Quasiparticle Calculations for Point Defects at Semiconductor Surfaces

Arno Schindlmayr[1,2] and Matthias Scheffler[2]

[1] Institut für Festkörperforschung, Forschungszentrum Jülich, 52425 Jülich, Germany
A.Schindlmayr@fz-juelich.de
[2] Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4–6, 14195 Berlin-Dahlem, Germany
scheffler@fhi-berlin.mpg.de

Abstract. We present a quantitative parameter-free method for calculating defect states and charge-transition levels of point defects in semiconductors. It combines the strength of density-functional theory for ground-state total energies with quasiparticle corrections to the excitation spectrum obtained from many-body perturbation theory. The latter is implemented within the G_0W_0 approximation, in which the electronic self-energy is constructed non-self-consistently from the Green's function of the underlying Kohn–Sham system. The method is general and applicable to arbitrary bulk or surface defects. As an example we consider anion vacancies at the (110) surfaces of III–V semiconductors. Relative to the Kohn–Sham eigenvalues in the local-density approximation, the quasiparticle corrections open the fundamental band gap and raise the position of defect states inside the gap. As a consequence, the charge-transition levels are also pushed to higher energies, leading to close agreement with the available experimental data.

1 Introduction

The electric properties of semiconductors, and hence their applicability in electronic devices, are to a large degree governed by defects that are either intrinsic or incidentally or intentionally introduced impurities. Considerable efforts, therefore, focus on determining the factors that lead to the formation of point defects and their influence on a material's electric properties.

The progress in the understanding of the atomic structure of point defects at cleaved III–V semiconductor surfaces, which serve as an illustration in this work, was recently reviewed by *Ebert* [1]. The possibility to image individual defects using scanning tunneling microscopy (STM) with atomic resolution, in particular, has yielded a wealth of data, but as STM provides a somewhat distorted picture of the electronic states close to the Fermi energy, these results cannot (and should not) be identified directly with the atomic geometry. Electronic-structure calculations have hence turned out to be an indispensable tool for the interpretation of the experimental findings. A deeper discussion of this point together with an example for the practical structure determination of a semiconductor surface is given in [2]. For point

D. A. Drabold, S. K. Estreicher (Eds.): Theory of Defects in Semiconductors,
Topics Appl. Physics **104**, 165–192 (2007)

defects at the (110) surfaces of III–V semiconductors several calculations were reported [3–8]. They are based on density-functional theory [9], where the exchange-correlation energy is typically treated in the local-density approximation (LDA) [10] or generalized gradient approximations (GGA) [11]. The agreement with experimental STM data appears to be very good. For example, the enhanced contrast of the empty p_z orbitals of the two Ga atoms nearest to an anion vacancy in p-type GaAs(110) observed under positive bias and initially interpreted as an outward relaxation [12] could thus be understood to result, instead, from a downward local band bending accompanied by an *inward* relaxation [3]. The band bending itself is caused by the positive charge of the defect. Another controversy centers on the lateral relaxation of the positively charged anion vacancy. While STM images show a density of states preserving the mirror symmetry of the surface at the defect site [12], early theoretical studies of the lattice geometry based on total-energy minimization produced conflicting evidence for [3] and against [4, 5] a possible breaking of the mirror symmetry. Well-converged electronic-structure calculations later confirmed that the distortion is indeed asymmetric [6–8] and that the apparently symmetric STM image results from the thermally activated flip motion between two degenerate asymmetric configurations.

In contrast, theoretical predictions of the electronic properties of point defects have been less successful and still show significant quantitative deviations from experimental results. The principal quantities of interest are the location of defect states in the fundamental band gap as well as the charge-transition levels. We will carefully distinguish in this chapter between these two quantities: "defect states" or "defect levels" on the one hand and "charge-transition levels" on the other. The former are part of the electronic structure and can, in principle, be probed by photoemission spectroscopy, although standard spectroscopic techniques are often not applicable due to the low density of the surface defects. The Franck–Condon principle is typically well justified, because the rearrangement of the atoms happens on a much slower timescale than the photoemission process. Nevertheless, the coupling of the electrons to the lattice may be visible in the linewidths and lineshapes. The defect levels thus contain the full *electronic* relaxation in response to the created hole in direct photoemission or the injected extra electron in inverse photoemission, but no atomic relaxation. Although investigations of electronic properties frequently rely on the Kohn–Sham eigenvalues from density-functional theory, these only provide a first approximation to the true band structure, and quantitative deviations from experimental results must be expected. In particular, for many materials the eigenvalue gap both in the LDA and the GGA underestimates the fundamental band gap significantly. Likewise, the position of defect states in the gap cannot be determined without systematic errors. With these words of warning we note, however, that the Kohn–Sham eigenvalues constitute a good and well-justified starting point for calculating band structures and defect states [13]. Therefore, a

perturbative approach starting from the Kohn–Sham eigenvalue spectrum is indeed appropriate.

While the measurement of the defect levels (see the discussion on the page previous of this Chapter) probes the geometry before the electron is added or removed, the charge-transition levels are thermodynamic quantities and specify the values of the Fermi energy where the stable charge state of the defect changes. Therefore, the charge-transition levels are affected noticeably by the atomic relaxation taking place upon the addition or removal of an electron. A quantitative analysis must hence be able to accurately compare the formation energies of competing configurations with different numbers of electrons. Density-functional theory at the level of the LDA or GGA is capable of giving a good account of the atomic geometries for many materials. A critical feature is the nonlinearity of the exchange-correlation functional. The pseudopotential approximation, which effectively removes the inner shells from the calculation by modeling their interaction with the valence electrons in terms of a modified potential, linearizes the core–valence interaction and thus does not treat this contribution correctly. In particular, freezing the d electrons in the core of a pseudopotential leads to poor lattice constants and a distorted electronic structure for some III–V semiconductors, such as GaN, where the Ga $3d$ states resonate strongly with the N $2s$ states [14]. For GaAs and InP this is a lesser problem, because the cation d states are energetically well below the anion $2s$ states and thus are relatively inert. As a result, the LDA and the GGA yield only small deviations from the experimental lattice constants and provide a good starting point for quasiparticle band-structure calculations. However, when competing configurations with a different number of electrons are compared, the relevant energy differences lack the required quantitative accuracy. As we will show in more detail below, the reason is that these jellium-based approximations of the exchange-correlation energy ignore important features of the exact functional, such as the discontinuity of the exchange-correlation potential with respect to a change in the particle number. As a consequence, previous calculations of charge-transition levels based on the LDA exhibit systematic errors, for example for anion vacancies at InP(110) [6].

As an alternative approach to the electronic structure of point defects, we employ techniques adapted from many-body perturbation theory that we have found to be very fruitful in the past [15]. Exchange and correlation effects are here described by a nonlocal and frequency-dependent self-energy operator. The solutions of the ensuing nonlinear eigenvalue equations have a rigorous interpretation as excitation energies and can be identified with the electronic band structure. We discuss the calculation of defect states as well as charge-transition levels within this framework and show that the results improve significantly upon earlier values obtained from the Kohn–Sham scheme in the LDA. As an example we consider anion vacancies at GaAs(110) and InP(110), but the method is general and can also be applied to other defects at surfaces as well as in the bulk. The (110) surfaces are not only

the natural cleavage planes of III–V semiconductors, but they have several characteristics that make them particularly interesting for defect studies. As no surface states exist inside the fundamental band gap [16, 17], the Fermi energy of a system that is clean and free from intrinsic defects is not pinned but controlled by the doping of the crystal. Only imperfections, such as anion vacancies or antisite defects, can introduce gap states and pin the Fermi energy at the surface. STM can probe filled and empty surface states by reversing the bias voltage [12], and both the GaAs(110) and the InP(110) surface are well characterized experimentally.

This Chapter is organized as follows. We start in Sect. 2 by reviewing the computational methods. In Sect. 3 we then explain the physics of the defect-free GaAs(110) and InP(110) surfaces. The calculation of defect levels is discussed in Sect. 4 and that of charge-transition levels in Sect. 5, together with a comparison with the available theoretical and experimental data. Finally, Sect. 6 summarizes our conclusions. Unless explicitly indicated otherwise, we use Hartree atomic units.

2 Computational Methods

A quantitative analysis of the electronic properties of point defects requires computational schemes that describe not only the ground state but also the excitation spectrum. While density-functional theory with state-of-the-art exchange-correlation functionals can be used to determine ground-state atomic geometries, many-body perturbation theory is the method of choice for excited states. In this work we take the Kohn–Sham eigenvalues in the LDA as a first estimate and then apply the G_0W_0 approximation for the electronic self-energy as a perturbative correction. The latter provides a good account of the discontinuity as well as other shortcomings of the LDA. For this reason we first review both schemes, emphasizing their strengths as well as limitations.

