided the values of V_{Ga} migration energies ranging from 1.4 to 1.7 eV [42]. This leads to E_f ranging from 2.0 to 2.3 eV that is somewhat lower than the value of $E_f = 2.4\,eV$ needed to explain the free electron concentration limits. The difference can easily be accounted for by the entropy contribution that has been neglected in the present considerations. At 900 K the entropy of 3.5 k leads to an effective formation energy difference of about 0.27 eV.

The success in explaining the doping limitations in n-type GaAs raises the question whether a similar mechanism is responsible for doping limits in p-type GaAs. As is shown in Fig. 2, V_{Ga} is an unstable defect for $E_F < E_{FS}$. It relaxes to the $V_{As} + As_{Ga}$ donor like configuration with the formation energy $E_f = E_{f0} + 3(E_F - E_{FS})$. With $E_{f0} = E_f(E_{FS}) = 3.1\,eV$ one finds that at a temperature of $T = 900\,K$ for E_F located at the valence band edge E_v, the formation energy, $E_f = 1.8\,eV$. This large formation energy gives a very low value of less than $10^{13}\,cm^{-3}$ for the concentration of the defect donors. Since for $E_F = E_v$ the concentration of free holes is equal to about $4 \times 10^{19}\,cm^{-3}$, it is evident that $(V_{As} + As_{Ga})$ donors are not expected to play any role in the compensation of intentionally introduced acceptors. This is consistent with experiments that indicate that rather high hole

concentrations can be relatively easily achieved in p-type GaAs.

However, it has also been shown that in GaAs doped with column II acceptors the hole concentration saturates at the doping levels slightly above $10^{20}\,cm^{-3}$ [43]. This saturation has been attributed to the fact that column II atoms can act either as acceptors, when they substitute Ga atom sites or as donors when they occupy interstitial sites. The concentration ratio of substitutional to interstitial atoms depends on the location of the Fermi energy. At low concentrations all dopant atoms substitute Ga sites acting as acceptors. With increasing doping level the Fermi energy shifts down towards the valence band and more and more dopants occupy interstitial sites acting as donors. As has been shown before, [16,40] this mechanism leads to a saturation of the position of the Fermi energy level and thus also of the concentration of free holes in the valence band. In the case of GaAs, with the maximum hole concentration of $\sim 10^{20}\,cm^{-3}$, the Fermi energy saturates at about $E_v - 0.2\,eV$ or at $E_{FS} - 0.67\,eV$ when measured with respect to E_{FS} as a common energy reference.

4. Group III-Nitrides

Recent years have witnessed an unprecedented growth of interest in the Group III-Nitrides as a new distinct class of III–V compounds with strongly ionic bonds, smaller lattice constants and large band gaps. These materials form the foundation of a new technology for short wavelength optoelectronics [4] and high power, high-speed electronic devices [44]. Group III-Nitrides have by now been studied for many years. As with many other wide-gap materials the main impediment for practical applications was their propensity to exhibit only one type of conductivity. Typically, as grown GaN or InN are found to be highly n-type conducting and for a long time it was impossible to dope them with acceptors. In the case of AlN neither type of doping appeared to be possible.

These trends in the doping behavior of the nitrides can be understood again in terms of the ADM. Fig. 4 shows the band offsets of group III-

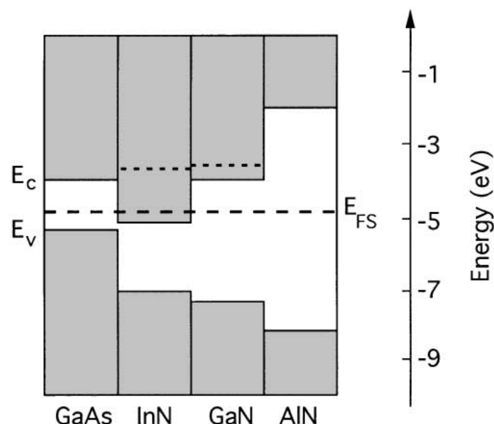

Fig. 4. Band offsets for group III-Nitrides. The dashed lines represent the Fermi energy for the maximum achievable free electron concentration in GaN and InN.

