Calculation of surface optical properties: from qualitative understanding to quantitative predictions

W.G. Schmidt\(^a\),\(^*\), K. Seino\(^a\), P.H. Hahn\(^a\), F. Bechstedt\(^a\), W. Lu\(^b\), S. Wang\(^b\), J. Bernholc\(^b\)

\(^a\)Institut für Festkörpertheorie und Theoretische Optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany
\(^b\)Department of Physics, North Carolina State University, Raleigh, NC 27695-8202, USA

Abstract

In the last couple of years there has been much methodological and computational progress in the modeling of optical properties from \textit{first principles}. While the calculation of non-linear optical coefficients is still hampered by numerical limitations—demonstrated here for the case of bulk GaAs—linear optical spectra can now be calculated accurately and with true predictive power, even for large and complex surface structures. This allows on one hand for a much better understanding of the origin of specific features such as surface optical anisotropies. We find that in particular microscopic electric fields at the surface induce slight deformations of bulk-like wavefunctions and thus give rise to optical anisotropies even from sub-surface layers. On the other hand, from the comparison of measured and calculated spectra, one can now confidently reach conclusions on the surface geometry. This short review focuses on the simulation of reflectance anisotropy spectroscopy. The clean, hydrogenated and uracil-covered Si\((001)\) surface is used to illustrate the microscopic origin of surface optical anisotropies and the present state-of-the-art in computational modeling of optical spectra.

© 2003 Elsevier B.V. All rights reserved.

PACS: 78.68.+m; 78.66.Fd; 78.66.Db; 78.66.Qn

Keywords: Reflectance anisotropy spectroscopy; Density-functional calculations; Semiconductor surfaces; Second harmonic generation

1. Reflectance anisotropy spectroscopy

Optical spectroscopies are extremely valuable for the in situ, non-destructive and real-time surface monitoring under challenging conditions as may be encountered, e.g., during epitaxial growth. Reflectance anisotropy spectroscopy (RAS) is one of the plethora of surface sensitive spectroscopies probing linear optical coefficients, which has been particularly successful in exploring surface structures.

Already in 1966 Cardona, Pollak and Shaklee observed that rotating a Si\((110)\) sample around the surface normal changed the reflectivity as a function of the azimuthal angle \([1]\). However, it took about twenty years before the phenomenon of surface-induced optical anisotropies started to be investigated systematically \([2–5]\). The high potential of RAS for growth-monitoring was demonstrated by Kamiya et al. \([6]\) by measurements on GaAs\((001)\) in ultra-high vacuum as well as in atmospheric pressure of H\(_2\), He and N\(_2\). It was noted that each of the well-known GaAs\((001)\) surface reconstructions such as \((4\times4)\), \((2\times4)\) and \((4\times2)\) gives rise to a characteristic, fingerprint-like RAS spectrum. The correlation between reflection high-energy electron diffraction (RHEED) patterns and the RAS spectra allowed them to establish a database for GaAs\((001)\) surface reconstructions \([7]\). The surface reconstruction can thus be determined in situ, even at ambient conditions. For many semiconductors a change of the surface reconstruction during layer growth occurs. RAS can, therefore be easily used to monitor growth and layer completion. This application has proven highly successful \([8,9]\) and is now exploited commercially.

However, surface optical spectroscopies give only indirect information about the surface structure and chemistry. The intuitive interpretation of the spectra in terms of bond polarizabilities is difficult and often leads to wrong results even for relatively simple surfaces such
as Si(001) (see discussion in Ref. [10]). Reliable calculations of surface optical spectra are needed to fully exploit their diagnostic potential.

One of the first major attempts to calculate surface optical properties was made in 1971 by McIntyre and Aspnes [11]. They approximated the inhomogeneity of the surface dielectric tensor by a two-step function, i.e. modeled the surface by a three-layer system consisting of vacuum, thin film and substrate. This model was later extended to include surface anisotropy [12] and remains very popular because of its simplicity.

