Microscopic Origin of the Phenomenological Equilibrium “Doping Limit Rule” in \textit{n}-Type III-V Semiconductors

S. B. Zhang, S.-H. Wei, and Alex Zunger
National Renewable Energy Laboratory, Golden, Colorado 80401
(Received 21 May 1999)

The highest equilibrium free-carrier doping concentration possible in a given material is limited by the “pinning energy” which shows a remarkable universal alignment in each class of semiconductors. Our first-principles total energy calculations reveal that equilibrium \textit{n}-type doping is ultimately limited by the spontaneous formation of close-shell acceptor defects: the \((3-)\)-charged cation vacancy in AlN, GaN, InP, and GaAs and the \((1-)\)-charged DX center in AlAs, AIP, and GaP. This explains the alignment of the pinning energies and predicts the maximum equilibrium doping levels in different materials.

PACS numbers: 61.72.Vv, 61.72.Ji, 71.15.Nc

Semiconductor-based high technology owes its existence, in large, to the fact that these materials can be doped, i.e., produce free charge carriers. Indeed, failure to dope a class of materials is often the single most important bottleneck for advancing semiconductor-based electronic or optoelectronic technology. Recent examples of difficult-to-dope systems include \textit{p}-type doping of wide-gap II-VI compounds [1] and nitrides [2], \textit{n}-type doping of diamond [3], and \textit{p}-type doping of transparent oxides [4]. In addition to these cases one notices that even if certain materials can be doped, there is a maximum equilibrium doping limit that cannot be exceeded. This maximum dopability is vanishingly low for certain “undopable” systems (\textit{p}-ZnO, \textit{p}-ZnS, \textit{n}-ZnS, \textit{n}-ZnTe), and is finite for others. For example, GaP can be doped \textit{n}-type only up to \(10^{18}\) \text{ cm}^{-3} [5], while GaAs can be doped even to a level of \(10^{20}\) \text{ cm}^{-3} [6]. A recent discovery provides a certain insight into the nature of this “doping limit” [7–10]: it showed that the failure to dope is related not merely to the existence of a large band gap, but to the position of the valence band maximum (VBM) with respect to a phenomenological \textit{p}-like pinning energy \(\epsilon_{\text{pin}}^{(p)}\) and the position of the conduction band minimum (CBM) with respect to the \textit{n}-like pinning energy \(\epsilon_{\text{pin}}^{(n)}\). Remarkably, in a given material class the values of \(\epsilon_{\text{pin}}^{(n)}\) and separately \(\epsilon_{\text{pin}}^{(p)}\) tend to line up, on an “absolute” energy scale obtained by aligning the VBM’s of different materials according to their respective band offsets [11]. This is illustrated by the dashed lines in Fig. 1 for III-V compounds (see also Ref. [10] for II-VI and I-III-VI\textsubscript{2} compounds). The emerging phenomenological “equilibrium doping limit” rules are as follows: (a) A material for which \(\epsilon_{\text{pin}}^{(n)} < \epsilon_{\text{CBM}}\) cannot be doped \textit{n}-type. (b) A material for which \(\epsilon_{\text{pin}}^{(p)} > \epsilon_{\text{VBM}}\) cannot be doped \textit{p}-type. This remarkable rule permits one to guess rather accurately if a material can be doped or not at equilibrium (and roughly by how much), merely by positioning its band edge energies on a diagram such as Fig. 1. However, the origin of this generally successful rule remains unclear.

In this paper we investigate, using first-principles total energy calculations, the microscopic origin of the \textit{n}-type pinning energy and the lineup revealed in Fig. 1. We note that in general intentional \textit{n}-type doping via donors raises the Fermi level in the gap. This lowers the formation enthalpy of intrinsic acceptor defects. The spontaneous formation of such specific acceptor “killer defects,” when \(\epsilon_{F} = \epsilon_{\text{pin}}^{(n)}\), is the cause for limiting \textit{n} doping. (Of course, if the formation of the “killer defects” is kinetically inhibited, one can dope in excess of such limits.) We find that the killer defects in III-V compounds are electronically close-shell structures, such as the \textit{triply} charged cation vacancy or the \textit{singly} charged DX center. In these close-shell

