Theory of second-harmonic generation from quantum well states in ultrathin metal films on semiconductors

Thomas Garm Pedersen, Kjeld Pedersen, and Thomas Brun Kristensen
Institute of Physics, Aalborg University, Pontoppidanstræde 103, DK-9220 Aalborg Øst, Denmark
(Received 13 May 1999; revised manuscript received 9 November 1999)

Second-harmonic generation from quantum well states in metal films deposited on a semiconducting substrate is discussed theoretically within a microscopic approach. The thickness dependence of electronic eigenstates as well as band-structure effects related to motion in the plane of the film are taken into account and semianalytical quantum well states are obtained at the perturbation level. Based on a general treatment of the second-harmonic response of inhomogenous systems an effective second-order response tensor is obtained. Agreement between theory and experiment is demonstrated from a comparison of calculated and measured thickness dependence of the nonlinear response. In the case of *p* to *p* curves the agreement is over the entire range of film thicknesses, whereas the agreement in the *p* to *s* case is restricted to relatively thin films due to the applied perturbation approximation. The transitions that are responsible for experimentally observed resonances are identified.

I. INTRODUCTION

The formation of quantum well (QW) states in layered structures is of great importance for applications as well as basic solid-state physics. Semiconductor QW's play an increasingly large role in fabrication of devices such as QW lasers and in fundamental studies of the behavior of quasi-two-dimensional electron systems (see, e.g., Ref. 1). Similarly, QW structures formed in purely metallic systems are emerging as important tools for studies of, e.g., magnetic properties.[2,3] In comparison, investigations of QW effects in mixed metal/semiconductor structures are relatively few. This is mainly because of difficulties in producing high-quality metallic overlayers in these systems. A few reports using various linear optical techniques such as photoemission[4] and differential reflectance measurements[5] have demonstrated the presence of QW states formed in metallic layers deposited on Si, however.

Recently, we have applied the unique surface sensitivity of second-harmonic generation (SHG) to follow the formation of QW states during growth of atomically flat Au (Refs. 6 and 7) and Ag (Ref. 8) films on Si(111). The sensitivity of SHG derives from the fact that electric dipole contributions vanish in media with inversion symmetry. Hence, in structures composed from such media, SHG mainly arises from interfaces, which break the symmetry. The interface sensitivity has previously been applied in studies of the metal/semiconductor contact.[9] In the case of an ultrathin overlayer, however, the electronic coherence extends throughout the layer and SHG cannot be discussed in terms of separate contributions from the vacuum/metal and metal/semiconductor interfaces. The added metallic overlayer modifies the electronic states of the semiconductor and, furthermore, introduces new electronic states that are localized by the QW potential. Consequently, SHG can be used as a highly sensitive probe of the formation of new states during deposition of the overlayer. In particular, resonant transitions between overlayer states will show up as resonances in the second-harmonic signal. Hence, SHG can directly probe the level structure of the electronic system. Furthermore the quality of the metallic film can be inferred from such a measurement since any roughness will tend to smear out the resonances. It follows that the presence of distinct resonances is a clear indication of high-quality film growth. An additional advantage of SHG is that different interface properties can be investigated by measuring different tensor elements of the nonlinear susceptibility. For instance, taking the *z* axis as the surface normal, the *zzz* element of the susceptibility provides information about the electronic motion perpendicular to the surface. In particular, the effects of size quantization along the growth direction will be seen in this tensor element. Conversely, the *xxx* element is entirely determined by the motion of electrons in the surface plane. It follows that by measuring this tensor element while rotating the sample around the surface normal, the in-plane symmetry of the electronic states can be deduced. In fact, the *xxx* element vanishes in an amorphous material and so its magnitude is a direct measure of the degree of structural order.

The major incentive for the construction of a theory describing the second-harmonic response of metal/semiconductor QW structures is the need for reliable methods of interpreting experimental optical data in terms of electronic and structural properties. In our previous work,[8] preliminary modelling results for the *p* to *s* response of Ag on Si(111) were presented. In the present paper we wish to present the detail of the complete theoretical framework describing SHG from QW's formed by depositing ultrathin metallic films on a semiconducting substrate. In particular, noble metals deposited on Si(111) will be investigated and results for all elements of the second harmonic response tensor will be presented. Previously, Petukhov and Liebsch[10] and Luce and Bennemann[11] have performed detailed microscopic calculations of SHG from metal surfaces. In addition, a qualitative theoretical treatment of SHG from QW states in purely metallic systems has been presented.[12] A microscopic theory of SHG from QW's formed in metal/semiconductor structures is lacking, however, making the interpretation of

SHG data from such structures difficult. In a typical experiment, the reflected second-harmonic signal is monitored during film growth using a fixed set of polarization directions for the second-harmonic and pump fields and a fixed pump photon energy. In order to directly compare experimental and theoretical results we consequently have to calculate the relevant tensor elements as a function of film thickness. As mentioned above, certain tensor elements are a hallmark of the presence of structural order, i.e., they vanish if the crystallinity of the material is destroyed. In terms of their theoretical calculation, this property means that the assumption of free-electron-like motion in the film plane cannot be applied. Hence, a quantum mechanical description of the electronic states including band structure effects will be required. In this work, such a quantum description is used and it is demonstrated how band structure effects may be taken into account at both quasiexact and perturbational levels. Subsequently, the induced nonlinear current density is calculated using the density matrix formalism taking into account the spatial variation of the electric fields. For simplicity, we only consider the contribution to the nonlinear current due to the added electronic states. Thus, the background signal from the substrate will be neglected. This simplification is not a serious one since the background signal is easily substracted from the experimental curves. Consequently, theoretical and experimental results are directly comparable. The remaining part of the paper is organized as follows: In the next section, quasiexact and perturbational calculations of QW states are presented. These states are then applied in a description of the nonlinear response in Sec. III and analytical expressions for the elements of the effective nonlinear susceptibility are derived. In Sec. IV, numerical calculations of the thickness dependence of the nonlinear response are presented for a number of different cases and a comparison between experimental and theoretical results is given. Finally, a summary is presented in Sec. V.

II. QUANTUM WELL STATES

In this section, we present the theoretical framework for the calculation of quantum well states including band structure effects. Highly accurate methods, such as self-consistent density-functional calculations,[13–14] exist for this purpose. We wish, however, to obtain semianalytical wave functions, which will allow us to evaluate the nonlinear response without extensive numerical work. Therefore, an approximate and simplified theoretical framework is applied below. In general, the in-plane motion of the high-lying *s-p* states is not expected to deviate much from free electron-like behavior. Hence, these states are well described by a standard truncated Fourier expansion of the potential. In contrast, the deeper *d* states are highly localized and a tight-binding treatment is better suited for their description. This is certainly the case in noble metals for which the bulk *d*-band is well below the Fermi level, e.g., 2.4 and 4.0 eV for Au and Ag, respectively. In the present study, however, we will concentrate on the nonlinear response to optical fields of low- or moderate-photon energy for which *d*-band excitations can be ignored. Moreover, this case is well studied experimentally due to the availability of Nd:YAG lasers with $\hbar\omega = 1.17$ eV for which neither pump nor second-harmonic photons can reach the *d*-band. Hence, in the following a simple Fourier expanded potential will be assumed. When fitted to match the experimental band structure, however, the Fourier expansion coefficients of the *s-p* potential will indirectly include the influence of the *d*-bands. Consequently, with the *z* axis perpendicular to the interfaces, the total potential can be written

$$V(\mathbf{r}) = \sum_{\mathbf{G}_\parallel} V_{\mathbf{G}_\parallel}(z) e^{i\mathbf{G}_\parallel \cdot \mathbf{r}}, \tag{1}$$

where $\mathbf{G}_\parallel$ is the projection of a reciprocal lattice vector $\mathbf{G}$ onto the surface plane. A potential of the form given by Eq. (1) leads to electronic states of the general form

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{2\pi} \sum_{\mathbf{K}_\parallel} \varphi_{n\mathbf{k}}(\mathbf{K}_\parallel, z) e^{i(\mathbf{k}+\mathbf{K}_\parallel)\cdot \mathbf{r}}, \tag{2}$$

where $\mathbf{k}$ is a surface Bloch vector, i.e., of the form $\mathbf{k} = k_x\hat{x} + k_y\hat{y}$, and the summation over $\mathbf{K}_\parallel$ runs over the different values of $\mathbf{G}_\parallel$. The factor $(2\pi)^{-1}$ is introduced for later convenience. Consequently, the Schrödinger equation reduces to a set of coupled one-dimensional equations

$$\left\{ \frac{\hbar^2}{2m} |\mathbf{k}+\mathbf{K}_\parallel|^2 - \frac{\hbar^2}{2m}\frac{d^2}{dz^2} - E_{n\mathbf{k}} \right\} \varphi_{n\mathbf{k}}(\mathbf{K}_\parallel, z)$$

$$= -\sum_{\mathbf{G}_\parallel} V_{\mathbf{G}_\parallel}(z) \varphi_{n\mathbf{k}}(\mathbf{K}_\parallel - \mathbf{G}_\parallel, z). \tag{3}$$

As mentioned above, band structure effects are expected to be minor corrections since we are dealing with *s-p* states exclusively. It follows that the solutions to the simple one-dimensional quantum well potential $V_0(z)$ are excellent zeroth-order approximations to the full solutions. These zeroth-order solutions are of the form $\psi^{(0)}_{n\mathbf{k}}(\mathbf{r}) = (2\pi)^{-1}\varphi_n(z)e^{i\mathbf{k}\cdot\mathbf{r}}$, where the real-valued function $\varphi_\alpha(z)$ is a solution to

$$\left\{ -\frac{\hbar^2}{2m}\frac{d^2}{dz^2} + V_0(z) - E_\alpha \right\} \varphi_\alpha(z) = 0. \tag{4}$$

