Theory of surface second-harmonic generation for semiconductors including effects of nonlocal operators

Sean M. Anderson,[1] Nicolas Tancogne-Dejean,[2,3] Bernardo S. Mendoza,[1,*] and Valérie Véniard[2,3]
[1]*Centro de Investigaciones en Optica, León, Guanajuato, México*
[2]*Laboratoire des Solides Irradiés, École Polytechnique, CNRS, CEA/DSM, 91128 Palaiseau, France*
[3]*European Theoretical Spectroscopy Facility (ETSF), Palaiseau, France*
(Received 4 December 2014; revised manuscript received 20 January 2015; published 5 February 2015)

We formulate a theoretical approach of surface second-harmonic generation from semiconductor surfaces based on the length gauge and the electron density operator. Within the independent particle approximation, the nonlinear second-order surface susceptibility tensor $\chi^{abc}(-2\omega;\omega,\omega)$ is calculated, including in one unique formulation (i) the scissors correction, needed to have the correct value of the energy band gap, (ii) the contribution of the nonlocal part of the pseudopotentials, routinely used in *ab initio* band-structure calculations, and (iii) the derivation for the inclusion of the cut function, used to extract the surface response. The first two contributions are described by spatially nonlocal quantum-mechanical operators and are fully taken into account in the present formulation. As a test case of the approach, we calculate $\chi^{xxx}(-2\omega;\omega,\omega)$ for the clean Si(001)2 × 1 reconstructed surface. The effects of the scissors correction and of the nonlocal part of the pseudopotentials are discussed in surface nonlinear optics. The scissors correction shifts the spectrum to higher energies though the shifting is not rigid and mixes the 1ω and 2ω resonances, and has a strong influence in the line shape. The effects of the nonlocal part of the pseudopotentials keeps the same line shape of $|\chi^{xxx}_{2\times1}(-2\omega;\omega,\omega)|$, but reduces its value by 15%–20%. Therefore the inclusion of the three aforementioned contributions is very important and makes our scheme unprecedented and opens the possibility to study surface second-harmonic generation with more versatility and providing more accurate results.

DOI: 10.1103/PhysRevB.91.075302 PACS number(s): 78.68.+m, 42.65.An, 71.15.Mb, 42.65.Ky

I. INTRODUCTION

In recent years, surface nonlinear optical spectroscopies, particulary surface second-harmonic generation (SSHG), have evolved as useful nondestructive and noninvasive tools to study surface and interface properties. These properties include atomic structure, phase transitions, adsorption of atoms, and many others [1–19]. Nowadays, SSHG spectroscopy is a crucial tool for research and development in microelectronics [20], semiconductors [21], nanomaterials [22], and many more recent areas of great scientific and commercial interest [23]. The high surface sensitivity of SSHG spectroscopy is due to the fact that within the dipole approximation the bulk SHG signal of centrosymmetric materials is identically zero. However, the SHG process can occur only at the surface where the inversion symmetry is broken. The bulk quadrupole contribution for centrosymmetric materials is different from zero, but usually it is very small [19], and we neglect it. Much of the foundation of surface science has been built from experiments involving emission or scattering of electrons from surfaces. These require ultrahigh vacuum (UHV) environments and provide no access to buried interfaces. However, SSHG is compatible with non-UHV conditions and has access to interfaces buried beneath transparent overlayers. Even when applied to surfaces in UHV, the light source and detectors can be aligned and used outside the vacuum chamber.

The usefulness of SSHG could be limited by the lack of microscopic theoretical understanding of the nonlinear spectra. The macroscopic phenomenological theory of SSHG, which relates the intensity, phase, and polarization of detected fields to nonlinear and linear susceptibilities at the material interface is now fully developed [19]. However, a microscopic theory that relates electronic-level structure to the nonlinear source polarization is still being developed [12,24–30]. Within the independent particle approximation (IPA), some frameworks for bulk SHG have been developed to study the nonlinear optical response of bulk materials [12,24–29]. In this paper, we put forward an approach to calculate the microscopic second-harmonic surface susceptibility that encompasses several theoretical features not taken into account before in the case of a surface.

The most used framework for *ab initio* calculations, density functional theory (DFT) within the local density approximation (LDA) [31], underestimates the energy band gap of semiconductors. It is well understood that one has to include the many-body interaction to correct for this underestimation of the gap. In this context, the so-called *GW* approximation [32] is known to correct the electronic gap of most semiconductors [33]. However, this can be a very expensive calculation and thus one uses the much simpler scissors operator scheme [34–36]. This allows us to "open" the DFT-LDA gap to its correct experimental or *GW* value for most bulk semiconductors. This approximation has already been used in linear optical calculations for surfaces [37], thus improving the agreement with experimental results. In this context, to correct for the underestimation of the energy band gap of semiconductors, Nastos *et al.* [38] used the "length gauge" or "$\mathbf{r}\cdot\mathbf{E}$" gauge to show how to correctly include the many-body corrections through the scissors operator in the second-harmonic (SH) susceptibility. Later, Cabellos *et al.* [39] elaborated a derivation of the "velocity gauge" or "$\mathbf{A}\cdot\mathbf{v}$" gauge properly including the scissors operator and proved gauge invariance with respect to the length gauge. From these

*bms@cio.mx

works, it is clear the length gauge is a much better starting point to obtain the surface SH (SSH) susceptibility, as will be elaborated in this paper. However, these considerations are only valid for bulk semiconductors.

Concerning the optical response of surfaces and interfaces, Reining *et al.* [40] introduced the concept of a cut function in order to obtain the surface SH susceptibility tensor. This cut function is required since one usually uses a slab approach when treating semi-infinite surface systems [40]. If the slab is centrosymmetric, the susceptibility tensor will be identically zero. The cut function is such that it separates the nonlinear response for the two surfaces of the slab avoiding the destructive interference between them giving a finite value that one identifies with the SSH susceptibility tensor. If the slab is not centrosymmetric, the cut function can be used to separate the different signals coming from either surface of the slab. Indeed, one of the main results of this paper is to show that the SSH susceptibility tensor obtained by using the cut function is correctly extracted from the slab. After Reining *et al.* [40], Refs. [18,41–44] followed upon this work and in particular Ref. [41] went into a detailed analysis of the different contributions to the SHG spectra of a surface and the nuanced relationship between bulk, surface, interband, intraband, 1ω and 2ω terms, and Ref. [45] developed a layer-by-layer analysis for the nonlinear responses of semiconductor systems, within a tight-binding framework. This model allows for obtaining results from selected regions of a system including the surface. However, in these references, the scissors operator is either excluded or incorrectly implemented. In the works that include this operator, the velocity gauge was used to derive the expressions for $\chi^{\text{abc}}(-2\omega;\omega,\omega)$. Nevertheless, a term in the time-dependent perturbation scheme necessary to satisfy the gauge invariance of $\chi^{\text{abc}}(-2\omega;\omega,\omega)$ was omitted. This was demonstrated in Ref. [39] where a comparison between the velocity and the length gauge was carried out.

Finally, DFT-LDA calculations are often based on the use of pseudopotentials. As it will be discussed in this paper, the presence of a nonlocal part of the pseudopotential introduces corrections to the momentum operator of the electron that have to be included with care in the SSH susceptibility. For the bulk counterpart, see, for instance, Refs. [46,47].

Therefore, within the IPA, the most complete approach for the calculation of the SSH susceptibility is one which includes (i) the scissors correction, (ii) the contribution of the nonlocal part of the pseudopotential, and (iii) the cut function. Therefore the goal of this paper is to derive a new expression within the length gauge for the SSH susceptibility tensor $\chi^{\text{abc}}(-2\omega;\omega,\omega)$ that includes the aforementioned contributions. The inclusion of these three contributions makes our scheme unprecedented and opens the possibility to study surface SHG with more versatility and providing accurate results.

The paper is organized as follows. In Sec. II, we present the relevant steps for deriving the surface second-order susceptibility tensor $\chi^{\text{abc}}(-2\omega;\omega,\omega)$ within the length gauge formalism. This derivation includes the addition of the terms mentioned above that have been absent in previous works. Also, our $\chi^{\text{abc}}(-2\omega;\omega,\omega)$ can be used for a layer-by-layer analysis if desired. Nevertheless, such an analysis could depend on the choice of the basis used to expand the wave functions. In Sec. III, we show the results of this reformulation with a study of a clean Si(001) surface with a 2×1 reconstruction, proving the correctness of our approach with a special test case, never exploited before. We compare results from before and after adding the different nonlocal contributions. Finally, in Sec. IV, we give our conclusions.

II. THEORY

In this section, we present the scheme used to calculate the surface second-order nonlinear response using the length gauge formalism and the electron density operator. Some of the results presented in Secs. II A and II B have already been discussed in earlier studies and can be applied to bulk and surface studies [39,48]. We present them in order to have a self-contained derivation. The terms presented in these sections are used in Sec. II C, where we derive the expression for the surface $\chi^{\text{abc}}(-2\omega;\omega,\omega)$.