A more general approach was taken in 1979 by Bagchi, Barrera and Rajagopal [13]. They started by considering a semi-infinite electron gas. The truncation of the bulk leads to a modification of the optical properties in the surface region. Based on the assumption that this region is much thinner than the light wavelength, the light propagation equations were solved. A few years later Del Sole [14] generalized this method to the case of larger anisotropies in real crystals and obtained an expression for the surface contribution to the reflectance, \( \Delta R/R \), where \( R \) is the reflectance according to the Fresnel equation. For s-light polarized along \( i \) and normal incidence one obtains

\[
\frac{\Delta R}{R}(\omega) = \frac{4\omega}{c} \text{Im} \left\{ \frac{\Delta \varepsilon_i(\omega)}{\varepsilon_b(\omega) - 1} \right\}
\]  

(1)

where \( \varepsilon_b \) is the bulk dielectric function, and

\[
\Delta \varepsilon_i = \int dzdz' \left[ \varepsilon_i(\omega;z,z') - \delta_{ij}(z-z') \varepsilon_i(\omega;z) \right]
\]  

\[= \int dzdz'dz''dz''' \varepsilon_i(\omega;z,z';z'',z''') \varepsilon_i^{-1}(\omega;z',z'',z) \delta_{ij}(z-z').
\]  

(2)

Here \( \varepsilon_i(\omega;z,z') \) is the non-local macroscopic dielectric tensor of the solid-vacuum interface accounting for all many-body and local-field (LF) effects [15]. The second term in Eq. (2) is difficult to calculate because of the inversion required to obtain \( \varepsilon_i^{-1} \) and because of the four-fold integration. It contains off-diagonal terms \( \varepsilon_i \) of the susceptibility of the semi-infinite crystal. Manghi et al. [16] investigated the influence of these terms for GaAs(110) and GaP(110) surfaces. No relevant contributions of the off-diagonal terms to the reflectance were noticed and they are therefore, usually neglected. The first term of Eq. (2) can be evaluated by replacing the semi-infinite crystal by an artificial super-cell, large enough to represent the vacuum as well as the surface and bulk regions of the crystal under investigation. Provided that: (i) the slab is large enough to properly describe the surface region of the crystal, i.e. the surface as well as surface-modified bulk wavefunctions and (ii) the off-diagonal terms of the dielectric tensor are small compared to the diagonal ones, a simple expression for the surface contribution to the reflectivity can be derived [16]

\[
\frac{\Delta R}{R}(\omega) = \frac{4\omega}{c} \text{Im} \left\{ \frac{4\pi \alpha^0_{ii}(\omega)}{\varepsilon_i(\omega) - 1} \right\}
\]  

(3)

Here \( \alpha^0_{ii}(\omega) \) with \( i=x,y \) is the diagonal tensor component of the averaged half-slab polarizability. It is important to note that under the conditions mentioned above, Eq. (3) contains in principle all surface contributions to the optical reflectance. This includes contributions due to surface electronic states, atomic relaxations, the influence of the surface potential on the bulk wavefunctions as well as the surface LF effects, i.e. the influence of the surface-modified microscopic fluctuations of the electric field on the macroscopic dielectric response, and many-particle effects such as the electronic self-energy and the electron–hole attraction. Of course, it is not an easy task to include all these contributions in practical calculations of the slab polarizability \( \alpha \). Up to very recently, many approximations were required to calculate \( \alpha \). One possibility consists in combining the atomic structure optimization within density-functional theory (DFT) with electronic structure calculations done in the tight-binding (TB) approximation. This method has for example been applied to clean [17] and Sb-covered GaAs(001) surfaces [18] as well as the As/InP(110) interface [19]. Other encouraging results were obtained by combining DFT calculations with the scissors-operator approximation [20] to the electronic self-energy. This method has been applied to, among others, bare and As-terminated Si(001) surfaces [21,22]. Pulci et al. [23] used a linear parametrization of the quasiparticle shifts calculated within the GW approximation with respect to surface localization to calculate the RAS for the GaAs cleavage face.