2.1 Density-Functional Theory

Density-functional theory is based on the Hohenberg–Kohn theorem [9], which observes that the total energy $E_{N,0}$ of a system of N interacting electrons in an external potential $V_{\text{ext}}(\boldsymbol{r})$ is uniquely determined by the ground-state electron density $n_N(\boldsymbol{r})$. While the Hohenberg–Kohn theorem itself makes no statement about the mathematical form of this functional, it has inspired algorithms that exploit the reduced number of degrees of freedom compared to a treatment based on many-particle wavefunctions. In the Kohn–Sham scheme [10], which underlies all practical implementations, the

density is constructed from the orbitals of an auxiliary noninteracting system according to

$$n_N(\boldsymbol{r}) = 2\sum_{j=1}^{\infty} f_{N,j}\,|\varphi_{N,j}(\boldsymbol{r})|^2\ . \tag{1}$$

The occupation numbers $f_{N,j}$ are given by the Fermi distribution; at zero temperature they equal one for states below the Fermi energy and zero for states above. The factor 2 stems from the spin summation. Here we only consider nonmagnetic systems and thus assume two degenerate spin channels throughout, although the formalism can easily be generalized if necessary. The energy functional is decomposed as

$$E_{N,0} \equiv E[n_N] = T_\mathrm{s}[n_N] + \int V_\mathrm{ext}(\boldsymbol{r}) n_N(\boldsymbol{r})\,\mathrm{d}^3 r + E_\mathrm{H}[n_N] + E_\mathrm{xc}[n_N]\,, \tag{2}$$

where $T_\mathrm{s}[n_N]$ is the kinetic energy of the auxiliary noninteracting system and $E_\mathrm{H}[n_N]$ the Hartree energy. The last term incorporates all remaining exchange and correlation contributions and is not known exactly. In practical implementations it must be approximated, for example by the LDA, which replaces the exchange-correlation energy $E_\mathrm{xc}[n_N]$ by that of a homogeneous electron gas with the same local density [10]. A variational analysis finally shows that the total energy is minimized if the single-particle orbitals satisfy

$$\left[-\tfrac{1}{2}\nabla^2 + V_\mathrm{ext}(\boldsymbol{r}) + V_\mathrm{H}([n_N];\boldsymbol{r}) + V_\mathrm{xc}([n_N];\boldsymbol{r})\right]\varphi_{N,j}(\boldsymbol{r}) = \varepsilon^\mathrm{KS}_{N,j}\varphi_{N,j}(\boldsymbol{r})\,. \tag{3}$$

The Hartree potential $V_\mathrm{H}([n_N];\boldsymbol{r})$ and the exchange-correlation potential $V_\mathrm{xc}([n_N];\boldsymbol{r})$ are defined as functional derivatives of the corresponding energy terms with respect to the density. The eigenvalues $\varepsilon^\mathrm{KS}_{N,j}$ are Lagrange parameters that enforce the normalization of the orbitals.

We use the Kohn–Sham scheme to determine the equilibrium geometry of clean surfaces and surface defects by relaxing the atomic coordinates and allowing the system to explore its energetically most favorable configuration. Although the Kohn–Sham eigenvalues differ from the true excitation energies and constitute only an approximation to the quasiparticle band structure, they are often numerically close in practice. This follows from the transition-state theorem of *Slater* [18] and *Janak* [19] and allows a correction within perturbation theory. Here, we are interested in the position of the defect state, which may be occupied or unoccupied, relative to the surface valence-band maximum. As the defect state is separated from the valence-band maximum by a finite energy difference, the location obtained from the Kohn–Sham eigenvalue spectrum contains two sources of errors: In addition to the chosen approximation for the exchange-correlation functional, there is another systematic error that is due to fundamental limitations of the Kohn–Sham scheme and would also be present if the exact functional was employed. In order to understand the latter, we now briefly sketch its origin.

One rigorous result is that the eigenvalue of the highest occupied Kohn–Sham state matches the corresponding quasiparticle energy [20], which is in turn equal to the ionization potential that marks the threshold for photoemission, i.e., $\varepsilon^{\mathrm{KS}}_{N,N} = E_{N,0} - E_{N-1,0}$. If the highest occupied state corresponds to the valence-band maximum, then the unoccupied defect state equals the electron affinity $\varepsilon^{\mathrm{KS}}_{N+1,N+1} = E_{N+1,0} - E_{N,0}$, because it is the next to be populated by one extra electron added to the system. Unfortunately, this is not the same as the eigenvalue $\varepsilon^{\mathrm{KS}}_{N,N+1}$ of the first unoccupied state obtained from (3). The difference is due to the fact that the exchange-correlation potential of an insulator $V_{\mathrm{xc}}([n_{N+1}];\boldsymbol{r}) = V_{\mathrm{xc}}([n_N];\boldsymbol{r}) + \Delta_{\mathrm{xc}} + O(1/N)$ with $\Delta_{\mathrm{xc}} > 0$ changes discontinuously upon addition of an extra electron [21, 22]. A similar argument can be made if the defect state is occupied. As a consequence, the Kohn–Sham eigenvalue gaps differ systematically from the gaps in the true quasiparticle band structure. The magnitude of the discontinuity is still a matter of controversy but is believed to be significant. For pure *sp*-bonded semiconductors like GaAs the LDA underestimates the experimental band gap by about 50 %. The GGA, designed only to improve the total energy, yields a very similar eigenvalue spectrum as the LDA when applied to the same atomic geometry. An entirely different construction that permits a more systematic treatment of exchange and correlation is the optimized effective potential method [23]. When evaluated to first order in the coupling constant, this approach yields the exact exchange potential, which can be used in band-structure calculations [24]. Remarkably, for many semiconductors the resulting Kohn–Sham eigenvalue gaps are very close to the true quasiparticle band gaps [25]. The significance of this observation is under debate, however, because there are indications that it may not uphold if correlation is treated on the same footing. Preliminary results suggest that the eigenvalue gaps are again close to the LDA values if correlation is included within the random-phase approximation [26, 27], but there is currently too little data to make a definite statement, and all existing calculations at this level also contain additional simplifications, for example a shape approximation for the potential in the linearized muffin-tin orbital method in [27].

The reasoning above suggests that density-functional theory is still, in principle, applicable for calculating the difference in total energy between ground-state configurations with different electron numbers, which is needed to determine the energetically most favorable charge state of a point defect. However, all jellium-based functionals lack the derivative discontinuity Δ_{xc} of the exact exchange-correlation potential. This neglect reduces the electron affinity $E_{N+1,0} - E_{N,0}$ both in the LDA and the GGA if the additional electron occupies a state separated from the valence-band maximum by a finite energy difference. For systems in contact with an electron reservoir, such as defects in solids, it hence lowers the threshold for an increase of the electron population. This is consistent with the observation that the calculated charge-transition levels for materials like InP are significantly smaller than the available values deduced from experimental measurements [6]. The *exact exchange* potential,

an implicit functional of the density defined in terms of the Kohn–Sham orbitals, includes a derivative discontinuity [13], but the latter exceeds the experimental band gap significantly and must hence be partially canceled by a correlation contribution with similar magnitude and opposite sign [25]. As the total energies are by construction close to the corresponding Hartree–Fock values, the electron affinities $E_{N+1,0} - E_{N,0}$ are grossly overestimated, and no reliable charge-transition levels can be obtained in this way.

Our implementation of density-functional theory employs the plane-wave pseudopotential method in combination with the LDA. We use the parametrization by *Perdew* and *Zunger* [28], which is in turn based on the quantum Monte-Carlo data of *Ceperley* and *Alder* for the homogeneous electron gas [29]. The norm-conserving pseudopotentials are of the fully separable Kleinman–Bylander form [30]. We choose d as the local component for all pseudopotentials except for In, where p is used instead. The Kohn–Sham wavefunctions are expanded in plane waves with a cutoff energy of 15 Ry. Our calculations are performed with the FHImd code [31, 32]. The bulk lattice constants obtained in this way, 5.55 Å for GaAs and 5.81 Å for InP in the absence of zero-point vibrations, are in good agreement with previously published data [33] and slightly smaller than the experimental values at room temperature by 1.8 % and 1.1 %, respectively [34]. We use the theoretical lattice constants in order to prevent errors resulting from a nonequilibrium unit-cell volume during the relaxation of the surface geometries.

2.2 Many-Body Perturbation Theory

Many-body perturbation theory [35] provides powerful techniques to analyze the electronic structure in the gap region, because the framework is designed specifically to give access to excited states. Quasiparticle excitations created by the addition or removal of one electron are obtained from the one-particle Green's function

$$G(\boldsymbol{r},\boldsymbol{r}';t-t') = -\mathrm{i}\langle\Psi_{N,0}|\mathcal{T}\{\hat{\psi}(\boldsymbol{r},t)\hat{\psi}^{\dagger}(\boldsymbol{r}',t')\}|\Psi_{N,0}\rangle\,, \tag{4}$$

where $|\Psi_{N,0}\rangle$ denotes the ground-state wavefunction of the interacting electron system in second quantization, $\hat{\psi}^{\dagger}(\boldsymbol{r}',t') = \exp(\mathrm{i}\hat{H}t')\hat{\psi}^{\dagger}(\boldsymbol{r}')\exp(-\mathrm{i}\hat{H}t')$ and $\hat{\psi}(\boldsymbol{r},t) = \exp(\mathrm{i}\hat{H}t)\hat{\psi}(\boldsymbol{r})\exp(-\mathrm{i}\hat{H}t)$ are the electron creation and annihilation operators in the Heisenberg picture, respectively, and the symbol $\mathcal{T}$ sorts the subsequent list of operators according to ascending time arguments from right to left with a change of sign for every pair permutation. The Green's function can be interpreted as a propagator: For $t > t'$ it describes a process in which an extra electron is added to the system at time t'. The resulting wavefunction is, in general, no eigenstate of the Hamiltonian $\hat{H}$ but a linear combination of many eigenstates $|\Psi_{N+1,j}\rangle$. Between the times t' and t each projection evolves with its own characteristic phase, and a Fourier analysis of this oscillatory behavior immediately yields the energy spectrum