A. Perturbative approach

We assume the IPA, a classical electromagnetic field, and quantum-mechanical matter. We neglect local field and excitonic effects. We can describe the system using the one electron density operator ρ, with which we can calculate the expectation value of a single-particle observable $\mathcal{O}$ as $\langle\mathcal{O}\rangle = \text{Tr}(\rho\mathcal{O})$, with $\mathcal{O}$ the associated quantum mechanical operator and Tr the trace. The density operator satisfies $i\hbar(d\rho/dt) = [H(t),\rho]$, with $H(t)$ as the total single-electron Hamiltonian, written as

$$H(t) = H_0 + H_I(t),$$

where H_0 is the unperturbed time-independent Hamiltonian, and $H_I(t)$ is the time-dependent potential energy due to the interaction of the electron with the electromagnetic field. To proceed with the solution of ρ, it is convenient to use the interaction picture, where a unitary operator $U = \exp(iH_0t/\hbar)$ transforms any operator $\mathcal{O}$ into $\tilde{\mathcal{O}} = U\mathcal{O}U^\dagger$. Even if $\mathcal{O}$ is time-independent, $\tilde{\mathcal{O}}$ is time-dependent through the explicit time dependence of U. The dynamical equation for $\tilde{\rho}$ is given by

$$i\hbar\frac{d\tilde{\rho}}{dt} = [\tilde{H}_I(t),\tilde{\rho}],$$

with solution

$$i\hbar\tilde{\rho}(t) = i\hbar\tilde{\rho}_0 + \int_{-\infty}^{t} dt'[\tilde{H}_I(t'),\tilde{\rho}(t')], \tag{1}$$

where $\tilde{\rho}_0 = \tilde{\rho}(t=-\infty)$ is the unperturbed density matrix. We look for the standard perturbation series solution, $\tilde{\rho}(t) = \tilde{\rho}^{(0)} + \tilde{\rho}^{(1)} + \tilde{\rho}^{(2)} + \cdots$, where the superscript denotes the order (power) with which each term depends on the perturbation $H_I(t)$. From Eq. (1), the Nth-order term is

$$\tilde{\rho}^{(N)}(t) = \frac{1}{i\hbar}\int_{-\infty}^{t} dt'[\tilde{H}_I(t'),\tilde{\rho}^{(N-1)}(t')]. \tag{2}$$

The series is generated by the unperturbed density operator $\tilde{\rho}^{(0)} \equiv \tilde{\rho}_0$, assumed to be the diagonal Fermi-Dirac distribution, $\langle n\mathbf{k}|\tilde{\rho}_0|n\mathbf{k}\rangle = f(\hbar\omega_n(\mathbf{k})) \equiv f_n$. For a clean, cold semiconductor $f_n = 1$ when n is a valence (v) or occupied band, and zero when n is a conduction (c) or empty band. We assume

this for the remainder of the paper. As we neglect spin-orbit coupling, the final expression for $\chi^{abc}(-2\omega;\omega,\omega)$ has to be multiplied by a factor of 2 to account for spin-degeneracy. The expectation values must satisfy $\langle\mathcal{O}\rangle = \mathrm{Tr}(\rho\mathcal{O}) = \mathrm{Tr}(\tilde{\rho}\tilde{\mathcal{O}})$.

We will look for the expectation value of the microscopic current density $\mathbf{J}$, given by

$$\mathbf{J} = \langle\mathbf{J}\rangle = \frac{e}{A}\mathrm{Tr}(\rho\dot{\mathbf{r}}),$$

where $\dot{\mathbf{r}}$ is the time derivative of the position operator of the electron with charge e, defined as

$$\mathbf{v} \equiv \dot{\mathbf{r}} = \frac{1}{i\hbar}[\mathbf{r}, H_0], \tag{3}$$

with $\mathbf{v}$ the velocity operator of the electron, and A the area of the unit cell. We calculate the polarization density $\mathbf{P}$, related to $\mathbf{J}$ by $\mathbf{J} = d\mathbf{P}/dt$. For a perturbing electromagnetic field, $\mathbf{E}(t) = \mathbf{E}(\omega)e^{-i\tilde{\omega}t} + \text{c.c.}$, where $\tilde{\omega} = \omega + i\eta$, and $\eta > 0$ adiabatically switches on the interaction, we write the second-order nonlinear polarization as

$$\mathcal{P}^a(2\omega) = \chi^{abc}(-2\omega;\omega,\omega)E^b(\omega)E^c(\omega), \tag{4}$$

where $\chi^{abc}(-2\omega;\omega,\omega)$ is the nonlinear susceptibility responsible for surface second-harmonic generation (SSHG). The superscripts in Eq. (4) denote Cartesian components, and if repeated are to be summed over. Without loss of generality, we will define $\chi^{abc}(-2\omega;\omega,\omega)$ to satisfy intrinsic permutation symmetry, $\chi^{abc}(-2\omega;\omega,\omega) = \chi^{acb}(-2\omega;\omega,\omega)$.

The unperturbed Hamiltonian is used to solve the Kohn-Sham equations [31] of density functional theory (DFT). It is convenient to work within the local density approximation (LDA), so we label the Hamiltonian with the corresponding LDA superscript. Any other approximation can be used (like the generalized gradient approximation) and our derivation remains the same. Then,

$$H_0^{\mathrm{LDA}}(\mathbf{r},\mathbf{p}) = \frac{p^2}{2m_e} + V(\mathbf{r},\mathbf{p})$$

with m_e the mass of the electron, $\mathbf{p} = -i\hbar\nabla$ its canonical momentum, and V the periodic crystal potential, where we neglect spin-orbit terms. To be more general in our derivation of $\chi^{abc}(-2\omega;\omega,\omega)$, we assume a contribution as is customary for most pseudopotentials, and then we replace V with

$$V^{\mathrm{ps}}(\mathbf{r},\mathbf{p}) = V(\mathbf{r}) + V^{\mathrm{nl}}(\mathbf{r},\mathbf{p}),$$

where $V(\mathbf{r})$ and $V^{\mathrm{nl}}(\mathbf{r},\mathbf{p})$ are the local and nonlocal parts, respectively. The argument $(\mathbf{r},\mathbf{p})$ is equivalent to the explicit $(\mathbf{r},\mathbf{r}')$ nonlocal notation [46]. In case of a local potential, i.e., $V = V(\mathbf{r})$, like that of all-electron schemes, we simply omit the contribution of $V^{\mathrm{nl}}(\mathbf{r},\mathbf{p})$ from the results that we have derived.

It is well known that the use of the LDA leads to an underestimation of the band gap. A standard procedure to correct for this is to use the "scissors approximation," where the conduction bands are rigidly shifted in energy so that the band gap corresponds to the accepted experimental electronic band gap [34–36]. This is often in fairly good agreement with the GW band gap based on a more sophisticated calculation [49]. The LDA wave functions are used since they produce band structures with dispersion relations similar to those predicted by the GW. Mathematically, the scissors (nonlocal) operator S is added to the unperturbed or unscissored Hamiltonian H_0^{LDA},

$$H_0^{\Sigma}(\mathbf{r},\mathbf{p}) = H_0^{\mathrm{LDA}}(\mathbf{r},\mathbf{p}) + S(\mathbf{r},\mathbf{p}),$$

where

$$S(\mathbf{r},\mathbf{p}) = \hbar\Delta\sum_n\int d^3k(1-f_n)|n\mathbf{k}\rangle\langle n\mathbf{k}|, \tag{5}$$

with $\hbar\Delta$ the rigid ($\mathbf{k}$-independent) energy correction to be applied. The unscissored and scissored Hamiltonians satisfy

$$H_0^{\mathrm{LDA}}(\mathbf{r},\mathbf{p})\psi_{n\mathbf{k}}(\mathbf{r}) = \hbar\omega_n^{\mathrm{LDA}}(\mathbf{k})\psi_{n\mathbf{k}}(\mathbf{r}),$$
$$H_0^{\Sigma}(\mathbf{r},\mathbf{p})\psi_{n\mathbf{k}}(\mathbf{r}) = \hbar\omega_n^{\Sigma}(\mathbf{k})\psi_{n\mathbf{k}}(\mathbf{r}),$$

where the scissor-shifted energies $\omega_n^{\Sigma}(\mathbf{k})$ are given by

$$\omega_n^{\Sigma}(\mathbf{k}) = \omega_n^{\mathrm{LDA}}(\mathbf{k}) + (1-f_n)\Delta.$$

We emphasize that the scissored and unscissored Hamiltonian have the same eigenfunctions, where $\psi_{n\mathbf{k}}(\mathbf{r}) = \langle\mathbf{r}|n\mathbf{k}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$, are the real-space representations of the Bloch states $|n\mathbf{k}\rangle$ labeled by the band index n and the crystal momentum $\mathbf{k}$, and $u_{n\mathbf{k}}(\mathbf{r})$ are cell periodic.

B. Length gauge formalism

According to Ref. [46], we first start with the interaction Hamiltonian expressed in the velocity gauge, containing the nonlocal parts $V^{nl}(\mathbf{r},\mathbf{p})$ and $S(\mathbf{r},\mathbf{p})$. Within the dipole approximation and using a gauge transformation, it can be transformed into an effective Hamiltonian [50]

$$H_I(t) = -e\mathbf{r}\cdot\mathbf{E}(t). \tag{6}$$

The treatment of the position operator $\mathbf{r}$ for extended Bloch states is problematic and has been discussed in Refs. [51,52]. Following Ref. [48], we take the matrix elements of Eq. (2) with the $H_I(t)$ of Eq. (6), and we obtain $(\tilde{\rho}^{(1)}(t))_{nm} = B^b_{nm}E^b(\omega)e^{i(\omega^{\Sigma}_{nm}-\tilde{\omega})t}$, with

$$B^b_{nm} = \frac{e}{\hbar}\frac{f_{mn}r^b_{nm}}{\omega^{\Sigma}_{nm}-\tilde{\omega}}, \tag{7}$$

and

$$(\tilde{\rho}^{(2)}(t))_{nm} = \frac{e}{i\hbar}\frac{1}{\omega^{\Sigma}_{nm}-2\tilde{\omega}} \times\left[i\sum_q\left(r^b_{nq}B^c_{qm} - B^c_{nq}r^b_{qm}\right) - \left(B^c_{nm}\right)_{;k^b}\right] \times E^b(\omega)E^c(\omega)e^{i(\omega^{\Sigma}_{nm}-2\tilde{\omega})t}. \tag{8}$$

