12
SiO_2 in Density Functional Theory and Beyond

L. Martin-Samos, G. Bussi, A. Ruini, E. Molinari, and M.J. Caldas

12.1
Introduction

From a theoretical point of view, amorphous silica is considered one of the most typical strong glasses belonging to the category of disordered tetrahedral networks. From a technological point of view, silica is widely used in different fields such as microelectronic industry (for metal-oxide-semiconductor devices), optical fiber technologies, and nanoimprint lithography. During the manufacturing process, a large variety of defects may be generated in the samples, which modifies the performances of the silica-based devices. Furthermore, when used in harsh environments, these pre-existing defects can act as precursor sites for the generation of new defects, or new defects can directly arise through ionization or knock-on processes making the defect assortment even wider and the device performances unpredictable. In the recent past, with the advent of nanodevices, the reduced dimensions of the oxide layers and the required abruptness of the interface demand for an atomic-scale understanding of the microscopic processes governing electronic performances, such as the carrier mobility and energy levels.

At the atomic scale, defects influence the electrical and optical properties of materials by adding localized states into their band structure. The main impact of such defect states occur when they are located in the band gap. Therefore, it would be of fundamental relevance to provide a reliable description of the gap surroundings. However, even for crystalline phases, the *ab initio* evaluation of the band gap is a hard task. In fact, the most celebrated approach for the calculation of the electronic properties for the ground-state is based on the density functional theory (DFT), but it is in principle unqualified for the description of excited state properties, at least in its usual formulation, and consequently for a correct prediction of the band gap; on the other hand, the application of appropriate and sophisticated approaches based on the many-body-perturbation theory is often hindered by the huge computational effort that is required for realistic systems. Moreover, the analysis of the electronic structure of amorphous systems implies a deep understanding of the impact of disorder – in its different possible manifestations – on the electronic properties of the system.

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

In this chapter, we discuss the possibility of defining a consistent criterium to univocally define the electronic gap of amorphous systems, and we calculate it from first principles for silica using different theoretical schemes. The results of our systematic investigations allow us to trace back the electronic structure of the system to the specific kinds of disorder present in the target system. Our findings are also discussed in relation to the Anderson model, which predicts the formation of localized band tails at the band edges of disordered systems.

12.2 The Band Gap Problem

Let us first recall that the energy difference between photo-emission and inverse photo-emission signals is the difference between the $N + 1$ excitation (-electron affinity) and the $N - 1$ excitation (-ionization potential), where N is the number of electrons in the ground state. The most common first-principles theoretical approach applied to study the electronic structure of solids is the density DFT. In DFT [1, 2], the exact ground-state density and energy for a given system are obtained by minimization of an energy functional which depends on the external potential and some "Universal Functional" of the density. As such "Universal Functional" is unknown, one has to face its practical realization: the building of a functional is not unique, which may lead to uncontrolled approximations and transferability issues. In the Kohn–Sham formalism the functional becomes the exchange–correlation functional. The oldest and most widely used approximation is the local density approximation (LDA) and the classical method to correct LDA is the explicit inclusion of inhomogeneity effects via gradient expansions, i.e., generalized gradient approximation (GGA). It is well known that DFT, declined in any of its two flavors (LDA or GGA), has reached a great success for modeling material properties such as equilibrium cell parameters, phonon spectra, etc. However, experimental band gap values are underestimated by about 30–40%. Even if in principle every excited state energy can be considered as a functional of the ground-state density, there remains the question on how such functionals can be realized in more explicit terms. The main lacks of LDA and GGA are their locality, energy independence, continuity with the adding or removing of electrons, and wrong long-wavelength limit which does not cancel the self-interaction (SI) part in the Hartree potential. Recently, functionals with a fraction of Hartree–Fock exchange (hybrid-functionals) have been build [3–6], in order to compensate for part of the above mentioned lacks (i.e., SI and discontinuity issue).