![Experimental (-----) and ab-initio (---) \(\epsilon_{\text{pin}}^{(n)}\) values relative to CBM of III-V’s](image-url)

\(\text{FIG. 1. LDA-corrected (solid lines) vs the experimental (dashed lines) \textit{n}-type pinning energies in seven III-V compounds. (a) Cation vacancies. (b) DX centers. The shaded region denotes the conduction bands with band alignments taken from Ref. [11]. The energy zero is the VBM of GaAs.}

1232 0031-9007/00/84(6)/1232(4)$15.00 © 2000 The American Physical Society
defects, the energy required to form defect gap levels is approximately canceled by the energy needed to fully occupy them. This cancellation results in approximate material independence of $\epsilon_{\text{pin}}^{(n)}$, hence in the alignment of $\epsilon_{\text{pin}}^{(n)}$ across different materials. Our first-principles calculated $\epsilon_{\text{pin}}^{(n)}$ are in good agreement with those obtained from the phenomenological model (Fig. 1), and can thus be used to predict the ultimate equilibrium doping limits in other materials, given their band offsets with materials in Fig. 1.

The formation enthalpy of a defect $\alpha$ (either native or foreign) of charge $q$ is
\[
\Delta H^{(q,\alpha)} = \Delta E^{(q,\alpha)} + n_a \mu_a + q \epsilon_F ,
\]
where $\Delta E^{(q,\alpha)}$ is the difference in total energy with and without the defect:
\[
\Delta E^{(q,\alpha)} = E^{(q,\alpha)}(\text{defect + host}) - E(\text{host only}).
\]
Here, $n_a$ is the number of atoms being removed during the defect formation from the host to an atomic reservoir with chemical potential $\mu_a$. In the case of single cation vacancy, $n_a = 1$ and $\mu_a = E_{\text{cation}}$ which is the total energy of the group III cation in its elemental bulk form. The last term in Eq. (1) is the energy required to remove $q$ electrons from the defect, placing them (in analogy with the atomic reservoir) in the electron reservoir, i.e., the Fermi energy of the semiconductor.

We illustrate our results for $\alpha = \text{cation vacancy}$ and for $\alpha = DX$ center in zincblende III-V compounds. The $DX$ center that we consider is a substitutional Si donor (either $d^0$ or $d^1$) that, upon capturing one or two electrons, becomes an acceptor, $DX^-$. The $DX^-$ is also called interstitial-vacancy pair because the Si atom is displaced along the [111] direction by about 1 Å [12]. We seek the Fermi level positions at which either (i) a complex between the “killer defect” and the intentional donor or (ii) an isolated killer defect form spontaneously. In case (i), spontaneous formation of the complex of charge $q'$ from the donor of charge $q$ requires that
\[
\Delta H^{q'}(\text{complex}) = \Delta H^{q}(\text{donor}).
\]
By Eq. (1), we then have
\[
\epsilon_{\text{pin}}^{(n)} = \epsilon(q/q') = [(\Delta E(q, \alpha) - \Delta E(q', \alpha))/(q' - q)],
\]
where $\epsilon(q/q')$ is the “defect transition energy.” The $DX$ center belongs to this category. In case (ii), because for noninteracting complex, $\Delta H^{q'}(\text{complex}) = \Delta H^{q'}(\text{killer}) + \Delta H^{q}(\text{donor})$, the pinning energy is determined solely by the killer defect creation as
\[
\Delta H^{q'}(\epsilon_F = \epsilon_{\text{pin}}^{(n)}) = 0.
\]
Isolated cation vacancy belongs to this category, and its creation also involves the transfer of atoms between the host and the reservoir.