The position operator $\mathbf{r}$ is split into the *intraband* ($\mathbf{r}_i$) and *interband* ($\mathbf{r}_e$) parts, where $\mathbf{r} = \mathbf{r}_i + \mathbf{r}_e$. For $\mathbf{r}_e$, one uses

$$\langle n\mathbf{k}|\mathbf{r}_e|m\mathbf{k}'\rangle = (1-\delta_{nm})\delta(\mathbf{k}-\mathbf{k}')\boldsymbol{\xi}_{nm}(\mathbf{k}), \tag{9}$$

such that $\mathbf{r}_{e,nm} = 0$ for $n = m$. From Eq. (3) with $H_0 \to H_0^{\Sigma}$, we obtain

$$\mathbf{r}_{e,nm}(\mathbf{k}) = \boldsymbol{\xi}_{nm}(\mathbf{k}) \equiv \mathbf{r}_{nm}(\mathbf{k}) = \frac{\mathbf{v}^{\Sigma}_{nm}(\mathbf{k})}{i\omega^{\Sigma}_{nm}(\mathbf{k})}, \quad n \notin D_m, \tag{10}$$

where we defined $\omega^{\Sigma}_{nm}(\mathbf{k}) \equiv \omega^{\Sigma}_n(\mathbf{k}) - \omega^{\Sigma}_m(\mathbf{k})$, and D_m are all the possible degenerate m-states. For the intraband part, $\mathbf{r}_i$ only

appears in commutators during the derivation of the optical response. We use [28]

$$\langle n\mathbf{k}|[\mathbf{r}_i,\mathcal{O}]|m\mathbf{k}'\rangle = i\delta(\mathbf{k}-\mathbf{k}')(\mathcal{O}_{nm})_{;\mathbf{k}}, \tag{11}$$

where

$$(\mathcal{O}_{nm})_{;\mathbf{k}} = \nabla_{\mathbf{k}}\mathcal{O}_{nm}(\mathbf{k}) - i\mathcal{O}_{nm}(\mathbf{k})(\boldsymbol{\xi}_{nn}(\mathbf{k}) - \boldsymbol{\xi}_{mm}(\mathbf{k})) \tag{12}$$

is the generalized derivative of the operator $\mathcal{O}$. The vectors $\boldsymbol{\xi}_{nn}(\mathbf{k})$ are defined in Ref. [28] though they do not need to be calculated explicitly in what follows.

Before continuing, we derive a key result for the length gauge formulation. Again, using H_0^{Σ} in Eq. (3), we obtain

$$\mathbf{v}^{\Sigma} = \mathbf{v} + \mathbf{v}^{\mathrm{nl}} + \mathbf{v}^{S} = \mathbf{v}^{\mathrm{LDA}} + \mathbf{v}^{S}, \tag{13}$$

where we have defined

$$\begin{aligned} \mathbf{v} &= \frac{\mathbf{p}}{m_e}, \\ \mathbf{v}^{\mathrm{nl}} &= \frac{1}{i\hbar}[\mathbf{r}, V^{\mathrm{nl}}], \\ \mathbf{v}^{S} &= \frac{1}{i\hbar}[\mathbf{r}, S], \\ \mathbf{v}^{\mathrm{LDA}} &= \mathbf{v} + \mathbf{v}^{\mathrm{nl}}, \end{aligned} \tag{14}$$

and using $[r^a, p^b] = i\hbar\delta_{ab}$, where δ_{ab} is the Kronecker delta. Using Eq. (5), we obtain

$$\mathbf{v}^{S}_{nm} = i\Delta f_{mn}\mathbf{r}_{nm}, \tag{15}$$

with $f_{nm} \equiv f_n - f_m$, where we see that $\mathbf{v}^{S}_{nn} = 0$. From Eqs. (10) and (13), it follows that

$$\mathbf{r}_{nm}(\mathbf{k}) = \frac{\mathbf{v}^{\Sigma}_{nm}(\mathbf{k})}{i\omega^{\Sigma}_{nm}(\mathbf{k})} = \frac{\mathbf{v}^{\mathrm{LDA}}_{nm}(\mathbf{k})}{i\omega^{\mathrm{LDA}}_{nm}(\mathbf{k})}, \quad n \notin D_m. \tag{16}$$

The matrix elements of $\mathbf{r}_{nm}(\mathbf{k})$ are identical using either the LDA or scissored Hamiltonian, thus negating the need to label them. Of course, it is more convenient to calculate them through $\mathbf{v}^{\mathrm{LDA}}_{nm}(\mathbf{k})$, which includes only the contribution of $\mathbf{v}^{\mathrm{nl}}_{nm}(\mathbf{k})$. These can be readily calculated for fully separable nonlocal pseudopotentials in the Kleinman-Bylander form [53–56]. The advantage of using the electron density operator along with the length gauge formalism for calculating linear and nonlinear optical responses, for the scissored Hamiltonian, resides in the ease with which the scissors operator can be introduced into the calculation by simply using the unscissored LDA Hamiltonian, H_0^{LDA}, for the unperturbed system with $-e\mathbf{r}\cdot\mathbf{E}(t)$ as the interaction. We stress that within the length gauge, we need only replace ω_n^{LDA} with ω_n^{Σ} at the end of the derivation to obtain the scissored results for any susceptibility expression, whether linear or nonlinear [38].

We have used the fact that for a cold semiconductor $\partial f_n/\partial\mathbf{k} = 0$ and thus the intraband contribution to the linear term vanishes identically. Note that the indices in Eq. (8) are all different from each other. This is due to the f_{nm} factor in Eq. (7), and therefore $B^a_{nn} = 0$. The dependence on $\mathbf{k}$ of all quantities is implicitly understood from this point forward.

C. Layered current density

The approach we use to study the surface of a semi-infinite semiconductor crystal is as follows. Instead of using a semi-infinite system, we replace it by a supercell that consists of a finite slab of atomic layers and a vacuum region (see Fig. 1). This supercell is repeated to form a full three-dimensional crystalline structure. The slab itself consists of front, back, and subsurface regions, and in between these a region that is equivalent to the bulk of the system. In general, the surface of a crystal reconstructs or relaxes as the atoms move to find equilibrium positions. This is due to the fact that the otherwise balanced forces are disrupted when the surface atoms do not find their partner atoms that are now absent at the slab surface. To take the reconstruction or relaxation into account, we take "surface" to mean the true surface of the first layer of atoms and some of the atomic sublayers adjacent to it. Since the front and the back surfaces of the slab are usually identical the total slab is centrosymmetric. This would imply that $\chi^{\mathrm{slab,abc}} = 0$, so we must find a scheme in order to have a finite χ^{abc} representative of the surface. Even if the front and back surfaces of the slab are different, breaking the centrosymmetry and therefore giving an overall $\chi^{\mathrm{slab,abc}} \neq 0$; we still need a procedure to extract the front surface $\chi^{f,\mathrm{abc}}$ and the back surface $\chi^{b,\mathrm{abc}}$ from the slab susceptibility. We have omitted the frequency dependence of χ^{abc} for convenience of notation.

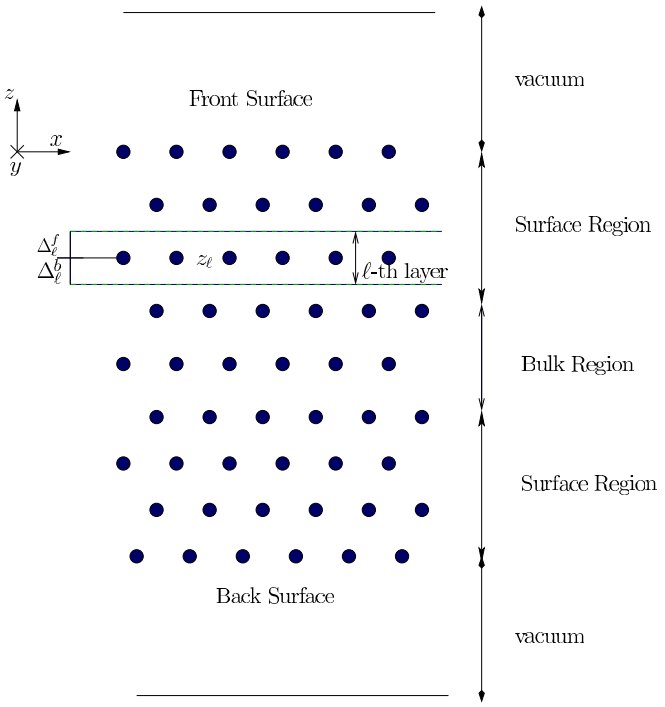

FIG. 1. (Color online) A sketch of the supercell. The atomic slab corresponds to the circles representing the atoms of the system.

A convenient way to accomplish the separation of the SH signal of either surface is to introduce a "cut function" $\mathcal{C}(z)$, which is usually taken to be unity over one half of the slab and zero over the other half [40]. In this case, $\mathcal{C}(z)$ will give the contribution of the side of the slab for which $\mathcal{C}(z) = 1$. As was done for the linear response [57], we can generalize this simple choice for $\mathcal{C}(z)$ by a top-hat cut function $\mathcal{C}^{\ell}(z)$ that selects a given layer,

$$\mathcal{C}^{\ell}(z) = \Theta\big(z - z_\ell + \Delta_\ell^b\big)\Theta\big(z_\ell - z + \Delta_\ell^f\big), \tag{17}$$

where Θ is the Heaviside function. Here, $\Delta_\ell^{f/b}$ is the distance that the ℓth layer extends towards the front (f) or back (b) from its z_ℓ position. We take z_ℓ to be at the center of an atom that belongs to layer ℓ, so the previous equation would give the ℓth atomic-layer contribution to the nonlinear optical response. $\Delta_\ell^f + \Delta_\ell^b$ is the thickness of layer ℓ (see Fig. 1).

