Title
Universal alignment of hydrogen levels in semiconductors and insulators

Permalink
https://escholarship.org/uc/item/3nw4493v

Author
Van de Walle, C G

Publication Date
2006-04-01

Peer reviewed
Universal alignment of hydrogen levels in semiconductors and insulators

Chris G. Van de Walle
Materials Department, University of California, Santa Barbara, CA 93106-5050, USA

Abstract
Hydrogen strongly affects the properties of electronic materials. It is always electrically active, and usually counteracts the prevailing conductivity of the semiconductor. In some materials, however, hydrogen acts as a source of doping. We have developed a model that enables us to predict the electrical activity of hydrogen in any material, based on its band alignment on an absolute energy scale. We discuss the underlying physics, as well as consequences for specific materials, including ZnO and InN.

Keywords: Hydrogen. First-principles calculations. Doping. Passivation.
1. Introduction

Hydrogen is the most abundant element in the universe. When it occurs in atomic form it is responsible for the 1420 MHz radiation that pervades the cosmos. When it occurs in molecular form, however, it is much more difficult to detect. The conversion between atomic and molecular hydrogen forms an important problem in astrophysics. It has recently become clear that this process takes place on the surface of dust grains [1], illustrating the importance of understanding the fundamentals of how hydrogen interacts with materials.

On earth, hydrogen is also quite abundant, and easily incorporated in materials. Hydrogen displays a range of complex behaviors when introduced as an impurity in solids. Interstitial hydrogen is a fast diffuser. It can bind to native defects or to other impurities, often eliminating their electrical activity—a phenomenon known as passivation. But hydrogen can also induce electrically active defects. Various technological developments are lending new urgency to a fuller understanding of the behavior of hydrogen in solids: the search for improved hydrogen storage systems and proton exchange membranes for fuel cells; the pursuit of novel dielectrics for integrated circuits (in which hydrogen may play an equally important role as it does at the Si/SiO₂ interface); attempts to better control the conductivity of transparent conducting oxides; and efforts to take advantage of the passivation of dopant impurities during growth to enhance doping levels.

It should not be surprising that semiconductors often contain significant concentrations of hydrogen, since many growth techniques include lots of hydrogen in the growth environment; these techniques include vapor-phase transport, hydrothermal growth, metal-organic chemical vapor deposition (MOCVD), laser ablation, and sputtering (which are often carried out in an H₂ atmosphere). Molecular beam epitaxy (MBE) is usually considered to be the growth technique that is most free of contaminants—but even there hydrogen tends to be the predominant residual gas, because of hydrogen desorbing from the stainless steel chamber. In addition, post-growth processing, such as forming-gas anneals, can also introduce hydrogen into the material.

The study of hydrogen in semiconductors started more than 50 years ago, when Mollwo et al. [2] observed that exposure of ZnO to hydrogen caused a marked increase in its conductivity. Then, in the 1980s, came a real explosion of activity. Haller and
coworkers [3,4] showed that hydrogen in germanium can *passivate* electrically active impurities—but also *activate* isoelectronic impurities which are normally electrically inactivate. In silicon, Sah *et al.* performed some clever detective work leading them to attribute the passivation of boron in metal-oxide-semiconductor (MOS) transistors to hydrogen released from the Al gate [5]. The role of hydrogen was soon confirmed in vibrational spectroscopy experiments by Pankove *et al.* [6]. At this point, it was clear that hydrogen in boron-doped silicon was behaving as a *donor* (since it was passivating the acceptor). Soon, however, Johnson *et al.* [7] demonstrated that hydrogen could *also* passivate phosphorus donors in Si, indicating that hydrogen had to behave as an *acceptor* in *n*-type material. In other words, hydrogen is *amphoteric*: it acts as a donor (H⁺) in *p*-type material, and as an acceptor (H⁻) in *n*-type material, always *counteracting* the prevailing conductivity. This behavior of hydrogen was confirmed and described in great detail with a variety of experimental techniques [8-11]. In parallel, theory and computation [12-15] elucidated the physics of hydrogen’s interactions with semiconductors, and established the correlation between its atomic and electronic structure. In a compound semiconductor, H⁺ wants to reside in the vicinity of the anion, which can be explained based on Coulomb attraction between the proton and the electronic charge density around the more electronegative anion. Conversely, H⁻ is found in the vicinity of the anion.