Once the set of states $\{\varphi_\alpha\}$ is obtained, each of the unknown functions $\varphi_{n\mathbf{k}}(\mathbf{K}_\parallel, z)$ can be expanded in this set according to

$$\varphi_{n\mathbf{k}}(\mathbf{K}_\parallel, z) = \sum_\alpha a_\alpha(\mathbf{K}_\parallel)\varphi_\alpha(z), \tag{5}$$

where we have suppressed the dependence on n and $\mathbf{k}$ of $a_\alpha(\mathbf{K}_\parallel)$ for notational convenience. An algebraic set of equations determining the coefficients $a_\alpha(\mathbf{K}_\parallel)$ can be constructed by rewriting Eq. (3) according to

$$\left\{ \frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{K}_\parallel|^2 - \frac{\hbar^2}{2m}\frac{d^2}{dz^2} + V_0(z) - E_{n\mathbf{k}} \right\} \varphi_{n\mathbf{k}}(\mathbf{K}_\parallel, z)$$

$$= -\sum_{\mathbf{G}_\parallel} \tilde{V}_{\mathbf{G}_\parallel}(z) \varphi_{n\mathbf{k}}(\mathbf{K}_\parallel - \mathbf{G}_\parallel, z), \tag{6}$$

where $\tilde{V}_{\mathbf{G}_\parallel}(z) = V_{\mathbf{G}_\parallel}(z) - V_0(z)\delta_{\mathbf{G}_\parallel,0}$. In this manner the coupled differential equations can be reduced to an algebraic system given by

$$\left\{\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{K}_\parallel|^2+E_\alpha-E_{n\mathbf{k}}\right\}a_\alpha(\mathbf{K}_\parallel)$$

$$=-\sum_\beta\sum_{\mathbf{G}_\parallel}V_{\alpha\beta}(\mathbf{G}_\parallel)a_\beta(\mathbf{K}_\parallel-\mathbf{G}_\parallel), \tag{7}$$

where

$$V_{\alpha\beta}(\mathbf{G}_\parallel)=\int_{-\infty}^{\infty}\varphi_\alpha(z)\tilde{V}_{\mathbf{G}_\parallel}(z)\varphi_\beta(z)dz. \tag{8}$$

Using the density-functional method of, e.g., Refs. 13 and 14 it is possible to obtain self-consistent solutions for the various components of the potential. This leads to potential functions which all decay rapidly to zero in the vacuum half-space except $V_0(z)$, which asymptotically approaches the image potential[13–14] [actually an exponential rather than a $(2z)^{-1}$ behavior is obtained]. In addition, $V_{\mathbf{G}_\parallel}(z)$ approaches the bulk potential a few lattice constants into the metallic region. For simplicity, however, we will neglect the effects of self consistency in the present study. Hence, the bulk potential is assumed to remain valid right up to the metal/vacuum and metal/Si interfaces, i.e., throughout the region $0\leqslant z\leqslant d$, where d is the thickness of the metal layer. In order to correctly reproduce the vacuum tail of the wave functions the image potential is added to $V_0(z)$ for $z<0$. Similarly, the coupling between QW states and the substrate valence band is neglected in that a constant potential $-V_S$, taken as the conduction band minimum of Si, is used inside the substrate. The metal/Si interface is modeled by joining the metallic potential to the Si potential over a distance Λ using a linear dependence. Hence, the simple quantum well potential $V_0(z)$, including the image potential, is taken as

$$V_0(z)=\begin{cases}e^2/8\pi\varepsilon_0(z-z_0) & z<0\\ -V_B & 0\leqslant z<d-\Lambda/2\\ -(V_B+V_S)/2+(V_B-V_S)(z-d)/\Lambda & d-\Lambda/2\leqslant z<d+\Lambda/2\\ -V_S & z\geqslant d+\Lambda/2\end{cases}. \tag{9}$$

The location of the image plane z_0 is adjusted so that $V_0(z)$ is continuous at $z=0$ and $-V_B$ is the spatially constant part of the metal potential. V_B is the sum of work function Φ and Fermi energy E_F of the metal and in line with the approximations used above, bulk values of these parameters are adopted. For both Ag and Au Φ and E_F are approximately[15] 4.3 and 5.5 eV, respectively so that $V_B=9.8$ eV and, consequently, $z_0=0.73$ Å. Notice that the bulk Fermi energy is only utilized to calculate the potential and that the correct thickness-dependent Fermi energy will be applied in the calculation of the optical response below. One last simplification concerns the treatment of quasifree states, i.e., states with an energy larger than $-V_S$ that are not localized in the quantum well region. This continuum is difficult to handle numerically and, hence, artificial barriers are introduced at a distance l inside the Si substrate and into the vacuum. Care is taken that l is sufficiently large that the final result is virtually unaffected by this simplification. This is ensured provided the spacing between two quasifree states is somewhat smaller than the width of these levels given by $\hbar/\tau$, where τ is the relaxation time. In practice, values around $l\approx 50$ Å are found to be sufficient. The approximate potential is illustrated in Fig. 1. It is seen from the figure that an approximate value of V_S can be determined from the fact that the conduction-band minimum of Si(111)-7×7 at the surface is located approximately 0.4 eV above the Fermi level[16] and, hence, $V_S\approx 3.9$ eV. From a fitting procedure described below the values $\Lambda=2\delta$(Au) and $\Lambda=3\delta$(Ag), where $\delta=2.3$ Å is the (111)-surface monolayer thickness, have been obtained for the width of the metal/Si interface.

For $\mathbf{G}_\parallel\neq 0$ the use of the bulk crystal potential allows us to write the components of the potential as

$$\tilde{V}_{\mathbf{G}_\parallel}(z)=\sum_{\mathbf{G}_\perp}V_{\mathbf{G}_\parallel+\mathbf{G}_\perp}\exp(iG_\perp z)[\theta(z)-\theta(z-d)], \tag{10}$$

where the $V_{\mathbf{G}}$'s are the Fourier components of the *bulk* potential and it is understood that $V_0=0$. The matrix elements defined by Eq. (8) then read

$$V_{\alpha\beta}(\mathbf{G}_\parallel)=\sum_{\mathbf{G}_\perp}V_{\mathbf{G}_\parallel+\mathbf{G}_\perp}\int_0^d\varphi_\alpha(z)\exp(iG_\perp z)\varphi_\beta(z)dz. \tag{11}$$

In the bulk limit the discrete index α is replaced by a continuous one ($k_\perp$) and Eq. (7) clearly reduced to the standard matrix problem of bulk band-structure calculations.[15] In fact,

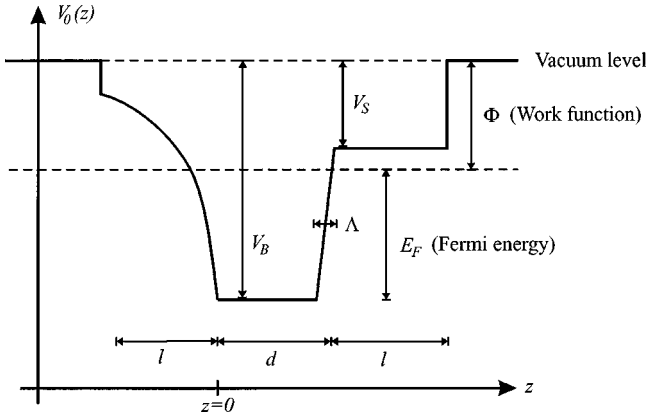

FIG. 1. The one-dimensional quantum well potential $V_0(z)$. The barriers at $z=-l$ and $z=d+l$ are introduced artificially in order to limit the number of levels.

solving the algebraic system in Eq. (7) is relatively straightforward. This follows from the fact that the dimension of the coefficient matrix is the number of different eigenstates in $\{\varphi_\alpha\}$ times the number of distinct $\mathbf{G}_\parallel$'s. In Refs. 17 and 18, the bulk band structure of different noble metals was fitted using 4 and 16 different reciprocal lattice vectors, respectively. These sets are (a): (0, 0, 0), (−1, −1, −1), (−1, −1, 1), (0, −2, 0) (Ref. 17) and (b): (0, 0, 0), eight vectors of type (1, 1, 1), six of type (2, 0, 0) and (−2, −2, 0).[18] For a (111) face, the parallel components of these vectors are found from $\mathbf{G}_\parallel=\mathbf{G}-(1/3)[(1,1,1)\cdot\mathbf{G}](1,1,1)$. Consequently, the number of distinct $\mathbf{G}_\parallel$'s in these sets is only 3 and 7, respectively. Typically, at $d\approx 30$ ML (monolayers) the set $\{\varphi_\alpha\}$ has about 50 members (counting only bound states) and, hence, the dimension of the coefficient matrix is approximately 150 and 350 in case (a) and (b). Thus, the $\mathbf{k}$ dependence of both energy eigenvalues and eigenstates can be obtained by implementation of simple numerical matrix diagonalization routines. Numerical problems arise, however, in the calculation of nonlinear susceptibilities. In order to evaluate the sum-over-states expressions presented below an integration over $\mathbf{k}$ is required and in the absence of an analytical expression for the $\mathbf{k}$ dependence of the integrand this task becomes extremely time consuming. As long as the Fourier components of the periodic crystal potential are much smaller than the separation between neighbouring unperturbed energy-eigenvalues, however, a perturbation approach can be applied. Naturally, the condition $|V_{\mathbf{G}}|\ll E_{\alpha+1}-E_\alpha$ is only fulfilled in thin QW's and, hence, the perturbation approach is restricted to this case. Solving Eq. (7) perturbatively produces analytical expressions for the $\mathbf{k}$ dependence of $E_{n\mathbf{k}}$ and $a_\alpha(\mathbf{k}_\parallel)$ and in fact, as will shown below, if only the first order correction to $a_\alpha(\mathbf{k}_\parallel)$ is retained the remaining integrations can be performed analytically. Using this approach, the wavefunctions can be written as

$$\psi_{n\mathbf{k}}(\mathbf{r})=\frac{1}{2\pi}e^{i\mathbf{k}\cdot\mathbf{r}}\left\{\varphi_n(z)+\sum_{\mathbf{G}_\parallel}\varphi_{n\mathbf{k}}(\mathbf{G}_\parallel,z)e^{i\mathbf{G}_\parallel\cdot\mathbf{r}}\right\}, \quad (12)$$

where the perturbative correction is given by

$$\varphi_{n\mathbf{k}}(\mathbf{G}_\parallel,z)=\sum_\alpha \frac{V_{n\alpha}(\mathbf{G}_\parallel)}{E_{n\alpha}+\frac{\hbar^2}{2m}k^2-\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}_\parallel|^2}\varphi_\alpha(z), \quad (13)$$

with $E_{\alpha\beta}=E_\alpha-E_\beta$ and it is understood that the diagonal term $\alpha=n$ is excluded from the summation. It is the relatively simple $\mathbf{k}$ dependence of this correction, which permits analytical $\mathbf{k}$ integrations.