To introduce the cut function $\mathcal{C}(z)$ in the calculation of $\chi^{\rm abc}$, we start from the operator for the electron current, $\mathbf{j}(\mathbf{r}) = \frac{e}{2}(\mathbf{v}^\Sigma|\mathbf{r}\rangle\langle\mathbf{r}| + |\mathbf{r}\rangle\langle\mathbf{r}|\mathbf{v}^\Sigma)$, that leads to

$$\mathbf{j}^{(N)}(\mathbf{r},t) = \mathrm{Tr}(\mathbf{j}(\mathbf{r})\rho^{(N)}(t)) = \int \frac{dk^3}{8\pi^3}\sum_{nm}\rho^{(N)}_{nm}(\mathbf{k};t)\mathbf{j}_{mn}(\mathbf{k};\mathbf{r}), \tag{18}$$

where

$$\mathbf{j}_{mn}(\mathbf{k};\mathbf{r}) = \frac{e}{2}(\langle m\mathbf{k}|\mathbf{v}^\Sigma|\mathbf{r}\rangle\langle\mathbf{r}|n\mathbf{k}\rangle + \langle m\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|\mathbf{v}^\Sigma|n\mathbf{k}\rangle). \tag{19}$$

Integrating the microscopic current $\mathbf{j}(\mathbf{r},t)$ over the entire slab gives the averaged microscopic current density, $\mathbf{J}(t)$. If we want the contribution from only one region of the unit cell towards the total current, we can integrate $\mathbf{j}(\mathbf{r},t)$ over the desired region. Then the contribution to the current density from the chosen region of the slab is given by

$$\frac{1}{A}\int d^3r\,\mathcal{C}(z)\,\mathbf{J}^{(N)}(\mathbf{r},t) \equiv \mathcal{J}^{(N)}(t),$$

where $\mathcal{J}^{(N)}(t)$ is the Nth-order current induced in the region specified by $\mathcal{C}(z)$. Therefore we define

$$e\boldsymbol{\mathcal{V}}^\Sigma_{mn}(\mathbf{k}) \equiv \int d^3r\,\mathcal{C}(z)\,\mathbf{j}_{mn}(\mathbf{k};\mathbf{r}), \tag{20}$$

to write the Fourier transform of Eq. (18) as

$$\mathcal{J}^{(N)}(2\omega) = \frac{e}{A}\int\frac{dk^3}{8\pi^3}\sum_{mn}\boldsymbol{\mathcal{V}}^\Sigma_{mn}(\mathbf{k})\rho^{(N)}_{nm}(\mathbf{k};2\omega), \tag{21}$$

that gives the induced microscopic current of the chosen region, to order N in the external perturbation. From Eqs. (20) and (19), we obtain

$$\begin{aligned}\boldsymbol{\mathcal{V}}^\Sigma_{mn}(\mathbf{k}) &= \frac{1}{2}\int \mathrm{d}^3r\,\mathcal{C}(z)[\langle m\mathbf{k}|\mathbf{v}^\Sigma|\mathbf{r}\rangle\langle\mathbf{r}|n\mathbf{k}\rangle \\ &\quad + \langle m\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|\mathbf{v}^\Sigma|n\mathbf{k}\rangle] \\ &= \frac{1}{2}\int \mathrm{d}^3r\,\mathcal{C}(z)[\psi_{n\mathbf{k}}(\mathbf{r})\mathbf{v}^{\Sigma *}\psi^*_{m\mathbf{k}}(\mathbf{r}) \\ &\quad + \psi^*_{m\mathbf{k}}(\mathbf{r})\mathbf{v}^\Sigma\psi_{n\mathbf{k}}(\mathbf{r})] \\ &= \int \mathrm{d}^3r\,\psi^*_{m\mathbf{k}}(\mathbf{r})\left[\frac{\mathcal{C}(z)\mathbf{v}^\Sigma + \mathbf{v}^\Sigma\mathcal{C}(z)}{2}\right]\psi_{n\mathbf{k}}(\mathbf{r}) \\ &= \int \mathrm{d}^3r\,\psi^*_{m\mathbf{k}}(\mathbf{r})\boldsymbol{\mathcal{V}}^\Sigma\psi_{n\mathbf{k}}(\mathbf{r}),\end{aligned}$$

where, we used the hermitian property of $\mathbf{v}^\Sigma$ and defined

$$\boldsymbol{\mathcal{V}}^\Sigma = \frac{\mathcal{C}(z)\mathbf{v}^\Sigma + \mathbf{v}^\Sigma\mathcal{C}(z)}{2}.$$

We see that the replacement

$$\mathbf{V} \to \boldsymbol{\mathcal{V}} = \frac{\mathcal{C}(z)\mathbf{V} + \mathbf{V}\mathcal{C}(z)}{2} \tag{22}$$

is all that is needed to change any of the electron velocity operators $\mathbf{V}$ to the new velocity operator $\boldsymbol{\mathcal{V}}$ that implicitly takes into account the contribution of the region of the slab given by $\mathcal{C}(z)$. We note that this modified operator is hermitian as it should [58]. The operator $\mathbf{V}$ could be any of those given by Eq. (13), thus

$$\begin{aligned}\boldsymbol{\mathcal{V}}^\Sigma &= \boldsymbol{\mathcal{V}}^{\rm LDA} + \boldsymbol{\mathcal{V}}^S, \\ \boldsymbol{\mathcal{V}}^{\rm LDA} &= \boldsymbol{\mathcal{V}} + \boldsymbol{\mathcal{V}}^{\rm nl}.\end{aligned}$$

To calculate $\boldsymbol{\mathcal{V}}^\Sigma_{nm}(\mathbf{k})$ we calculate the matrix elements of $\boldsymbol{\mathcal{V}}^{\rm LDA}$ and $\boldsymbol{\mathcal{V}}^S$ (separately) according to the expressions of Appendices A 2 and A 4. If not stated differently, calligraphic letters correspond to layer quantities.

To limit the SHG response to one surface, Eq. (22) for $\boldsymbol{\mathcal{V}}$ was proposed in Ref. [40] and later used in Refs. [18,42,44,45]. In this paper, we formally introduce the cut function $\mathcal{C}(z)$ for the second-harmonic optical response of semiconductor surfaces, from an average of the second-order polarization over the region of interest.

Using $\boldsymbol{\mathcal{J}} = d\boldsymbol{\mathcal{P}}/dt$ and Eq. (21), we obtain the SH polarization of a given region as

$$\boldsymbol{\mathcal{P}}^{(2)}(2\omega) = \frac{ie}{2A\tilde{\omega}}\int\frac{dk^3}{8\pi^3}\sum_{mn}\boldsymbol{\mathcal{V}}^\Sigma_{mn}(\mathbf{k})\rho^{(2)}_{nm}(\mathbf{k};2\omega), \tag{23}$$

and using Eqs. (4) and (8) leads to

$$\begin{aligned}&\chi^{\rm abc}(-2\omega;\omega,\omega) \\ &= \frac{e^2}{2A\hbar\tilde{\omega}}\int\frac{dk^3}{8\pi^3}\sum_{mn}\frac{\mathcal{V}^{\Sigma,{\rm a}}_{mn}(\mathbf{k})}{\omega^\Sigma_{nm\mathbf{k}} - 2\tilde{\omega}}\Big[-\big(B^{\rm c}_{nm}(\mathbf{k},\omega)\big)_{;k^{\rm b}} \\ &\quad + i\sum_q\left(r^{\rm b}_{nq}B^{\rm c}_{qm}(\mathbf{k},\omega) - B^{\rm c}_{nq}(\mathbf{k},\omega)r^{\rm b}_{qm}\right)\Big],\end{aligned} \tag{24}$$

which gives the susceptibility $\chi^{\rm abc}(-2\omega;\omega,\omega)$ of the layers of the slab specified by $\mathcal{C}(z)$. We mention that the units of $\chi^{\rm abc}(-2\omega;\omega,\omega)$ are m^2/V, as they should be for a surface SH susceptibility. Using Eq. (7), we split this equation into two contributions from the first and second terms on the right-hand side of Eq. (24):

$$\begin{aligned}\chi^{\rm abc}_i(-2\omega;\omega,\omega) &= -\frac{e^3}{A\hbar^2 2\tilde{\omega}}\int\frac{dk^3}{8\pi^3} \\ &\quad\times\sum_{mn}\frac{\mathcal{V}^{\Sigma,{\rm a}}_{mn}}{\omega^\Sigma_{nm} - 2\tilde{\omega}}\left(\frac{f_{mn}r^{\rm b}_{nm}}{\omega^\Sigma_{nm} - \tilde{\omega}}\right)_{;k^{\rm c}},\end{aligned} \tag{25}$$

related to intraband transitions, and

$$\begin{aligned}\chi^{\rm abc}_e(-2\omega;\omega,\omega) &= \frac{ie^3}{A\hbar^2 2\tilde{\omega}}\int\frac{dk^3}{8\pi^3}\sum_{qmn}\frac{\mathcal{V}^{\Sigma,{\rm a}}_{mn}}{\omega^\Sigma_{nm} - 2\tilde{\omega}} \\ &\quad\times\left(\frac{r^{\rm c}_{nq}r^{\rm b}_{qm}f_{mq}}{\omega^\Sigma_{qm} - \tilde{\omega}} - \frac{r^{\rm b}_{nq}r^{\rm c}_{qm}f_{qn}}{\omega^\Sigma_{nq} - \tilde{\omega}}\right),\end{aligned} \tag{26}$$

related to interband transitions. The generalized derivative in Eq. (25) is dealt with by the chain rule

$$\left(\frac{f_{mn}r^{\rm b}_{nm}}{\omega^\Sigma_{nm} - \tilde{\omega}}\right)_{;k^{\rm c}} = \frac{f_{mn}}{\omega^\Sigma_{nm} - \tilde{\omega}}\left(r^{\rm b}_{nm}\right)_{;k^{\rm c}} - \frac{f_{mn}r^{\rm b}_{nm}\Delta^{\rm c}_{nm}}{\left(\omega^\Sigma_{nm} - \tilde{\omega}\right)^2}, \tag{27}$$

where substituting H_0^Σ into Eq. (11) and then Eq. (16), we obtain

$$\left(\omega_{nm}^{\Sigma}\right)_{;k^{\mathrm{a}}} = \left(\omega_{nm}^{\mathrm{LDA}}\right)_{;k^{\mathrm{a}}} = v_{nn}^{\mathrm{LDA,a}} - v_{mm}^{\mathrm{LDA,a}} \equiv \Delta_{nm}^{\mathrm{a}}.$$