Nitrides. The band edges are placed relative to E_{FS} using the fact that there is a negligibly small conduction band offset between GaAs and GaN. Fig. 6 shows that E_{FS} is located in the upper part of the bandgap in GaN and slightly above the conduction band edge in InN. Therefore it is expected that it should be relatively easy to dope those materials with donors and much more difficult to dope with acceptors. Free electron concentrations exceeding $10^{20}\,cm^{-3}$ [45,46] and $10^{21}\,cm^{-3}$ [47] have been reported in unintentionally doped GaN and InN, respectively. These concentrations correspond to approximately the same Fermi energy at $E_{FS} + 1.3\,eV$ in both materials. The large difference in the maximum electron concentration is consistent with the large conduction band offset between GaN and InN. Also the very large conduction band offsets of 2 eV between GaN and AlN explains why the n-type doping efficiency decreases with increasing Al content [45].

A saturation of free electron concentration has been recently observed in GaN intentionally doped with Se [48]. At low doping levels all Se atoms form electrically active donors. With increasing doping level the electron concentration tends to saturate, showing the characteristic $(N_d)^{1/3}$ dependence on the donor concentration. Similarly, as in GaAs, the saturation of the electron concentration has been explained by incorporation of triply

ionized V_{Ga} acceptors. This result confirms the universal nature of the group III vacancies as compensating acceptor centers in n-type III–V compounds.

5. Group II–VI Semiconductors

Wide-gap group II–VI semiconductors are the group of materials that exhibit the most severe limitations on doping. Indeed, it is this family of materials for which the problem of doping has been recognized first [2]. Early studies have shown that all wide-gap II–VI compounds show a propensity for either n- or p-type conductivity. As grown ZnO, ZnS, HgSe, CdSe and CdS show n-type conductivity and p-type doping is very difficult if not impossible to achieve in these compounds. On the other hand ZnTe typically exhibits p-type conductivity only. It was recognized at that time that the doping limits could originate from compensating native defects that are formed when the Fermi energy shifts towards the band edges [2]. It was not clear, however, how within this picture one could explain differences between apparently similar materials exhibiting completely different doping behavior.

Recent advances in the utilization of II–VI compounds for short wavelength light emitting devices have brought the issue of the doping limitations to the forefront and led to intensive efforts aimed at understanding the mechanisms responsible for the limited dopability of these materials [49,50]. Because of its importance for the blue–green light emitters, ZnSe has been considered a prototypical material to study the doping limitations. It can be relatively easily doped n-type but p-type doping is very difficult to accomplish and only recently doping with reactive nitrogen was successful in achieving p-type conductivity. However, even in this case the free hole concentration is limited to $10^{18}\,cm^{-3}$ [49].

One explanation for this effect is based on the argument that it is energetically favorable for the dopant species to form new compounds with the host crystal atoms rather than substitute lattice sites and act as donors or acceptors [16]. In the case of N doped ZnSe the calculations suggested that Zn_3N_2 should be easily formed preventing N from acting as a substitutional acceptor [9]. Also, these first principle calculations seemed to indicate that the formation energies of native defects are too large and the concentrations are too small to explain low electrical activity of N atoms in ZnSe with compensation by native defects [9]. Later, improved calculations have shown that incorporation of lattice relaxation lowers the formation energy of native defects so that they are likely to play a role in the compensation of N acceptors in ZnSe [51].

Native defects were frequently invoked as the centers compensating electrical activity of intentionally introduced dopants. It was very difficult, however to identify the defects responsible for the compensation or to account for the trends in the doping behavior observed in different II–VI compounds and their alloys. There is evidence that in the specific case of ZnSe : N, V_{Se} or $V_{Se}-N$ defect complexes are responsible for the compensation of p-type conduction [52]. This finding however does not provide any guidance on how to identify the compensating defects in other II–VI compounds.

It has been shown that the trends in the doping behavior of different group II–VI compounds can be understood within the ADM without any need to know the specific identity of the compensating defects [53,54]. The conduction and valence bands for various II–VI semiconductors are shown in Fig. 5 [53]. The Fermi level stabilization energy is located again at about 4.9 eV below the vacuum level. As in the case of III–V compounds, it is assumed that there is a band of allowed Fermi energies $\Delta E_F = E_{Fmax} - E_{Fmin}$ determining the maximum electron and hole concentration that can be achieved in a given material.

