Energetics and metastability of the silicon vacancy in cubic SiC

Fabien Bruneval and Guido Roma

CEA, DEN, Service de Recherches de Métallurgie Physique, F-91191 Gif-sur-Yvette, France
(Received 14 March 2011; published 29 April 2011)

The silicon vacancy is a prominent intrinsic defect of cubic SiC (3C-SiC) to which much effort has been devoted so far, experimentally and theoretically. We calculate its properties using the $GW$ approximation which does not suffer from the band gap problem. The obtained formation and transition energies deviate significantly from the usual density functional theory evaluations and now compare favorably with experiment. An alternate assignment for the main line of photoluminescence is then proposed. We further perform $GW$ calculations for the saddle point of reaction paths. The resulting barrier energies explain the thermal annealing experiments thanks to an original mechanism mediated by a minority charge configuration.

DOI: 10.1103/PhysRevB.83.144116
PACS number(s): 71.55.–i, 61.72.Cc, 71.15.Mb, 78.55.–m

I. INTRODUCTION

Cubic silicon carbide (3C-SiC) is considered as a promising semiconductor for use in severe environments. The potential applications include hardened electronic devices and nuclear fuel coatings. In particular, its temperature and radiation resistance make 3C-SiC attractive for the next generation of high-temperature gas-cooled nuclear reactors. Because of these properties, irradiation-induced point defects of 3C-SiC have been widely studied both experimentally and theoretically. Among the intrinsic defects, the silicon vacancy plays a prominent role since it remains stable up to relatively high temperatures and it is easily identified by electron paramagnetic resonance (EPR) measurements due to its high-spin configuration. Unfortunately, a careful comparison between the experimental data and the $ab initio$ calculations obtained within density functional theory (DFT) is not satisfactory.

The usual approximations of DFT suffer from the infamous band gap problem. There is nowadays a common agreement that the band gap problem particularly plagues DFT predictions of defect properties in semiconductors and insulators. Point defects in nonmetallic solids generally produce additional electronic levels inside the band gap region. The precise location of these levels is critical for most defect properties: charge state transition energy, formation energy, photoluminescence (PL) lines. Predictive calculations absolutely need to be based on schemes that are devoid of the band gap problem. The $GW$ approximation of the many-body perturbation theory has been proven for several years to yield the correct band gaps. Still the $GW$ calculation of defects is a great numerical challenge.

In the present paper, we calculate, using the $GW$ approximation, the energetics and the stability of the silicon vacancy in 3C-SiC. In order to access these properties, we performed the $GW$ calculations not only for equilibrium structures, but also for saddle point positions. The $GW$ results were combined so as to reduce an error, named concavity, that was recently identified for the $GW$ approximation by one of us. The procedure, explained in Sec. II, allows us to calculate the formation energies of the metastable form $V_{Si}$ and of the stable configuration ($V_{CCSi}$) within the $GW$ framework in Sec. III. The results bring a quantitative agreement with various experimental measurements: EPR, PL, and thermal annealing. Noticeably, the calculations propose a reassignment of the main PL line in Sec. IV and highlight an original decay mechanism through a majority charge state in Sec. V.

II. METHOD

A. $GW$ approximation applied to the defects

Thanks to its accuracy in predicting band gaps and electronic levels, the $GW$ approximation appears to be the method of choice to deal with point defects in semiconductors and insulators. However some noticeable bottlenecks have inhibited the use of this scheme so far. Very recent works have lifted the theoretical and numerical problems as recapitulated in the following.

Deriving from Green’s function theory, the $GW$ approximation provides meaningful quasiparticle electronic levels. The $GW$ approximation is hence reliable in predicting charge changes with constant geometry. For instance, the energy of the highest occupied molecular orbital (HOMO) can be interpreted as a total energy difference at constant geometry, which in turn defines the ionization potential $I$:

$$\epsilon_{GW}^{HOMO}(V_{Si}^q) = E_0(V_{Si}^q) - E_0(V_{Si}^q + 1)$$

$$= -I(V_{Si}^q),$$

where the first argument refers to the geometry of the defect and the second shows the actual charge state $q$. $E_0$ stands for the ground-state total energy and $\epsilon$ for the quasiparticle energy. Symmetrically, the lowest unoccupied molecular orbital (LUMO) energy is a total energy difference, which can be understood as the opposite of the electron affinity: $-A$. In the context of defects, these quasiparticle energies are named vertical transition energies $\epsilon_{v}(q/q + 1)$. Note that this vertical transition energy $\epsilon_{v}(q/q + 1)$ could be obtained in principle either from $\epsilon_{GW}^{HOMO}(V_{Si}^q)$ or from $\epsilon_{LUMO}^{GW}(V_{Si}^q)$. One of us showed recently that these two values slightly differ in practice due to the concavity of the $GW$ approximation. The most consistent way to handle this discrepancy is hence to approximate the vertical transition energy by the mean value of the two mentioned quasiparticle energies.

The calculation of defect formation energies also requires one to relax the geometries, since the structure of a defect changes according to the charge state. Unfortunately the $GW$ total energies and forces are out of reach even in state-of-the-art...
implementations. That is why Rinke and co-workers recently introduced a scheme that combines DFT and GW calculations. DFT is used to deal with the structural changes at constant charge state. The GW approximation is employed for charge changes at constant geometry. When traveling on a given Born-Oppenheimer surface, the band gap underestimation is expected to be harmless and therefore DFT can be safely used. Then the GW approximation allows one to address the vertical transitions from one Born-Oppenheimer surface to another one. The combination of the two moves finally permits one to change both the geometry and the charge states. The thermodynamic charge transition energies \( \epsilon_{\text{th}}(q/q + 1) \) can thus be obtained by introducing the energy of an intermediate point:

\[
\epsilon_{\text{th}}(q/q + 1) = E_0(V^q_{\text{Si}}, q) - E_0(V^{q+1}_{\text{Si}}, q) + E_0(V^{q+1}_{\text{Si}}, q) - E_0(V^{q+1}_{\text{Si}}, q + 1).
\]

The first two terms account for a structural change at constant charge \( q \), which is obtained accurately from two DFT calculations. The last two terms account for a change in geometry at constant structure, which is calculated from GW quasiparticle energies as explained above.

With the present combined DFT/GW approach, the absolute formation energy of the defect for any charge state can be obtained, provided one reference formation energy. In Ref. 13 the reference formation energy was chosen to be the DFT total energy with all defect states in the band gap empty, so that the band gap problem does not enter in the total energy. Then the neighboring charge state formation energy can be obtained as exemplified here:

\[
E_f(V^+_{\text{Si}}) = E_f(V^{2+}_{\text{Si}}) + \epsilon_{\text{th}}(+1/2) - \epsilon^{GW}_{\text{VBM}} - \mu_e.
\]

where \( E_f \) stands for the usual formation energy and \( \mu_e \) stands for the Fermi level with its zero set at the GW valence band maximum, \( \epsilon^{GW}_{\text{VBM}} \).

B. Technical details

The main numerical bottleneck of the GW approximation comes from the dependence of the GW exchange-correlation self-energy onto the empty states. This problem was given a partial answer in Ref. 15, which allows us to run GW calculations for supercells as large as 215 atoms using a very low 3:1 ratio between empty and occupied states.

We employ the usual perturbative GW method (\( G_0W_0 \)) with norm-conserving pseudopotentials and the Godby-Needs plasmon-pole model, using a 30 hartree cutoff for the wave functions and a 6 hartree cutoff for the dielectric matrix. The \( k \)-point grid is a \( 2 \times 2 \times 2 \) Monkhorst-Pack grid for the DFT force calculations. It is reduced to the \( \Gamma \) point only when turning to GW calculations. The calculations are spin polarized and the neutral vacancy is approximated by its triplet spin state, which may overestimate its formation energy by less than 0.1 eV. The obtained band gap is 1.35 eV within the local density approximation (LDA) and 2.19 eV within GW to be compared to the experimental value, 2.37 eV. The discrepancy between the GW band gap and the experimental band gap can be considered as the error bar of the results presented in the following.