The apparent divergence as $\tilde{\omega} \to 0$ in Eqs. (25) and (26), is removed by a partial fraction expansion over $\tilde{\omega}$. Using time-reversal symmetry, an integration by parts to remove the square in the denominator of the second term of Eq. (27), and taking the limit of $\eta \to 0$, we obtain the following expressions for the imaginary parts of Eqs. (25) and (26),

$$\mathrm{Im}\left[\chi_{e,\omega}^{\mathrm{abc}}\right] = \frac{\pi|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{vc}\sum_{q\neq(v,c)}\frac{1}{\omega_{cv}^{\Sigma}}\left[\frac{\mathrm{Im}\left[\mathcal{V}_{qc}^{\Sigma,\mathrm{a}}\left\{r_{cv}^{\mathrm{b}}r_{vq}^{\mathrm{c}}\right\}\right]}{\left(2\omega_{cv}^{\Sigma}-\omega_{cq}^{\Sigma}\right)} - \frac{\mathrm{Im}\left[\mathcal{V}_{vq}^{\Sigma,\mathrm{a}}\left\{r_{qc}^{\mathrm{c}}r_{cv}^{\mathrm{b}}\right\}\right]}{\left(2\omega_{cv}^{\Sigma}-\omega_{qv}^{\Sigma}\right)}\right]\delta\left(\omega_{cv}^{\Sigma}-\omega\right), \tag{28a}$$

$$\mathrm{Im}\left[\chi_{i,\omega}^{\mathrm{abc}}\right] = \frac{\pi|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{cv}\frac{1}{\left(\omega_{cv}^{\Sigma}\right)^2}\left[\mathrm{Re}\left[\left\{r_{cv}^{\mathrm{b}}\left(\mathcal{V}_{vc}^{\Sigma,\mathrm{a}}\right)_{;k^{\mathrm{c}}}\right\}\right] + \frac{\mathrm{Re}\left[\mathcal{V}_{vc}^{\Sigma,\mathrm{a}}\left\{r_{cv}^{\mathrm{b}}\Delta_{cv}^{\mathrm{c}}\right\}\right]}{\omega_{cv}^{\Sigma}}\right]\delta\left(\omega_{cv}^{\Sigma}-\omega\right), \tag{28b}$$

$$\mathrm{Im}\left[\chi_{e,2\omega}^{\mathrm{abc}}\right] = -\frac{\pi|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{vc}\frac{4}{\omega_{cv}^{\Sigma}}\left[\sum_{v'\neq v}\frac{\mathrm{Im}\left[\mathcal{V}_{vc}^{\Sigma,\mathrm{a}}\left\{r_{cv'}^{\mathrm{b}}r_{v'v}^{\mathrm{c}}\right\}\right]}{2\omega_{cv'}^{\Sigma}-\omega_{cv}^{\Sigma}} - \sum_{c'\neq c}\frac{\mathrm{Im}\left[\mathcal{V}_{vc}^{\Sigma,\mathrm{a}}\left\{r_{cc'}^{\mathrm{c}}r_{c'v}^{\mathrm{b}}\right\}\right]}{2\omega_{c'v}^{\Sigma}-\omega_{cv}^{\Sigma}}\right]\delta\left(\omega_{cv}^{\Sigma}-2\omega\right), \tag{28c}$$

$$\mathrm{Im}\left[\chi_{i,2\omega}^{\mathrm{abc}}\right] = \frac{\pi|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{vc}\frac{4}{\left(\omega_{cv}^{\Sigma}\right)^2}\left[\mathrm{Re}\left[\mathcal{V}_{vc}^{\Sigma,\mathrm{a}}\left\{\left(r_{cv}^{\mathrm{b}}\right)_{;k^{\mathrm{c}}}\right\}\right] - \frac{2\mathrm{Re}\left[\mathcal{V}_{vc}^{\Sigma,\mathrm{a}}\left\{r_{cv}^{\mathrm{b}}\Delta_{cv}^{\mathrm{c}}\right\}\right]}{\omega_{cv}^{\Sigma}}\right]\delta\left(\omega_{cv}^{\Sigma}-2\omega\right), \tag{28d}$$

where we have split the interband and intraband 1ω and 2ω contributions and supressed the ω arguments for convenience of notation. The factor of 2 for spin degeneracy is not included in Eq. (28). The real part of each contribution can be obtained through a Kramers-Kronig transformation [59] and $\chi^{\mathrm{abc}} = \chi_{e,\omega}^{\mathrm{abc}} + \chi_{e,2\omega}^{\mathrm{abc}} + \chi_{i,\omega}^{\mathrm{abc}} + \chi_{i,2\omega}^{\mathrm{abc}}$. To fulfill the required intrinsic permutation symmetry, the {} notation symmetrizes the bc Cartesian indices, i.e., $\{u^{\mathrm{b}}s^{\mathrm{c}}\} = (u^{\mathrm{b}}s^{\mathrm{c}} + u^{\mathrm{c}}s^{\mathrm{b}})/2$, and thus $\chi^{\mathrm{abc}} = \chi^{\mathrm{acb}}$. The full expressions for $\chi^{\mathrm{abc}}(-2\omega;\omega,\omega)$, along with the various quantities involved in Eq. (28) are given in the Appendix A. We mention that if we take $\mathcal{C}(z) = 1$ through out, the layered matrix elements $\boldsymbol{\mathcal{V}}_{nm}^{\Sigma}$ become standard bulk-like $\mathbf{v}_{nm}^{\Sigma}$ matrix elements. We mention that in this case, Eq. (28) is equivalent to the expressions of Ref. [39], valid for bulk semiconductors.

Finally, we could also calculate the nonlinear surface susceptibility as

$$\boldsymbol{\chi}(-2\omega;\omega,\omega) = \sum_{\{\ell\}}\boldsymbol{\chi}^{\ell}(-2\omega;\omega,\omega), \tag{29}$$

where ℓ would denote a particular layer chosen through $\mathcal{C}^{\ell}(z)$ of Eq. (17) and $\{\ell\}$ is meant to be a chosen set of layers. For instance, one can take a single layer encompassing half of the slab, or take each atomic layer individually to the middle of the slab. For the first case, there is a single summand in Eq. (29). For the second case, there is a sum from $\ell = 1$, denoting the first layer right at the surface, to $\ell = N$, denoting the layer at the middle of the slab that behaves like a bulk layer. We remark that the value of N is not universal and the slab needs to have enough atomic layers in order to give converged results for $\boldsymbol{\chi}(-2\omega;\omega,\omega)$. We can use Eq. (29) for either the front or the back surface.

III. RESULTS

In this section, we present a relevant test case to check the consistency of our approach. We have selected a clean Si(001) surface with a 2×1 surface reconstruction. The slab for such a surface could be chosen to be centrosymmetric by creating the front and back surfaces with the same 2×1 reconstruction. However, we choose to terminate one of the surfaces with hydrogen producing an ideal terminated bulk Si surface. The H atoms simply saturate the dangling bonds of the bulklike Si atoms at the surface, as seen in Fig. 2. We take the z coordinate pointing out of the surface and the x coordinate along the crystallographic [011] direction is parallel to the dimmers. The idea behind this slab configuration is that the crystalline symmetry of the H terminated surface imposes that $\chi_{\mathrm{H}}^{xxx} = 0$. The 2×1 surface has no such restrictions, so $\chi_{2\times1}^{xxx} \neq 0$. This

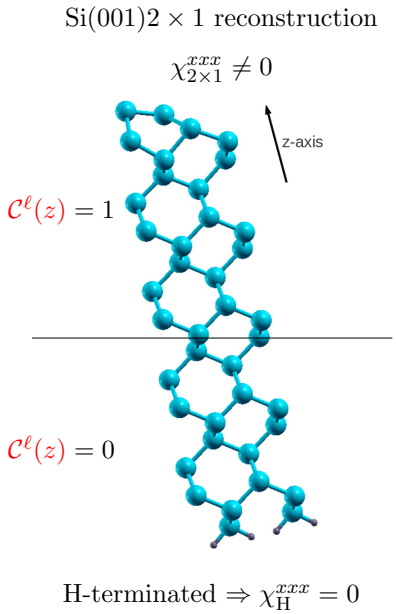

FIG. 2. (Color online) The slab shows a clean Si(001)2×1 front surface with an ideal terminated Si bulk back surface. The dangling bonds are H (small balls) saturated. This image depicts 12 Si atomic layers with one H atomic layer.

is due to the fact that along the y direction there is a mirror plane for the H-saturated surface, whereas for the 2×1 surface this mirror is lost as the dimers are asymmetric along x. Thus, calculating χ^{xxx} for the full slab, or the half-slab containing the 2×1 surface [60] should yield the same result since the contribution from the H saturated surface is zero regardless. We must check that the following relationship is satisfied for this particular slab:

$$\chi^{xxx}_{\text{half-slab}}(-2\omega;\omega,\omega) = \chi^{xxx}_{\text{full-slab}}(-2\omega;\omega,\omega),$$

where $\chi^{xxx}_{\text{half-slab}}(-2\omega;\omega,\omega)$ is calculated using $\mathcal{C}(z) = 1$ from the upper half containing the 2×1 surface reconstruction, as seen in Fig. 2, and $\chi^{xxx}_{\text{full-slab}}(-2\omega;\omega,\omega)$ is calculated using $\mathcal{C}(z) = 1$ through the full slab. We show the results for this comparison in the remainder of this section. Also, we checked that for the dihydride surface $\chi^{xxx}_{\text{half-slab}}(-2\omega;\omega,\omega) = 0$.