The GW approximation offers a valuable parameter-free alternative to such *ad hoc* building of new exchange–correlation functionals. GW has its roots in the work by Hedin and Lundqvist [7, 8]. In its first formulation, an approximated form for the electron self-energy (exchange–correlation self-energy) was obtained through an expansion in terms of the screened Coulomb potential. This procedure can be interpreted as a generalization of the Hartree–Fock method, where the expansion is based on the screened rather than the bare Coulomb potential. In more recent times,

see among others [9–16], the GW method has been used on top of DFT calculations, often within the LDA for the exchange–correlation functional. This combined approach (called one-shot GW or G^0W^0) allows for great improvement on the agreement of electronic band structures with experimental results, and is presently becoming the state-of-the-art for the *ab initio* prediction of electronic properties in extended systems. The main ingredient that makes GW so successful is related to the fact that it contains most of the SI cancellation, and that the self-energy is non-local and energy dependent. Moreover, in the GW framework, as in Hartree–Fock through Koopman's Theorem, the HOMO (and each occupied band) is by definition the $N-1$ electronic excitation (where N is the number of electrons) and the LUMO (and each unoccupied band) is the $N+1$ excitations, i.e., the self-energy is discontinuous.

Let us concentrate on the self-energy. Within the Hartree–Fock approximation the self-energy can be written as:

$$\Sigma \equiv \Sigma_x(12) = iv(1+2)G(12), \tag{12.1}$$

where v is the bare Coulomb potential and G is the Green function. Indexes 1 and 2 are generalized coordinates plus spin and time ($1 \equiv x_1, t_1; 2 \equiv x_2, t_2$). As the poles of the Green function are the $N+1$ and $N-1$ excitations, the Hartree–Fock self-energy contains the discontinuity. In the long wavelength limit it is easy to prove that the $q=0$ term compensates the SI coming from the Hartree potential. However, as v depends only on $|r-r'|$, it is still a local and energy independent function.

In the GW approximation the self-energy takes the form of a "dressed" Hartree–Fock:

$$\Sigma(12) = iW(1+2)G(12), \tag{12.2}$$

where W is the screened Coulomb interaction that is calculated through a Dyson-like equation with a random-phase approximation (RPA) irreducible polarizability. The W term contains the response of the electronic system when non-interacting electron–hole pairs are created, which is a function of $(r-r')$, $(r''-r''')$ and the energy. The polarizability is also a function of $(r-r')$ and $(r''-r''')$, and its calculation usually needs a summation on all possible transitions (from occupied to unoccupied states). It is easy to understand why even if the equations are known since the sixties, only recently, with some additional approximations, we manage to computationally afford systems with tens of atoms. One of the approximations of great success has been the Coulomb hole and screened exchange approximation [17] (COHSEX). The COHSEX formula is obtained from the general GW self-energy by imposing a static W before the energy convolution. The COHSEX approximation has the advantage of being independent of the unoccupied states summation in the self-energy. The first term (SEX) is a screened Hartree–Fock, where the screening is the static W, while the second (COH) is just a static local potential.

After COHSEX, different multipole expansions have been proposed to include, without explicitly performing the energy convolution, the energy dependence in W. The most popular is the plasmon pole approximation [17–19], where the energy dependence in W is replaced by an interpolation through a single pole (at some plasmon energy).

Table 12.1 Fundamental/Homo–Lumo gap (eV) for alpha-quartz and amorphous SiO_2 within different approximations, DFT–LDA, one-shot HF, one-shot COHSEX and one-shot GW on top of DFT–LDA.

	DFT–LDA	HF	COHSEX	*GW*
alpha-quartz	5.9	16.9	10.1	9.4
a-SiO_2	5.6	16.2	10.1	9.3

In order to highlight the effect of inclusion of off-diagonal elements, also called local-field effects (with explicit dependency on $(r - r')$ and $(r'' - r''')$) on the description of the gap, we have performed four calculations on alpha-quartz and on an a-SiO_2 model (for details on the generation of the amorphous sample we refer the reader to Ref. [20]) within the four different approximations DFT–LDA, HF, COHSEX (from LDA), and G^0W^0 (from LDA), see Table 12.1. For more details on the GW calculation see Ref. [21, 22].