The calculations of charged defects in replicated supercells notoriously converge slowly as a function of the supercell size. The supercell technique gives rise to important finite-size effects, among which the electrostatic interaction between the charged defects is arguably the largest. In order to address this issue, a Madelung correction can be designed. However, it has been shown in recent years that a straightforward Madelung correction does more harm than benefit. We clarified this issue for our specific system by performing a careful convergence study about the silicon vacancy \( V^+_\text{Si} \), which should be the most dramatic case of our study.

In Fig. 1, the formation energy of \( V^+_\text{Si} \) is provided for different cubic supercells ranging from 63 atoms to 999 atoms. The structure is frozen with all the atoms in the perfect crystal positions except for the four nearest neighbors to the vacancy, whose positions were relaxed in a 63-atom supercell. This explains why the formation energy is slightly larger than the fully relaxed energy reported elsewhere in the text. In the present case, the Madelung correction clearly overestimate the correction. In Fig. 1, we provide the curve corrected by the popular Lany-Zunger correction scheme, which consists in a potential alignment together with a reduced Madelung correction (precisely 2/3 of the usual Madelung term). The Lany-Zunger scheme is also not satisfactory as shown by the figure. Note that the dielectric constant has been calculated \( \text{ab initio} \) consistently using the same parameters as in the defect calculations. We then found it more reliable to perform a mere potential alignment. Using a 215-atom supercell with no charge correction allows us to evaluate within 0.1 eV the formation energy of \( V^+_\text{Si} \). The values compared to experiments are always differences of formation energies, which are expected to be even more accurate, since the errors compensate to some extent.
III. FORMATION ENERGY OF THE METASTABLE SILICON VACANCY AND THE STABLE COMPLEX

The silicon vacancy is a high-energy defect that can only be observed in heavily irradiated SiC. The numerous silicon vacancies created by irradiation are a metastable configuration: the silicon vacancy experiences a large energy drop when transforming into the complex made of a carbon vacancy and a carbon antisite (VCCSi). Furthermore, the experimentally observed silicon vacancies possess a peculiar high-spin configuration ($S = 3/2$) with three aligned spins resulting in a $-1$ charge state. This particular spin configuration gives rise to an unambiguous EPR signal, named the T1 center. The T1 center can be observed in irradiated samples, whatever the initial doping conditions and irradiating particle. After irradiation, due to the quantity and the variety of point defects introduced in the material, the donor and acceptor defects give rise to a compensated system and the Fermi level remains pinned in the vicinity of the midgap.

In Fig. 2 we provide the formation energy of the two competing configurations for the silicon vacancy: the original silicon vacancy $V_{Si}$ and the complex (VCCSi). The LDA clearly suffers from the band gap problem so that all the transition energies are constrained to be below the LDA conduction edge, symbolized by a vertical dotted line. In addition, it would be doubtful to determine the midgap region: would it be the middle of the Kohn-Sham band gap or the middle of the experimental band gap? Conversely, the $GW$ results offer a much clearer view of the system. First, the $GW$ transition energies are correctly placed within the full range of the $GW$ band gap, which matches reasonably well the experimental band gap. Second, the silicon vacancy adopts a correct $-1$ charge state in a high-spin configuration for a wide region around midgap. The silicon vacancy appears as metastable with respect to the complex (VCCSi) for any Fermi level in the band gap. The energy differences are sizable, ranging from 4.4 eV for the $p$-type region to 0.7 eV for the $n$-type region. The $GW$ results show that the complex carries a $+2$ charge state in the major part of the band gap.

IV. PROPOSAL FOR AN INTERPRETATION OF THE $E$ LINE OF PHOTOLUMINESCENCE

The zero-phonon line of PL spectroscopy measures the vertical transition energy down from an excited defect, as schematically shown in the left panels of Fig. 3. In irradiated samples, additional features appear that can be linked to the intrinsic defects. In particular, a line, labeled $E$ by Itoh and coworkers, was convincingly attributed to the silicon vacancy. Indeed the $E$ line and the above mentioned T1 center in EPR have precisely the same behavior upon thermal annealing: the three annealing stages. This clearly demonstrates that the photons of the $E$ line with an energy of 1.91 eV arise from electronic transitions occurring at the silicon vacancy.