The self-consistent ground state and the Kohn-Sham states were calculated in the DFT-LDA framework using the plane-wave ABINIT code [61]. We used Troullier-Martins pseudopotentials [62] that are fully separable nonlocal pseudopotentials in the Kleinman-Bylander form [55]. The contribution of $\mathbf{v}^{\text{nl}}$ and $\boldsymbol{\mathcal{V}}^{\text{nl}}$ to Eq. (28) is carried out using the DP code [63]. The surfaces have been studied with the experimental lattice constant of 5.43 Å. Structural optimizations were performed with the ABINIT code [61]. The geometry optimization has been carried out in slabs of 12 atomic layers where the central four layers where fixed at the bulk positions. The structures were relaxed until the Cartesian force components were less than 5 meV/Å. The geometry optimization for the clean surface gives a dimer buckling of 0.721 Å, and a dimer length of 2.301 Å. For the Si(001)1 × 1:2H dihydride surface, we have obtained a Si-H bond distance of 1.48 Å. This results are in good agreement with previous theoretical studies [57,64]. The vacuum size is equivalent to one quarter the size of the slab, avoiding the effects produced by possible wave-function tunneling from the contiguous surfaces of the full crystal formed by the repeated supercell scheme [57].

Spin-orbit, local field, and electron-hole attraction [30] effects on the SHG process are all neglected. Although these are important factors in the optical response of a semiconductor, their efficient calculation is still theoretically and numerically challenging and under debate. This merits further study but is beyond the scope of this paper. For a given slab size, we find the converged spectra to obtain the relevant parameters. The most important of these are: an energy cutoff of 10 Ha for the 16, 24, and 32 layered slabs and 13 Ha for the 40 layer slab, an equal number of conduction and valence bands, and a set of 244 $\mathbf{k}$ points. The $\mathbf{k}$ points are used for the linear analytic tetrahedron method for evaluating the 3D Brillouin-zone (BZ) integrals where special care was taken to examine the double resonances of Eq. (28) [38]. Note that the Brillouin zone for the slab geometry collapses to a 2D zone, with only one $\mathbf{k}$ point along the z axis. All spectra were calculated with a Gaussian smearing of 0.15 eV.

We must evaluate $T^{\text{ab}}_{nm} = (i/\hbar)[r^{\text{b}}, v^{\text{nl,a}}]_{nm}$ in order to obtain Eqs. (A4) and (A7), which are required for Eq. (28). Computing second-order derivatives is required thus making the numerical procedure very time consuming. This adds significantly to the already lengthy time needed for the calculation of the $\mathbf{v}^{\text{nl}}$ contribution that is proportional only to the first order derivatives. Memory requirements are also increased for both $\mathbf{v}^{\text{nl}}$ and $[\mathbf{r},\mathbf{v}^{\text{nl}}]$. However, the contribution from $[\mathbf{r},\mathbf{v}^{\text{nl}}]$ is very small [50] and therefore we neglect it in this work.

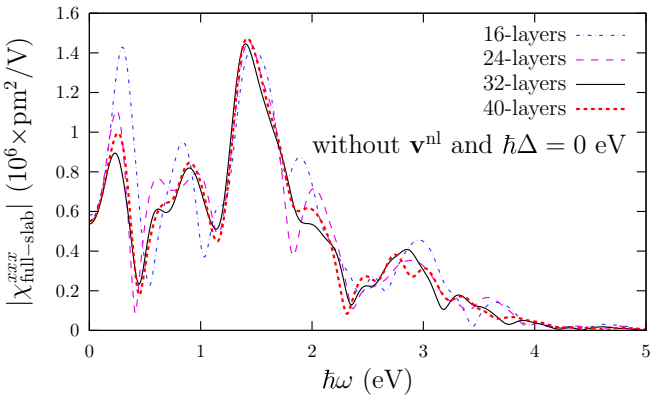

FIG. 3. (Color online) $|\chi^{xxx}_{\text{half-slab}}|$ vs $\hbar\omega$ for the slab with 16, 24, 32, and 40 atomic Si layers. The front surface is in a clean 2×1 reconstruction and the back surface is an ideal terminated bulk H-saturated dangling bonds (see Fig. 2).

A. Full-slab results

In Fig. 3, we show $|\chi^{xxx}_{\text{full-slab}}|$ for the slab with 16, 24, 32, and 40 Si atomic layers, without the contribution of $\mathbf{v}^{\text{nl}}$ and with no scissors correction. Since the clean Si(001) surface is 2×1, there are two atoms per atomic layer, thus the total number of atoms per slab is twice the number of atomic layers of the slab. In making the slabs larger, we add steps of 8 layers of bulk-like atomic positions. We note that the response differs substantially for 16 and 24 layers but is quite similar for 32 and 40 layers. As explained above, the calculation of the $\mathbf{v}^{\text{nl}}$ contribution is computationally expensive. A good compromise between the accuracy in the convergence of $\chi^{xxx}_{\text{full-slab}}$ as a function of the number of layers in the slab, and the computational expense is to consider the slab with 32 Si atomic layers as an accurate representation of our system.

B. Half slab versus full slab

In Fig. 4, we compare $\chi^{xxx}_{\text{half-slab}}$ versus $\chi^{xxx}_{\text{full-slab}}$ for the four different possibilities between including or not including the effects of $\mathbf{v}^{\text{nl}}$ or the scissors correction $\hbar\Delta$. For these results, we chose $\hbar\Delta = 0.5$ eV, that is, the GW gap reported in Refs. [65,66]. This is justified by the fact that the surface states of the clean Si(001) surface are rigidly shifted and maintain their dispersion relation with respect to LDA according to the GW calculations of Ref. [65]. We see that for all four instances the difference between responses is quite small. Indeed, when the value $|\chi^{xxx}|$ is large, the difference between the two is very small; when the value is small the difference increases only slightly, but the spectra is so close to zero that it is negligible. These differences would decrease as the number of atomic layers increases. We remark that 32 layers in the slab is more than enough to confirm that the extraction of the surface second-harmonic susceptibility from the 2×1 surface is readily possible using the formalism contained in Eq. (28). We have confirmed that for the dihydride surface $|\chi^{xxx}_{\text{half-slab}}| \approx 0$

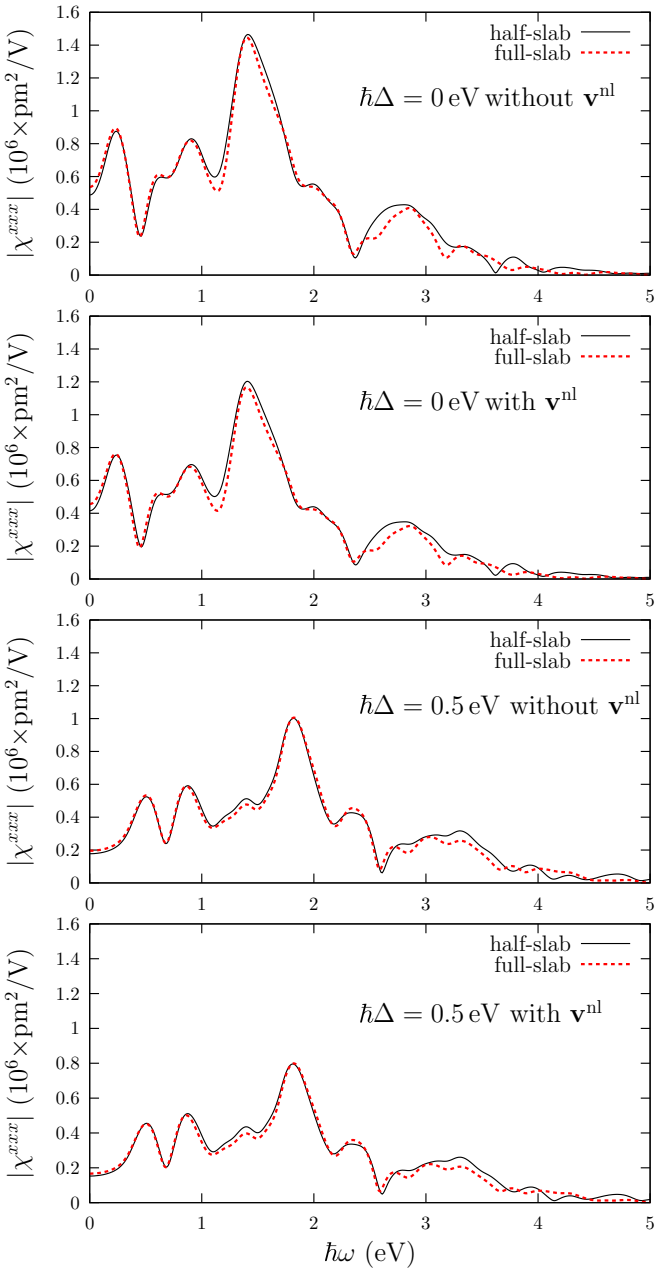

FIG. 4. (Color online) $\chi^{xxx}_{\text{half-slab}}$ and $\chi^{xxx}_{\text{full-slab}}$ vs $\hbar\omega$ for a slab with 32 atomic Si layers plus one H layer.