As expected, DFT underestimates the gap, HF overestimates it, while COHSEX and, in particular, G^0W^0 give the closest values to experimental outcomes – that range from 8.8 to 11.5 eV [23–25]. From a careful comparison between the gap values for quartz and a-SiO_2 (see Table 12.1) we can note that the gap size for quartz is larger than for the a-SiO_2 phase within DFT and HF, while the experiments suggest that, if a difference exists, it is smaller than the experimental accuracy. Local-field effects taken into account by W in the COHSEX and GW self-energy enhances the opening of the gap due to the disorder, bringing the a-SiO_2 gap closer to that of quartz. This enhancement is not related to the SI and discontinuity issues, already corrected (or not inserted) in HF. This behavior implies that, for disordered materials, it is fundamental to treat correlations through non-diagonal objects. Indeed, it is unlikely that any *ad hoc* choice of local functionals, whatever fraction of Hartree–Fock exchange it contains, would be able to reproduce the whole effect of disorder on the electronic structure.

12.3 Which Gap?

In a crystalline semiconductor/insulator the fundamental gap/band gap is a well defined quantity, i.e., the energy difference between the bottom of the conduction band and the top of the valence band. When one deals with a disordered material, the definition of "gap" is not straightforward. Indeed, following the ideas presented in the pioneering works of Anderson, Mott and Cohen [26–29], disorder induces the formation of localized band tails (flat bands) at each band edge, creating boundary regions (mobility edges) and opening the so-called "mobility gap", see Figure 12.1. Conceptually, the mobility gap for a disordered system is equivalent to the band gap for a crystal. In addition to the mobility gap, one could also define a "HOMO–LUMO" gap (following simply an occupied/unoccupied criterium), which is smaller, by definition, than the mobility gap.

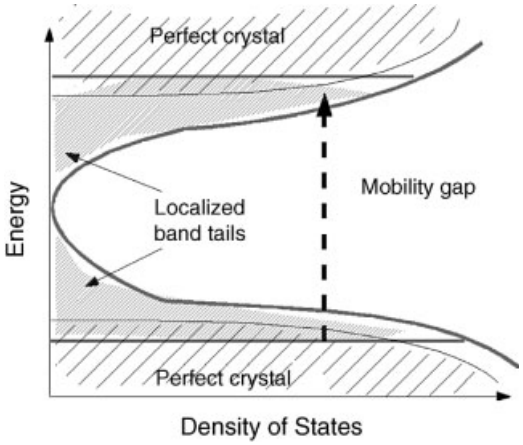

Figure 12.1 Schematic representation of the density of states close to the top of the valence band and the bottom of the conduction band, in a disordered semiconductor/insulator compared to the representative crystal.

For semiconductors such as a-Si/H, the density of states in the mobility gap region has been intensively investigated [30, 31]. However, in the SiO_2 case, there is no experimental evidence of localized band tails between valence and conduction bands. Moreover, experiments do not see any marked difference between the quartz and the a-SiO_2 electronic structure. The effect of disorder has been theoretically studied mostly by addressing model systems through simplified Hamiltonians, that contain parametrized terms to account for the disorder contribution (see, e.g., in Ref. [32]). With respect to the band gap of a perfectly ordered system, a low degree of disorder produces a closure of the mobility gap, while strong disorder opens it. The electronic structure of amorphous SiO_2 obtained within *ab initio* DFT is usually compared to the electronic properties of crystalline alpha quartz, and suggest a small closure of the mobility gap [33]. Amorphous SiO_2 and other low pressure crystalline phases of SiO_2, such as quartz and cristobalite, are constituted by "well-connected" SiO_4 networks: at the short range scale differences between the crystals and the amorphous samples come from bond-angle and bond-length fluctuations, while the medium-range structure is governed by the connectivity of the SiO_4 network itself [34] which is different in all these phases. Given the fact that a low degree of disorder acts by producing bond angle and bond length variation while a topology change is expected to be produced by strong disorder, the reason why DFT results in a small gap closure of the amorphous versus the crystalline system, instead of an opening, is to be clarified. This could be traced back to several reasons, such as to the specific amorphous model used in the simulations, or to the specific theoretical approach, or to the absence "strong" disorder for a-SiO_2, or to the fact that the mobility gap has not been compared to the correct reference crystal. Indeed, the density of a-SiO_2 is around 2.2 g/cm^3 while the density of quartz is around 2.6 g/cm^3. It is well known that the gap increases with the density in tetrahedral networks. Therefore the