While the examples discussed above allow for a meaningful interpretation of the experimental spectra, the quantitative agreement is often unsatisfactory. This has caused a considerable amount of confusion and speculation about the origin of surface optical anisotropies in the literature. Even the validity of the slab approach for the calculation of surface optical spectra was questioned. In particular the fact that surface optical spectra often show resonances near the energies of bulk critical point (CP) energies has been considered puzzling. It has led to the development of a series of models which—by choosing suitably fitted parameters—were often successful in reproducing specific features of the measured spectra:

1. The surface local-field effect was one of the first explanations for optical anisotropies. Although exact expressions for the LF contributions to the surface optical response have been derived decades ago [15,24], the complexity of the problem required until
recently an approximate treatment based on modeling the crystal surface by a lattice of polarizable entities, which obey a Clausius-Mossotti-like relation [25–27].

2. A photon-induced localization of the electron wave packets constituting the initial and final states of the optical excitation was made responsible for derivative-like features in the optical spectra of passivated Si(001) and Si(113) surfaces [28–32].

3. Optical anisotropies of the GaAs(001) and (111) surfaces near the \( E_1 \) and \( E_0 \) CP energies of bulk GaAs were attributed to a so-called surface termination effect, i.e. the termination of the electron wavefunctions at the crystal surface [33].

4. In addition to those models there were studies that emphasized the influence of the surface electric field and strain induced dichroism at the bulk CPs of the dielectric function on the surface optical response [34–37].

The surface termination effect has already been shown by Del Sole and Onida [38] to be irrelevant for RAS. However, a comprehensive, quantitative picture on how strongly the surface optical anisotropies are related to surface electric fields, surface induced strain, surface defects, local fields and many-body effects such as self-energy and electron–hole attraction effects could only be obtained very recently, thanks to the advent of supercomputing and advanced numerical algorithms.

2. Origin of surface optical anisotropies

The optical anisotropy of the Si(001) surface has been intensively studied by both experiment [22,28,30,40–44] and theory [10,21,22,45–47], due to its technological importance and its model character for semiconductor surface science. In order to determine the origin of the spectral features of the Si(001)c(2×4) surface, Schmidt et al. [39] performed ab initio calculations based on a massively parallel real-space finite-difference implementation [48] of the DFT. Self-energy effects were taken into account using the scissors operator. A linear cut-off function was used to separate the contributions to the RAS from electronic transitions within the uppermost four atomic layers and from the bulk-like layers underneath. Fig. 1 shows that surface-state related transitions are mainly responsible for the optical anisotropies below the \( E_1 \) CP energy. This holds in particular for the features T1 and T2. The appearance of optical anisotropies due to transitions between surface states is not surprising: In a number of instances such as InP, GaP and SiC surfaces, numerical modeling succeeded in relating characteristic anisotropy peaks to specific surface states [49–53].

More difficult to understand is the appearance of optical anisotropies such as T3, which are not related to transitions between surface states. It has led to the suggestion that subsurface stress due to the dimerization of the surface atoms may be largely responsible for RAS contributions from deeper layers [36]. In order to test this hypothesis additional calculations were performed for a geometry, where only the uppermost two atomic layers were allowed to relax, and the remaining atoms occupy their ideal bulk positions. Fig. 1 shows the RAS spectrum of the configuration with surface-restricted relaxation (dashed lines) in comparison with the results for the fully relaxed geometry (solid lines). The surface-related RAS minima T1 and T2 become more pronounced. Only small modifications occur for the RAS arising from electronic transitions below the uppermost four atomic layers.

A second possibility to probe the influence of surface stress on the optical spectra consists in slightly changing the lateral dimensions of the surface unit cell. The dimerized surface is stretched along the dimer bonds and compressed across the dimer rows [54]. Accordingly, the Si(001)c(2×4) surface was relaxed in a unit cell whose lateral dimensions are shrunk/enlarged by 0.5% along [110]/[110]. The RAS spectra calculated for this configuration are shown by dotted lines in Fig. 1. The optical anisotropies arising from subsurface transitions are slightly affected. In particular the T3 feature is enhanced. It may thus be partially attributed to reconstruction-induced stress in the layers beneath the surface. The calculations thus show that both the structural relaxation in the bulk and anisotropic strain fields...
modify the bulk-related optical anisotropy. Its salient features, however, are not affected. This indicates that the surface induces optical anisotropies in the bulk predominantly via an electronic coupling.