In the case of ZnSe the highest electron concentration of about $2 \times 10^{19}\,cm^{-3}$ [55] defines $E_{Fmax} = E_{FS} + 1.3\,eV$ as the upper limit of allowed Fermi energies. The lower limit at $E_{FS} - 1.3\,eV$ corresponds to a maximum free hole concentration of $10^{18}\,cm^{-3}$ [49]. Transferring the same limits to other compounds we find that in ZnTe E_{Fmin} is located deep in the valence band, confirming the experimental observation that it is very easy to dope this material with acceptors. Indeed, free hole

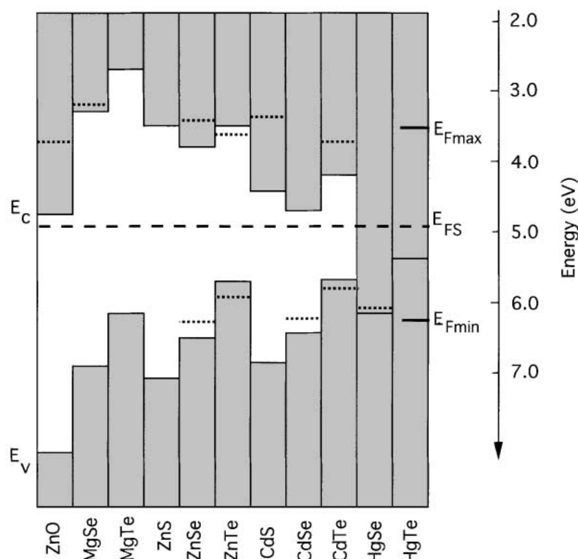

Fig. 5. Band offsets and the Fermi level stabilization energy, E_{FS}, in II–VI compounds. The dashed lines represent positions of the Fermi energy corresponding to the highest hole and electron concentrations reported for the given material.

concentrations as high as $10^{20}\,cm^{-3}$ were reported in ZnTe [56]. On the other hand, since E_{Fmax} is located below the conduction band edge, it is expected that n-type conductivity will be much more difficult to achieve. In fact, it was only recently that n-type conduction with a low electron concentration of $4\times10^{17}\,cm^{-3}$ was reported in ZnTe [57].

As can be seen in Fig. 5 for CdSe and CdS, the upper Fermi energy limit is located in the conduction band in agreement with the observation that both materials are very good n-type conductors. As expected p-type conductivity is much more difficult to realize in these materials. A maximum hole concentration of only $10^{17}\,cm^{-3}$ was reported in CdSe [58]. It is not surprising that in CdS with its very low position of the valence band no p-type doping was ever achieved.

ZnO represents a case of a material with the band edges shifted to very low energies. The conduction band edge is located very close to E_{FS} at $E_{FS}+0.2\,eV$ and the valence band edges lies at the very low energy of $E_{FS}-3.1\,eV$. Such an alignment strongly favors n-type conductivity. Existing experimental data indicate that undoped ZnO can exhibit free electron concentrations as large as $1.1\times10^{21}\,cm^{-3}$ [59]. However the extremely low position of the valence band edge indicates that it will be very difficult, if not impossible, to achieve any p-type doping of this material.

6. Group III–N–V alloys

An excellent example for the predictive power of the ADM has been the recently discovered high activation efficiency of shallow donors in GaInNAs alloys. It has been shown several years ago that alloying of group III–V compounds with group III-nitrides leads to dramatic change of the electronic properties of the resulting group III–N–V alloys [60]. For example GaNAs with only 1% of N has its band gap reduced by 0.18 eV [61]. We have shown recently that the reduction of the band gap results from an interaction between highly localized nitrogen states and the extended states of the host semiconductor matrix [62,63]. The interaction splits the conduction band into two subbands with highly non-parabolic dispersion

relations. It has been shown that the dispersion relation for the lower E_- and upper E_+ conduction subbands are given by

$$E_\pm(k) = \{(E_\mathrm{N} + E_\mathrm{M}(k)) \pm [(E_\mathrm{N} - E_\mathrm{M}(k))^2 + 4(V_\mathrm{NM})^2]^{1/2}\}/2, \quad (3)$$