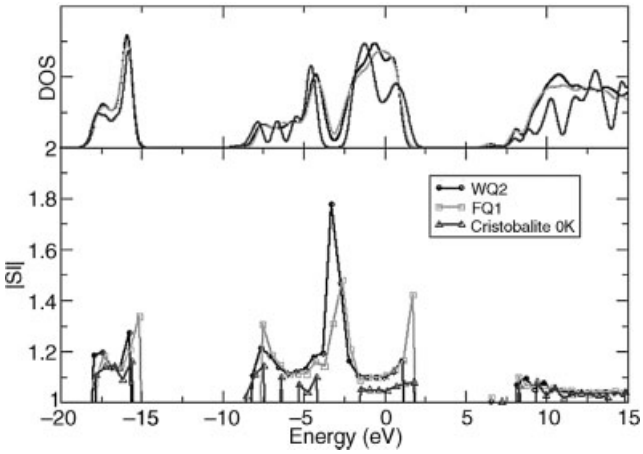

Figure 12.2 (online color at: www.pss-b.com) Upper panel: DFT–LDA energy density of states for two a-SiO_2 models WQ2 and FQ1, obtained by quenching from a melt with two different quench rates, 2.6×10^{13} and 1.1×10^{15} K/s, respectively, and a perfectly ordered crystobalite (Cristobalite 0 K). Lower panel: states localization by means of their |SI|. Each |SI| point has been calculated by averaging within an energy interval of 0.09 eV. The DOS have been aligned maximizing the overlap between deep valence levels.

electronic structure of a-SiO_2 has to be compared with the crystal phase that exhibits the most similar density, i.e., cristobalite.

We have performed DFT calculations on two 108 atoms amorphous SiO_2 models, WQ2 and FQ1 generated at two different quench rate 2.6×10^{13} K/s and 1.1×10^{15} K/s, respectively (where WQ and FQ stay for "well-quenched" and "fast-quenched" to distinguish between the two quench rates, see Ref. [20] for further details), and on two 192 atoms cristobalite models (Fd3-m), one at $T = 0$ K (C0) and one at $T = 300$ K (C300). Here, we use the temperature as a way to add to the system a small stochastic disorder. Calculations have been performed at the Γ point, with a wave function cut-off of 70 Ry and norm-conserving pseudopotentials. Hundred and eight atoms supercell is big enough for having a system-size convergency within 0.1 eV at the self-energy level. In Figure 12.2, we have plotted the density of states and the localization of each state as a function of the energy for the two amorphous models as well as for the perfectly ordered cristobalite, which we use as crystal reference. The localization is described by means of the normalized SI, obtained by dividing the SI, i.e., the Coulomb interaction between an electronic state and itself, as generally defined by the following equation:

$$\mathrm{SI} = \frac{1}{V}\int\int \frac{\varphi_s^*(r)\varphi_s^*(r)\varphi_s^*(r')\varphi_s^*(r')}{||\, r-r' \,||}\mathrm{d}^3r \quad \mathrm{d}^3r', \tag{12.3}$$

by the SI of a plane wave normalized in the corresponding cell (which generalizes the SI tool, extensively used in the quantum chemical community for finding maximally localized basis sets, for the description of extended systems that are modeled through periodic boundary conditions [21]): normalizing the SI allows for a quantitative,