$\varepsilon_{N,j} = E_{N+1,j} - E_{N,0}$ of the accessible excited states. Likewise, for $t' > t$ the Green's function describes the propagation of an extra hole between t and t', yielding the energies $\varepsilon_{N,j} = E_{N,0} - E_{N-1,j}$. In mathematical terms, we insert a complete set of eigenstates between the field operators in (4) and Fourier transform to the frequency axis. The resulting expression

$$G(\boldsymbol{r},\boldsymbol{r}';\omega) = \lim_{\eta\to+0}\sum_{j=1}^{\infty} f_{N,j}\frac{\psi_{N,j}(\boldsymbol{r})\psi_{N,j}^*(\boldsymbol{r}')}{\omega-\varepsilon_{N,j}-\mathrm{i}\eta} + \lim_{\eta\to+0}\sum_{j=1}^{\infty}(1-f_{N,j})\frac{\psi_{N,j}(\boldsymbol{r})\psi_{N,j}^*(\boldsymbol{r}')}{\omega-\varepsilon_{N,j}+\mathrm{i}\eta} \tag{5}$$

shows that the poles of the Green's function correspond directly to the quasiparticle energies $\varepsilon_{N,j}$. Significantly, these not only yield the true band structure, but the highest occupied state also equals the exact ionization potential $E_{N,0} - E_{N-1,0}$ and the lowest unoccupied state the exact electron affinity $E_{N+1,0} - E_{N,0}$. Therefore, many-body perturbation theory provides a convenient way to analyze both defect states and the energetics of charge transitions. The wavefunctions $\psi_{N,j}(\boldsymbol{r}) = \langle\Psi_{N-1,j}|\hat{\psi}(\boldsymbol{r})|\Psi_{N,0}\rangle$ for occupied and $\psi_{N,j}(\boldsymbol{r}) = \langle\Psi_{N,0}|\hat{\psi}(\boldsymbol{r})|\Psi_{N+1,j}\rangle$ for unoccupied states are obtained from the quasiparticle equations

$$\left[-\tfrac{1}{2}\nabla^2 + V_{\mathrm{ext}}(\boldsymbol{r}) + V_{\mathrm{H}}(\boldsymbol{r})\right]\psi_{N,j}(\boldsymbol{r}) + \int \Sigma_{\mathrm{xc}}(\boldsymbol{r},\boldsymbol{r}',\varepsilon_{N,j})\psi_{N,j}(\boldsymbol{r}')\,\mathrm{d}^3r' = \varepsilon_{N,j}\psi_{N,j}(\boldsymbol{r})\,. \tag{6}$$

The self-energy $\Sigma_{\mathrm{xc}}(\boldsymbol{r},\boldsymbol{r}',\varepsilon)$ incorporates all contributions from exchange and correlation processes. In contrast to the exchange-correlation potential of density-functional theory, it is nonlocal, energy dependent and has a finite imaginary part, which is proportional to the damping rate resulting from electron–electron scattering. Together with other relevant decay channels, such as scattering from phonons or impurities, this mechanism is responsible for a finite lifetime of the excitations.

For real materials the self-energy can only be treated approximately. Like the majority of practical applications, we use the G_0W_0 approximation [36]

$$\Sigma_{\mathrm{xc}}(\boldsymbol{r},\boldsymbol{r}';t-t') = \lim_{\eta\to+0}\mathrm{i}G_0(\boldsymbol{r},\boldsymbol{r}';t-t')W_0(\boldsymbol{r},\boldsymbol{r}';t-t'+\eta)\,. \tag{7}$$

The Fourier transform of $G_0(\boldsymbol{r},\boldsymbol{r}';t-t')$ on the frequency axis is constructed in analogy to (5) from the eigenstates of an appropriate mean-field system, in our case from the Kohn–Sham orbitals $\varphi_{N,j}(\boldsymbol{r})$ and eigenvalues $\varepsilon_{N,j}^{\mathrm{KS}}$. The dynamically screened Coulomb interaction $W_0(\boldsymbol{r},\boldsymbol{r}';t-t')$ can be modeled in different ways. Many implementations use plasmon-pole models, in which the frequency dependence is described by an analytic function whose parameters are determined by a combination of known sum rules and asymptotic limits [37,38]. This simplification has the advantage that it facilitates an analytic

treatment, but it ignores details of the dynamic screening processes in a material and thus constitutes a potential source of errors. Here we take the full frequency dependence of the dielectric function into account by employing the random-phase approximation

$$W_0(\boldsymbol{k},\omega) = v(\boldsymbol{k}) + v(\boldsymbol{k})P_0(\boldsymbol{k},\omega)W_0(\boldsymbol{k},\omega) \tag{8}$$

with the bare Coulomb potential $v(\boldsymbol{k}) = 4\pi/|\boldsymbol{k}|^2$ and the polarizability

$$P_0(\boldsymbol{r},\boldsymbol{r}';t-t') = -2\mathrm{i}G_0(\boldsymbol{r},\boldsymbol{r}';t-t')G_0(\boldsymbol{r}',\boldsymbol{r};t'-t)\,. \tag{9}$$

The inclusion of dynamic screening derives from the concept of quasiparticles, which comprise an electron or hole together with its surrounding polarization cloud. The composite is called a quasiparticle because it behaves in many ways like a single entity. The polarization cloud is created by the repulsive Coulomb potential and reduces the effective charge of the quasiparticle compared to that of the bare particle at its center. The G_0W_0 expression (7) constitutes the leading term in the expansion of the self-energy and is of first order in the dynamically screened interaction. Finally, we exploit the formal similarity between (6) and the Kohn–Sham equations (3) by evaluating the quasiparticle energies within first-order perturbation theory as

$$\varepsilon_{N,j} = \varepsilon^{\mathrm{KS}}_{N,j} + \langle\varphi_{N,j}|\Sigma_{\mathrm{xc}}(\varepsilon_{N,j}) - V_{\mathrm{xc}}[n_N]|\varphi_{N,j}\rangle\,. \tag{10}$$

The treatment within first-order perturbation theory is justified if the eigenvalues of the underlying Kohn–Sham system are already sufficiently close to the expected quasiparticle band structure. This is guaranteed by the transition-state theorem [18, 19]. Besides, the orbitals $\varphi_{N,j}(\boldsymbol{r})$ are usually a good approximation to the true quasiparticle wavefunctions. For the homogeneous electron gas both are plane waves and thus coincide exactly. Numerical calculations for bulk semiconductors indicate an overlap close to unity between the Kohn–Sham orbitals in the LDA and the quasiparticle wavefunctions obtained from (6) with the G_0W_0 approximation for states near the band edges [39]. Larger effects have been observed for surfaces, especially for image states, because the latter are located outside the surface in a region where the LDA potential is qualitatively wrong [40, 41]. Changes in the wavefunctions of other surface states can also be identified but have only a minor influence on the quasiparticle energies in the gap region. For GaAs(110) this has been confirmed explicitly [42].

In principle, the equations (4) to (9) could be solved self-consistently by successively updating the self-energy with the quasiparticle orbitals derived from it. This approach is appealing on formal grounds because it makes the results independent of the original mean-field approximation. In addition, the self-consistent Green's function satisfies certain sum rules, including particle-number conservation [43]. In practice, however, this procedure produces a poor excitation spectrum. The reason lies in the mathematical structure of

Hedin's equations, which describe the recipe for constructing the self-energy from the Green's function by means of functional-derivative techniques [36]. The G_0W_0 approximation is the result of a single iteration. Further iterations would not only ensure self-consistency in the Green's function but would also introduce higher-order self-energy terms, so-called vertex corrections. If the latter are neglected, then the spectral features deteriorate, as was first demonstrated for the homogeneous electron gas [44]. For bulk semiconductors self-consistency appears to lead to a gross overestimation of the fundamental band gap [45] in combination with poor spectral weights and linewidths. We hence follow the established procedure for practical applications and terminate the cycle at the G_0W_0 approximation. Of course, the final results depend on the input Green's function in this case.

Another point that has been raised in the same context concerns possible errors resulting from the pseudopotential approach, which has traditionally dominated applications of the G_0W_0 approximation. Numerical deviations must be expected because the matrix elements of the self-energy in (10) are influenced by the pseudoization of the wavefunctions – that is the neglect of core states plus appropriate smoothening of the valence states in the core region – and the treatment of the core–valence interaction in pseudopotential calculations. Indeed, a number of early all-electron implementations of the G_0W_0 approximation, based on the linearized augmented plane-wave or the linearized muffin-tin orbital method, found quasiparticle band gaps for prototype semiconductors that were significantly smaller than previously reported values and blamed the discrepancy on the pseudopotential approximation [46,47]. However, this claim was later refuted by all-electron calculations using plane waves [48,49]. The issue is currently still under debate, but there is mounting evidence that a large part of the observed discrepancy is due to insufficient convergence of the early calculations in combination with deficiencies of the linearized basis sets for the description of high-lying unoccupied states [48,50]. If these factors are properly accounted for, then the deviation is significantly reduced, and the all-electron results are again in good agreement with experimental measurements. A certain discrepancy with respect to pseudopotential calculations still remains, although the difference is small compared to that from the Kohn–Sham eigenvalues. While some errors of the pseudopotential approach can be partially suppressed, for example through a better description of the core–valence interaction [51], many others are inherent. In particular, the pseudopotential construction also entails an incorrect description of high-lying states. A careful analysis of the quantitative implications of the pseudopotential approximation is the subject of active research.