III. NONLINEAR RESPONSE

The nonlinear response of a metallic QW system is different from that of a semiconductor system in several aspects. First of all, the absence of a band gap in metals allows for efficient excitations even with a relatively small photon energy. Hence, there is no wavelength threshold for the nonlinear interactions in metallic quantum wells. In addition, semiconductor heterostructures are usually fabricated by varying the content of a substitutional element added to the host material along the growth direction, such as, e.g., Al in $Al_xGa_{1-x}As/GaAs/Al_xGa_{1-x}As$. Hence, the confining barriers in semiconductor structures are relatively low and, furthermore, the variation in the dielectric constant is relatively small. Conversely, in the vacuum/metal/semiconductor structure, tall barriers are formed and the spatial variations of the dielectric properties are clearly important. Due to the tall barriers the occupied QW states are highly localized and the large confinement energy yields intraband resonances in the eV range. The spatial variations of the dielectric constant, on the other hand, render the usual dipole selection rules invalid since the dipole approximation is inapplicable for field components that are rapidly varying in space. In addition, the significant absorption in metals produce electric fields that vary on a scale set by the penetration depth (~10–50 nm) rather than the wavelength.

In Ref. 19, a framework for the second-order nonlinear response to spatially varying fields is constructed based on the density-matrix formalism. This framework relies on the jellium approximation for the electronic states. The inclusion of band-structure effect, however, is straightforward. Three sources contribute to the nonlinear response.[19] The first of these terms is due to the 2ω part of the interaction Hamiltonian $(\mathbf{A}_\omega\cdot\mathbf{A}_\omega)$. A second contribution can be traced to the product of the ω part of the interaction Hamiltonian $(\mathbf{p}\cdot\mathbf{A}_\omega)$ times the field-induced first harmonic of the density matrix ρ_ω. Finally, the third term arises from the ω part of the current density operator (diamagnetic part) times ρ_ω. Among these terms, however, only the second one survives in the limit where the dipole approximation is applied to both first- and second-harmonic fields. Thus, this term is generally expected to dominate in the nonlinear response and in the present work only this term is included. As a result, the nonlinear current density $\mathbf{j}_{2\omega}(\mathbf{r})$ produced by the electric field $\mathbf{E}_\omega(\mathbf{r})$ will be calculated from[19]

$$\mathbf{j}_{2\omega}(\mathbf{r})=-\frac{1}{\omega^2}\sum_{\lambda\mu\nu}\frac{1}{2\hbar\omega+i\hbar/\tau-E_{\nu\mu}}\left\{\frac{f_\mu-f_\lambda}{\hbar\omega+i\hbar/\tau-E_{\lambda\mu}}+\frac{f_\nu-f_\lambda}{\hbar\omega+i\hbar/\tau-E_{\nu\lambda}}\right\}\mathbf{j}_{\mu\nu}(\mathbf{r})\int\mathbf{j}_{\nu\lambda}(\mathbf{r})\cdot\mathbf{E}_\omega(\mathbf{r})d^3r\times\int\mathbf{j}_{\lambda\mu}(\mathbf{r})\cdot\mathbf{E}_\omega(\mathbf{r})d^2r. \quad (14)$$

In this expression, summation over spin degrees of freedom has already been performed and composite indices $\lambda=\{l,\mathbf{k}''\}$, $\mu=\{m,\mathbf{k}'\}$ and $\nu=\{n,\mathbf{k}\}$ have been introduced for notational convenience. In addition, τ denotes the intraband relaxation time, f_α is the Fermi distribution function evaluated at the energy E_α and $\mathbf{j}_{\alpha\beta}(\mathbf{r})$ is the transition current density element given by

$$\mathbf{j}_{\alpha\beta}(\mathbf{r})=-\frac{e\hbar}{2im}\{\psi_\alpha^*(\mathbf{r})\nabla\psi_\beta(\mathbf{r})-\psi_\beta(\mathbf{r})\nabla\psi_\alpha^*(\mathbf{r})\}. \quad (15)$$

The electric field $\mathbf{E}_\omega(\mathbf{r})$, which drives the nonlinear current density in Eq. (14) is the *local* field and, in general, quite different from the incident field $\mathbf{E}_\omega^0(\mathbf{r})$. In a layered structure displaying full rotational symmetry in the linear optical properties the spatial dependence of $\mathbf{E}_\omega(\mathbf{r})$ is limited to the form $\mathbf{E}_\omega(\mathbf{r})=\mathbf{E}_\omega(z)e^{i\mathbf{q}_\parallel\cdot\mathbf{r}}$, where $\mathbf{q}_\parallel$ is the parallel component of

the optical wave vector. In addition, the nonlinear current density averaged over a surface unit cell will vary according to $\mathbf{j}_{2\omega}(\mathbf{r})=\mathbf{j}_{2\omega}(z)e^{i2\mathbf{q}_\parallel\cdot\mathbf{r}}$. Since $q_\parallel=2\pi/\lambda_0\sin(\theta)$, where λ_0 is the vacuum wavelength and θ is the angle of incidence, the (x,y) variations of both electric field and nonlinear current density occur on a scale set by the wavelength. Consequently, this dependence can be safely disregarded in comparison with the z dependence. The amplitude of the incident field is denoted $\mathbf{E}^0_\omega$ and the x,y and z components of $\mathbf{E}_\omega(z)$ can be formally related to those of $\mathbf{E}^0_\omega$ via the introduction of field distribution functions according to

$$E_{\omega,i}(z)=f_i^\omega(z)E^0_{\omega,i}. \tag{16}$$

In the case of s and p polarization the components of the incident field are $\mathbf{E}^0_\omega=E^0_\omega(0,1,0)$ and $\mathbf{E}^0_\omega=E^0_\omega(\cos\theta,0,\sin\theta)$, respectively. The field distribution functions $f_i^\omega(z)$ include (i) the discontinuity of the z components at the material boundaries determined by the dielectric constants 1, $\varepsilon(\Omega)$, and $\varepsilon_{Si}(\Omega)$ of vacuum, metal and substrate, respectively, (ii) phase variation within each medium, and (iii) a spatially varying local field enhancement factor. For simplicity, we will replace the spatially varying enhancement factor by an average one. Hence, assuming a vacuum/metal dielectric boundary at $z=a$ and a metal/Si dielectric boundary at $z=b$ the field distribution function for the z component at frequency Ω is given by

$$f_z^\Omega(z)=\begin{cases} L_\Omega\{e^{iq_0(z-a)}+r_p e^{-iq_0(z-a)}\} & z<a \\ \varepsilon^{-1}(\Omega)L_\Omega(1+r_{01})/(1+r_{01}r_{12}e^{i2q_1\tilde{d}})\{e^{iq_1(z-a)}+r_{12}e^{i2q_1\tilde{d}}e^{-iq_1(z-a)}\} & a\leqslant z\leqslant b \\ \varepsilon_{Si}^{-1}(\Omega)L_\Omega(1+r_{01})(1+r_{12})/(1+r_{01}r_{12}e^{i2q_1\tilde{d}})e^{iq_1\tilde{d}}e^{iq_2(z-b)} & z>b \end{cases}. \tag{17}$$

In this expression, $\tilde{d}\equiv b-a, L_\Omega$ is the average local field factor at frequency Ω and q_0, q_1, and q_2 are the z-components of the wave vectors in the three regions. Also, r_{01} and r_{12} are the p-polarized coefficients of reflection for the vacuum/metal and metal/Si interfaces, respectively and $r_p=(r_{01}+r_{12}e^{i2q_1\tilde{d}})/(1+r_{01}r_{12}e^{i2q_1\tilde{d}})$ is the p-polarized coefficient of reflection for the entire structure. It is understood that all these quantities are to be evaluated at the frequency Ω. Similarly, $f_x^\Omega(z)$ and $f_y^\Omega(z)$ denote tangential field distribution functions at frequency Ω. These functions are quite analogous to Eq. (17) in that $f_x^\Omega(z)$ and $f_y^\Omega(z)$ are obtained from $f_z^\Omega(z)$ simply by removing $\varepsilon^{-1}(\Omega)$ and $\varepsilon_{\mathrm{Si}}^{-1}(\Omega)$ and making the substitutions $r_{ij}^p\rightarrow -r_{ij}^p$ and $r_{ij}^p\rightarrow r_{ij}^s$, respectively, where $r_{ij}^{p,s}$ is the reflection coefficient between layers i and j, the superscript indicating p or s polarization. Due to the absence of $\varepsilon^{-1}(\Omega)$ and $\varepsilon_{\mathrm{Si}}^{-1}(\Omega)$ the z dependence of $f_x^\Omega(z)$ and $f_y^\Omega(z)$ is only via phase factors of the form $e^{\pm iq_i z}$ and in the limit $z/\lambda\approx 0$ these functions reduce to the constants $L_\Omega(1-r_p)$ and $L_\Omega(1+r_s)$, respectively. The discontinuities of the normal component of the electric field, however, yield $f_z^\Omega(z)\approx L_\Omega(1+r_p)\varepsilon^{-1}(\Omega;z)$ in this limit, where $\varepsilon^{-1}(\Omega;z)$ is the inverse dielectric function at z. The appearance of two different tangential field distribution functions $f_x^\Omega(z)$ and $f_y^\Omega(z)$ will complicate matters as they give rise to different matrix elements. It is easily seen, however, that to a high degree of accuracy the relation $f_y^\Omega(z)=\gamma_\Omega f_x^\Omega(z)$ with $\gamma_\Omega=[1+r_s(\Omega)]/[1-r_p(\Omega)]$ is valid. Hence, the y-matrix elements are simply proportional to the x-matrix elements in this approximation.