(not shown). This confirms the validity of our theory and is the main result of this paper; through the proposed layer formalism we can calculate the surface SH $\chi^{\text{abc}}(-2\omega;\omega,\omega)$ including the contribution of the nonlocal part of the pseudopotentials and the part of the many-body effects through the scissors correction. Our scheme should work for any slab.

C. Results for $\chi^{xxx}_{2\times1}(-2\omega;\omega,\omega)$

We proceed to explain some of the features seen in $|\chi^{xxx}_{2\times1}|$ that, as explained above, are obtained by calculating $|\chi^{xxx}_{\text{half-slab}}|$. First, from Fig. 4, we note a series of resonances that derive from 1ω and 2ω terms in Eq. (28). Notice that the 2ω resonances start below $E_g/2$ where E_g is the band gap

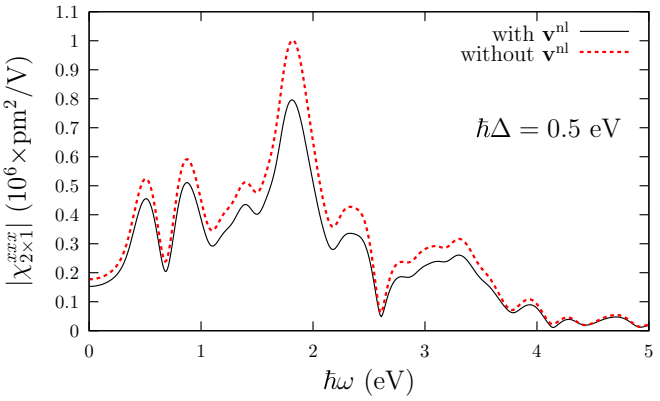

FIG. 5. (Color online) $\chi^{xxx}_{2\times1}$ vs $\hbar\omega$ for a slab with 32 atomic Si layers plus one H layer, with and without the contribution from $\mathbf{v}^{\text{nl}}$.

(0.53 eV for LDA and 1.03 eV if the scissor is used with $\hbar\Delta = 0.5$ eV). These resonances come from the electronic states of the 2×1 surface, that lie inside the bulk band gap of Si and are the well known electronic surface states [65]. In Fig. 5, we see that the effect of $\mathbf{v}^{\text{nl}}$ reduces the value of $|\chi^{xxx}_{2\times1}|$ by 15%–20% showing the importance of this contribution for a correct calculation of SSHG, in agreement with the analysis for bulk semiconductors [47]. However, the inclusion of $\mathbf{v}^{\text{nl}}$ does not changes the spectral shape of $|\chi^{xxx}_{2\times1}|$; this also can be confirmed from the cases of zero scissors correction from Fig. 4.

To see the effect of the scissors correction, we take two different finite values for $\hbar\Delta$. The first one with a value of $\hbar\Delta = 0.5$ eV, used in the above results, is the "average" GW gap taken from Ref. [65] that is in agreement with Ref. [66]. The second one with a value of $\hbar\Delta = 0.63$ eV is the "average" gap taken from Ref. [67], where more $\mathbf{k}$ points in the Brillouin zone were used to calculate its GW value. From Fig. 6, we note that the scissors correction shifts the spectra from its LDA value to higher energies as expected. However, contrary to the case of linear optics [39] the shift introduced by the scissors correction is not rigid, as pointed out in Ref. [38]. This is because the second-harmonic optical response mixes 1ω and 2ω transitions [see Eq. (28)], and accounts for the nonrigid shift. The reduction of the spectral strength is in agreement

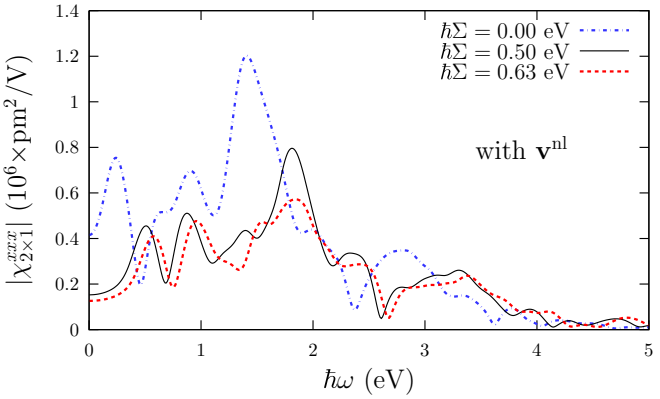

FIG. 6. (Color online) $\chi^{xxx}_{2\times1}$ vs $\hbar\omega$ for a slab with 32 atomic Si layers plus one H layer, for two different values of the scissors correction $\hbar\Delta$.

with previous calculations for bulk systems [38,68,69]. When we compare $|\chi^{xxx}_{2\times1}|$ for the two finite values of $\hbar\Delta$, we see that the first two peaks are almost rigidly shifted with a small difference in height while the rest of the peaks are modified substantially. This behavior comes from the fact that the first two peaks are almost exclusively related to the 2ω resonances of Eq. (28). The other peaks are a combination of 1ω and 2ω resonances and yield a more varied spectrum. We mention that for large gap materials, the 1ω and 2ω would be split showing a small interference effect, but still the 2ω would strongly depend on the surface states. This way we see that small changes in the value of the scissors shift can in general affect the SSH susceptibility spectrum quite dramatically. In Ref. [70], the authors already remarked that nonlinear optical response of bulk materials is more influenced by the electronic structure of the material than the linear case. In the case of semiconducting surfaces, the problem is even more intricate due to the presence of electronic surface states. The high sensitivity of SSHG to the energy position of surface states, as seen in Fig. 6, makes SSHG a good benchmark spectroscopical tool for testing the validity of the inclusion of many-body effects and, in particular, the quasiparticle correction to the electronic states.

Although local fields are neglected they should, in principle, be small parallel to the interface as the electric field is continuous. So, we would expect that the xxx component of $\boldsymbol{\chi}(-2\omega;\omega,\omega)$ would have a small influence from the local fields. Also, the excitonic effects ought to be explored, but their efficient calculation is theoretically and numerically challenging [30] and beyond the scope of this paper. Unfortunately, the experimental measurement of the xxx component of $\boldsymbol{\chi}(-2\omega;\omega,\omega)$ is not possible as the SH radiated intensity would be proportional not only to this component but also to the other components of $\boldsymbol{\chi}(-2\omega;\omega,\omega)$. However, in a forthcoming publication we will present a study of SSHG from several Si surfaces with comparison to experimental results.

IV. CONCLUSIONS

We have presented a formulation to calculate the surface second-harmonic (SSH) susceptibility tensor $\boldsymbol{\chi}(-2\omega;\omega,\omega)$, using the length gauge formalism and within the independent particle approximation (IPA). It includes on equal footing: (i) the scissors correction, (ii) the contribution of the nonlocal part of the pseudopotentials, and (iii) the cut function. We have used a Si(001)2×1 surface to confirm that our scheme correctly obtains the surface response as we confirm that $\chi^{xxx}_{\text{half-slab}}(-2\omega;\omega,\omega)\approx\chi^{xxx}_{\text{full-slab}}(-2\omega;\omega,\omega)$. Although one can in principle increase the number of atomic layers, **k** points, etc., to improve even further on the similarity of the half-slab and full-slab results, we have chosen a good compromise between accuracy and the burden and time of the computations. We describe the effect of the independent inclusion of the three effects mentioned above in the calculation of $\boldsymbol{\chi}(-2\omega;\omega,\omega)$. The scissors correction shifts the spectrum to higher energies though the shifting is not rigid and mixes the 1ω and 2ω resonances, and has a strong influence in the line shape, as for the case of bulk semiconductors [68,69,71]. The cut function allows us to extract unequivocally $\chi^{xxx}_{2\times1}(-2\omega;\omega,\omega)$. The effects of the nonlocal part of the pseudopotentials keeps the same lineshape of $|\chi^{xxx}_{2\times1}(-2\omega;\omega,\omega)|$, but reduces the value of by 15%–20%. The xxx component of $\boldsymbol{\chi}_{2\times1}(-2\omega;\omega,\omega)$, can not be experimentally isolated, however, in a forthcoming publication we will compare our formulation against experimental results. We have neglected local field and excitonic effects. Although these are important factors in the optical response of a semiconductor, their efficient calculation is theoretically and numerically challenging and still under debate [30]. This merits further study but is beyond the scope of this paper. Nevertheless, the inclusion of aforementioned contributions in our scheme opens the unprecedented possibility to study surface SHG with more versatility and more accurate results.

ACKNOWLEDGMENTS

B.S.M. acknowledges the Laboratoire des Solides Irradiés (Ecole Polytechnique, Palaiseau, France) for the support and hospitality during a sabbatical year. B.S.M. acknowledges partial support from CONACYT-México Grant 153930. S.M.A. gratefully acknowledges full support from CONACYT-México scholarship 349278.

APPENDIX

We give explicit expressions for the quantities used in the evaluation of Eq. (28); when appropriate, some intermediate steps are given for their derivation.