In Fig. 2 the local part of the effective potential entering the DFT Kohn–Sham equations

\[ V(r) = V_\text{local}^{\text{ps}}(r) + V_H(r) + V_{\text{XC}}^{\text{GGA}}(r) \]  

(4)

is shown in two planes perpendicular to the surface normal which cut the uppermost and seventh atomic layer of Si(001). Here \( V_\text{local}^{\text{ps}} \) is the local part of the ionic pseudopotentials and \( V_H \) and \( V_{\text{XC}}^{\text{GGA}} \) denote the Hartree and exchange and correlation potential in generalized gradient approximation (GGA), respectively. As can be seen from the figure, both the absolute magnitude and the ’shape’ of the potential differ strongly between the first and seventh atomic layer. In order to quantify this effect we average the potential in the planes perpendicular to the surface normal. The resulting quantity

\[ \langle V \rangle(z) = \frac{1}{A} \int_A dxdy V(x,y,z), \]  

(5)

where \( A \) is the area of the surface unit cell, is shown in Fig. 3. The averaged potential \( \langle V \rangle \) in the uppermost two atomic layers differs strongly from the value in the deeper layers. In order to quantify the anisotropy of the potential we consider the layer-averaged microscopic electric fields perpendicular to the surface normal. In detail, we calculate

\[ \langle \nabla V \cdot (\nabla V) \rangle(z) = \frac{1}{L} \int_{z-L/2}^{z+L/2} dz' \frac{1}{A} \int_A dxdy \left[ \frac{\partial V(x,y,z')}{\partial x} \right] \left[ \frac{\partial V(x,y,z')}{\partial y} \right] , \]  

(6)

where \( L \) is the bilayer thickness and \( x/y \) are parallel to the 110/110 direction. The difference Eq. (6) approaches zero only below the seventh atomic layer beneath the surface (cf. Fig. 3). Obviously, the surface-induced anisotropy of the microscopic electric fields reaches far into the bulk. It therefore affects wavefunctions, which cannot be categorized as electronic surface states, but should rather be addressed as electric-field modified bulk wavefunctions. The surface-induced deformations of these bulk-like wavefunctions are weighted differently by the \( x \) and \( y \) components of the optical transition operator, leading to an anisotropic optical response from layers beneath the surface. Although these anisotropies might thus to be thought to be related to surface local fields, this is not true in a strict sense, because they already show up in the \((0,0)\)-element of the microscopic dielectric tensor in reciprocal space.

Superimposed on the surface-induced anisotropies of the electron wavefunctions may be deformations due to electric fields induced by, e.g. the pinning of the Fermi level at the sample surface. These deformations may be anisotropic too, as has been shown by first-principles calculations for GaAs(001) surfaces (cf. Fig. 6 in Ref. [55]). The linear electro-optic effect modifies mainly the optical anisotropy from the bulk atomic layers, resulting in changes of the RAS signal, which are strongly reconstruction dependent, however. The electric field affects the atomic lattice as well as the electron wavefunctions and their eigenvalues. It has been found, however, that the changes of the combined density of states due to the electric field as well as the field-induced lattice relaxations are negligible compared to the deformations of the wavefunctions, resulting in modified transition matrix elements [55].

The surface potential may not only be altered by electric fields, but also by surface defects such as steps.
RAS spectra measured for different Si experiment anisotropies (steps of type $S$) on Si to the three most commonly observed step configurations. We have plotted the calculated optical anisotropies due to the self-energy within the GW approach for the monohydride Si(001) surface. From Ref. [39]

The modified potential may on one hand lead to defect-localized electronic states, which are directly involved in electronic transitions, thus modifying the surface optical response. On the other hand, additional optical signals can be expected from perturbed bulk-like wavefunctions localized far beneath the uppermost substrate layers. Defect-induced optical anisotropies may be of similar magnitude as the signals from the perfectly ordered surface. This is demonstrated in Fig. 4, where we have plotted the calculated optical anisotropies due to the three most commonly observed step configurations on Si(001) surfaces [56]: $S_A$, $S_B$, and $D_B$ [39]. The calculated step-related anisotropies (which are in excellent agreement with experiment [41]) explain the distinct differences in the RAS spectra measured for different Si(001) surface samples [22,40–42].