Our calculation indicates, in agreement with Ref. [13], that for cation vacancy the $(2-3)$ transition level is near the VBM. For $n$ doping, $\epsilon_F$ is near the CBM. Hence, the cation vacancy has a charge of $q = -3$ and Eq. (5) gives
\[
\epsilon_{\text{pin}}^{(n)} = [(\Delta E(V_{\text{cation}}^{1-}) + E_{\text{cation}})/3].
\]
The $DX$ center (and its precursor donor) may exist in three charge states $q = (+, 0, -)$ with the shallow donor $(+/0)$ level near the CBM. Depending on the host, the $(+/-)$ level can be either above the $(+/0)$ level (thus, a positive U system) or below (a negative U system). For the negative U system (where two electrons in the same defect orbital attract each other)
\[
\epsilon_{\text{pin}}^{(n)} = \epsilon(+/-) = [\Delta E(DX^-) - \Delta E(Si_{\text{Ga}}^0)]/2.
\]
For the positive U system (where the two electrons, instead, repel each other)
\[
\epsilon_{\text{pin}}^{(n)} = \epsilon(0/-) = [\Delta E(DX^-) - \Delta E(Si_{\text{Ga}}^0)].
\]

We calculated $\epsilon_{\text{pin}}^{(n)}$ [Eqs. (6)–(8)] using $ab$ initio pseudopotentials [14] in a plane wave basis set [15] and the local density approximation (LDA) [16]. We used the Ceperley-Alder exchange correlation [17] as parametrized by Perdew and Zunger [18]. A 64-atom supercell was used to mimic the isolated point defects with two $k$ points sampling in the irreducible Brillouin zone of the supercell. The basis set cutoff is 25 Ry for all the nitrides and 10 Ry otherwise. For charged defects, a jellium background was used, and the background error was corrected according to Makov and Payne [19], to O(L$^{-3}$), where L is the dimension of the cubic supercell. We relax all the atoms to their equilibrium positions by force minimizations.

Table I lists (in parentheses) the LDA calculated $\epsilon_{\text{pin}}^{(n)}$ for the cation vacancies [from Eq. (6)] for AlN, GaN, AIP, GaP, InP, AlAs, and GaAs, and for the $DX$ centers [from Eq. (8)] for AIP, InP, and GaAs and [from Eq. (7)] for GaP and AlAs. Note that no metastable Si-derived $DX^-$ configuration was found for the nitrides, in agreement with Park and Chadi [20].

<table>
<thead>
<tr>
<th>Material</th>
<th>$\epsilon_{\text{pin}}^{(n)}$ Cation vacancy</th>
<th>$\epsilon_{\text{pin}}^{(n)}$ DX center</th>
<th>$\epsilon_{\text{CBM}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIN</td>
<td>1.97 (1.60)</td>
<td>NE</td>
<td>3.30</td>
</tr>
<tr>
<td>GaN</td>
<td>1.65 (1.28)</td>
<td>NE</td>
<td>1.32</td>
</tr>
<tr>
<td>AIP</td>
<td>2.02 (1.65)</td>
<td>1.52 (1.48)</td>
<td>1.54</td>
</tr>
<tr>
<td>GaP</td>
<td>1.90 (1.53)</td>
<td>1.82* (1.78*)</td>
<td>1.92</td>
</tr>
<tr>
<td>InP</td>
<td>1.82 (1.45)</td>
<td>1.96 (1.92)</td>
<td>1.10</td>
</tr>
<tr>
<td>AlAs</td>
<td>1.93 (1.56)</td>
<td>1.59* (1.55*)</td>
<td>1.72</td>
</tr>
<tr>
<td>GaAs</td>
<td>1.71 (1.34)</td>
<td>1.76 (1.72)</td>
<td>1.52</td>
</tr>
</tbody>
</table>
The LDA values of $\varepsilon^{(n)}_{\text{pin}}$ depend on $\Delta H^{(q,a)}_{\text{LDA}}$ [e.g., Eq. (5)], but the latter can be in error due to the well-known [21] LDA error in the band gap $\varepsilon^{\text{LDA}}_g$; e.g., according to the LDA values in Table I, GaAs cannot be doped n-type because $\varepsilon^{(n)}_{\text{pin}} < \varepsilon_{\text{CBM}}$. Here, we present a simple method to approximately correct this LDA error. Imagine that $\lambda$ is a parameter that alters the LDA band gap $\varepsilon^{\text{LDA}}_g(\lambda)$. The enthalpy $\Delta H_{\text{cor}}$, corrected due to fixing the band gap error by tuning $\lambda$ is given, in a Taylor expansion as