The good quantitative agreement between the G_0W_0 approximation and experimental band structures has been demonstrated for a wide range of bulk materials, including III–V semiconductors [52–54]. Due to the high computational cost in present implementations, which stems from the evaluation of the nonlocality of the propagators, their frequency dependence, and the

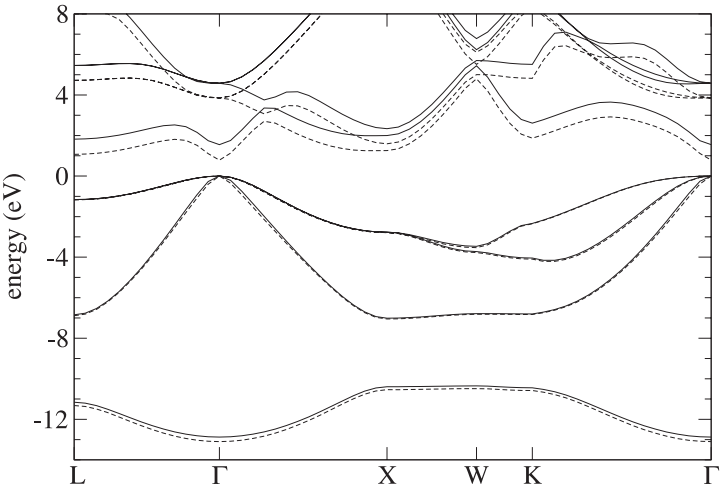

Fig. 1. Bulk band structure of GaAs. The G_0W_0 approximation (*straight lines*) opens the too small Kohn–Sham eigenvalue gap in the LDA (*dashed lines*) and is in very good agreement with the experimental band gap. The calculation is carried out at the theoretical lattice constant 5.55 Å, and the valence-band maximum is set to zero in both schemes

need to include a large number of unoccupied states, there have been relatively few applications to more complex systems so far, however. Furthermore, these often contain additional simplifications: the two available studies of the quasiparticle band structure of GaAs(110) both employed plasmon-pole models instead of the more accurate random-phase approximation [55, 56], and the only published results for InP(110) were obtained within an even more restrictive tight-binding formalism [57].

For a quasiparticle band-structure calculation the matrix elements of the self-energy in (6) must be evaluated in the frequency domain for states with a given wavevector $\boldsymbol{k}$. Therefore, most practical implementations of the G_0W_0 approximation choose a reciprocal-space representation for all propagators. The cell-periodic part of the wavefunctions is often expanded in plane waves [52, 55], although localized basis sets like Gaussian orbitals have also been used [54]. The disadvantage of the representation in reciprocal space is that the products (7) and (9) turn into numerically expensive multidimensional convolutions. Therefore, we employ a representation in real space and imaginary time [58, 59], in which the self-energy and the polarizability can be calculated by simple multiplications. The projection on wavevectors and imaginary frequencies used to solve the Dyson-type equation (8) can be done efficiently by exploiting fast Fourier transforms. The imaginary time and frequency arguments are chosen because the functions are smoother on these axes and can be sampled with fewer grid points, although they contain

exactly the same information. The physical self-energy on the real frequency axis is eventually recovered by an analytic continuation.

As an illustration of the G_0W_0 approximation we show the calculated bulk band structure of GaAs in Fig. 1. The band gap is direct and located at Γ in the center of the Brillouin zone. While the LDA underestimates the fundamental band gap and yields a Kohn–Sham eigenvalue gap of only 0.78 eV at a lattice constant of 5.55 Å, the subsequent addition of the self-energy correction opens the band gap to 1.55 eV, which is in very good agreement with the experimental value of 1.52 eV [34]. As we measure all energies relative to the valence-band maximum, we set the latter to zero and align the two sets of curves at this point. The principal effect of the G_0W_0 approximation is a rigid upward shift of the conduction bands, although the dispersion is also slightly modified, as can be seen by the reduced bandwidth at the bottom of the valence band. The band structure of InP looks very similar. In this case we obtain a Kohn–Sham eigenvalue gap of 0.76 eV at the theoretical lattice constant 5.81 Å and a quasiparticle band gap of 1.52 eV that is again close to the experimental value 1.42 eV [34]. Incidentally, the band gap depends sensitively on the lattice constant. For GaAs it decreases at a rate of $-4.07\,\mathrm{eV/Å}$ in the LDA and $-4.59\,\mathrm{eV/Å}$ in the G_0W_0 approximation when the lattice constant increases. For InP the values are $-3.13\,\mathrm{eV/Å}$ and $-3.68\,\mathrm{eV/Å}$.

3 Electronic Structure of Defect-Free Surfaces

Before focusing on defect states we first briefly discuss the electronic structure of the defect-free GaAs(110) and InP(110) surfaces. The nonpolar (110) surface, illustrated in Fig. 2, is the natural cleavage plane of the zincblende lattice, because it cuts the smallest number of bonds per unit area. Both the cations and anions in the terminating layer are threefold coordinated, and each possess one dangling bond extending into the vacuum. Due to the different electron affinities of the two species, charge is transfered from the dangling bonds of the cations to those of the anions. This charge transfer is the driving force for a structural relaxation, which consists of an outward movement of the anion atoms and a corresponding inward movement of the cation atoms [33]. As a result, the orbitals of the latter rehybridize from an sp^3 towards an energetically more favorable sp^2 bonding situation in a nearly planar environment; the empty p_z-like orbitals perpendicular to this plane are pushed to higher energies and form an unoccupied surface band. At the same time, the three bonds between the anions and their neighboring group-III atoms are rearranged at almost right angles and become more p-like in character; the nonbonding electron pairs in the fourth orbital pointing away from the surface are in turn lowered in energy and give rise to an occupied surface band. The relaxation preserves the $C_{1\mathrm{h}}$ point-group symmetry with a single mirror plane perpendicular to the $[\bar{1}10]$ direction.

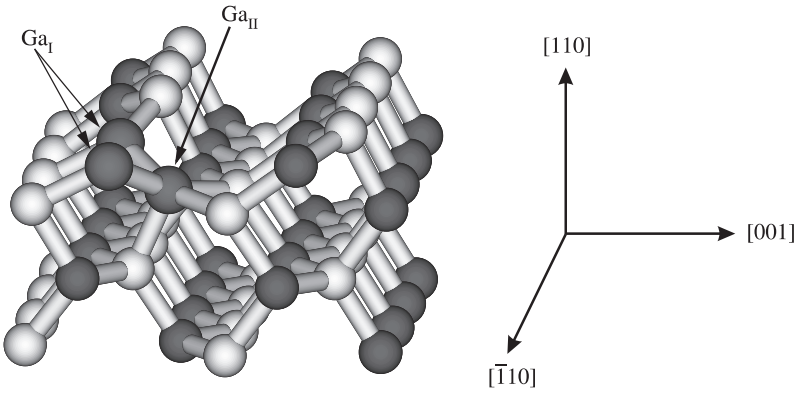

Fig. 2. Geometry of the anion vacancy at the GaAs(110) surface. The As atoms are shown in *light* and the Ga atoms in *dark gray*. The three Ga atoms nearest to the vacancy are indicated by *arrows*

In Fig. 3 we show the calculated band structure of the clean and defect-free GaAs(110) surface at the theoretical lattice constant 5.55 Å, modeled as a slab element placed in a supercell with periodic boundary conditions in all directions. The slab consists of six atomic layers, of which the top three layers are allowed to relax, while the three base layers are kept fixed at their ideal bulk positions. The dangling bonds at the bottom of the slab are passivated by pseudoatoms with noninteger nuclear charges of 0.75 and 1.25 for anion and cation termination, respectively. The slabs are separated by a vacuum region equivalent to four atomic layers. The gray-shaded regions in the figure mark the projection of the G_0W_0 bulk bands onto the two-dimensional surface Brillouin zone, shown in the inset. The dashed lines indicate the occupied and unoccupied surface bands obtained from the LDA, the straight lines are the corresponding G_0W_0 results. For the calculation of the ground-state density and the Kohn–Sham eigenvalues we used four Monkhorst–Pack $\boldsymbol{k}$-points [60] in the irreducible part of the Brillouin zone, while the self-energy was evaluated at the four high-symmetry points $\overline{\Gamma}$, $\overline{\mathrm{X}}'$, $\overline{\mathrm{M}}$, and $\overline{\mathrm{X}}$. In our implementation the $\boldsymbol{k}$-point set enters merely as the reciprocal grid of the real-space mesh used to describe the nonlocality of $G_0(\boldsymbol{r}, \boldsymbol{r}'; t - t')$ and $\Sigma_{\mathrm{xc}}(\boldsymbol{r}, \boldsymbol{r}'; t - t')$ [58]; the four selected $\boldsymbol{k}$-points correspond to a real-space mesh that extends over four surface unit cells. This is sufficient, because the correlation length is of the order of the interatomic distance [61]. The position of the Kohn–Sham surface bands relative to the corresponding bulk bands was determined by aligning the electrostatic potential in the central part of the slab with that of the bulk. We then applied the self-energy correction independently to surface and bulk bands and again chose the valence-band maximum as the energy zero. From the figure it is evident that the G_0W_0 approximation has only a small effect on the dispersion of the surface bands.