This basic understanding of *isolated interstitial hydrogen* formed the basis for explaining the interaction between hydrogen and other impurities. For instance, in GaN hydrogen passivates Mg acceptors, and needs to be removed in a post-growth anneal in order to activate the acceptors [16]. Many experiments and first-principles calculations for a number of semiconductors seemed to confirm the amphoteric behavior as a general feature of hydrogen’s interactions with semiconductors. Meanwhile, little or no attention was paid to the fact that hydrogen in some cases is *not* amphoteric, but actually acts as a source of doping. The difference between these two cases will be explained in Section 3, after a brief description of methods in Section 2. In Section 4 we will describe the model that allows us to *predict* which type of behavior hydrogen will exhibit in a specific material, and in Section 5 we will discuss some specific examples.
2. Methods

Our studies of the behavior of hydrogen in a wide range of materials are based on first-principles calculations. The approach is based on density-functional theory (DFT) within the local density approximation (LDA) and the pseudopotential-plane-wave method. The calculations for interstitial H were carried out in supercells containing either 64 atoms for zinc-blende or 96 atoms for wurtzite. For hydrogen the Coulomb potential is used. Energy differences for hydrogen-containing systems are well converged at a 40 Ry plane-wave cutoff. For each charge state of H, many possible sites in the lattice were explored and the global minimum was identified. Further details can be found in Refs. [16] and [17].

The key quantities that determine the properties of hydrogen are the formation energy, i.e., the energy needed to incorporate H in the host, and the electronic transition levels, which define the electrical behavior. We obtain the formation energy of interstitial hydrogen in charge state \( q \) (where \( q = -1, 0, \) or \( +1 \)) by placing the hydrogen impurity in a volume of host material, calculating the total energy \( E_{\text{tot}}(H^q) \) of this structure, and subtracting the energy \( E_{\text{tot}}(\text{bulk}) \) of a corresponding volume of pure host material \([15,16]\):

\[
E_f^q(H^q) = E_{\text{tot}}(H^q) - E_{\text{tot}}(\text{bulk}) - \frac{1}{2} E_{\text{tot}}(H_2) + q E_F.
\]

Here the reference for the hydrogen energy is given by an H\(_2\) molecule at \( T=0 \). The last term in the formation energy accounts for the fact that H\(^+\) donates an electron, and H\(^-\) accepts an electron; the energy of the reservoir with which these electrons are exchanged is the electron chemical potential or Fermi level, \( E_F \).

The donor level \( \varepsilon(+/0) \) is defined as the Fermi-level position where H\(^+\) and H\(^0\) have equal formation energies; i.e., when the Fermi level \( E_F \) is below \( \varepsilon(+/0) \), H\(^+\) is stable, and when \( E_F \) is above \( \varepsilon(+/0) \), H\(^0\) is stable. Similarly, the acceptor level \( \varepsilon(0/-) \) is defined as the Fermi-level position where H\(^0\) and H\(^-\) have equal formation energies; i.e., when the Fermi level \( E_F \) is below \( \varepsilon(0/-) \), H\(^0\) is stable, and when \( E_F \) is above \( \varepsilon(0/-) \), H\(^-\) is stable. Interestingly, in most semiconductors and insulators the hydrogen donor level \( \varepsilon(+/0) \) is located above the acceptor level \( \varepsilon(0/-) \), an unusual ordering which is characteristic of a “negative-\( U \)” impurity. In this situation, the transition level \( \varepsilon(+/-) \) between the
positive and negative charge states is actually a better descriptor of the electronic behavior of the impurity. It reflects the fact that the neutral charge state, $H^0$, is never thermodynamically stable. $H^0$ can of course still form and be observed at low enough temperatures or in transient situations, allowing observation of the individual donor and acceptor levels; however, in equilibrium (i.e., at high enough temperature and/or after long enough times) no $H^0$ will be present.

It is appropriate to comment on error bars. The use of supercells, finite basis sets, and $k$-point sampling introduces some uncertainties, but we believe the corresponding numerical error bar is smaller than 0.2 eV. A larger uncertainty, however, arises from the use of density-functional theory with its associated band-gap problem. Calculations for impurities with defect levels in the gap (such as hydrogen) will reflect this uncertainty. We believe that, in the specific case of the hydrogen $\varepsilon(+/-)$ level, the uncertainty is relatively small, due to the specifics of the electronic structure: in the case of $H^+$, the defect level is unoccupied and hence the formation energy should not be affected when the band gap is adjusted. In the case of $H^-$, the microscopic origin of the level is mostly anion-related, i.e., it has mostly valence-band character, and again we do not expect a significant correction when the band gap is adjusted. However, to the extent that a correction is necessary it will shift the formation energy of $H^-$ to a higher value, resulting in a shift of the $\varepsilon(+/-)$ level to higher energies.