The induced nonlinear polarization $\mathbf{p}_{2\omega}(\mathbf{r})=\mathbf{p}_{2\omega}(z)e^{i2\mathbf{q}_\parallel\cdot\mathbf{r}}$ is related to the nonlinear current density through $\mathbf{j}_{2\omega}(z)=-2i\omega\mathbf{p}_{2\omega}(z)$ and, using Eq. (16), this quantity can be related to the incident field. In this manner, averaging Eq. (14) over a surface unit cell produces an effective z-dependent nonlinear susceptibility, i.e.,

$$\mathbf{p}_{2\omega}(z)=\varepsilon_0\vec{\chi}(z):\mathbf{E}^0_\omega\mathbf{E}^0_\omega, \tag{18}$$

where the elements of $\vec{\chi}(z)$ are given by

$$\chi_{ijk}(z)=\frac{1}{2i\varepsilon_0\omega^3 A_{UC}}\sum_{\lambda\mu\nu}\frac{1}{2\hbar\omega+i\hbar/\tau-E_{\nu\mu}} \times\left\{\frac{f_\mu-f_\lambda}{\hbar\omega+i\hbar/\tau-E_{\lambda\mu}}+\frac{f_\nu-f_\lambda}{\hbar\omega+i\hbar/\tau-E_{\nu\lambda}}\right\} \times\int_{UC}\mathbf{j}_{\mu\nu}(\mathbf{r})\cdot\hat{e}_i d^2r\int\mathbf{j}_{\nu\lambda}(\mathbf{r})\cdot\hat{e}_j f_j^\omega(z)d^3r \times\int\mathbf{j}_{\lambda\mu}(\mathbf{r})\cdot\hat{e}_k f_k^\omega(z)d^3r, \tag{19}$$

where A_{UC} is the area of a surface unit cell and the first integral is over this area. It is noted that taking the $\mathbf{q}_\parallel\rightarrow 0$ limit has reduced the problem of calculating nonlinear fields and polarizations to a one-dimensional one. In fact, taking the $\mathbf{q}_\parallel\rightarrow 0$ limit has the additional advantage of greatly simplifying the rotational symmetry properties of the nonlinear response since these are now solely determined by the material properties. For a (111) face the surface symmetry operations are reflections in three mirror planes and rotations by a multiple of 120° around the surface normal. By requiring invariance under these operations the number of independent tensor elements is reduced to four.[20] In a coordinate system with the x and y axes fixed by the requirement that the y axis be perpendicular to one of the mirror planes, these tensor elements are[20] χ_{xxx}, χ_{zxx}, χ_{xzx} and χ_{zzz}. In general, the plane of incidence is rotated by an angle ϕ around the surface normal with respect to this coordinate system. In terms of this angle the nonlinear polarization is given by[20]

$$\mathbf{p}_{2\omega}(z)=-\varepsilon_0\chi_{xxx}(z)|\gamma_\omega|^2|E^0_\omega|^2\sin 3\phi\hat{y}\quad s_\omega\ \text{to}\ s_{2\omega}$$

$$\mathbf{p}_{2\omega}(z)=\varepsilon_0\chi_{xxx}(z)|E^0_\omega|^2\cos^2\theta\sin 3\phi\hat{y}\quad p_\omega\ \text{to}\ s_{2\omega}$$

$$\mathbf{p}_{2\omega}(z) = -\varepsilon_0 \chi_{xxx}(z) |\gamma_\omega|^2 |E_\omega^0|^2 \cos 3\phi \hat{x} + \varepsilon_0 \chi_{zxx}(z) |\gamma_\omega|^2 |E_\omega^0|^2 \hat{z} \quad s_\omega \text{ to } p_{2\omega}.$$

$$\mathbf{p}_{2\omega}(z) = \varepsilon_0 [\chi_{xxx}(z) \cos 3\phi \cos^2\theta + \chi_{xzx}(z) \sin 2\theta] |E_\omega^0|^2 \hat{x} \quad p_\omega \text{ to } p_{2\omega} + \varepsilon_0 [\chi_{zxx}(z) \cos^2\theta + \chi_{zzz}(z) \sin^2\theta] |E_\omega^0|^2 \hat{z}. \tag{20}$$

The reflected second-harmonic field is of the form $\mathbf{E}_{2\omega}(z) = \hat{e}_{2\omega} E_{2\omega} e^{-iq_0 z}$, where $\hat{e}_{2\omega}$ is a polarization vector and the scalar amplitude $E_{2\omega}$ can be related to the nonlinear polarization via a vectorial Green's function according to

$$E_{2\omega} = \int_{-\infty}^{\infty} \mathbf{G}^{2\omega}(z) \cdot \mathbf{p}_{2\omega}(z) dz. \tag{21}$$

In the case of the *s*-polarized second-harmonic field only the *y* element of the Green's function is relevant, whereas both *x* and *z* components contribute to the magnitude of the reflected field in the *p*-polarized case. Using results from Ref. 21 the explicit expression for these functions can be written

$$G_i^{2\omega}(z) = -i\omega \sqrt{\frac{\mu_0}{\varepsilon_0}} f_i^{2\omega}(z) \times \begin{cases} 1 & i = x \\ 1 & i = y, \\ \tan\theta & i = z \end{cases} \tag{22}$$

where $f_i^{2\omega}(z)$ is, in fact, the *i* component of the field distribution of a field at frequency 2ω. By utilizing the various Green's functions we finally arrive at the following expressions for the second-harmonic reflectivity $\eta = I_{2\omega}/(I_\omega)^2$:

$$\eta_{s \text{ to } s} = 2\omega^2(\mu_0/c) |\chi_{xxx}|^2 |\gamma_\omega|^4 |\gamma_{2\omega}|^2 \sin^2 3\phi$$

$$\eta_{p \text{ to } s} = 2\omega^2(\mu_0/c) |\chi_{xxx}|^2 |\gamma_{2\omega}|^2 \cos^2\theta \sin^2 3\phi$$

$$\eta_{s \text{ to } p} = 2\omega^2(\mu_0/c) |\chi_{xxx} \cos 3\phi - \chi_{zxx} \tan\theta|^2 |\gamma_\omega|^4$$

$$\eta_{p \text{ to } p} = 2\omega^2(\mu_0/c) |\chi_{xxx} \cos 3\phi \cos^2\theta + \chi_{xzx} \sin 2\theta + \chi_{zxx} \cos^2\theta \tan\theta + \chi_{zzz} \sin^2\theta \tan\theta|^2. \tag{23}$$

Here, effective *z*-independent tensor elements have been introduced according to

$$\chi_{ijk} = \frac{1}{2i\varepsilon_0 \omega^3 A_{UC}} \sum_{\lambda\mu\nu} \frac{1}{2\hbar\omega + i\hbar/\tau - E_{\nu\mu}} \times \left\{ \frac{f_\mu - f_\lambda}{\hbar\omega + i\hbar/\tau - E_{\lambda\mu}} + \frac{f_\nu - f_\lambda}{\hbar\omega + i\hbar/\tau - E_{\nu\lambda}} \right\} \times \int_{UC} \mathbf{j}_{\mu\nu}(\mathbf{r}) \cdot \hat{e}_i f_i^{2\omega}(z) d^3r \int \mathbf{j}_{\nu\lambda}(\mathbf{r}) \cdot \hat{e}_j f_j^{\omega}(z) d^3r \times \int \mathbf{j}_{\lambda\mu}(\mathbf{r}) \cdot \hat{e}_k f_k^{\omega}(z) d^3r. \tag{24}$$

In evaluating this expression it should be remembered that the threefold summation over $\{\lambda,\mu,\nu\}$ actually entails a summation over $\{l,m,n\}$ as well as a threefold integral over $\{\mathbf{k}'',\mathbf{k}',\mathbf{k}\}$. Explicit expressions for the effective tensor elements are presented in the Appendix. From these results it is seen that two different types of matrix elements denoted $X_{\alpha\beta}^{\Omega}$ and $Z_{\alpha\beta}^{\Omega}$ [defined in Eq. (A3)] are involved. These matrix elements describe electron motion parallel and perpendicular to the surface, respectively. In the case of an electric field with a slow spatial variation, the selection rules implied by these matrix elements are quite different, however. This follows from the fact that in the dipole approximation the field distribution functions $f_x^{\Omega}(z)$ and $f_z^{\Omega}(z)$ reduce to the constants $L_\Omega(1-r_p)$ and $L_\Omega(1+r_p)$, respectively. Hence, using Eq. (A3) it is seen that $X_{\alpha\beta}^{\Omega} = L_\Omega(1-r_p)\delta_{\alpha\beta}$ and $Z_{\alpha\beta}^{\Omega} = L_\Omega(1+r_p)\langle\varphi_\alpha|z|\varphi_\beta\rangle$ in the dipole approximation. The Kronecker delta obtained for $X_{\alpha\beta}^{\Omega}$ is readily seen to imply that $\chi_{xxx} = \chi_{xzx} = \chi_{zxx} = 0$. It is concluded that only χ_{zzz} survives in this limit and, in fact, this conclusion holds even if only field variations of the form $e^{\pm iq_i z}$ are neglected. It follows that the magnitude of χ_{zzz} is expected to be much larger than those of χ_{xxx}, χ_{xzx}, and χ_{zxx}.