$$\Delta H_{\text{cor}}(\lambda) = \Delta H^{\text{LDA}}_{\text{LDA}}(\lambda_0) + \frac{\partial \Delta H^{\text{LDA}}_{\text{LDA}}(\lambda_0)}{\partial \varepsilon_g(\lambda_0)} [\varepsilon^{\text{LDA}}_g(\lambda) - \varepsilon^{\text{LDA}}_g(\lambda_0)] + \ldots.$$ \hspace{1cm} (9)

We calculate the first term, $\Delta H^{\text{LDA}}_{\text{LDA}}(\lambda_0)$ exactly. We then evaluate the smaller, second term in Eq. (9) for $\lambda$ that satisfies $\varepsilon_g(\lambda) = \varepsilon^{\text{exp}}_g$. Ideally, one should use for $\lambda$ a physical, LDA-correcting effect (e.g., a GW correction [22], or a self-interaction correction (SIC) [18]). Here, however, we assume that one can apply a fictitious band-gap-modifying parameter $\lambda$ such that the scaling of the second term in Eq. (9) (e.g., the derivative) approximately holds. In this study, we select for $\lambda$ the basis set cutoff energy $E_1$. Figure 2 plots $\varepsilon^{(n)}_{\text{pin}}$ for the cation vacancy and the $DX$ center in GaAs against $\varepsilon^{\text{LDA}}_g(E_1)$. By equating $\varepsilon^{\text{LDA}}_g(E_1)$ with the experimental bulk band gap $\varepsilon^{\text{exp}}_g$ in Fig. 2, we obtain a correction of 0.37 eV for $\varepsilon^{(n)}_{\text{pin}}(V_{\text{Ga}})$ and a correction of 0.04 eV for $\varepsilon^{(n)}_{\text{pin}}(DX)$ relative to the respective LDA $\varepsilon^{(n)}_{\text{pin}}$’s at $E_1 = 10$ Ry.

Our correction procedure makes sense because (a) when the band gap is corrected, the bulk formation enthalpy $\Delta H(\text{GaAs})$ is correct. (b) The trends in correction are physical, e.g., $\Delta H_{\text{cor}}(\lambda) - \Delta H^{\text{LDA}}_{\text{LDA}}(\lambda_0)$ is positive for neutral defects (0.7 eV for $V^{0}_{\text{Ga}}$ and 0.9 eV for $V^{0}_{\text{As}}$), more positive for negatively charged defects (1.1 eV for $V^{0}_{\text{Ga}}$) but negative for positively charged defects ($-0.25$ eV for $V^{3+}$). (c) After correction, $\varepsilon^{(n)}_{\text{pin}} > \varepsilon_{\text{CBM}}$ for GaAs in agreement with experiment.

We find that LDA corrections for $\varepsilon^{(n)}_{\text{pin}}$ between GaAs and InP differ by less than 0.1 eV despite the lack of common element between the two. Hence, we have corrected the LDA pinning energies of all III-V compounds in Table I using the GaAs values in Fig. 2. The solid lines in Fig. 1 show the LDA-corrected pinning energies compared with those derived from experimental data (dashed lines). The latter are obtained by using the measured maximum free-electron densities and electron effective masses in each material, and calculating the corresponding $\varepsilon^{(n)}_{\text{pin}}$, as in Ref. [10]. We see the following:

(i) There is a good quantitative agreement between the ab initio $\varepsilon^{(n)}_{\text{pin}}$ and those deduced from experiment. This means that our calculated $\varepsilon^{(n)}_{\text{pin}}$ for the specific defects considered here can be used to predict the maximum n-type carrier density. For example, this explains why $n$ doping in GaP is much less than InP [$\varepsilon^{(n)}_{\text{pin}}(V^{3+}_{\text{Ga}})$ is considerably higher than the CBM in InP] and why $n$ doping is impossible in AlN [$\varepsilon^{(n)}_{\text{pin}}(V^{3+}_{\text{cat}})$ is in the gap], but possible in GaN.