where $E_\mathrm{M}(k)$ is the conduction band energy of the host semiconductor matrix, E_N is the energy of the localized nitrogen levels and V_NM is the coupling parameter. For a random distribution of N atoms $V_\mathrm{NM} = C_\mathrm{NM}x^{1/2}$ where x is the molar fraction of N atoms in the alloy and C_NM is a constant dependent on the host semiconductor material only. The downward shift of the lower conduction subband given by Eq. (2) accounts for the large reduction of the fundamental band gap [63]. Also, the interaction leads to a large enhancement of the electron effective mass and thus also to an increased density of states of the lower conduction subband [64].

Fig. 6 shows the location of the conduction band edge as function of the N content in $\mathrm{GaN}_x\mathrm{As}_{1-x}$. The maximum Fermi energy that can be achieved by doping is also shown in this figure. It is seen that both the downward shift of the conduction band edge and the increase in the density of states of the lower subband should result in a higher maximum electron concentration in GaNAs alloys [63]. Studies of Se doped GaInNAs alloys have fully confirmed these predictions [65]. As is shown in Fig. 7, GaInNAs alloys with a relatively small N content exhibit a large enhancement of the maximum electron concentration. A more than one order of magnitude improvement of donor activation efficiency has been found in GaInNAs with only 3.3% N.

Our most recent studies indicate that the band anticrossing model well describes the electronic structure of a broad class of highly mismatched semiconductor alloys. We have shown that in addition to III–N–V alloys large downward shifts of the conduction band are also found in group II–VI alloys such as $\mathrm{ZnS}_x\mathrm{Te}_{1-x}$ or $\mathrm{ZnSe}_y\mathrm{Te}_{1-y}$ in which metallic Te is partially replaced by much more electronegative S or Se [66]. It is therefore expected that one could significantly improve the donor activation efficiency by alloying ZnTe with ZnS or ZnSe. For example a more than one order of magnitude higher maximum electron concentration is expected in ZnSTe with only few % of S.

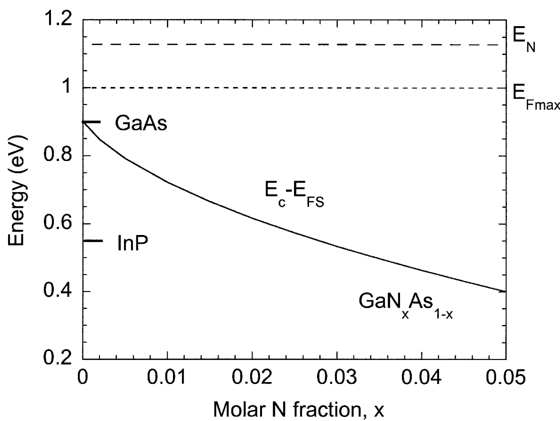

Fig. 6. Energy of the lowest conduction band edge relative to E_FS as function of the N content in $\mathrm{Ga}_x\mathrm{As}_{1-x}$. The energies of the localized nitrogen level E_N and the maximum Fermi energy E_Fmax are also shown.

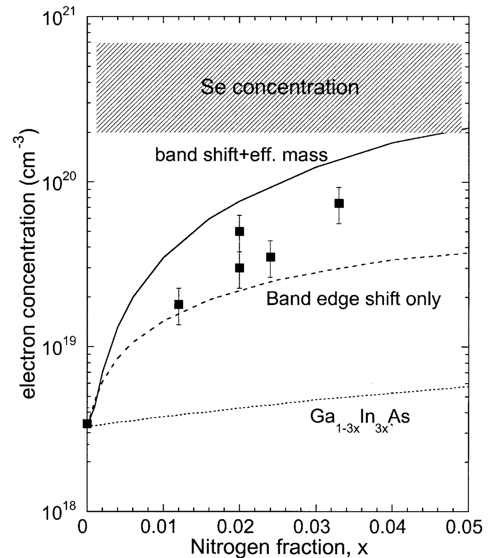

Fig. 7. A comparison of the measured and calculated maximum free electron concentrations as functions of the N content in $\mathrm{Ga}_{1-3x}\mathrm{In}_{3x}\mathrm{N}_x\mathrm{As}_{1-x}$. Two different cases for calculated n_max are shown: one includes effects of the downward shift of the conduction band edge only (dashed curve) and the other includes both the band shift and the enhancement of the density of states effective mass (solid curve). The dotted line shows the increase in n_max expected in $\mathrm{Ga}_{1-3x}\mathrm{In}_{3x}\mathrm{As}$ alloys. The shaded area indicates the of Se concentration in the studied samples.