3. Many-body effects in surface optical spectra

The inclusion of many-body effects—beyond those already contained in the approximations for exchange and correlation, which enter the Kohn–Sham equations—in the calculation of surface optical spectra is a fairly recent achievement. In the late 1990s it has become possible to correct the single-particle excitation energies for quasiparticle effects, using approximations to the self-energy within the GW approach [23,49,57]. The solution of the Bethe–Salpeter equation (BSE) allows electron–hole attraction, i.e. excitonic, and LF contributions to the surface optical response to be taken into account. While the first applications of the BSE to surfaces were restricted to the description of few strongly localized surface states [58,59], advanced algorithms together with a massive parallelization of the computations made it recently possible to model the excitonic and LF effects in a wide spectral range [60,61].

In Fig. 5 the influence of many-body effects on the calculated optical anisotropy of the monohydride Si(001)$(2 \times 1)$ surface [61] is shown. The spectrum calculated in the independent-particle approximation, i.e. the DFT-LDA spectrum shows in principle the features observed experimentally [43]. However, the calculated $A'$ and $B'$ peaks are redshifted in comparison with the experiment by 0.5 eV and 0.8 eV, respectively. Self-energy corrections calculated in GW approximation shift the excitation energies to larger values. Furthermore, the $B'$ optical anisotropy is much increased. This is not only related to the non-uniform self-energy corrections of the DFT-LDA eigenvalues, but also a consequence of the $\omega$ scaling of the surface contribution to the reflectance (cf. Eq. (1)). Surface LF effects can be expected from both the microscopic fluctuations of the electric field within the bulk, and from the truncation of the bulk itself. The numerical calculation, however, shows that LF effects lead to surprisingly small changes of the spectrum in the particular case studied here. A distinct change of the RAS spectrum, however, results from the inclusion of the attractive electron–hole interaction. A strong reduction of the $B'$ optical anisotropy and an increase of the integrated $A'$ peak area occurs. At the same time, the $A'$ and $B'$ optical anisotropies are shifted to lower energies by approximately 0.2 eV.

The stepwise inclusion of many-particle effects in the calculation leads to a considerable and systematic improvement of the agreement with the experiment. The remaining discrepancies are mainly due to the numerical limitations of the study [61], in particular with respect to the slab thickness and Brillouin zone sampling.

The full inclusion of many-body effects in the calculation of surface optical spectra leads to the limits of today’s computer capacity. However, numerically well-converged spectra calculated within the independent-particle approximation and using the scissors-operator approach to account for the band-gap underestimation...
are in many instances sufficient for the reliable prediction of surface optical anisotropies. This is because RAS spectra are difference spectra, which are furthermore normalized to the bulk dielectric function (cf. Eq. (1)). Many systems, for example surface defects or very large reconstructions such as the ones formed by In nanowires on Si(111) surfaces [62] can presently only be described within the single-particle approximation.

4. Outlook

Organic functionalization of semiconductors has become important for the development of new semiconductor-based devices. The physical and chemical properties of hybrid organic/inorganic materials depend crucially on the structural and chemical details of the interface. In principle it should be possible to obtain such information from optical spectroscopies such as RAS. However, at present there are only relatively few experimental and theoretical studies that explore optical anisotropies of organically modified semiconductors (see, e.g. Refs. [63–65]).

The adsorption of uracil on Si(001), investigated experimentally by scanning tunneling microscopy and high-resolution electron energy-loss spectroscopy [66], leads to what can be considered a prototypical interface between a polyfunctional molecule and a semiconductor surface. The (001) surface of silicon is the starting point for the fabrication of most microelectronic devices. Uracil (C$_4$H$_4$N$_2$O$_2$) is a small molecule featuring one C=C double bond, two amino and two carbonyl groups and may thus bond to the surface in various ways. In addition, its tautomerism and electrostatic effects have been found to be important for the interface formation [66,67]. In order to probe the sensitivity of RAS with respect to different uracil bonding configurations we calculate spectra for the most relevant dative (D1, D2 in Fig. 6) and covalently bonded (C1, C2) interface configurations (details will be published elsewhere).