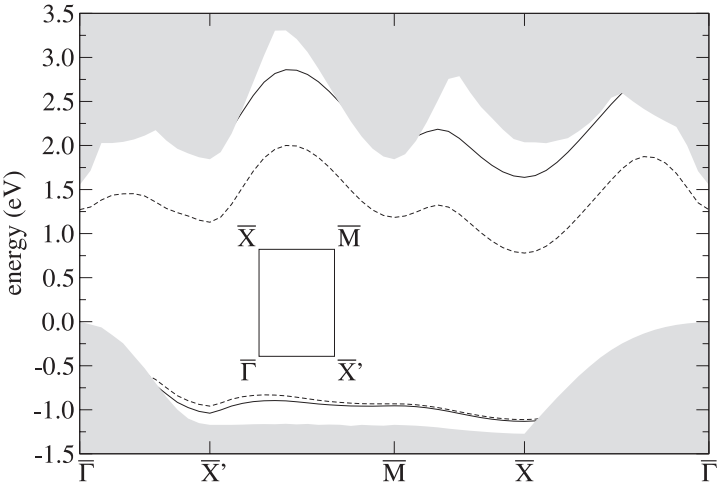

Fig. 3. Surface band structure of GaAs(110) in the LDA (*dashed lines*) and the G_0W_0 approximation (*straight lines*) at the theoretical lattice constant 5.55 Å. The *gray-shaded regions* mark the projected G_0W_0 bulk bands. The *inset* indicates the two-dimensional surface Brillouin zone

While the position of the occupied surface band relative to the valence-band edge at $\overline{\Gamma}$ remains almost unchanged, the upward shift of the unoccupied surface band (0.86 eV) slightly exceeds that of the bulk conduction bands (0.77 eV). The larger impact of the self-energy correction on the surface gap was noticed before [55, 56]; our result for the gap correction lies between the two previously published values. Small deviations of about 0.1 eV are due to differences in the implementations, for example the reliance on plasmon-pole models in [55, 56] compared to the random-phase approximation for the dynamically screened Coulomb interaction in this work. Both the occupied and the unoccupied surface bands are in close proximity to the projected bulk bands and, in fact, overlap with them in large parts of the Brillouin zone. As neither extends into the fundamental gap between the bulk valence and conduction band edges at $\overline{\Gamma}$, they do not pin the surface Fermi level.

Although prevalent in electronic-structure calculations for surfaces, the supercell approach is a drastic alteration of the system's geometry whose influence must be carefully monitored, because the occurrence of electric multipole moments may lead to artificial long-range interactions between the periodic slabs. Static dipoles, if present, can be eliminated in density-functional theory [62]. With this correction the limit of isolated slabs is quickly reached. Dynamic dipoles are always created in dielectric media, however, and contribute to the polarizability in the G_0W_0 approximation. Actually, there are two contributions that must be distinguished. The first is the dynamic polarization between the slabs, which gives rise to an additional slowly varying potential that reduces the band gap. This effect can be understood and even

quantitatively modeled in terms of classical image charges [63]; for the geometry used in this work it amounts to about 0.1 eV. The model thus allows an a-posteriori extrapolation to the limit of an isolated slab. The second contribution is the finite width of the slab, which increases the gap due to quantization effects. At present there is no obvious cure for this problem within the G_0W_0 approximation, and as the two effects counteract each other, we have not applied any partial correction to eliminate the dynamic polarization between the slabs either. In principle, both problems could be avoided by studying systems comprised of semi-infinite matter and vacuum regions. Within density-functional theory several methods have indeed been proposed for this purpose [64–67]. Their efficiency relies on the fact that a perturbation breaking the translational symmetry, such as a surface or defect, modifies the effective potential only in its immediate vicinity. In the G_0W_0 formalism this cannot be exploited to the same degree, because all propagators are explicitly nonlocal, and a much larger simulation cell must hence be taken into account. So far, only one G_0W_0 calculation for an effectively one-dimensional system, a semi-infinite jellium surface, has been reported [68].

In preparation for later applications to larger supercells, we repeated the G_0W_0 calculation with lower cutoff energies and found that the self-energy correction to the surface gap remained stable up to 10 Ry. The reason for the rapid convergence, which is well known and can be exploited to reduce the computational expense considerably, is that the kinetic energy and the electrostatic Hartree potential, the two largest contributions to the quasiparticle energies, are already included in the Kohn–Sham eigenvalues in (10); the matrix elements of the self-energy are less sensitive to the number of plane waves. Besides, we obtained essentially the same surface gap when reducing the width of the slab to four layers. For this geometry and a cutoff energy of 10 Ry, we performed test calculations in which we included up to 1049 unoccupied bands in the Green's function. With 379 bands the results are already converged within 0.02 eV, sufficient for our purpose.

4 Defect States

The geometry of the anion vacancy is illustrated in Fig. 2 for the GaAs(110) surface. The removal of the As atom leaves each of the three Ga atoms surrounding the vacancy with a dangling bond. As a consequence, the two Ga_I atoms in the first layer move downwards while the Ga_{II} atom in the second layer moves upwards and forms two new bonds with the Ga_I atoms across the void. Its coordination number thus increases from four to five, while the threefold coordination of the Ga_I atoms remains unchanged. The relaxation preserves the C_{1h} point-group symmetry, except in the positive charge state where an asymmetric distortion that pushes the unoccupied defect level in the band gap to higher energies is more favorable [6, 8]. We find that the distortion lowers the total energy by 0.17 eV for GaAs and 0.11 eV for InP.

The anion vacancy gives rise to three electronic states, all localized at the Ga_I–Ga_{II} bond pair. They are labeled as 1a′, 1a″ and 2a′, where a′ denotes states that are even with respect to the mirror plane and a″ denotes states that are odd. Although the asymmetric relaxation in the positive charge state destroys this symmetry and deforms the orbitals slightly, it leaves the order of the states intact, and we continue to use the same notation for simplicity. The 1a′ state is located several eV below the valence-band maximum and thus always filled with two electrons, while the 2a′ state is too high in energy to become populated. Only the 1a″ state falls inside the fundamental band gap. Depending on the doping, it can be occupied either by zero, one or two electrons, which corresponds to the positive, neutral or negative charge state, respectively. It is important to note that the charge state influences the defect geometry, as the Ga_I–Ga_{II} bonds contract with increasing electron occupancy, reflecting the bonding character of the 1a″ state. In the following we examine the position of this defect level for a given charge state; the question of which charge state is preferred under specific conditions, such as doping, is answered in the next section.

For the density-functional calculations we choose a supercell consisting of 2×4 surface unit cells and six atomic layers. For charged systems we include a uniform charge density with opposite sign in order to compensate the extra electron or hole and restore overall neutrality in the supercell. Instead of a well-defined defect state, the supercell periodicity gives rise to an artificial dispersion as illustrated in Fig. 4. At the surface of each slab the defects form a rectangular grid whose lattice parameters a_x and a_y equal the dimensions of the supercell. Since the 1a″ state is odd with respect to the mirror plane, one can regard it as a p orbital that exhibits π-type bonding along the [001] and σ-type bonding along the $[\bar{1}10]$ direction. We hence consider a tight-binding model

$$\begin{aligned}\varepsilon_{\boldsymbol{k}}^{\mathrm{KS}} = {} & \varepsilon_{1\mathrm{a}''}^{\mathrm{KS}} + 2V_{1\pi}^{\mathrm{KS}} \cos(k_x a_x) + 2V_{1\sigma}^{\mathrm{KS}} \cos(k_y a_y) \\ & + V_2^{\mathrm{KS}} \cos(k_x a_x) \cos(k_x a_x) + 2V_{3\pi}^{\mathrm{KS}} \cos(2k_x a_x) + 2V_{3\sigma}^{\mathrm{KS}} \cos(2k_y a_y)\end{aligned} \tag{11}$$

with parameters fitted to the calculated Kohn–Sham band, where $\varepsilon_{1\mathrm{a}''}^{\mathrm{KS}}$ equals the eigenvalue of a single defect and the other parameters have the meaning of hopping integrals. The above expression includes interactions up to third-nearest neighbors and reproduces the dispersion with a correlation coefficient close to 0.999 9 for all systems under consideration.

As the G_0W_0 formalism involves nonlocal propagators, the amount of data that must be processed grows rapidly with the system size. In order to limit the computational expense we use a smaller (2×2) supercell and four atomic layers instead of the (2×4) cell to determine the self-energy correction of the 1a″ state. The stronger defect–defect interaction along the $[\bar{1}10]$ direction increases the dispersion but does not change it qualitatively. As the presence of the defect does not modify the range of the nonlocal propagators appre-

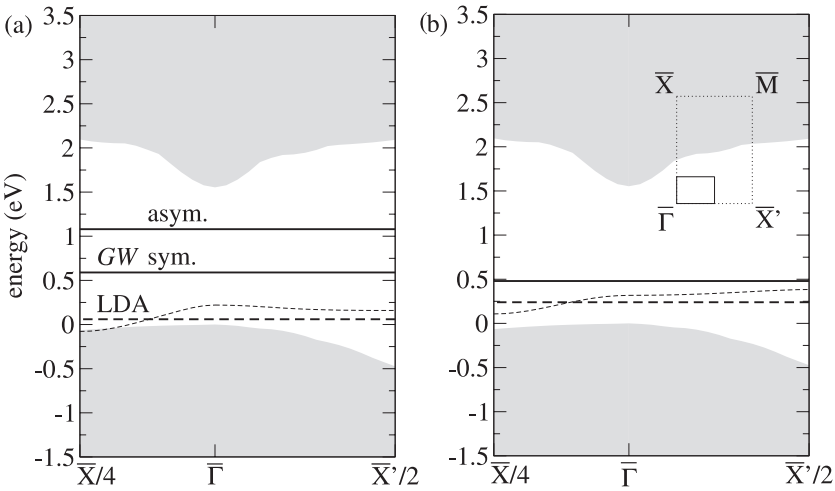

Fig. 4. Calculated 1a″ defect level of the As vacancy at GaAs(110) in the (**a**) positive and (**b**) negative charge state (*dashed line*: LDA; *solid line*: G_0W_0). The artificial dispersion in the LDA is due to the supercell periodicity; the *horizontal lines* mark the actual defect level $\varepsilon^{\mathrm{KS}}_{1a''}$ and the corresponding G_0W_0 result. As the calculations in (**a**) refer to the constrained symmetric relaxation, the additional shift due to the asymmetric distortion is shown separately. The *inset* in (**b**) indicates the downfolded Brillouin zone for the (2×4) supercell