7. Other wide-gap semiconductors

As has been shown above, the ADM works well in explaining the doping limitations in a number of compound semiconductor systems. The obvious question arises whether it can provide any guidance in evaluating doping limits in elemental semiconductors. Fig. 8 shows the band offsets for group IV materials with energy gaps ranging from zero-gap in grey tin (α-Sn) to 5.5 eV in diamond. Since both Si and Ge have relatively small energy gaps it is not surprising that there are no serious limitations for the doping of these materials. An implantation study has shown that it is relatively easy to activate acceptors in Ge [67]. It has been found that activation of B acceptors does not require any thermal annealing. Such a behavior is understandable in view of the fact that in Ge E_{FS} is located very close to the valence band edge.

Obviously the most intriguing case is represented by diamond. Although there is some uncertainty regarding the actual position of the diamond conduction band it is now well established that it is located close to the vacuum level [68,69]. Consequently, as is shown in Fig. 8, the conduction band is separated from E_{FS} by almost 5 eV and the valence band is located rather close to E_{FS} at $E_{FS} - 0.6$ eV. This vast asymmetry in the location of the band edges with respect to E_{FS} has significant consequences for the behavior of dopants in diamond. It explains why, despite years of intense efforts, there is still no convincing evidence of n-type conductivity in diamond [70]. It further indicates that there is a fundamental reason why it will not be possible to have a thermally stable, well conducting n-type diamond. On the other hand the close location of the valence band to E_{FS} accounts very well for the relatively high p-type conductivity that can be achieved by doping diamond with boron. It has been reported that the concentrations of electrically active boron in excess of 10^{20} cm^{-3} are possible [71].

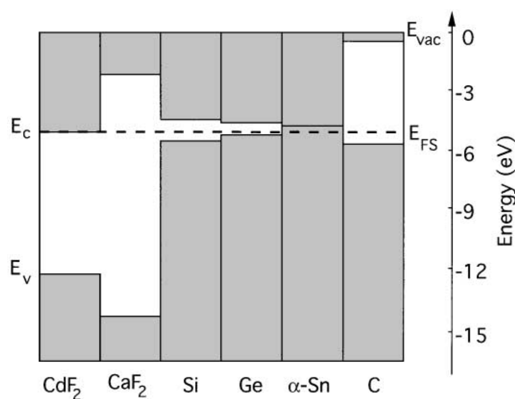

Fig. 8. Band offsets of the group IV and group II-F$_2$ materials. The distinct asymmetry of the location of the band edges of diamond and CdF_2 with respect to E_{FS} determines the doping characteristics of these materials.

Interestingly enough the ADM doping rule applies also to more exotic materials with extremely wide gaps. Despite, as is shown in Fig. 8, the very large energy gap of about 8 eV of CdF_2 [72], it can be doped quite efficiently with donors. Doping with trivalent impurities followed by annealing converts CdF_2 into conducting material [73] and free electron concentrations as high as 4×10^{18} cm^{-3} have been reported. (see e.g. [74]). Again, the reason for this unusual behavior is the extreme asymmetry in the location of the band edges with respect to E_{FS}. As is shown in Fig. 8, E_c of CdF_2 is located almost exactly at E_{FS} [75] which, not only explains a strong propensity towards n-type conductivity but also indicates that p-type doping will not be possible in this material. In contrast CaF_2 with both conduction and the valence band edges far removed from E_{FS}, cannot be doped at all and always acts as a good insulator.