IV. NUMERICAL CALCULATIONS

At a first glance, it would seem that the expressions given in Eq. (23) are readily applicable for a comparison between theory and experiment. A number of subtleties must be considered, however. One difficulty is related to the fact that Eq. (23) is an expression for the second-harmonic signal generated by a perfectly smooth QW of arbitrary thickness *d*. Even under ideal growing conditions, i.e., layer-by-layer growth, the physical QW will display a minimum roughness of one monolayer, however, since the topmost atomic layer will cover only a certain fraction of the surface area. Hence, a given surface will contain a mixture of areas with, e.g., *n* and $n+1$ ML's, respectively. The fraction of the total surface area with precisely *n* ML's will be denoted P_n and it may be considered as the probability of encountering a thickness of precisely *n* ML's in random sampling. More generally, deviations from layer-by-layer growth will be found and, hence, a binomial distribution of the form

$$P_n = \binom{N}{n} p^n (1-p)^{N-n} \tag{25}$$

will be appropriate. In using this expression, *d* should now be regarded as the *mean* thickness and the parameters *N* and *p* must be expressed in terms of *d* and the width of the distribution $\sigma(d)$, which may, in general, depend on *d*. The appropriate relations are $p = 1 - \sigma^2/d$ and $N = d^2/(d-\sigma^2)$. The requirement that *N* must be an integer is only met at certain values of *d* and, hence, quadratic interpolation is used between these values. Next, an open question is whether the contributions from areas of different thickness should be added coherently, i.e., by adding the generated currents, or incoherently by adding the intensities. Obviously, two contributions which are imaged onto different points in the detection system (photomultiplier tube) should be added incoherently and for simplicity we will assume that this case is

valid generally. Consequently, for a sample with mean thickness d, the second-harmonic reflectivity $\eta(d)$ should be calculated from

$$\eta(d)=\sum_n \eta(n\delta)P_n, \tag{26}$$

where $\delta=2.3$ Å is the thickness of a monolayer for a (111) surface. An attractive feature of the present microscopic approach is that the calculations are based on either microscopic quantities, e.g., Fourier components of the electronic potential, or quantities that can be determined independently such as the dielectric constant of the Si substrate used to evaluate field distribution functions. An exception, however, is the relaxation time τ, which determines the width of the resonances in Eq. (24) and, in fact, also influences the screening of the fields inside the metallic layer. The difficulty encountered in the present context is that τ cannot be expected to attain its bulk value in a thin film due to surface scattering. Hence, the relaxation time varies with thickness in a manner that is unknown except that a monotonically increasing dependence is expected on physical grounds (see Ref. 5, however). If the rate of scattering events is the sum of surface and other contributions, τ^{-1} will consist of a constant term τ_0^{-1} and monotonically decreasing surface term, i.e.,

$$\tau^{-1}(d)=\tau_0^{-1}+\tau_s^{-1}(d). \tag{27}$$

The thickness dependent functions $\tau^{-1}(d)$ and $\sigma(d)$ will be determined from a fitting procedure. In fact, these quantities may serve as measures of the film quality since they mainly derive from the density of scatterers and the surface roughness. In addition, it should be noted that the location of resonances in the thickness or frequency dependence of the calculated second-harmonic signal is insensitive to the values of these parameters since they only determine the width of the resonances.

The linear optical properties of the structure determine the field distribution functions. Thus, these properties actually play an essential role in determining the nonlinear optical properties as well. As shown above, only χ_{zzz} survives if field variations through the structure are neglected and so the magnitudes of χ_{xxx}, χ_{xzx}, and χ_{zxx} will be critically dependent on the actual field variations and thereby on the linear optical properties. The dielectric properties of the Si substrate can be taken from experimental bulk data.[22] For an ultrathin metallic layer, however, the bulk dielectric function cannot be applied directly primarily because of the modified relaxation time given by Eq. (27). Generally, the dielectric function $\varepsilon(\omega)$ contains an interband term $\varepsilon'(\omega)$ as well as an intraband contribution given by the Drude formula:

$$\varepsilon(\omega)=\varepsilon'(\omega)-\frac{\omega_p^2}{\omega(\omega+i/\tau)}, \tag{28}$$

where ω_p is the unscreened plasma frequency. The interband contribution can be approximated by its bulk value and so the thickness dependence of the expression above is only via τ. In the low frequency range far from interband resonances $\varepsilon'(\omega)$ is approximately constant and by fitting the experimental bulk data[22] to Eq. (28) the following bulk values are found: $\hbar\omega_p=8.72$ eV and $\hbar/\tau_{\rm bulk}=0.074$ eV for Ag and $\hbar\omega_p=9.39$ eV and $\hbar/\tau_{\rm bulk}=0.079$ eV for Au. As usual, the optical relaxation frequency is slightly larger than the corresponding dc resistivity value.[15] Using these values $\varepsilon'(\omega)$ is calculated by subtracting the bulk intraband contribution from the experimental values of $\varepsilon(\omega)$. In turn, the values of $\varepsilon'(\omega)$ and ω_p allow us to compute the correction to the dielectric constant due to Eq. (27). Given $\varepsilon(\omega)$, the average local field factor is simply taken as $L_\omega=\{2+[\varepsilon(\omega)+\varepsilon_{Si}(\omega)]/2\}/3$. The remaining problem concerning the linear optical properties lies in determining the location of the dielectric boundaries a and b. In a simple picture, these would coincide with 0 and d. However, the delocalized metal electron density extends into the vacuum and Si regions, thus effectively broadening the screening region. The metal electron density is a weighted summation of contribution due to the different *bound* electron states. Hence, the spatial extent of the bound states approximately coincides with the screening region. The more delocalized excited states may extend beyond the screening region, however. For simplicity, though, we will assume that all wave functions at the metal/vacuum interface are sufficiently well localized that the metal dielectric function applies throughout this region. Similarly, the location of the dielectric boundary b is expected to lie somewhat beyond the ''electronic'' boundary d. From the fitting procedure described below, the values $b=d+2\delta$ (Au) and $b=d+1.5\delta$ (Ag) have been found to yield satisfactory agreement with experiment.

Our fitting procedure relies on the fact that χ_{zzz} is the dominant tensor element. Hence, the *xxx*, *xzx*, and *zxx* contributions to $\eta_{p\,\text{to}\,p}$ are safely neglected. Accordingly, by fitting the experimental p to p conversion efficiency to the simplified expression for $\eta_{p\,\text{to}\,p}$ found in this approximation a simple procedure for determining the parameters $\tau(d)$ and $\sigma(d)$ as well as the remaining parameters is obtained. It should be noticed that this method is only weakly dependent on band structure effects, since χ_{zzz} is governed by the motion of electrons perpendicular to the interfaces. In fact, as argued in the Appendix, band-structure effects may be neglected altogether in the calculation of χ_{zzz} in the case of low incident photon energies. Using the result for χ_{zzz} derived in the Appendix we find that

$$\hbar/\tau(d)=\begin{cases}0.15+0.3\exp(-d/45\,\text{Å})+0.9\exp[-(d/15\,\text{Å})^2] & \text{for Au}\\ 0.10+0.2\exp(-d/35\,\text{Å})+0.8\exp[-(d/12\,\text{Å})^2] & \text{for Ag}\end{cases} \tag{29}$$

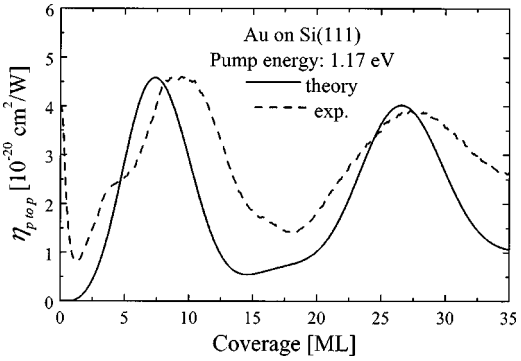

FIG. 2. Comparison between theoretical and experimental p to p second-harmonic reflectivity $\eta_{p\,to\,p}$ as a function of film coverage. The curves are for Au on Si(111).

in units of eV. Similarly, the fitting procedure yields $\sigma^2(d) = d \cdot 0.25\,\text{Å}$ for Au and $\sigma^2(d) = d \cdot 0.13\,\text{Å}$ for Ag. At a thickness of 10 ML these values correspond to a roughness of approximately 1 ML (Au) and 0.7 ML (Ag).

We now turn to a comparison between theoretical and experimental results. All experimental curves are obtained using the procedure described in Refs. 7 and 8. A comparison between experiments and the theoretical curves obtained using the above relations is shown in Figs. 2 and 3. A quite convincing agreement between experimental and theoretical curves is noted. The location of theoretical and experimental resonances agree within a few monolayers except for a small resonance in the Ag/Si curve at $d \approx 1$ ML, which is seen experimentally but not reproduced in the theoretical curves. Most likely, this resonance cannot be attributed to QW transitions but rather to a modified surface state[4] or a localized plasmon resonance in isolated Ag islands.[23] In addition to the agreement between the experimental and theoretical thickness dependence, the magnitudes of the calculated and measured signals are estimated to be in rough agreement. In order to confirm that χ_{zzz} dominates the p to p response the ratios $|\chi_{zxx}/\chi_{zzz}|$ and $|\chi_{xzx}/\chi_{zzz}|$ (not averaged over different thicknesses) are calculated as illustrated in Fig. 4. A large value of these ratios is seen at low coverage. This feature is a result of small χ_{zzz} values rather than large χ_{zxx} or χ_{xzx} values, however. On an absolute scale, all tensor elements are small at low coverage. At larger film thickness, χ_{zxx} and χ_{xzx} contribute on average only about 10–20% and, hence, can be neglected as assumed. The close agreement between experimental and theoretical results is a clear indication that

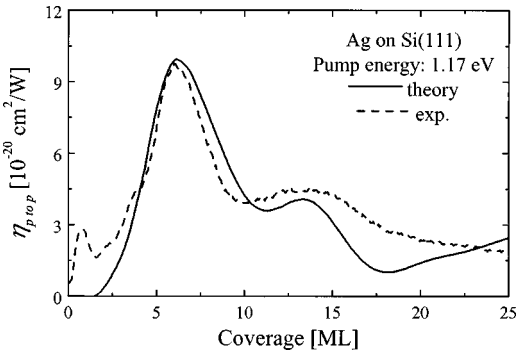

FIG. 3. Theoretical and experimental p to p second-harmonic reflectivity for Ag on Si(111).