3. Hydrogen as a source of doping

In the Introduction we noted that hydrogen is amphoteric in most semiconductors and insulators; i.e., it can occur in positive as well as negative charge states. Let’s examine the consequences using GaN as an example. Figure 1 shows the formation energy of hydrogen in its various charge states in GaN. $H^0$ is clearly never the lowest-energy state (characteristic of a “negative-U” system), causing the donor level $\varepsilon(+/0)$ to lie above the acceptor level $\varepsilon(0/-)$. When $E_F$ moves through the band gap, the stable charge state thus changes directly from positive (for $E_F$ below 2.4 eV) to negative (for $E_F$ above 2.4 eV). This implies that in $p$-type GaN [$E_F$ close to the valence-band maximum (VBM)] $H^+$ is favored, while in $n$-type GaN [$E_F$ close to the conduction-band minimum (CBM)] $H^-$ is stable. As a consequence, hydrogen always counteracts the prevailing conductivity. This also means that hydrogen can never act as a source of doping in GaN. Indeed, imagine that hydrogen is the only impurity
present in GaN and that there is no compensation by other defects or impurities. If we start with intrinsic material ($E_F$ near midgap) and slowly add hydrogen, it will first act as a donor (H$^+$). This will drive the Fermi level closer to the conduction band, but when it reaches $\varepsilon(+/−)$ hydrogen starts acting as an acceptor (H$^−$). This will tend to drive the Fermi level back down. Hydrogen thus effectively “self-compensates”, and $E_F$ ends up “pinned” at the $\varepsilon(+/−)$ level. Only when the $\varepsilon(+/−)$ level is very close to or above the conduction-band minimum (CBM) (or very close to or below the valence-band maximum (VBM)), would doping with hydrogen actually succeed in driving the Fermi level close to the respective band edge, and having an observable effect on electron or hole concentrations.

Figure 1 shows that the formation energy of H$^+$ is negative when the Fermi level is very close to the VBM. This is not a cause for concern, because Fermi-level positions for which such a negative formation occurs can in practice never be achieved. Indeed, we are treating the Fermi level as a variable quantity here, but in practice the Fermi level is determined by dopants that are added to the semiconductor. Pushing $E_F$ close to the VBM would be accomplished by incorporating acceptors. We have already pointed out that under such conditions hydrogen acts as a donor, and would therefore compensate the acceptors. In equilibrium, charge neutrality would cause the Fermi level to be pinned at the point where the concentrations of hydrogen and acceptor impurities are equal. At that point $E_F$ would be a significant distance away from the VBM, and the formation energy of H$^+$ would still be positive.

Note that it is not sufficient for the donor level, $\varepsilon(+/0)$, to lie close to or above the CBM in order for hydrogen to act as a donor. Indeed, $\varepsilon(+/0)$ is close to the CBM in GaN, but in spite of that we saw above that the Fermi level would be pinned at $\varepsilon(+/−) \approx 2.4$ eV. Care must thus be taken in the interpretation of experiments that effectively only determine $\varepsilon(+/0)$: the donor level being close to or above the CBM is a necessary, but not a sufficient condition for hydrogen to act as a shallow donor (i.e., a source of electrons). This concern applies in particular to non-equilibrium experiments, in which hydrogen is observed on a time scale that is not long enough to allow full equilibration. Muon-spin rotation experiments, in which muonium is used as a pseudo-isotope of hydrogen, have produced very useful information about the electronic structure of muonium (and by inference of hydrogen) in semiconductors. However, because of the short time scales (the muon lifetime being 2.2 μs) this experiment does
not allow for equilibration, and thus provides only information about \( \varepsilon(+/0) \) or \( \varepsilon(0/-) \) individually, not about \( \varepsilon(+/-) \).