8. Conclusions

It has been shown that native defects in a semiconductor crystal lattice exhibit amphoteric behavior. Depending on the location of the Fermi energy they can act either as acceptors or donors. The demarcation energy separating donor- from acceptor-like behavior plays an important role of the energy at which the Fermi level is stabilized in the presence of large concentrations of native defects. It also serves as a convenient energy reference to evaluate the Fermi energy dependent part of the defect formation energy. Based on these

observations a model has been developed that addresses the issue of the relationship between the native defects and intentionally introduced dopants. It is shown that the maximum free electron or hole concentration that can be achieved by doping is an intrinsic property of a given semiconductor and is fully determined by the location of the semiconductor band edges with respect to the Fermi level stabilization energy. The ADM provides a simple phenomenological rule that explains experimentally observed trends in free carrier saturation in semiconductors. It correctly predicts the maximum attainable concentrations of free electrons and free holes in a variety of semiconductor materials systems. It has been also used successfully in addressing other issues including impurity segregation and interdiffusion in semiconductor heterostructures and doping induced suppression of dislocation formation.

Use of complex, layered structures of different semiconductor materials plays an increasingly important role in the design of modern optoelectronic devices. Such structures allow not only to tune the emitted light energy but also to control the confinement and separation of free electron and hole systems. This is achieved by the proper tuning of the conduction and the valence band offsets between different component layers of the devices. The problems of the maximum doping and impurity redistribution within such device structures have always been treated as entirely separate issues. The ADM unifies those two apparently unrelated aspects of optoelectronic devices by providing a simple rule relating the maximum doping levels and dopant diffusion and redistribution to the same conduction and the valence band offsets that control the distribution of free electrons and holes in optoelectronic devices.

Acknowledgements

The author would like to thank Eugene Haller for a critical reading of the manuscript. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, Division of Materials Sciences, of the US Department of Energy under Contract No. DE-AC03-76SF00098.

References

[1] R.L. Longini, R.F. Greene, Phys. Rev. 102 (1956) 992.
[2] F.A. Kroger, H.J. Vink, Solid State Phys. III (1956) 307.
[3] E.F. Schubert, Doping in III–V Semiconductors, Cambridge University Press, Cambridge, 1993.
[4] S. Nakamura, J. Vac. Sci. Technol. A 13 (1995) 705.
[5] I. Akasaki, S. Sota, H. Sakai, T. Tanaka, M. Koike, H. Amano, Electron. Lett. 32 (1996) 1105.
[6] P. Perlin, I. Gorczyca, N.E. Christansen, I. Grzegory, H. Teisseyre, T. Suski, Phys. Rev. B 45 (1992) 13307.
[7] R.M. Park, M.B. Troffer, C.M. Rouleau, J.M. DePuydt, M.A. Haase, Appl. Phys. Lett. 57 (1990) 2127.
[8] A. Ishibashi, Proceedings of seventh International Conference on II–VI Comp. and Devices, Edinburgh, Scotland, UK, North Holland, Amsterdam, August 1995, p. 555.
[9] D.B. Laks, C.G. Van de Walle, G.F. Neumark, P.E. Blochl, S.T. Pantelides, Phys. Rev. B 45 (1992) 10.
[10] E.E. Haller, Semicond. Sci. Technol. 6 (1991) 73.
[11] I. Akasaki, H. Amano, Jpn. J. Appl. Phys. 36 (1997) 5393.
[12] S. Nakamura, M. Senoh, T. Mukai, Jpn. J. Appl. Phys. 30 (1991) L1708.
[13] T. Yamamoto, H.K. Yoshida, Jpn. J. Appl. Phys. 38 (1999) L166.
[14] M. Joseph, H. Tabata, T. Kawai, Jpn. J. Appl. Phys. 38 (1999) L1205.
[15] W. Walukiewicz, Phys. Rev. B 39 (1989) 8776.
[16] W. Walukiewicz, Mater. Res. Soc. Symp. Proc. 300 (1993) 421.
[17] W. Walukiewicz, Inst. Phys. Conf. Ser. No 141 (1994) 259.
[18] W. Walukiewicz, Appl. Phys. Lett. 54 (1989) 2094.
[19] A. Zunger, Ann. Rev. Mater. Sci. 15 (1985) 411.
[20] J.M. Langer, H. Heinrich, Phys. Rev. Lett. 55 (1985) 1414.
[21] D.D. Nolte, W. Walukiewicz, E.E. Haller, Phys. Rev. Lett. 59 (1987) 501.
[22] V.N. Brudnyi, V.A. Novikov, Fiz. Tekh. Poluprovodn. 19 (1985) 747 [Sov. Phys. Semicon. 19 (1985) 460].
[23] V.N. Brudnyi, M.A. Krivov, A.I. Potapov, V.I. Shakhovostov, Fiz. Tekh. Poluprovodn 16 (1982) 39 [Sov. Phys. Semicon. 16 (1982) 21].
[24] J.W. Cleland, J.H. Crawford, Phys. Rev. 100 (1955) 1614.
[25] V.N. Brudnyi, V.A. Novikov, Fiz. Tkh. Poluprovodn. 16 (1982) 1880 [Sov. Phys. Semicon. 16 (1983) 1211].
[26] N.P. Kekelidze, G.P. Kekelidze, Radiation Damage and Defects in semiconductors, Vol. 16, Inst. Phys. Conf. Series, IOP, London, 1972, p. 387.
[27] T.V. Mashovets, R.Yu. Khansevarov, Fiz. Tverd. Tela 8 (1966) 1690 [Sov. Phys-Solid State 8 (1966) 1350].
[28] W. Walukiewicz, Phys. Rev. B 37 (1988) 4760.
[29] W. Walukiewicz, J. Vac. Sci. Technol. B 6 (1988) 1257.
[30] W. Walukiewicz, J. Vac. Sci. Technol. B 5 (1987) 1062.