These spectra (Fig. 6) indicate that the adsorption of organic molecules strongly modifies the optical response of the Si(001) surface (cf. Fig. 1) and leads to a signal that is very sensitive to the details of the interface. In particular for photon energies of approximately 3.3 eV, i.e. slightly below the $E_1$ CP, a strong dependence of the surface optical anisotropy on the adsorption configuration can be stated: oxygen insertion into the uppermost Si dimers (D2, C2) leads to a strong negative anisotropy. However, also the structures where uracil remains intact, and is dative (D1) or covalently bonded to the substrate (C1) can be clearly distinguished by means of RAS.

RAS spectra simulated for cyclopentene adsorption configurations on Si(001) [65] indicate that measurements of reflectance difference data can even discriminate between ‘cis’ and ‘trans-type’ adsorption configurations. The high sensitivity of RAS with respect to the details of the surface chemical bonding in conjunction with the accuracy achieved in the numerical prediction of such spectra thus suggests RAS as a powerful tool also for monitoring the formation of organic thin films.

Thus far we have focused on the modeling of linear optical spectra. However, non-linear optical processes such as second harmonic generation (SHG) are also increasingly being used for surface and interface analysis [71]. Unfortunately, the microscopic interpretation of the measured spectra is severely hampered by numerical difficulties. This is demonstrated in Fig. 7, where we compare recently calculated values for the second-harmonic susceptibility $\chi^{(2)}_{xy}(\omega, -2\omega)$ of bulk GaAs with experiment [68]. Obviously, the theoretical results obtained within a full-potential linearized augmented plane-wave method [70] and from plane-wave pseudopotential calculations [69] neither agree well with each other nor with the present calculations utilizing a real-space finite-difference approach [48]. The agreement between theory and experiment is also unsatisfactory.

Fig. 6. Left: Uracil/Si(001) adsorption configurations. White, gray, dark gray, dark, and small symbols indicate C, N, Si, O, and H atoms, respectively. Right: RAS $[\Re\{\epsilon_{001} - \epsilon_{110}\}/\epsilon_0]$ for the respective Uracil/Si(001) adsorption configurations, calculated within DFT-GGA. A scissors operator has been used to account for self-energy effects.

The reasons for the discrepancies are on the one hand related to the extreme sensitivity of the calculations to nearly all technical parameters, in particular the sampling of the Brillouin zone (in the present calculations a set of 32,000 randomly chosen points was needed to obtain converged results). On the other hand, non-linear optical spectra depend on many-body effects in a complex and not easily predictable manner [69,72] (in the present calculations a scissors-operator approach was used). Both, the technical aspects as well as the correct modeling of many-body effects become much more difficult to control in the case of surface calculations [73], where it seems presently that only empirical models using fitted values for the bond polarizabilities allow for the satisfactory modeling of surface SHG spectra [74]. Thus, for non-trivial cases neither a qualitative understanding nor a quantitative prediction of non-linear surface optical spectra seems presently possible.

5. Conclusions

The difference between the optical response of the surface compared to that of the bulk may be related to the possible presence of surface electronic states, but can also arise from surface induced anisotropies in the crystal potential extending far below the uppermost surface layer, surface electric fields, surface defects, and surface induced strain fields. It has now become possible to include these effects highly accurately in parameter-free calculations of linear surface optical spectra. Even many-body effects such as self-energy and excitonic contributions to the optical response and the influence of local fields can now be studied from first principles. Quantitative predictions of linear surface optical properties have thus become possible. The accuracy of calculated non-linear optical properties, however, still suffers from numerical limitations.

Acknowledgments

Grants of computer time from the DoD Challenge Program, the Leibniz-Rechenzentrum München, and the Höchstleistungsrechenzentrum Stuttgart are gratefully acknowledged.

References