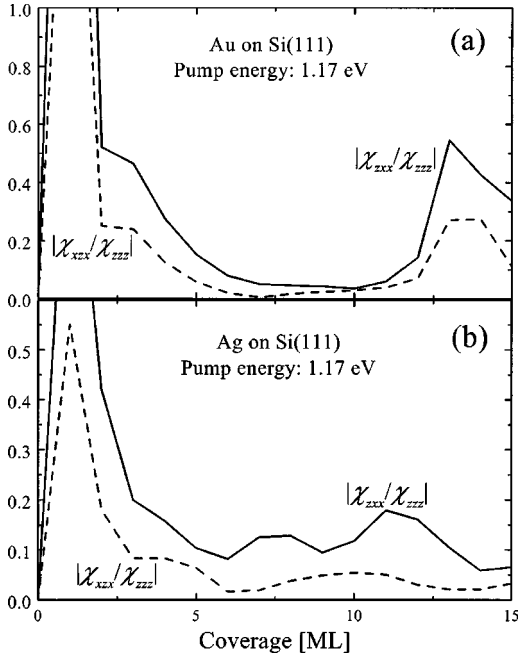

FIG. 4. Illustration of the relative contribution to the nonlinear response due to the *xzx* and *zxx* tensor elements. (a): Au on Si(111), (b) Ag on Si(111).

QW states are indeed responsible for the nonlinear response. Our fitted values for the width of the height distribution indicate, however, that the structures display a substantial variation in film thickness along the surface. Unfortunately, the lateral size distribution of the plateaus cannot be deduced from the present analysis. The fact that QW resonances are seen, however, demonstrates that the typical lateral dimension of a plateau is much larger than the height. This conclusion is supported by atomic force microscope scans, which for an Ag film of $d = 11$ ML have revealed triangular plateaus with typical sizes of several hundred nanometers.

We will now turn to the theoretical evaluation of χ_{xxx}. We will only investigate the thin film case for which the perturbation calculation of band-structure effects is sufficient. Expressions for the relevant matrix elements are presented in the Appendix. In order to obtain a reliable result a symmetrical set of $\mathbf{G}_\parallel$ vectors must be used. For the set of bulk-$\mathbf{G}$ vectors we may take that of Ref. 18 excluding the element $(-2, -2, 0)$, which breaks the symmetry. The nonvanishing parallel components of the remaining 15 $\mathbf{G}$ vectors can be written

$$\mathbf{G}_\parallel^{(k)} = \frac{2\pi}{a}\sqrt{\frac{8}{3}}[\cos(k\pi/3)\hat{x} + \sin(k\pi/3)\hat{y}],$$

$$k \in \{0,1,2,3,4,5\}, \tag{30}$$

where a is the lattice constant (4.08 Å). These vectors are seen to be of equal length and oriented at an angle k times 60° with respect to the x axis. The corresponding potential matrix elements given by Eq. (11) read

$$V_{\alpha\beta}(\mathbf{G}_\parallel^{(k)}) = \int_0^d \varphi_\alpha(z)\{V_{111}\exp(iG_\perp^{(1)}z) + V_{200}\exp(iG_\perp^{(2)}z)\}\varphi_\beta(z)dz \equiv W_{\alpha\beta}, \tag{31}$$

$$k \in \{0,2,4\},$$

where $G_{\perp}^{(1)} = -2\pi/(\sqrt{3}a)$ and $G_{\perp}^{(2)} = -2G_{\perp}^{(1)}$. For $k \in \{1,3,5\}$ the result $V_{\alpha\beta}(\mathbf{G}_{\parallel}^{(k)}) = W_{\alpha\beta}^{*}$ is found. These simple relations allow us to perform the summation over $\mathbf{G}_{\parallel}$ analytically. The $\mathbf{G}_{\parallel}$ dependence of χ_{xxx} is via the potential matrix elements and, as shown in the Appendix, the $T_a^{(p)}(Z)$ functions defined by Eq. (A14). As exemplified by Eq. (A17) in the case $p=2$ the $T_a^{(p)}(Z)$ functions have a simple dependence on the azimuthal angle of $\mathbf{G}_{\parallel}$. In the present case the azimuthal angle of $\mathbf{G}_{\parallel}^{(k)}$ is simply $k\pi/3$ and all $\mathbf{G}_{\parallel}^{(k)}$ are of equal length. This allows us to express the result of the summation over $\mathbf{G}_{\parallel}$ through the quantities

$$S_a^{(p)}(Z) \equiv \frac{\Sigma_{\mathbf{G}_{\parallel}} V_{l\alpha}^{*}(\mathbf{G}_{\parallel}) V_{m\beta}(\mathbf{G}_{\parallel}) T_a^{(p)}(Z)}{i\,\mathrm{Im}\{W_{l\alpha}^{*} W_{m\beta}\}}, \tag{32}$$

where, for instance,

$$S_a^{(3)}(Z) = 6K_a^{(33)}(Z) - \frac{9}{2}K_a^{(31)}(Z). \tag{33}$$

The $K_a^{(pq)}(Z)$ functions are defined in Eq. (A16) and similarly simple results are obtained for $p=0$, 1 and 2. To illustrate the importance of these results we consider the summation over $\mathbf{G}_{\parallel}$ for the particular component of χ_{xxx} considered in the Appendix. The final contribution from this part can be expressed as

$$\chi_{xxx}^{(1)} = -\frac{e^3\hbar^3}{8\pi^2\varepsilon_0\omega^3 m^3} \sum_{lmn,\alpha\beta} \frac{X_{\alpha\beta}^{\omega} X_{mn}^{2\omega} X_{nl}^{\omega}\,\mathrm{Im}\{W_{l\alpha}^{*} W_{m\beta}\}}{(E_{m\beta}-E_{l\alpha})(2\hbar\omega + i\hbar/\tau - E_{nm})}$$
$$\times \left\{ \frac{S_m^{(2)}(E_{l\alpha}) - S_l^{(2)}(E_{l\alpha})}{\hbar\omega + i\hbar/\tau - E_{lm}} + \frac{S_n^{(2)}(E_{l\alpha}) - S_l^{(2)}(E_{l\alpha})}{\hbar\omega + i\hbar/\tau - E_{nl}} - \frac{S_m^{(2)}(E_{m\beta}) - S_l^{(2)}(E_{m\beta})}{\hbar\omega + i\hbar/\tau - E_{lm}} - \frac{S_n^{(2)}(E_{m\beta}) - S_l^{(2)}(E_{m\beta})}{\hbar\omega + i\hbar/\tau - E_{nl}} \right\}. \tag{34}$$

The fact that the $\mathbf{G}_{\parallel}$ summation can be performed analytically in the present case leads to a large (sixfold) reduction in computation time.

The *xxx* element is probed in *s* to *s* and *p* to *s* SHG experiments and in the present work we focus on $\eta_{p\,\mathrm{to}\,s}$ for which experimental curves exist. A comparison is only attempted in the range $d \leqslant 7$ ML since the steeply rising theoretical curve beyond this range indicates a breakdown of the perturbation approach. At coverages below 7 ML, however, the agreement is reasonable as seen in Fig. 5 in which the theoretical *p* to *s* curves (not averaged) are compared to experimental data. The agreement is obviously less convincing than in the *p* to *p* case, which we partially attribute to the use of the perturbation approach. In addition, the assumption of perfect structural order may be questioned. The differences between experiment and theory cannot be explained by our neglect of χ_{zxx} and χ_{xzx} since these elements do not contribute to the *p* to *s* response. As in the *p* to *p* case, an additional resonance, which is not accounted for in the QW model is seen in the experimental Ag curve around 1 ML. The resonance around 4–5 ML is reproduced by theory for both Au and Ag, however. We believe that this fact demonstrates the soundness of the model and indicates that the entire coverage range can be modeled by going beyond the perturbation approach.

The relatively good agreement between theory and experiment allow us to use our model to locate the physical origin of the nonlinear response. The calculated response may be broken down into contributions stemming from a particular combination of occupied and excited levels. For Au on Si the coverage dependence of the energy eigenvalues of these levels is illustrated in Fig. 6. The topmost part of this graph is the pseudocontinuum of Si conduction-band states. Below this region the bound QW states are seen. The resonant transitions are mainly between the bound states and the Si conduction-band continuum due to the close proximity between the Fermi level and the conduction band edge. In fact, if the states above the conduction-band edge are ignored the calculated value of $\eta_{p\,\mathrm{to}\,p}$ is reduced to approximately 10–20% and the location of the resonances is shifted significantly. In order to determine the nonlinear response due to a particular occupied state we rewrite Eq. (A5) in the form $\chi_{zzz} = \Sigma_{mocc}\chi_{zzz}(m)$, where $\chi_{zzz}(m)$ is the contribution from the *m*th occupied state. This form of the equation is obtained by relabeling indices according to $l \leftrightarrow m$ and $n \leftrightarrow m$ in the parts proportional to H_l and H_n, respectively. The *m*th term of the resulting expression is then proportional to $\theta(E_F$

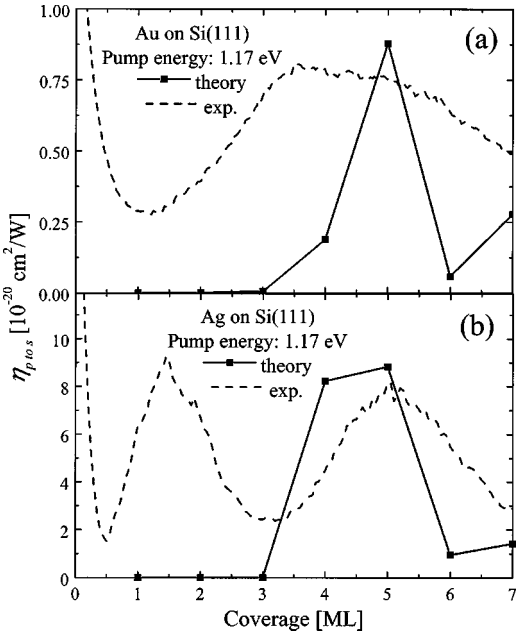

FIG. 5. Theoretical and experimental *p* to *s* second-harmonic reflectivity for Au (a) and Ag (b) on Si(111). The theoretical curve is obtained using the perturbation approach explained in the text.