In Ref. 4, the authors proposed that the $E$ line comes from the recombination of a conduction electron with a hole on a silicon vacancy, i.e., $V^{0}_{Si}$. Thus they tentatively assigned it to the transition level $E_{v}(0/-1)$ at an energy of $E_{c} - 1.91$ eV. This assignment was mainly supported by calculations: for instance, our LDA results place the transition $E_{v}(0/-1) = E_{c} + 0.54$ eV. By combining the LDA transition level together with the experimental band gap, 2.37 eV, the resulting emitted photon matches the experimentally measured photon.

However, this assignment is brought into question by our calculations. When performing higher-accuracy $GW$ calculations with the fair $GW$ band gap value, the obtained vertical transition lies at $E_{v}(0/-1) = E_{c} + 0.79$ eV = $E_{c} - 1.58$ eV, as represented in Fig. 3. In this improved framework, the obtained transition energy does not fit the experimental
interpretation any longer. The disagreement is definitely larger than the calculation uncertainties, like the slight underestimation of the band gap by the $GW$ approximation (2.19 eV instead of 2.37 eV) and the neglect of the excitonic effects.

From our calculations, another interesting scenario can, however, be drawn that challenges the previous interpretation of the $E$ line. In fact, the PL technique is not capable of distinguishing donor-valence recombinations from acceptor-conduction ones, so that the observed $E$ line may rather arise from the recombination of an extra electron on the silicon vacancy (i.e., $V_{\text{Si}}$) with a hole in the valence bands. Then, the electronic transition that emits the PL photon will be a signature for the vertical transition $\epsilon_i(-2/-1)$. Previous evaluations for this transition were biased by the band gap problem: within the LDA, $\epsilon_i(-2/-1) = E_v + 1.15$ eV, close to the LDA conduction edge (1.35 eV), and therefore it was disregarded for the $E$ line. Within the $GW$ approximation, this transition is located interestingly close to the experimental $E$ line, $\epsilon_i(-2/-1) = E_v + 1.85$ eV. A small error of the photon energy is in fact expected, since our single-electron picture does not take into account the exciton binding energy and since our calculated band gap is slightly too small. Note that the two sources of error have different signs. Indeed, the electron-hole attraction always lowers the emitted photon energy. Exciton binding energies comprised between 0.18 and 0.24 eV were recently calculated for the carbon vacancy in 4H-SiC.22 It is likely that an increase of the band gap would push the defect level up, since it is close to the conduction edge. Our data clearly advocate for a reinterpretation of the nature of the $E$ line: the $E$ line is a signature of the transition from charge state $-2$ to charge state $-1$, with an experimental value of $\epsilon_i(-2/-1) = E_v + 1.91$ eV against a $GW$ calculated value of $\epsilon_i(-2/-1) = E_v + 1.85$ eV.

V. ANNEALING OF THE SILICON VACANCIES THROUGH A MINORITY CHARGE STATE

Along with the low-temperature EPR and PL studies, the thermal annealing of the silicon vacancy signatures has been widely studied.2,4,23 The annealing of the silicon vacancy related signals occurs in three stages. The recovery stages I and II at rather low temperature (400–700 K) anneal about half of the signal; then stage III at high temperature (1050 K) marks the complete anihilation of $V_{\text{Si}}$. The activation energy of the last stage was determined to be $E_A = 2.2 \pm 0.3$ eV.2,23 The thermal annealing curve yields accurate effective activation energies, but does not give any clue about the mechanisms at the atomic scale. In the following we propose a picture of the recovery stages of $V_{\text{Si}}$ thanks to the published ab initio data and our $GW$ results for a reaction path.

It is reasonable to assume that the first two recovery stages are associated with the elimination of silicon vacancies by recombination with carbon and silicon interstitials, whose migration and recombination barriers are in the range 0.5–1.5 eV.24,25 We have to exclude from this process the +4 charged silicon interstitials, whose migration energy is expected to be large. Vacancies themselves can also be considered immobile.7 The extinction of the T5 center in EPR, assigned to $C_{\text{Si}}$ at the same temperature as the first stage suggests that this is related to the $V_{\text{Si}} + C_{\text{Si}} \rightarrow C_{\text{Si}}$ recombination. The antisite is a very low-energy defect that is very stable once formed.