The behavior that we described for hydrogen in GaN is actually the most common behavior for hydrogen in semiconductors: the \( \varepsilon(+/-) \) level usually resides well within the band gap [16]. The fact that hydrogen would always counteract the prevailing conductivity of a semiconductor had actually become accepted as a general feature of the behavior of hydrogen—until calculations for hydrogen in ZnO [18] provided a vivid illustration that hydrogen can actually act as a source of doping. I.e., in ZnO the \( \varepsilon(+/-) \) level is above the CBM, and therefore H\(^+\) is the only stable charge state. Hydrogen will always give up its electron, thus increasing the concentration of free electrons, and since H\(^-\) cannot form there is no self-compensation. This behavior is consistent with Mollwo's early observations about the increase in conductivity when ZnO is exposed to hydrogen [2]; it has also been confirmed in a large number of experiments that were triggered by the theoretical work [19-40].

Knowing that hydrogen can act as a source of doping is obviously very important since, as pointed out in the introduction, it is easily introduced during growth and processing and should therefore always be expected to be present as a contaminant. This is particularly relevant for a material such as ZnO, where high levels of unintentional \( n \)-type conductivity are routinely observed. The knowledge that hydrogen may play a role in this conductivity provides a direct avenue for controlling the conductivity of the material.

We note that the experimentally observable level for H in ZnO corresponds to a shallow donor level. Since the \( \varepsilon(+/-) \) level is above the CBM, hydrogen's electron can lower its energy by being transferred to the CBM. However, this electron can still be bound to the ionized donor (i.e., the proton) in a “hydrogenic effective mass state”. Note that it is somewhat ironic that in most materials hydrogen does not give rise to a “hydrogenic state” (i.e., a shallow dopant state); but in ZnO it actually does.

As mentioned above, the finding that hydrogen acts as a shallow donor in ZnO was unexpected, particularly in light of the fact that ZnO is in many ways very similar to GaN, in which the hydrogen \( \varepsilon(+/-) \) level is definitely well within the band gap. A general explanation and model is described in the next Section.

### 4. Universal alignment
At first sight it is very puzzling that in ZnO the $\varepsilon(+/−)$ level occurs at a much higher energy (above the CBM) than it does in GaN. The puzzle is resolved, however, by taking the band alignment of these materials into account. Indeed the valence-band (as well as conduction-band) offset between GaN and ZnO is very large, about 1.3 eV, with the band of ZnO occurring at a much lower energy (see Fig. 2). Therefore the $\varepsilon(+/−)$ level can be considered as essentially constant (on an “absolute energy scale”) when going from GaN to ZnO.

These band alignments between semiconductors can be calculated using first-principles calculations, using the same techniques as described in Section 2. The methodology is addressed in Refs. [41-44]. We have performed such calculations for a large number of interfaces. The resulting alignments can be combined with our calculated results for the hydrogen $\varepsilon(+/−)$ level in each of the materials. The results [45] show that the $\varepsilon(+/−)$ values exhibit remarkable consistency in terms of their absolute energy across a wide range of materials. The spread in the values around this “universal alignment level” is very narrow in light of the fact that the band edges of the materials (including oxides such as SiO2) span a range of more than 9 eV. An example of the resulting diagram, that focuses on some of the materials we wish to discuss here, is shown in Fig. 3. The alignment of the band structures of the materials shown on the left-hand side of Fig. 3 (up to SiO2) is based on explicit first-principles calculations. For the materials on the right-hand side, we have not yet performed explicit first-principles calculations (neither for band alignments nor for the hydrogen level); we include them here to illustrate the type of predictions and insights that can be obtained from the hydrogen alignment model. For these materials we have used experimental electron affinities to position the conduction-band minimum with respect to the vacuum level [46]. Since electron affinities are surface quantities, they are not necessarily reliable for predicting band alignments. First-principles calculations will be necessary to obtain more accurate band alignments for the oxides.

5. Predictions for specific materials

Figure 3 illustrates that in most materials the electronic level of hydrogen falls within the band gap, allowing H to occur either in the + or − charge states, always acting as a source of compensation. But in ZnO, the level lies about the CBM, causing hydrogen to behave exclusively as a donor and to act as source of doping. Figure 3 allowed us to
predict that this would also happen for hydrogen in InN, a prediction that has already been confirmed [47,48]. The Figure also shows that H in Ge and GaSb should act as a shallow acceptor. Of course, these predictions are subject to the error bars discussed in Section 2. There we noted that the band-gap problem in density-functional theory is likely to introduce a small systematic error that pushes the hydrogen $\varepsilon(+/−)$ level to higher energies. This means that our predictions for cases where hydrogen will act as a shallow donor [$\varepsilon(+/−)$ above the CBM] would become more robust. However, predictions for hydrogen acting as a shallow acceptor [$\varepsilon(+/−)$ below the VBM] have a higher degree of uncertainty.