ciably, we use one $\boldsymbol{k}$-point and 1500 unoccupied bands for the construction of the Green's function G_0, which corresponds to the same Brillouin-zone sampling as the four $\boldsymbol{k}$-points and 379 bands that we found satisfactory for the (1×1) unit cell of the defect-free surface. For the positively charged As vacancy at GaAs(110) we evaluated the self-energy correction in the entire Brillouin zone [15]. By relating the calculated quasiparticle dispersion to a tight-binding model equivalent to (11), we found very similar values for the parameters $V_{3\pi}$ and $V_{3\sigma}$ as in the LDA, which implies that the influence of the third-nearest neighbors on the self-energy correction is negligible. This observation can be exploited as follows. At $\boldsymbol{k}' = (2\pi/4)(1/a_x, 1/a_y, 0)$ the contributions from the first- and second-nearest neighbors vanish, so that the Kohn–Sham eigenvalue is given by

$$\varepsilon^{\mathrm{KS}}_{\boldsymbol{k}'} = \varepsilon^{\mathrm{KS}}_{1a''} - 2V^{\mathrm{KS}}_{3\pi} - 2V^{\mathrm{KS}}_{3\sigma}\,, \tag{12}$$

and the corresponding quasiparticle energy by

$$\varepsilon_{\boldsymbol{k}'} = \varepsilon_{1a''} - 2V_{3\pi} - 2V_{3\sigma}\,. \tag{13}$$

If the surviving hopping integrals on the right-hand sides coincide, then the identity $\varepsilon_{1a''} - \varepsilon^{\mathrm{KS}}_{1a''} = \varepsilon_{\boldsymbol{k}'} - \varepsilon^{\mathrm{KS}}_{\boldsymbol{k}'}$ holds. On the other hand, the self-energy correction is defined through the relation

$$\varepsilon_{\boldsymbol{k}'} - \varepsilon^{\mathrm{KS}}_{\boldsymbol{k}'} = \langle \varphi^{\mathrm{KS}}_{\boldsymbol{k}'} | \Sigma(\varepsilon^{\mathrm{KS}}_{\boldsymbol{k}'}) - V_{\mathrm{xc}} | \varphi^{\mathrm{KS}}_{\boldsymbol{k}'} \rangle \,. \tag{14}$$

Instead of a scan over the whole Brillouin zone, we can hence determine the self-energy correction for an isolated defect from a single calculation at $\boldsymbol{k}'$. The quasiparticle results are obtained by adding this correction to the Kohn–Sham value $\varepsilon^{\mathrm{KS}}_{1a''}$ of the larger (2×4) supercell, where the position of the latter relative to the valence-band maximum can be established more accurately. We estimate that the uncertainty of the final quasiparticle energies, which results from the discrete $\boldsymbol{k}$-point sampling, the approximate treatment of the core–valence interaction in the pseudopotential approach, the finite size of the supercell, and other convergence factors amounts to 0.1 eV to 0.2 eV.

We performed explicit G_0W_0 calculations for the positive and negative charge states, in the first case using a constrained symmetric relaxation. The results for GaAs(110) are displayed in Fig. 4. For the neutral anion vacancy the supercell contains an odd number of electrons; in combination with the requirement of spin degeneracy this leads to fractional occupation numbers in each spin channel. At present we cannot treat such systems. For the positive charge state with the proper asymmetric distortion the defect level in the G_0W_0 approximation is not calculated directly but deduced as follows. As the lowest unoccupied state, the 1a″ defect level equals the electron affinity, i.e., $\varepsilon_{1a''} = E^{\mathrm{vac}}(0, Q^{\mathrm{asym}}_+) - E^{\mathrm{vac}}(+, Q^{\mathrm{asym}}_+)$, where $E^{\mathrm{vac}}(q, Q)$ denotes the total energy of a surface featuring an anion vacancy with the actual electron population $q \in \{+, 0, -\}$ and a geometric structure optimized for the charge state $Q \in \{Q^{\mathrm{asym}}_+, Q^{\mathrm{sym}}_+, Q_0, Q_-\}$. This expression can be rewritten as

$$\begin{aligned}\varepsilon_{1a''} = {} & \left[E^{\mathrm{vac}}(0, Q^{\mathrm{asym}}_+) - E^{\mathrm{vac}}(0, Q^{\mathrm{sym}}_+)\right] \\ & + \left[E^{\mathrm{vac}}(0, Q^{\mathrm{sym}}_+) - E^{\mathrm{vac}}(+, Q^{\mathrm{sym}}_+)\right] \\ & + \left[E^{\mathrm{vac}}(+, Q^{\mathrm{sym}}_+) - E^{\mathrm{vac}}(+, Q^{\mathrm{asym}}_+)\right]\end{aligned} \tag{15}$$

by adding and subtracting intermediate configurations. The term in the second line on the right-hand side equals the quasiparticle energy for the corresponding symmetric relaxation, which can be calculated with less computational cost by exploiting the C_{1h} point-group symmetry. The other two terms are simple total-energy differences between the symmetric and the asymmetric geometry for a constant number of electrons; both are positive and can be obtained from density-functional theory.

The calculated defect levels are summarized in Table 1 for GaAs(110) and Table 2 for InP(110). Our LDA results are consistent with most of the previously published calculations [3, 5, 7]. A notable exception are the values by *Zhang* and *Zunger* [3] for the negative charge state of the As vacancy at GaAs(110) and for the constrained symmetric geometry of the positive charge

Table 1. Calculated defect levels for the As vacancy at GaAs(110) in eV relative to the valence-band maximum. For the positive charge state the first column refers to the asymmetric and the second to the constrained symmetric relaxation. The quasiparticle band gap of 1.55 eV in this work, calculated at the theoretical lattice constant 5.55 Å, is close to the experimental value of 1.52 eV

Charge state	(+1)		(0)	(−1)
LDA (this work)	0.70	(0.06)	0.13	0.24
LDA [3]	0.73	(0.41)		0.5
LDA [5]		(0.06)	0.23	0.24
G_0W_0 (this work)	1.08	(0.59)		0.48

Table 2. Calculated defect levels for the P vacancy at InP(110) in eV relative to the valence-band maximum. For the positive charge state the first column refers to the asymmetric and the second to the constrained symmetric relaxation. The quasiparticle band gap of 1.52 eV in this work, calculated at the theoretical lattice constant 5.81 Å, slightly exceeds the experimental value of 1.42 eV

Charge state	(+1)		(0)	(−1)
LDA (this work)	0.89	(0.39)	0.49	0.60
LDA [7]		(0.326)	0.479	0.580
G_0W_0 (this work)	1.36	(0.91)		1.01

state in Table 1. After a full relaxation of the asymmetric distortion in the latter case, the discrepancy vanishes, however. The agreement with the results of *Kim* and *Chelikowsky* [5] for the same system is very good, except for a difference of 0.1 eV for the neutral charge state. The origin of these deviations cannot be traced, because both groups of authors give very little information about their computational details, and the size of the corresponding Kohn–Sham eigenvalue gap is not stated. For all configurations with symmetric geometries the defect level moves steadily upwards with increasing electron population. The asymmetric distortion has a large effect on the unoccupied defect level and pushes it to significantly higher energies. Concomitant with the opening of the fundamental band gap, the G_0W_0 approximation adds a positive quasiparticle correction to all defect levels and predicts larger values than the LDA for all charge states. The size of the self-energy shift depends both on the charge state and the geometry, although the differences are of the same order as the error bar of the calculation. We note that our numerical values differ slightly from those reported earlier in [15]. The discrepancy is due to an improved description of the anisotropic screening in the slab geometry in this work but lies within the estimated overall error bar of the calculation.

For the P vacancy at InP(110) we obtain a similar picture, although the defect levels are at higher energies both in the LDA and the G_0W_0 approximation.

Unfortunately, due to experimental difficulties, no direct measurements of the defect states by photoemission are available. In this situation one can only try to extract values from indirect methods like surface photovoltage imaging with STM. One study using this technique claimed a value of $0.62 \pm 0.04\,\mathrm{eV}$ for the As vacancy in the positive charge state, based on the known position of the sample's Fermi energy 0.09 eV above the valence-band maximum and a local band bending of 0.53 eV [69]. An STM measurement on a different sample found a band bending of only 0.1 eV, however [12]. The origin of this discrepancy is unclear but points to a strong influence of experimental conditions and/or sample quality. Because of this uncertainty, no experimental values are included in the tables.