The third recovery stage shows a first-order kinetics according to the isothermal annealing of Ref. 2. This stage was tentatively proposed as a transformation of the metastable $V_{\text{Si}}$ into the lower-energy complex ($V_{\text{C}}C_{\text{Si}}$).24 but our calculated barriers do not agree with the experimental activation energy as shown in the following discussion. Again we argue that the calculations suffer from the band gap problem and that the mechanism involved is more complex than was thought previously.

We therefore performed reaction path calculations within LDA for the relevant charge states. The LDA configurations for the saddle points were then used to build up the $GW$ formation energies as described above. The LDA paths were obtained thanks to the climbing image nudged elastic band method26 (CI-NEB) with five images for 215-atom supercells as displayed in the upper panel of Fig. 4. In order to compare the different charge states, we set the Fermi level to midgap. One can safely assume that the Fermi level is pinned by the numerous deep donors and acceptors, which somehow compensate each other in heavily irradiated materials. In focusing on the Fermi level we implicitly assume that the thermodynamic equilibrium between charge states is always satisfied. This hypothesis is not true in general. However, in the present case, the annealing experiment is carried out at very high temperature (1050 K) and the kinetics of charge equilibration is a matter of nanoseconds. We will discuss this point in further detail later on. The obtained LDA transition paths have a minimal barrier of 2.9 eV for the $-1$ charge state. This value deviates greatly from the experimental activation energy.

In the lower panel of Fig. 4 we provide the $GW$ energies for saddle points and for stable points. The $GW$ barriers

![FIG. 4. (Color online) Energy path of the transformation of $V_{\text{Si}}$ into ($V_{\text{C}}C_{\text{Si}}$) for different charge states. The carbon-rich conditions were set. The Fermi level has been set to the experimental midgap value. The upper panel shows the five-image CI-NEB path obtained for the 215-atom supercell within LDA. The lower panel provides the $GW$ evaluation for the energy of the three critical points of the path. The points actually calculated are depicted with the symbols and the lines are guides to the eye.](144116-4)
shine light onto a very peculiar diffusion mechanism. The vast majority of the silicon vacancies possess a $-1$ charge state; however, the direct transformation into the complex ($V_{CCSi}$)$^-$ is blocked by a large barrier of $2.75$ eV. The very few positively charged silicon vacancies $V_{Si}^+$ experience a low energy barrier of $1.50$ eV. However the total effective activation energy of the reaction in the $+1$ charge state is the sum of the barrier energy for charge $+1$ and of the energy needed to turn the vacancy from $-1$ to $+1$. We finally obtain an activation energy of $2.32$ eV, which lies within the experimental uncertainty. Some lowering of this figure by temperature effects can be further expected. The transformation of the metastable silicon vacancy into the stable complex is mediated by a double charge change.

The annealing experiment usually lasts for several minutes or hours. As a consequence, the charge reequilibration is many orders of magnitude faster than the annealing experiment itself. The instantaneous charge equilibration can be safely assumed as we did to draw Fig. 4.

VI. CONCLUSIONS

As a conclusion, the use of the state-of-art GW approximation for large supercells of $215$ atoms allowed us to completely clarify the experimental observations concerning the silicon vacancy in $3C$-SiC. The calculated properties deviate noticeably from the previous LDA studies, owing much to the absence of the band gap problem. In addition to the nice agreement with the experimental data, we brought deeper understanding to the physics of the silicon vacancy: we proposed an assignment for the main PL line, attributing it to the $-2/-1$ charge transition; we confirmed that the third release stage of $V_{Si}$ is indeed a transformation into ($V_{CCSi}$) and proposed a mechanism that channels through the very rare $+1$ charge state. This last application was possible only by using GW calculations for the reaction saddle point.

ACKNOWLEDGMENTS

We acknowledge useful discussions with M. Bockstedte, J.-P. Crocombette, and Y. Limoge. The present calculations were based on the ABINIT code (Ref. 29) for the GW calculations and on the QUANTUM-ESPRESSO code (Ref. 30) for the CI-NEB paths. This work was performed using HPC resources from GENCI-CINES and GENCI-CCRT (Grant No. 2010-gen6018).