energy-independent estimate of localization for systems with different unit-cell volumes (the boundaries being $|SI| = 1$ for a fully delocalized state, i.e., a plane wave, and divergent $|SI|$ for a completely localized state, i.e., a Delta function)]. We observe signatures of localization almost exclusively at valence edges, while the bottom of the conduction band exhibits a perfectly delocalized character. As expected, the increase of the quench speed increases the localization, and we see at the neighborhood of the valence band top for the FQ1 model states highly localized, that can be considered as localized band tail states. It is also easy to see that the cristobalite gap is smaller than the *mobility* gap of the disordered systems. If we compare the fundamental gap of the perfectly ordered cristobalite, 5.4 eV, with the system perturbed just by thermal disorder, 5.3 eV [21], it is clear that a small degree of disorder induces a small closure of the gap. Therefore, we recover completely the Anderson, Mott, and Cohen picture [26–29], i.e., small disorder degree close the gap while a strong disorder degree widen it.

The widening of the gap due to strong disorder is particularly enhanced when moving to *GW*, as it includes non-diagonal screening, as explained above. Indeed, the *GW* gap for cristobalite is 8.9 eV while the mobility gap for a-SiO_2 is 9.4–9.2 eV for WQ2 and FQ1, respectively (for a more extended discussion see Ref. [21]).

12.4 Deep Defect States

Defect levels (donor or acceptor) and formation energies of charged defects are difficult to describe quantitatively within DFT [35]. Indeed, they are related to ionization potentials ($N-1$ excitation)/electron affinities ($N+1$ excitation) of the defect state [36, 37]. Some attempts have been made to try to circumvent the need of going beyond DFT [38, 39]. However, it is unlikely that one can find semi-empirical rules that work in a general case, and for defects in the already disordered structure of an amorphous system this is even more critical. To illustrate the complexity of the problem, we can compare results for the well-quenched model we showed above, to the results of a model with connectivity defects.

We have performed DFT–LDA and G^0W^0 (on top of LDA) on an amorphous SiO_2 model with two point defects: a non-bridging oxygen NBO (coordination 1) and a tri-coordinated silicon (the simulation parameters are the same as in Ref. [21]). These two atomic defects have produced five strongly localized defect states, a dangling bond (Figure 12.3a), two occupied oxygen 2p non-bonding orbitals (Figure 12.3b and c), and two unoccupied silicon non-bonding orbitals (Figure 12.3d and e).

The corresponding defect levels are located inside the gap (single lines in Figure 12.4). Comparing the level alignment in the neighborhood of the gap, DFT, and *GW*, it is evident that neither an *ad hoc* band stretching nor a rigid energy shift can locate correctly the states from their DFT–LDA relative position, even if one tries to treat separately the occupied and unoccupied states. In the case of the highest localized states, such as the O dangling bond of the NBO, the many-body corrections even alter their position relative to the band tails, precluding scissor shift approximations. Even if

Figure 12.3 (online color at: www.pss-b.com) Localized defect states due to a non-bridging oxygen [(a), (b), and (c)] and a tri-coordinated silicon [(d) and (e)] in an a-SiO_2 model (in white oxygen atoms and in black silicon atoms).