[31] G.A. Baraff, M.S. Schluter, Phys. Rev. Lett. 55 (1985) 1327.
[32] S.B. Zhang, J.E. Northrup, Phys. Rev. Lett. 67 (1991).
[33] D.J. Chadi, Mater. Sci. Forum 258–263 (1997) 1321.
[34] E.F. Schubert, J.M. Kuo, R.F. Kopf, H.S. Luftman, L.C. Hopkins, N.J. Sauer, J. Appl. Phys. 67 (1990) 1969.
[35] T. Yamada, E. Tokumitsu, K. Saito, T. Akatsuka, M. Miyauchi, M. Konagai, K. Takahashi, J. Crystal Growth 95 (1989) 145.
[36] L.J. Veiland, I. Kudman, J. Phys. Chem. Solids 24 (1963) 437.
[37] M.G. Milvidskii, V.B. Osvenskii, V.I. Fistul, E.M. Omelyanovskii, S.P. Grishina, Sov. Phys. Semicond. 1 (1968) 813.
[38] O.V. Emelyanenko, T.S. Lagunova, D.B. Nasledov, Sov. Phys.-Solid State 2 (1960) 176.
[39] J.M. Whelan, J.D. Struthers, J.A. Ditzenberger, Proceedings of International Conference on Semiconductors in Physics, Prague, 1960, Academic Press, New York and London, 1961, p. 943.
[40] W. Walukiewicz, Phys. Rev. B 50 (1994) 5221.
[41] H. Bracht, E.E. Haller, K. Eberl, M. Cardona, Appl. Phys. Lett. 74 (1999) 49.
[42] D.E. Bliss, W. Walukiewicz, J.W. Ager III, E.E. Haller, K.T. Chan, S. Tanigawa, J. Appl. Phys. 71 (1992) 1699.
[43] L.Y. Chan, K.M. Yu, M. Ben-Tzur, E.E. Haller, J.M. Jaklevic, W. Walukiewicz, C.M. Hanson, J. Appl. Phys. 69 (1991) 2998.
[44] Y.-F. Wu, B.P. Keller, S. Keller, D. Kapolnek, P. Kozodoy, S.P. Denbaars, U.K. Mishra, Solid-State Electron. 41 (1997) 1569.
[45] S. Yoshida, S. Misawa, S. Gonda, J. Appl. Phys. 53 (1982) 6844.
[46] Y. Koide, H. Itoh, N. Sawaki, I. Akasaki et al., J. Elechtrochem. Soc 133 (1986) 1956.
[47] Y. Sato, S. Sato, J. Crystal Growth 144 (1994) 15.
[48] Gyu-Chul Yi, B.W. Wessels, Appl. Phys. Lett. 69 (1996) 3028.
[49] J. Qiu, J.M. DePuydt, H. Cheng, M.A. Haase, Appl. Phys. Lett. 59 (1989) 2992.
[50] D.J. Chadi, Phys. Rev. Lett. 72 (1994) 534.
[51] A. Garcia, J.E. Northrup, Phys. Rev. Lett. 74 (1995) 1131.
[52] A.L. Chen, W. Walukiewicz, K. Duxstad, E.E. Haller, Appl. Phys. Lett. 68 (1996) 1522.
[53] W. Walukiewicz, J. Crystal Growth 159 (1996) 244.
[54] S.B. Zhang, Su-Huai Wei, A. Zunger, J. Appl. Phys. 83 (1998) 3192.