5 Charge-Transition Levels

The occurrence of different charge states is a direct consequence of the position of the defect level inside the fundamental band gap: in p-doped materials the defect state lies above the Fermi energy and is hence depopulated, whereas it is fully occupied in n-doped materials with a higher Fermi energy close to the conduction-band edge. Indeed, a charge state of $(+1)$ has been confirmed experimentally for anion vacancies at p-GaAs(110) [70] and p-InP(110) [71] and a charge state of (-1) at n-GaAs(110) [72] and n-InP(110) [73]. In a theoretical treatment the stable charge state can be identified by comparing the different formation energies

$$E^{\mathrm{form}}(q, \mu_{\mathrm{A}}, \varepsilon_{\mathrm{F}}) = E^{\mathrm{vac}}(q, Q_q) + \mu_{\mathrm{A}} + q\varepsilon_{\mathrm{F}} - E^{\mathrm{surf}}\,, \tag{16}$$

where we use the same notation for the total energy of the vacancy as in the previous section. The chemical potential μ_{A} of the anion atoms is controlled by the partial pressure and temperature, ε_{F} denotes the Fermi level and E^{surf} is the total energy of the defect-free surface. The qualitative behavior of the formation energies is illustrated in Fig. 5: due to their different slopes the stable charge state with the lowest formation energy changes from positive to negative as the Fermi energy varies between the valence-band maximum and the conduction-band minimum. In between there may be a region where the neutral vacancy is stable. The charge-transition levels are defined as the values of the Fermi energy where the curves intersect and the stable charge state changes. They are given explicitly by $\varepsilon^{+/0} = E^{\mathrm{vac}}(0, Q_0) - E^{\mathrm{vac}}(+, Q_+)$ for the transition from $q = +1$ to $q = 0$ and $\varepsilon^{0/-} = E^{\mathrm{vac}}(-, Q_-) - E^{\mathrm{vac}}(0, Q_0)$ for the transition from $q = 0$ to $q = -1$.

All quantities in (16) are ground-state energies and can, in principle, be calculated within density-functional theory. As explained above, however, the parametrizations most commonly used in practical implementations lack

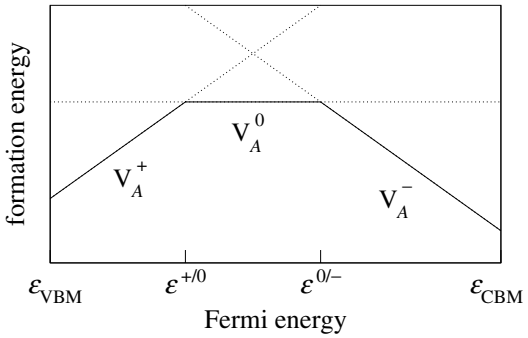

Fig. 5. Qualitative behavior of the formation energies for anion vacancies in the positive (V_A^+), neutral (V_A^0), and negative (V_A^-) charge state. The Fermi energy is limited by the valence-band maximum and the conduction-band minimum. The charge-transition levels $\varepsilon^{+/0}$ and $\varepsilon^{0/-}$ mark the values of the Fermi energy where the stable charge state with the lowest formation energy changes

essential properties of the exact exchange-correlation functional, so that the charge-transition levels obtained in this way suffer from systematic errors. For a more accurate quantitative description we use the same trick as in (15) and rewrite $\varepsilon^{+/0}$ as

$$\varepsilon^{+/0} = [E^{\mathrm{vac}}(0, Q_0) - E^{\mathrm{vac}}(0, Q_+)] + [E^{\mathrm{vac}}(0, Q_+) - E^{\mathrm{vac}}(+, Q_+)] \quad (17)$$

by adding and subtracting the total energy $E^{\mathrm{vac}}(0, Q_+)$ of a configuration with the atomic geometry of the positively charged vacancy but with one extra electron. In this way the charge-transition level is naturally decomposed into two distinct contributions. The first term on the right-hand side is purely structural and describes the relaxation energy of the neutral system from the atomic structure optimized for the positive charge state to its own equilibrium geometry. It is always negative and can be calculated within density-functional theory. We obtain $-0.59\,\mathrm{eV}$ for GaAs(110) and $-0.54\,\mathrm{eV}$ for InP(110) when taking the asymmetric distortion into account. The second term on the right-hand side is purely electronic and equals the electron affinity of the positively charged vacancy, which in a many-body framework corresponds to the lowest unoccupied state, i.e., the empty $1a''$ defect level in the band gap. This was already calculated within the G_0W_0 approximation in the preceding section and can be taken from Tables 1 and 2. In the same spirit we rewrite $\varepsilon^{0/-}$ as

$$\varepsilon^{0/-} = [E^{\mathrm{vac}}(-, Q_-) - E^{\mathrm{vac}}(0, Q_-)] + [E^{\mathrm{vac}}(0, Q_-) - E^{\mathrm{vac}}(0, Q_0)] \; . \quad (18)$$

The first term now equals the ionization potential of the negatively charged vacancy, corresponding to the highest occupied quasiparticle state. Again this is the $1a''$ defect level, which in this case is filled with two electrons, and the G_0W_0 results can be taken from the tables in the previous section.

The second term is the energy difference between the electrically neutral system in the atomic structure optimized for the positive charge state and its own relaxed geometry. This contribution is always positive and can be obtained from density-functional theory. We obtain 0.12 eV for GaAs(110) and 0.08 eV for InP(110). The structural component is much smaller than for $\varepsilon^{+/0}$ because there is no symmetry-breaking distortion in this case, only a minor reduction of the Ga–Ga bond length across the vacancy.

Incidentally, the charge-transition levels within the LDA, which are usually obtained according to the definitions $\varepsilon^{+/0} = E^{\mathrm{vac}}(0, Q_0) - E^{\mathrm{vac}}(+, Q_+)$ and $\varepsilon^{0/-} = E^{\mathrm{vac}}(-, Q_-) - E^{\mathrm{vac}}(0, Q_0)$ by evaluating the total-energy differences directly, can also be decomposed into structural and electronic energy contributions. The latter are *not* given by the Kohn–Sham eigenvalues in Tables 1 and 2, however. Instead, they must be calculated with the help of the transition-state theorem [18] from intermediate configurations with half-integer occupation numbers of 1/2 and 3/2 electrons, respectively.

Table 3. Calculated charge-transition levels for the As vacancy at GaAs(110) in eV relative to the valence-band maximum. For $\varepsilon^{+/0}$ the first column refers to the asymmetric and the second to the constrained symmetric relaxation

Transition level	$\varepsilon^{+/0}$		$\varepsilon^{0/-}$
LDA (this work)	0.24	(0.07)	0.15
LDA [3]	0.32		0.4
LDA [5]		(0.10)	0.24
G_0W_0 (this work)	0.49	(0.32)	0.60

In Table 3 we summarize the results for the As vacancy at GaAs(110). In contrast to earlier studies [3, 5] that found a small energy window in which the neutral charge state is stable, our own calculation at the level of the LDA predicts $\varepsilon^{+/0} > \varepsilon^{0/-}$ if the correct asymmetric distortion is taken into account and hence a direct transition from the positive to the negative charge state, but the small energy difference is within the uncertainty of the calculation. The G_0W_0 approximation, on the other hand, reverses this ordering and simultaneously moves all charge-transition levels to higher energies. The values for the constrained symmetric relaxation are merely shown for the purpose of comparison with earlier work. The deviations from earlier LDA studies, especially by *Zhang* and *Zunger* [3], are related to differences of similar magnitude in the defect states, which were already mentioned above.

The results for the P vacancy at InP(110) are given in Table 4. For this material an indirect experimental measurement of $\varepsilon^{+/0}$, obtained with a combination of scanning tunneling microscopy and photoelectron spectroscopy, is available [6]. Consistent with previous studies [6, 7], we find that the LDA significantly underestimates the experimentally deduced value. The larger

Table 4. Calculated charge-transition levels for the P vacancy at InP(110)

Transition level	$\varepsilon^{+/0}$		$\varepsilon^{0/-}$
LDA (this work)	0.47	(0.39)	0.54
LDA [6]	0.52	(0.45)	
LDA [7]	0.388		0.576
G_0W_0 (this work)	0.82	(0.79)	1.09
Expt. [6]	0.75 ± 0.1		

G_0W_0 result, however, lies within the error bar of the experimental measurement. We take this as a confirmation of our approach and an indicator that the other defect states and charge-transition levels calculated within the same framework are also meaningful. Nevertheless, further calculations for different systems are necessary to establish the general validity of this scheme.

6 Summary

We have presented a parameter-free method for calculating defect states and charge-transition levels of point defects in semiconductors. Compared to previous studies that extracted these quantities directly from the self-consistent iteration of the Kohn–Sham equations, it apparently corrects important errors that are inherent in all jellium-based exchange-correlation functionals and employs a separation of structural and electronic energy contributions. While the former are accurately obtainable in density-functional theory, we use many-body perturbation theory and the G_0W_0 approximation for the self-energy to calculate the latter with proper quasiparticle corrections. The scheme is general and can be applied to arbitrary bulk or surface defects. As an example we examined the electronic structure of anion vacancies at the (110) surfaces of III–V semiconductors. For the As vacancy at GaAs(110) our calculation indicates that all three charge states including the neutral configuration are stable, in contrast to the LDA that predicts a direct transition from the positive to the negative charge state. Due to a general lack of experimental data, a direct comparison between theoretical and experimental values is only possible for the charge-transition level $\varepsilon^{+/0}$ of the P vacancy at InP(110). In this case our calculation is in good agreement with the experimentally deduced result and constitutes a clear improvement over previous LDA treatments. Nevertheless, besides an improvement of experimental techniques, further developments in theoretical and computational procedures are highly desirable. Our method only opens the door to novel approaches; with present implementations the results are derived at a cost that may be infeasible for more complex systems, and the numerical uncertainty of 0.1 eV to 0.2 eV is often of the same order as the relevant energy differences. The

study of excitons, which, e.g., play an important role at GaAs(110) [74], further requires an extension of the mathematical framework beyond single quasiparticles.

Acknowledgements

We thank Magnus Hedström, Günther Schwarz, and Jörg Neugebauer for fruitful collaborations in the course of this work and Philipp Ebert for useful discussions. This work was funded in part by the EU through the Nanophase Research Training Network (Contract No. HPRN-CT-2000-00167) and the Nanoquanta Network of Excellence (Contract No. NMP-4-CT-2004-500198).