(ii) There is a clear tendency of $\varepsilon^{(n)}_{\text{pin}}$ to line up with respect to a common energy reference for $V^{3+}$ – and separately for $DX$ (uncorrected LDA produces exactly the same lineups in Fig. 1 except that the positions of $\varepsilon^{(n)}_{\text{pin}}$’s for $V^{3+}$ are too low). In the case of the cation vacancy, the variance $\sigma_p$ of the seven calculated $\varepsilon^{(n)}_{\text{pin}}$ values is less than 0.4 eV. The variance would be an order of magnitude larger, 3.3 eV, should one choose to line up instead the VBM’s without the valence band offset. For the five $DX$ centers, $\sigma_p$ is also about 0.4 eV. Remarkably, we find that lineups exist only for the charge state $q$ that corresponds to an electronic closed shell of defect $\alpha$. For $\alpha = \text{cation vacancies this occurs at } q = -3$ for the III-V compounds and at $q = -2$ for the II-VI compounds, while for the $\alpha = DX$ center this occurs at $q = -1$. As we will see shortly, these results provide a microscopic model for the phenomenological “doping limit rule” [8–10].

(iii) The “killer defect” with the lowest $\varepsilon^{(n)}_{\text{pin}}$ value will pin the Fermi energy for n-type materials. Thus, according to Table I, DX centers are the bottleneck for $n$-doping of AlAs, AlP, and GaP, while for AlN, GaN, InP, and GaAs, the cation vacancy is the “killer defect.” Since the DX center involves specific impurities (e.g., Si), it could be possible to significantly increase the free-electron concentration in AlAs and AlP by choosing a different type of dopant to eliminate the DX centers. For example, for

FIG. 2. The n-type pinning energies as a function of the calculated LDA band gap for GaAs. The LDA band gap is a function of the kinetic energy cutoff as indicated in the top part of the figure.
and, from Eq. (1) with \( n_a = 1 \) and \( \mu_a = E_{\text{cation}} \),

\[
\Delta H(V_N^{q=0}) = (6 - N)\epsilon_t - 6e_{\text{VBM}} + F(V_N^{q}) - F(\text{host}) + E_{\text{cation}}. \tag{11}
\]

For a charge \( q \) cation vacancy, \( \Delta H(V_N^{q}) = \Delta H(V_N^{q=0}) - q\epsilon_t + q\epsilon_F \), thus

\[
\Delta H(V_N^{q}) = [6(\epsilon_t - e_{\text{VBM}}) + F(V_N^{q}) - F(\text{host}) + E_{\text{cation}}] - (q + N)\epsilon_t + q\epsilon_F. \tag{12}
\]

Depending on \( q \) and \( N \), the enthalpy \( \Delta H(V_N^{q}) \) can vary from material to material, as shown by the \( \sim 7 \) eV spread of the horizontal lines for \( q = 0 \) in Fig. 3. However, for closed shell vacancies (\( q = -N \)), the \( (q + N)\epsilon_t \) term in Eq. (12) vanishes. This considerably reduces the large material dependence of \( \Delta H(V_N^{q}) \), as shown by

The pinning energy \( \epsilon_{\text{pin}}^{(n)} \) is the Fermi energy \( \epsilon_F \) at which \( \Delta H(V_N^{q=-3}) = 0 \). We see that the spread in \( \epsilon_{\text{pin}}^{(n)} \) (on the horizontal line in Fig. 3) is only \( 0.4 \) eV, thus nearly independent of the semiconductors.

In summary, we have identified the microscopic origin of the Fermi level pinning in \( n \)-type III-V compounds as the spontaneous formation of the cation vacancies, as well as the DX centers. We have calculated from first principles the pinning energies \( \epsilon_{\text{pin}}^{(n)} \), in quantitative agreement with those of the phenomenological model and found a simple physical explanation for the hitherto elusive “doping limit rule.”

We thank W. Walukiewicz for suggesting this study, and T. Mattila for valuable comments. This work was supported by the U.S. DOE-SC-BES under Contract No. DE-AC36-99GO10337 and NERSC for Cray T3E time.