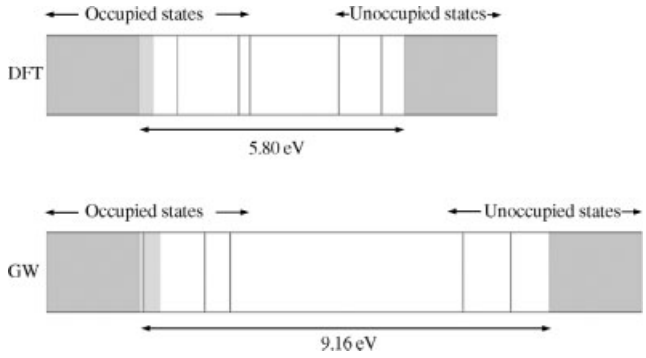

Figure 12.4 (online color at: www.pss-b.com) Electronic structure of a 108 atoms a-SiO_2 system with defects introduced by one non-bridging oxygen and one tri-coordinated silicon, within DFT (upper panel) and G^0W^0 (lower panel). Shaded blue and red bands represent, respectively, delocalized states and localized tails due to disorder. Calculation for the defect system was also performed through norm-conserving pseudopotentials and Γ-point sampling; an energy cutoff of 70 Ry was used for the wave-functions and Fock operator. Single gray lines stay for the strongly localized defect states (occupied oxygen non-bonding and unoccupied silicon non-bonding states). The single red line shows the position of the oxygen dangling bond level. Energies have been aligned to the top of the valence mobility edge.

this kind of approximation has been successfully applied to ideal crystalline systems, it is fundamental to include the exact many-body correction for quantitative studies when treating states with markedly different characters. A qualitatively correct ordering of localized states with a different nature is already provided at the Hartree–Fock level, indicating that the inclusion of exact-exchange is mandatory for a sound description of dangling bonds. However, the weight of the exchange part with respect to the correlation part is different from state to state and no semi-empirical rule can be extrapolated from the results. It is also interesting to note the increase in the energy difference between the two 2p non-bonding orbitals, which are seen almost degenerate within DFT–LDA. The sensitivity to local field effects in *GW* enhances the energy difference, which is due to the different orientation of the orbitals. As for the two unoccupied (and more delocalized) defect states, they maintain almost the same relative position, and are just deeper in the gap, more detached from the mobility edge. We also observe that the sensitivity of defect states to a proper treatment of many-body effects is in agreement with a very recent paper [40], where the application of the *GW* scheme to the analysis of a carbon vacancy in 4H-SiC turned out to be decisive to correctly account for electron–electron correlations.

12.5 Conclusions

We have shown that the inclusion of *local field effects* may be relevant even just for a correct quantitative evaluation of the gap size. The example of the deep defects in

amorphous silica shows how it could be difficult to try to extract general semi-empirical models to circumvent the needs of going beyond DFT. Unfortunately, the way to a bi-univocal modeling of the electronic properties in the gap neighborhood is still under an active debate [36–38]. Indeed, the computational effort needed by a full *GW* calculation, or *GW* including vertex-corrections, free from pseudo-potential effects, precludes its application to realistic systems [41, 42], forcing the research community to find shortcuts, that need to be more extensively tested: pseudo-potentials, approximations to the screening, one-shot *GW* starting from different wave-functions and energies [42]. It is of fundamental relevance to systematically explore all this shortcuts and give complete benchmarks.