[55] S.O. Ferreira, H. Sitter, W. Faschinger, Appl. Phys. Lett. 66 (1995) 1518.
[56] S.O. Ferreira, H. Sitter, W. Faschinger, G. Brunthaler, J. Crystal Growth 140 (1994) 282.
[57] H. Ogawa, G.S. Ifran, H. Nakayama, M. Nishio, A. Yoshida, Jpn. J. Appl. Phys. 33 (1994) L980.
[58] T. Ohtsuka, J. Kawamata, Z. Zhu, T. Yao, Appl. Phys. Lett. 65 (1994) 466.
[59] M. Sondergeld, Phys. Stat. Sol. B 81 (1977) 253.
[60] M. Weyers, M. Sato, H. Ando, Jpn. J. Appl. Phys. 31 (1992) L853.
[61] K. Uesugi, N. Marooka, I. Suemune, Appl. Phys. Lett. 74 (1999) 1254.
[62] W. Shan, W. Walukiewicz, J.W. Ager III, E.E. Haller, J.F. Geisz, D.J. Friedman, J.M. Olson, S.R. Kurtz, Phys. Rev. Lett. 82 (1999) 1221.
[63] W. Walukiewicz, W. Shan, J.W. Ager III, D.R. Chamberlin, E.E. Haller, J.F. Geisz, D.J. Friedman, J.M. Olson, S.R. Kurtz, Proceedings of the 195th Meeting. Electrochem. Soc. Vol. 99–11, 1999, p. 190.
[64] C. Skierbiszewski, P. Perlin, P. Wisniewski, W. Knap, T. Suski, W. Walukiewicz, W. Shan, K.M. Yu, J.W. Ager, E.E. Haller, J.F. Geisz, J.M. Olson, Appl. Phys. Lett. 76 (2000) 2409.
[65] M. Yu, W. Walukiewicz, W. Shan, J.W. Ager III, J. Wu, E.E. Haller, J.F. Geisz, D.J. Friedman, J.M. Olson, Phys. Rev. B 61 (2000) R1337.
[66] W. Walukiewicz, W. Shan, K.M. Yu, J.W. Ager III, E.E. Haller, I. Miotkowski, M.J. Seong, H. Alawadhi, A.K. Ramdas, Phys. Rev. Lett. 85 (2000) 1552.
[67] K.S. Jones, E.E. Haller, J. Appl. Phys. 61 (1987) 2469.
[68] F.J. Himpsel, J.A. Knapp, J.A. Van Vechten, D.E. Eastman, Phys Rev. B 20 (1979) 624.
[69] J. van der Weide, Z. Zhang, P.K. Baumann, M.G. Wensell, J. Bernholc, R.J. Nemanich, Phys. Rev. B 50 (1994) 5803.
[70] A.T. Collins, Properties and growth of diamond, Emis, Data Rev. Ser. 9 (1994) 281.
[71] K. Nishimura, K. Das, J.T. Glass, J. Appl. Phys. 69 (1991) 3142.
[72] B.A. Orlowski, J.M. Langer, Phys. Stat. Sol. B 91 (1979) K53.
[73] J.D. Kingsley, J.S. Prener, Phys. Rev. Lett. 8 (1962) 315.
[74] R.P. Khosla, Phys Rev. 183 (1969) 695.
[75] A. Izumi, Y. Hirai, K. Tsutsui, N.S. Sokolov, Appl. Phys. Lett. 67 (1995) 2792.