Regarding oxides, Fig. 3 includes two types of materials. The first is the category of transparent conducting oxides. These are essentially highly $n$-type-doped semiconductors, that are transparent because they are wide-band-gap semiconductors (i.e., they absorb little or no visible light), but they exhibit metallic conductivity. Transparent conductors are used as metallic contacts in devices that are designed for emitting or absorbing light, such as light-emitting diodes, active matrix liquid-crystal displays, or solar cells. The most widely used transparent conductor is indium tin oxide (ITO). The mechanisms that govern the conductivity in these materials are generally not well understood. Oxygen vacancies have frequently been invoked as shallow donors, but recent experience with ZnO (where these vacancies are very deep rather than shallow donors [49-52]) indicates that this hypothesis may not necessarily be correct. It is interesting to contemplate other potential sources of doping, and hydrogen could definitely be a candidate in some of these materials, just like in ZnO. For instance, Fig. 3 indicates that hydrogen in SnO$_2$ will act as a shallow donor.

Another category of materials that is of high interest are the novel dielectrics with high values of dielectric constant that are being investigated as replacements for SiO$_2$ in silicon microelectronics. It is well known that hydrogen is essential for passivating defects at the SiO$_2$/Si interface, and there is reason to expect it may play a similar role at interfaces with these so-called high-k dielectrics. However, the behavior of hydrogen at the interface will depend on its behavior within the oxide. As a first step, interstitial hydrogen in various candidate oxides should be examined, and Fig. 3 includes ZrO$_2$ and HfO$_2$, which are among the most frequently discussed candidate oxides. The electronic level of hydrogen falls within the upper part of the band gap, rendering the positive charge state more likely. This situation is less “symmetrical” than in the case of SiO$_2$,
where the \( \varepsilon(+/−) \) is close to the middle of the gap. This asymmetry may have consequences for the ability of hydrogen to passivate interface states.

Finally we comment on II-VI materials, several of which are included in Fig. 3. We see that ZnO is the only II-VI compound in which the \( \varepsilon(+/−) \) level falls above the CBM; in the other II-VI compounds the level lies well within the band gap. Based on muon spin rotation experiments in which they observed a shallow donor state for muonium in CdS, Gil et al. [53] argued that hydrogen would act as a shallow donor in this material. Both the alignment model and the explicit results for the \( \varepsilon(+/−) \) level in CdS [Fig. 3] make it very unlikely that hydrogen would act as a source of doping in CdS. Instead, we suggest (as discussed in Section 3) that the muon spin rotation experiments actually probe the donor level \( \varepsilon(+/0) \), which may very well be above the CBM; however, due to their nonequilibrium nature they cannot establish whether hydrogen would be a stable dopant, which requires \( \varepsilon(+/−) \) to lie above the CBM.

6. Summary
We have discussed the behavior of hydrogen in semiconductors and oxides in general, and shown how first-principles calculations can provide qualitative and quantitative information about this behavior. A “universal alignment” model allows predicting the electronic behavior of hydrogen in a wide range of materials. Specific examples were discussed, including nitride semiconductors and transparent oxides.

Acknowledgements

References
List of figure captions

Fig. 1 Calculated formation energy \textit{versus} Fermi level for various charge states of hydrogen in GaN.

Fig. 2 Calculated band alignment between GaN and ZnO.

Fig. 3 Band alignments and position of the hydrogen $\varepsilon(+/−)$ level for selected semiconductors and oxides. For each semiconductor, the lower line indicates the position of the valence-band maximum, the upper line the position of the conduction-band minimum, and the thick bar the position of $\varepsilon(+/−)$ (calculated with density functional theory) with respect to the valence-band maximum. The dashed line at $−4.5$ eV indicates the “universal alignment” level. Materials in which hydrogen is predicted to act as a shallow dopant are indicated by shading. The alignment of the band structures of the materials on the left-hand side of the diagram (up to SiO$_2$) is based on explicit first-principles calculations, while the alignments for the other materials are based on electron affinities [46].
Figure 1

Formation Energy (eV)

$E_F$ (eV)

$\varepsilon(0/-)$

$\varepsilon(+/-)$

$\varepsilon(+/0)$

$H^0$

$H^+$

$H^-$
Figure 2