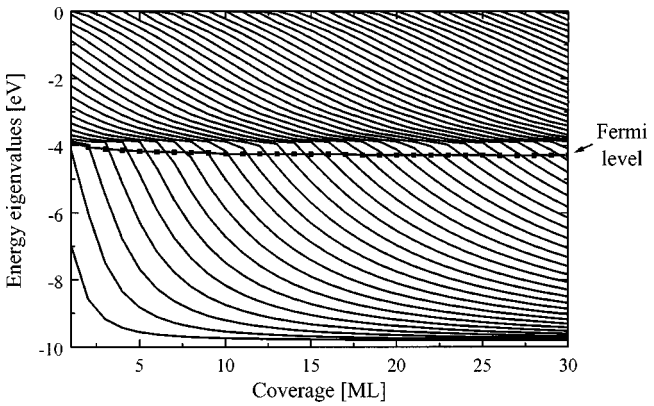

FIG. 6. The thickness dependence of the energy levels for an Au quantum well on Si(111). The thickness-dependent Fermi level is indicated by the additional line. The dramatic change around $\div 3.9$ eV marks the Si conduction band edge.

$-E_m$). In this manner, we obtain the contribution to the nonlinear response due the mth occupied state in the form $\eta_{p\,\mathrm{to}\,p}(m) \propto |\chi_{zzz}(m)|^2$. The thickness dependence of $\eta_{p\,\mathrm{to}\,p}(m)$ for Au on Si shown in Fig. 7. Note that $\eta_{p\,\mathrm{to}\,p} \neq \Sigma_{mocc}\eta_{p\,\mathrm{to}\,p}(m)$ since interference terms between different levels are absent. Hence, $\eta_{p\,\mathrm{to}\,p}(m)$ can be regarded as the nonlinear response of a fictitious QW containing only the mth occupied state. From Fig. 7 it is seen that $\eta_{p\,\mathrm{to}\,p}(m)$ peaks in a particular coverage range. This range corresponds approximately to the interval $\hbar\omega \leqslant E_F - E_m \leqslant 2\hbar\omega$. The figure clearly shows that the first resonance, obtained theoretically around $d \sim 7$ ML, contains contribution from levels in the range $m \in \{2,...,10\}$. Similarly, the second resonance at $d \sim 27$ ML is composed from contributions in the range $m \in \{12,...,24\}$. The range in which a particular level contributes is seen to increase with m. This is due to the fact that the magnitude of the slope of E_m versus d is decreasing with m. Hence, a particular level is resonantly excited in a broader range. Based on these results it is concluded that the resonances observed experimentally cannot be attributed to a single electronic transition but rather to simultaneously excited transitions between several highly localized occupied states and the unoccupied states of the conduction-band continuum.

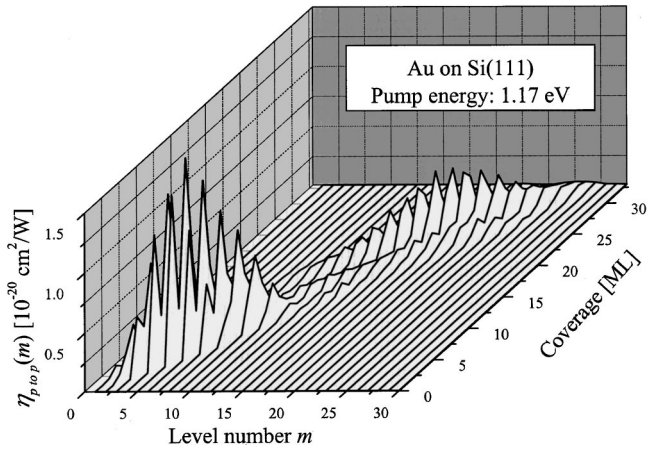

FIG. 7. Plot of the contribution to the nonlinear response due to the mth occupied quantum well level in the case of Au on Si(111).

SUMMARY

In summary, a complete microscopic framework describing second-harmonic generation from quantum well states formed in thin-metal layers deposited on semiconductors has been formulated. The major problem of obtaining sufficiently accurate electronic eigenstates taking band-structure effects into account has been solved by means of a truncated expansion technique. Furthermore, analytical solutions at the perturbation level have been obtained. Based on these electronic wave functions, explicit expressions for all relevant elements of an effective second-order response tensor have been derived. Theoretical and experimental curves of p to p and p to s second-harmonic signals versus film thickness have been compared for Au on Si(111) and Ag on Si(111). For both structures, convincing agreement between theoretical and measured p to p curves was found over the entire range of film thicknesses under study. The experimentally observed resonances are reproduced by theory and the absolute magnitude of the signal is roughly correct. In the case of p to s second-harmonic signals the perturbation calculation was shown to yield satisfactory agreement with experiments in a limited thickness range (0–7 monolayers) beyond which the perturbation approach is found to break down. From a comparison of the various contributions to the nonlinear response it is concluded that the resonances cannot be ascribed to transitions involving only a single occupied quantum well state. Rather, the resonances are composed from simultaneously excited transitions between several localized quantum well states and the unoccupied states of the conduction-band continuum.

APPENDIX: EFFECTIVE TENSOR ELEMENTS

In this Appendix, explicit expressions for the various effective tensor elements are derived. Among these elements, χ_{xxx} is a special one since it vanishes in amorphous materials. Hence, even in free-electron-like materials band structure effects must be included in order to calculate χ_{xxx}. Contrariwise, the remaining tensor elements (χ_{zxx}, χ_{xzx}, and χ_{zzz}) can be obtained without taking band structure effects into account. These effects are expected to be minor corrections as long as only states belonging to the s-p band are excited and therefore such a treatment presumably suffices. Hence, for these tensor elements wavefunctions of the simple form $\psi^{(0)}_{n\mathbf{k}}(\mathbf{r}) = (2\pi)^{-1}\varphi_n(z)e^{i\mathbf{k}\cdot\mathbf{r}}$ apply and the x- and z-matrix elements needed in Eq. (24) give

$$\begin{aligned}\int \mathbf{j}_{\nu\lambda}(\mathbf{r})\cdot\hat{x} f_x^{\Omega}(z)d^3r &= -\frac{e\hbar}{m}k_x\int_{-\infty}^{\infty}\varphi_n(z)\varphi_l(z)f_x^{\Omega}(z)dz\,\delta(\mathbf{k}-\mathbf{k}'') \\ &= -\frac{e\hbar}{m}k_x X_{nl}^{\Omega}\delta(\mathbf{k}-\mathbf{k}''),\end{aligned} \tag{A1}$$

and

$$\int \mathbf{j}_{\nu\lambda}(\mathbf{r})\cdot\hat{z}f_x^{\Omega}(z)d^3r = -\frac{e\hbar}{2im}\int_{-\infty}^{\infty}\left\{\varphi_n(z)\frac{d\varphi_l(z)}{dz} - \varphi_l(z)\frac{d\varphi_n(z)}{dz}\right\}f_z^{\Omega}(z)dz\,\delta(\mathbf{k}-\mathbf{k}'') = \frac{e}{i\hbar}E_{nl}Z_{nl}^{\Omega}\delta(\mathbf{k}-\mathbf{k}''), \tag{A2}$$

where we have introduced the abbreviations

$$X_{\alpha\beta}^{\Omega} = \int_{-\infty}^{\infty}\varphi_\alpha(z)\varphi_\beta(z)f_x^{\Omega}(z)dz$$

and

$$Z_{\alpha\beta}^{\Omega} = \int_{-\infty}^{\infty}\varphi_\alpha(z)\varphi_\beta(z)\int_0^z f_z^{\Omega}(z')dz'dz. \tag{A3}$$

In the last equality in Eq. (A2) integration by parts has been applied. In the low-temperature limit, the Fermi functions $f_{n\mathbf{k}}$ are simple step functions $f_{n\mathbf{k}}=\theta(E_F-E_n-\hbar^2k^2/2m)$. Due to the simple $\mathbf{k}$-dependence of the matrix elements above only the following two integrals involving $f_{n\mathbf{k}}$ are needed:

$$\frac{1}{4\pi^2}\int f_{n\mathbf{k}}d^2k = \frac{m}{2\pi\hbar^2}H_n$$

and

$$\frac{1}{4\pi^2}\int f_{n\mathbf{k}}k_x^2d^2k = \frac{m^2}{4\pi\hbar^4}H_n^2, \tag{A4}$$

where $H_n=(E_F-E_n)\theta(E_F-E_n)$. It should be noted that the Bloch wave vector $\mathbf{k}$ is not restricted to the first Brillouin zone in the present treatment. When these results are utilized, the final expressions for the effective tensor elements read

$$\chi_{zzz} = +\frac{e^3m}{4\pi\varepsilon_0\hbar^5\omega^3}\sum_{lmn}\frac{E_{lm}E_{mn}E_{nl}}{2\hbar\omega+i\hbar/\tau-E_{nm}}\times\left\{\frac{H_m-H_l}{\hbar\omega+i\hbar/\tau-E_{lm}}+\frac{H_n-H_l}{\hbar\omega+i\hbar/\tau-E_{nl}}\right\}Z_{mn}^{2\omega}Z_{nl}^{\omega}Z_{lm}^{\omega}, \tag{A5}$$

$$\chi_{zxx} = -\frac{e^3}{8\pi\varepsilon_0\hbar^3\omega^3}\sum_{lmn}\frac{E_{mn}}{2\hbar\omega+i\hbar/\tau-E_{nm}}\times\left\{\frac{H_m^2-H_l^2}{\hbar\omega+i\hbar/\tau-E_{lm}}+\frac{H_n^2-H_l^2}{\hbar\omega+i\hbar/\tau-E_{nl}}\right\}Z_{mn}^{2\omega}X_{nl}^{\omega}X_{lm}^{\omega}, \tag{A6}$$

$$\chi_{xzx} = -\frac{e^3}{8\pi\varepsilon_0\hbar^3\omega^3}\sum_{lmn}\frac{E_{nl}}{2\hbar\omega+i\hbar/\tau-E_{nm}}\times\left\{\frac{H_m^2-H_l^2}{\hbar\omega+i\hbar/\tau-E_{lm}}+\frac{H_n^2-H_l^2}{\hbar\omega+i\hbar/\tau-E_{nl}}\right\}X_{mn}^{2\omega}Z_{nl}^{\omega}X_{lm}^{\omega}. \tag{A7}$$