References

[1] P. Ebert: Curr. Opin. Solid State Mater. Sci. **5**, 211 (2001) 165
[2] V. P. LaBella, H. Yang, D. W. Bullock, P. M. Thibado, P. Kratzer, M. Scheffler: Phys. Rev. Lett. **83**, 2989 (1999) 165
[3] S. B. Zhang, A. Zunger: Phys. Rev. Lett. **77**, 119 (1996) 166, 182, 183, 186
[4] H. Kim, J. R. Chelikowsky: Phys. Rev. Lett. **77**, 1063 (1996) 166
[5] H. Kim, J. R. Chelikowsky: Surf. Sci. **409**, 435 (1998) 166, 182, 183, 186
[6] P. Ebert, K. Urban, L. Aballe, C. H. Chen, K. Horn, G. Schwarz, J. Neugebauer, M. Scheffler: Phys. Rev. Lett. **84**, 5816 (2000) 166, 167, 170, 179, 186, 187
[7] M. C. Qian, M. Göthelid, B. Johansson, S. Mirbt: Phys. Rev. B **66**, 155326 (2002) 166, 182, 183, 186, 187
[8] M. C. Qian, M. Göthelid, B. Johansson, S. Mirbt: Phys. Rev. B **67**, 035308 (2003) 166, 179
[9] P. Hohenberg, W. Kohn: Phys. Rev. **136**, B864 (1964) 166, 168
[10] W. Kohn, L. J. Sham: Phys. Rev. **140**, A1133 (1965) 166, 168, 169
[11] J. P. Perdew, K. Burke, M. Ernzerhof: Phys. Rev. Lett. **77**, 3865 (1996) 166
[12] G. Lengel, R. Wilkins, G. Brown, M. Weimer: Phys. Rev. Lett. **72**, 836 (1994) 166, 168, 184
[13] A. Görling: Phys. Rev. A **54**, 3912 (1996) 166, 171
[14] V. Fiorentini, M. Methfessel, M. Scheffler: Phys. Rev. B **47**, 13353 (1993) 167
[15] M. Hedström, A. Schindlmayr, M. Scheffler: phys. stat. sol. (b) **234**, 346 (2002) 167, 181, 183
[16] L. Sorba, V. Hinkel, H. U. Middelmann, K. Horn: Phys. Rev. B **36**, 8075 (1987) 168
[17] B. Reihl, T. Riesterer, M. Tschudy, P. Perfetti: Phys. Rev. B **38**, 13456 (1988) 168
[18] J. C. Slater: Adv. Quant. Chem. **6**, 1 (1972) 169, 173, 186
[19] J. F. Janak: Phys. Rev. B **18**, 7165 (1978) 169, 173
[20] C.-O. Almbladh, U. von Barth: Phys. Rev. B **31**, 3231 (1985) 170
[21] J. P. Perdew, M. Levy: Phys. Rev. Lett. **51**, 1884 (1983) 170
[22] L. J. Sham, M. Schlüter: Phys. Rev. Lett. **51**, 1888 (1983) 170
[23] A. Görling, M. Levy: Phys. Rev. A **50**, 196 (1994) 170

[24] A. Görling: Phys. Rev. B **53**, 7024 (1996) 170
[25] M. Städele, J. A. Majewski, P. Vogl, A. Görling: Phys. Rev. Lett. **79**, 2089 (1997) 170, 171
[26] R. W. Godby, M. Schlüter, L. J. Sham: Phys. Rev. B **37**, 10159 (1988) 170
[27] T. Kotani: J. Phys.: Condens. Matter **10**, 9241 (1998) 170
[28] J. P. Perdew, A. Zunger: Phys. Rev. B **23**, 5048 (1981) 171
[29] D. M. Ceperley, B. J. Alder: Phys. Rev. Lett. **45**, 566 (1980) 171
[30] L. Kleinman, D. M. Bylander: Phys. Rev. Lett. **48**, 1425 (1982) 171
[31] M. Bockstedte, A. Kley, J. Neugebauer, M. Scheffler: Comput. Phys. Commun. **107**, 187 (1997) 171
[32] M. Fuchs, M. Scheffler: Comput. Phys. Commun. **119**, 67 (1999) 171
[33] J. L. A. Alves, J. Hebenstreit, M. Scheffler: Phys. Rev. B **44**, 6188 (1991) 171, 176
[34] O. Madelung, U. Rössler, M. Schulz (Eds.): *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology – New Series*, vol. III/41A (Springer, Berlin, Heidelberg 2001) 171, 176
[35] G. D. Mahan: *Many-Particle Physics*, 3rd ed. (Springer, Berlin, Heidelberg 2000) 171
[36] L. Hedin: Phys. Rev. A **139**, 796 (1965) 172, 174
[37] W. von der Linden, P. Horsch: Phys. Rev. B **37**, 8351 (1988) 172
[38] G. E. Engel, B. Farid: Phys. Rev. B **47**, 15931 (1993) 172
[39] M. S. Hybertsen, S. G. Louie: Phys. Rev. B **34**, 5390 (1986) 173
[40] I. D. White, R. W. Godby, M. M. Rieger, R. J. Needs: Phys. Rev. Lett. **80**, 4265 (1998) 173
[41] M. Rohlfing, N.-P. Wang, P. Krüger, J. Pollmann: Phys. Rev. Lett. **91**, 256802 (2003) 173
[42] O. Pulci, F. Bechstedt, G. Onida, R. Del Sole, L. Reining: Phys. Rev. B **60**, 16758 (1999) 173
[43] A. Schindlmayr, P. García-González, R. W. Godby: Phys. Rev. B **64**, 235106 (2001) 173
[44] B. Holm, U. von Barth: Phys. Rev. B **57**, 2108 (1998) 174
[45] W.-D. Schöne, A. G. Eguiluz: Phys. Rev. Lett. **81**, 1662 (1998) 174
[46] W. Ku, A. G. Eguiluz: Phys. Rev. Lett. **89**, 126401 (2002) 174
[47] T. Kotani, M. van Schilfgaarde: Solid State Commun. **121**, 461 (2002) 174
[48] M. L. Tiago, S. Ismail-Beigi, S. G. Louie: Phys. Rev. B **69**, 125212 (2004) 174
[49] K. Delaney, P. García-González, A. Rubio, P. Rinke, R. W. Godby: Phys. Rev. Lett. **93**, 249701 (2004) 174
[50] C. Friedrich, A. Schindlmayr, S. Blügel, T. Kotani: Phys. Rev. B **74**, 045104 (2006) 174
[51] E. L. Shirley, X. Zhu, S. G. Louie: Phys. Rev. B **56**, 6648 (1997) 174
[52] R. W. Godby, M. Schlüter, L. J. Sham: Phys. Rev. B **35**, 4170 (1987) 174, 175
[53] X. Zhu, S. G. Louie: Phys. Rev. B **43**, 14142 (1991) 174
[54] M. Rohlfing, P. Krüger, J. Pollmann: Phys. Rev. B **48**, 17791 (1993) 174, 175
[55] X. Zhu, S. B. Zhang, S. G. Louie, M. L. Cohen: Phys. Rev. Lett. **63**, 2112 (1989) 175, 178
[56] O. Pulci, G. Onida, R. Del Sole, L. Reining: Phys. Rev. Lett. **81**, 5374 (1998) 175, 178
[57] S. J. Jenkins, G. P. Srivastava, J. C. Inkson: Surf. Sci. **254**, 776 (1996) 175

[58] M. M. Rieger, L. Steinbeck, I. D. White, H. N. Rojas, R. W. Godby: Comput. Phys. Commun. **117**, 211 (1999) 175, 177
[59] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. D. White, R. W. Godby: Comput. Phys. Commun. **125**, 105 (2000) 175
[60] H. J. Monkhorst, J. D. Pack: Phys. Rev. B **13**, 5188 (1976) 177
[61] A. Schindlmayr: Phys. Rev. B **62**, 12573 (2000) 177
[62] J. Neugebauer, M. Scheffler: Phys. Rev. B **46**, 16067 (1992) 178
[63] P. Eggert, C. Freysoldt, P. Rinke, A. Schindlmayr, M. Scheffler: Verhandl. DPG (VI) **40**, 2/491 (2005) 179
[64] N. D. Lang, A. R. Williams: Phys. Rev. B **18**, 616 (1978) 179
[65] P. J. Braspenning, R. Zeller, A. Lodder, P. H. Dederichs: Phys. Rev. B **29**, 703 (1984) 179
[66] M. Scheffler, C. Droste, A. Fleszar, F. Máca, G. Wachutka, G. Barzel: Physica B **172**, 143 (1991) 179
[67] J. Bormet, J. Neugebauer, M. Scheffler: Phys. Rev. B **49**, 17242 (1994) 179
[68] G. Fratesi, G. P. Brivio, L. G. Molinari: Phys. Rev. B **69**, 245113 (2004) 179
[69] S. Aloni, I. Nevo, G. Haase: Phys. Rev. B **60**, R2165 (1999) 184
[70] K. J. Chao, A. R. Smith, C. K. Shih: Phys. Rev. B **53**, 6935 (1996) 184
[71] P. Ebert, X. Chen, M. Heinrich, M. Simon, K. Urban, M. G. Lagally: Phys. Rev. Lett. **76**, 2089 (1996) 184
[72] C. Domke, P. Ebert, K. Urban: Phys. Rev. B **57**, 4482 (1998) 184
[73] P. Ebert, K. Urban, M. G. Lagally: Phys. Rev. Lett. **72**, 840 (1994) 184
[74] O. Pankratov, M. Scheffler: Phys. Rev. Lett. **75**, 701 (1995) 188

Index