References

1 Hohenberg, P. and Kohn, W. (1964) *Phys. Rev.*, **136**, B864.
2 Kohn, W. and Sham, L.J. (1965) *Phys. Rev.*, **140**, A1133.
3 Perdew, J.P., Ernzerhof, M., and Burke, K. (1996) *J. Chem. Phys.*, **105**, 9982.
4 Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2003) *J. Chem. Phys.*, **118**, 8207.
5 Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2006) *J. Chem. Phys.*, **124**, 219906.
6 Stephens, P.J., Devlin, F.J., Chabalowski, C.F., and Frisch, M.J. (1994) *J. Phys. Chem.*, **98**, 11623.
7 Hedin, L. (1965) *Phys. Rev.*, **139**, A796
8 Hedin, L. and Lundqvist, S. (1969) *Solid State Phys.*, **23**, 1.
9 Hybertsen, M.S. and Louie, S.G. (1984) *Phys. Rev. B*, **30**, 5777.
10 Hybertsen, M.S. and Louie, S.G. (1985) *Phys. Rev. Lett.*, **55**, 1418,
11 Godby, R.W., Schluter, M., and Sham, L.J. (1988) *Phys. Rev. B*, **37**, 10159.
12 Farid, B., Daling, R., Lenstra, D., and van Haeringen, W. (1988) *Phys. Rev. B*, **38**, 7530.
13 Engel, G.E., Farid, B., Nex, C.M.M., and March, N.H. (1991) *Phys. Rev. B*, **44**, 13356.
14 Shirley, E.L. and Martin, R.M. (1993) *Phys. Rev. B*, **47**, 15404.
15 Rojas, H.N., Godby, R.W., and Needs, R.J. (1995) *Phys. Rev. Lett.*, **74**, 1827.
16 Rohlfing, M., Kruger, P., and Pollmann, J. (1995) *Phys. Rev. B*, **52**, 1905.
17 Hybertsen, M.S. and Louie, S.G. (1986) *Phys. Rev. B*, **34**, 5390.
18 Godby, R.W. and Needs, R.J. (1989) *Phys. Rev. Lett.*, **62**, 1169.
19 Engel, G.E. and Farid, B. (1993) *Phys. Rev. B*, **47**, 15931.
20 Martin-Samos, L., Limoge, Y., Crocombette, J.P., and Roma, G. (2005) *Phys. Rev. B*, **71**, 014116.
21 Martin-Samos, L., Bussi, G., Ruini, A., Caldas, M.J., and Molinari, E. (2010) *Phys. Rev. B*, **81**, 081202(R).
22 Martin-Samos, L. and Bussi, G. (2009) *Comput. Phys. Commmun.*, **180**, 1416.
23 van den Keybus, P. and Grevendonk, W. (1986) *Phys. Rev. B*, **33**, 8540.
24 Evrard, R. and Trukhin, A.N. (1982) *Phys. Rev. B*, **25**, 4102.
25 Weinberg, Z.A., Rubloff, G.W., and Bassous, E. (1979) *Phys. Rev. B*, **19**, 3107.
26 Anderson, P.W. (1958) *Phys. Rev.*, **109**, 1492.
27 Cohen, M.H., Fritzsche, H., and Ovshinsky, S.R. (1969) *Phys. Rev. Lett.*, **22**, 1065.
28 Anderson, P. (1975) *Phys. Rev. Lett.*, **34**, 953.
29 Mott, N.F. (1978) *Rev. Mod. Phys.*, **50**, 203.
30 Wronski, C.R., Lee, S., Hicks, M., and Kumar, S. (1989) *Phys. Rev. Lett.*, **63**, 1420.
31 Biryukov, A.V. et al., (2005) *Semiconductors*, **39**, 351.

32 Fazileh, F., Chen, X., Gooding, R.J., and Tabunshchyk, K. (2006) *Phys. Rev. B*, **73**, 035124.

33 Sarnthein, J., Pasquarello, A., and Car, R. (1995) *Phys. Rev. Lett.*, **74**, 4682.

34 King, S.V. (1967) *Nature (London)*, **47**, 3053.

35 Fazzio, A., Caldas, M., and Zunger, A. (1984) *Phys. Rev. B*, **29**, 5999.

36 Rinke, P., Janotti, A., Scheffler, M., and Van de Walle, C.G. (2009) *Phys. Rev. Lett.*, **102**, 026402.

37 Martin-Samos, L., Roma, G., Rinke, P., and Limoge, Y. (2010) *Phys. Rev. Lett.*, **104**, 075502.

38 Xiong, K., Robertson, J., Gibson, M.C., and Clark, S.J. (2005) *Appl. Phys. Lett.*, **87**, 183505.

39 Alkauskas, A., Broqvist, P., and Pasquarello, A. (2008) *Phys. Rev. Lett.*, **101**, 046405.

40 Bockstedte, M., Marini, A., Pankratov, O., and Rubio, A. (2010) *Phys. Rev. Lett.*, **105**, 026401.

41 Bruneval, F. (2009) *Phys. Rev. Lett.*, **103**, 176403.

42 Rostgaard, C., Jacobsen, K.W., and Thygesen, K.S. (2010) *Phys. Rev. B*, **81**, 085103.