The calculation of χ_{xxx} proceeds along similar lines as those above. Band structure effects cannot be ignored, however, and the mathematical details are complicated by the corresponding modification of the wave functions. The advantage of the perturbational approach lies in the fact that all three $\mathbf{k}$-integrations can be performed analytically. In order to demonstrate this, the general expression for the electronic states, given by Eq. (2), is taken as a starting point. Using this result, the second integral in the Eq. (24) yields

$$\int \mathbf{j}_{\nu\lambda}(\mathbf{r})\cdot\hat{x}f_x^{\omega}(z)d^3r = -\frac{e\hbar}{m}\sum_{\mathbf{G}_\parallel,\mathbf{K}_\parallel}(k_x+G_x)\int_{-\infty}^{\infty}\varphi_{n\mathbf{k}}^*(\mathbf{G}_\parallel,z)\times\varphi_{l\mathbf{k}''}(\mathbf{K}_\parallel,z)f_x^{\omega}(z)dz\,\delta(\mathbf{k}+\mathbf{G}_\parallel-\mathbf{k}''-\mathbf{K}_\parallel). \tag{A8}$$

A similar expression is obtained for the last integral. Due to the delta functions in these expressions the integrals over $\mathbf{k}'$ and $\mathbf{k}''$ are readily carried out. The resulting form of Eq. (24) then contains a sixfold summation over the set of $\mathbf{G}_\parallel$'s. One of these summations can be eliminated using conservation of crystal momentum, which is required by the integral over the surface unit cell in Eq. (24). In the remaining five-fold summation a zeroth-order term in which all $\mathbf{G}_\parallel$'s equal zero exists. As expected, it may readily be shown that this contribution to χ_{xxx} vanishes. This follows from the fact that in this case all three integrals in Eq. (24) contribute a factor k_x. Hence, the integrand is an odd function of k_x and the integral vanishes. Due to the absence of a zeroth-order term it can be shown that the leading contribution to χ_{xxx} contains 15 terms, in each of which two $\mathbf{G}_\parallel$'s are nonzero. The calculations are rather tedious and in order to avoid unnecessary details we choose as an example to present only the results for the first among the 15 terms. This contribution can be written

$$\chi_{xxx}^{(1)} = \frac{ie^3\hbar^3}{8\pi^2\varepsilon_0\omega^3m^3}\sum_{\mathbf{G}_\parallel}\int\sum_{lmn}F_{lmn}(\mathbf{k},\mathbf{k},\mathbf{k})k_x^2\times(k_x+G_x)X_{mn}^{2\omega}X_{nl}^{\omega}X_{lm}^{\omega}(3)d^2k, \tag{A9}$$

where we have introduced the notation

$$F_{lmn}(\mathbf{k},\mathbf{k}',\mathbf{k}'') = \frac{1}{2\hbar\omega+i\hbar/\tau-E_{n\mathbf{k}}+E_{m\mathbf{k}'}}\times\left\{\frac{f_{m\mathbf{k}'}-f_{l\mathbf{k}''}}{\hbar\omega+i\hbar/\tau-E_{l\mathbf{k}''}+E_{m\mathbf{k}'}}+\frac{f_{n\mathbf{k}'}-f_{l\mathbf{k}''}}{\hbar\omega+i\hbar/\tau-E_{n\mathbf{k}}+E_{l\mathbf{k}''}}\right\}, \tag{A10}$$

and

$$X_{\alpha\beta}^{\Omega}(3) = \int_{-\infty}^{\infty}\varphi_{\alpha\mathbf{k}}^*(\mathbf{G}_\parallel,z)\varphi_{\beta\mathbf{k}}(\mathbf{G}_\parallel,z)f_x^{\Omega}(z)dz. \tag{A11}$$

By applying Eq. (13) and a partial fraction expansion of the $\mathbf{k}$-dependent denominator of the matrix element $X_{lm}^{\omega}(3)$ it is found that

$$\chi^{(1)}_{xxx}=\frac{ie^3\hbar^3}{8\pi^2\varepsilon_0\omega^3m^3}\sum_{\mathbf{G}_\parallel}\sum_{lmn,\alpha\beta}\frac{V^*_{l\alpha}(\mathbf{G}_\parallel)V_{m\beta}(\mathbf{G}_\parallel)X^{\omega}_{\alpha\beta}X^{2\omega}_{mn}X^{\omega}_{nl}}{E_{m\beta}-E_{l\alpha}}\times[A^{(2)}_{lmn}(E_{l\alpha})-A^{(2)}_{lmn}(E_{m\beta})], \tag{A12}$$

where we have introduced the abbreviation

$$A^{(p)}_{lmn}(Z)\equiv\int F_{lmn}(\mathbf{k},\mathbf{k},\mathbf{k})\frac{k_x^p(k_x+G_x)^{3-p}}{Z+\frac{\hbar^2}{2m}k^2-\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}_\parallel|^2}d^2k. \tag{A13}$$

The $\mathbf{k}$ dependence of $F_{lmn}(\mathbf{k},\mathbf{k},\mathbf{k})$ is via the Fermi functions $f_{a\mathbf{k}}$ and, hence, integrals of the form

$$T^{(p)}_a(Z)=\int f_{a\mathbf{k}}\frac{k_x^p(k_x+G_x)^{3-p}}{Z+\frac{\hbar^2}{2m}k^2-\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}_\parallel|^2}d^2k, \tag{A14}$$

are needed. This notation allows us to write

$$A^{(p)}_{lmn}(Z)=\frac{1}{2\hbar\omega+i\hbar/\tau-E_{nm}}\left\{\frac{T^{(p)}_m(Z)-T^{(p)}_l(Z)}{\hbar\omega+i\hbar/\tau-E_{lm}}+\frac{T^{(p)}_n(Z)-T^{(p)}_l(Z)}{\hbar\omega+i\hbar/\tau-E_{nl}}\right\}. \tag{A15}$$

By writing $G_x=G_\parallel\cos\delta$ and using polar coordinates and the low temperature limit the $T^{(p)}_a(Z)$ functions can be further decomposed in terms of the integrals

$$K^{(pq)}_a(Z)=\left(\frac{2m}{\hbar^2}\right)^{5/2}E_\parallel^{3/2}\theta(E_F-E_a)\times\int_0^{k_a}\int_0^{2\pi}\frac{k^{p+1}\cos^q\varphi}{Z/E_\parallel-1-2k\cos\varphi}d\varphi dk, \tag{A16}$$

where $E_\parallel=\hbar^2G_\parallel^2/(2m)$, E_F is the thickness-dependent Fermi energy that can be calculated from the charge neutrality condition, $\theta(x)$ is the step function and $k_a=\sqrt{(E_F-E_a)/E_\parallel}$. For instance,

$$T^{(2)}_a(Z)=\{\cos\delta(\cos^2\delta-3\sin^2\delta)K^{(33)}_a(Z)+3\sin^2\delta\times K^{(31)}_a(Z)+\sin^2\delta K^{(20)}_a(Z)+(\cos^2\delta-\sin^2\delta)\times K^{(22)}_a(Z)\}. \tag{A17}$$

The $K^{(pq)}_a(Z)$ integrals are readily evaluated using tables[24] or analytical computer software. Finally, the remaining contributions to χ_{xxx} are evaluated in a similar fashion.

[1] C. Weisbuch and B. Vinter, *Quantum Semiconductor Structures. Fundamentals and Applications* (Academic, New York, 1991).
[2] J. E. Ortega, F. J. Himpsel, G. J. Mankey, and R. F. Willis, Phys. Rev. B **47**, 1540 (1993).
[3] A. Kirilyuk, Th. Rasing, R. Megy, and P. Beauvillain, Phys. Rev. Lett. **77**, 4608 (1996).
[4] A. L. Wachs, A. P. Shapiro, T. C. Hsieh, and T. C. Chiang, Phys. Rev. B **33**, 1460 (1986).
[5] M. Jalochowski, M. Strozak, and R. Zdyb, Electron Technol. **31**, 291 (1998).
[6] K. Pedersen and P. Morgen, J. Phys.: Condens. Matter **9**, 9497 (1997).
[7] K. Pedersen, T. G. Pedersen, T. B. Kristensen, and P. Morgen, Appl. Phys. B: Lasers Opt. **68**, 637 (1999).
[8] T. G. Pedersen, K. Pedersen, and T. B. Kristensen, Phys. Rev. B **60**, R13 997 (1999).
[9] T. Rasing, Appl. Phys. A: Solids Surf. **59**, 531 (1994).
[10] A. V. Petukhov and A. Liebsch, Surf. Sci. **294**, 381 (1993).
[11] T. A. Luce and K. H. Bennemann, Phys. Rev. B **58**, 15 821 (1998).
[12] T. A. Luce, W. Hubner, A. Kirilyuk, Th. Rasing, and K. H. Bennemann, Phys. Rev. B **57**, 7377 (1998).
[13] J. A. Appelbaum and D. R. Hamann, Phys. Rev. B **6**, 2166 (1972).
[14] J. A. Appelbaum and D. R. Hamann, Rev. Mod. Phys. **48**, 479 (1976).
[15] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).
[16] Winfried Mönch, *Semiconductor Surfaces and Interfaces* (Springer, Berlin, 1993).
[17] N. V. Smith and L. F. Mattheiss, Phys. Rev. B **9**, 1341 (1974).
[18] N. V. Smith, Phys. Rev. B **19**, 5019 (1979).
[19] O. Keller, Phys. Rev. B **33**, 990 (1986).
[20] J. E. Sipe, D. J. Moss, and H. M. van Driel, Phys. Rev. B **35**, 1129 (1987).
[21] J. E. Sipe, J. Opt. Soc. Am. B **4**, 481 (1987).
[22] *Handbook of Optical Properties of Solids*, edited by E. D. Palik (Academic, Orlando, 1985).
[23] Y. Borensztein, R. Alameh, and M. Roy, Phys. Rev. B **48**, 14 737 (1993).
[24] I. S. Gradsteyn and I. M. Ryznik, *Table of Integrals, Series and Products* (Academic, San Diego, 1994).