1. Expressions for $\chi^{\text{abc}}(-2\omega;\omega,\omega)$

Omitting the frequency dependence for convenience of notation, $\chi^{\text{abc}}=\chi^{\text{abc}}_{e,\omega}+\chi^{\text{abc}}_{i,\omega}+\chi^{\text{abc}}_{e,2\omega}+\chi^{\text{abc}}_{i,2\omega}$, where

$$\chi^{\text{abc}}_{e,\omega}=\frac{|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{l\neq mn}\frac{f_{mn}}{\omega^{\Sigma}_{nm}}\left[\frac{\text{Im}\left[\mathcal{V}^{\Sigma,\text{a}}_{ml}\{r^{\text{c}}_{ln}r^{\text{b}}_{nm}\}\right]}{2\omega^{\Sigma}_{nm}-\omega^{\Sigma}_{lm}}-\frac{\text{Im}\left[\mathcal{V}^{\Sigma,\text{a}}_{ln}\{r^{\text{b}}_{nm}r^{\text{c}}_{ml}\}\right]}{2\omega^{\Sigma}_{nm}-\omega^{\Sigma}_{nl}}\right]\frac{1}{\omega^{\Sigma}_{nm}-\tilde{\omega}},$$

$$\chi^{\text{abc}}_{i,\omega}=\frac{|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{mn}\frac{f_{mn}}{\left(\omega^{\Sigma}_{nm}\right)^2}\left[\text{Re}\left[r^{\text{b}}_{nm}\left(\mathcal{V}^{\Sigma,\text{a}}_{mn}\right)_{;k^{\text{c}}}\right]+\frac{\text{Re}\left[\mathcal{V}^{\Sigma,\text{a}}_{mn}r^{\text{b}}_{nm}\right]\Delta^{\text{c}}_{nm}}{\omega^{\Sigma}_{nm}}\right]\frac{1}{\omega^{\Sigma}_{nm}-\tilde{\omega}},$$

$$\chi^{\text{abc}}_{e,2\omega}=\frac{|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{l\neq mn}4\left[\frac{f_{ln}\,\text{Im}\left[\mathcal{V}^{\Sigma,\text{a}}_{mn}\{r^{\text{b}}_{nl}r^{\text{c}}_{lm}\}\right]}{2\omega^{\Sigma}_{nl}-\omega^{\Sigma}_{nm}}-\frac{f_{ml}\,\text{Im}\left[\mathcal{V}^{\Sigma,\text{a}}_{mn}\{r^{\text{c}}_{nl}r^{\text{b}}_{lm}\}\right]}{2\omega^{\Sigma}_{lm}-\omega^{\Sigma}_{nm}}\right]\frac{1}{\omega^{\Sigma}_{nm}-2\tilde{\omega}},$$

$$\chi^{\text{abc}}_{i,2\omega}=\frac{|e|^3}{2\hbar^2}\int\frac{dk^3}{8\pi^3}\sum_{mn}\frac{4f_{mn}}{\left(\omega^{\Sigma}_{nm}\right)^2}\left[\text{Re}\left[\mathcal{V}^{\Sigma,\text{a}}_{mn}\left(r^{\text{b}}_{nm}\right)_{;k^{\text{c}}}\right]-\frac{2\,\text{Re}\left[\mathcal{V}^{\Sigma,\text{a}}_{mn}r^{\text{b}}_{nm}\right]\Delta^{\text{c}}_{nm}}{\omega^{\Sigma}_{nm}}\right]\frac{1}{\omega^{\Sigma}_{nm}-2\tilde{\omega}}.$$

2. Expressions for $\mathcal{V}_{nm}(\mathbf{k})$ and $\mathcal{C}_{nm}(\mathbf{k})$

Expanding the wave function in plane waves, we obtain

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} A_{n\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}},$$

where $\{\mathbf{G}\}$ are the reciprocal basis vectors satisfying $e^{i\mathbf{R}\cdot\mathbf{G}} = 1$, with $\{\mathbf{R}\}$ the translation vectors in real space, and $A_{n\mathbf{k}}(\mathbf{G})$ the expansion coefficients. Using $m_e\mathbf{v} = \mathbf{p} = -i\hbar\nabla$ into Eq. (22), we obtain [57]

$$\mathcal{V}_{nm}(\mathbf{k}) = \frac{\hbar}{2m_e} \sum_{\mathbf{G},\mathbf{G}'} A^*_{n\mathbf{k}}(\mathbf{G}') A_{m\mathbf{k}}(\mathbf{G})(2\mathbf{k} + \mathbf{G} + \mathbf{G}') \times \delta_{\mathbf{G}_\parallel \mathbf{G}'_\parallel} f(G_\perp - G'_\perp), \tag{A1}$$

where

$$f(G_\perp) = \frac{1}{L}\int \mathcal{C}(z) e^{iG_\perp z} dz,$$

with $f^*(G_\perp) = f(-G_\perp)$, and L is the length of the supercell. The reciprocal lattice vectors $\mathbf{G}$ are decomposed into components parallel ($\mathbf{G}_\parallel$), and perpendicular ($G_\perp\hat{z}$) to the surface, so that $\mathbf{G} = \mathbf{G}_\parallel + G_\perp\hat{z}$. The double summation over the $\mathbf{G}$ vectors can be calculated efficiently by creating a pointer array to identify all the plane-wave coefficients associated with the same $G_\parallel$.

Likewise we obtain that

$$\mathcal{C}_{nm}(\mathbf{k}) = \sum_{\mathbf{G},\mathbf{G}'} A^*_{n\mathbf{k}}(\mathbf{G}') A_{m\mathbf{k}}(\mathbf{G}) \delta_{\mathbf{G}_\parallel \mathbf{G}'_\parallel} f(G_\perp - G'_\perp). \tag{A2}$$

If $\mathcal{C}(z) = 1$, then $f(G_\perp) = \delta_{G_\perp 0}$ and we obtain the full-slab/bulk values, $\mathbf{v}_{nm}(\mathbf{k})$ and $\mathcal{C}_{nm}(\mathbf{k}) = \delta_{nm}$, from Eqs. (A1) and (A2).

3. Expressions for $(\mathcal{V}^{\mathrm{LDA},a}_{nm})_{;k^b}$, $(r^a_{nm})_{;k^b}$, and $(\mathcal{C}_{nm})_{;\mathbf{k}}$

Using Eqs. (9), (11), (12), and defining $\mathcal{T}^{\mathrm{ab}} \equiv [r^{\mathrm{b}}, \mathcal{V}^{\mathrm{LDA,a}}] \equiv [r^{\mathrm{b}}, \mathcal{V}^{\mathrm{a}}]$, one can show that

$$\left(\mathcal{V}^{\mathrm{LDA,a}}_{nm}\right)_{;k^{\mathrm{b}}} = \mathcal{T}^{\mathrm{ab}}_{nm} + i\sum_q \left(r^{\mathrm{b}}_{nq}\mathcal{V}^{\mathrm{LDA,a}}_{qm} - \mathcal{V}^{\mathrm{LDA,a}}_{nq} r^{\mathrm{b}}_{qm}\right) + i r^{\mathrm{b}}_{nm}\Delta^{\mathrm{a}}_{mn}, \tag{A3}$$

where

$$\Delta^{\mathrm{a}}_{mn} = \mathcal{V}^{\mathrm{LDA,a}}_{nn} - \mathcal{V}^{\mathrm{LDA,a}}_{mm}, \qquad \mathcal{T}^{\mathrm{ab}}_{nm} = \frac{\hbar}{m_e}\delta_{\mathrm{ab}}\mathcal{C}_{nm} - \hbar\sum_q L^{\mathrm{ab}}_{nq}\mathcal{C}_{qm}, \tag{A4}$$

and

$$L^{\mathrm{ab}}_{nm} = \frac{i}{\hbar}[r^{\mathrm{b}}, v^{\mathrm{nl,a}}]_{nm}. \tag{A5}$$

The matrix elements L^{ab}_{nm} are small as compared to the other terms, thus we neglect it throwout this work [50]. Notice that $(v^{\mathrm{LDA,a}}_{nm})_{;k^{\mathrm{b}}}$ is obtained from Eq. (A3) by taking $\mathcal{C}(z) = 1$ or $\mathcal{C}_{nm} = \delta_{nm}$.

To obtain $(r^{\mathrm{a}}_{nm})_{;k^{\mathrm{b}}}$, we use Eq. (16) to write $(r^{\mathrm{a}}_{nm})_{;k^{\mathrm{b}}} = (v^{\mathrm{LDA,a}}_{nm}/i\omega^{\mathrm{LDA}}_{nm})_{;k^{\mathrm{b}}}$ and simply apply the chain rule,

$$\left(r^{\mathrm{a}}_{nm}\right)_{;k^{\mathrm{b}}} = T^{\mathrm{ab}}_{nm} + \frac{r^{\mathrm{b}}_{nm}\Delta^{\mathrm{a}}_{mn} + r^{\mathrm{a}}_{nm}\Delta^{\mathrm{b}}_{mn}}{\omega^{\mathrm{LDA}}_{nm}} + \frac{i}{\omega^{\mathrm{LDA}}_{nm}}\sum_q \left(\omega^{\mathrm{LDA}}_{qm} r^{\mathrm{b}}_{nq} r^{\mathrm{a}}_{qm} - \omega^{\mathrm{LDA}}_{nq} r^{\mathrm{a}}_{nq} r^{\mathrm{b}}_{qm}\right), \tag{A6}$$

where

$$\Delta^{\mathrm{a}}_{mn} = v^{\mathrm{LDA,a}}_{nn} - v^{\mathrm{LDA,a}}_{mm}$$

and

$$T^{\mathrm{ab}}_{nm} = \frac{\hbar}{m_e}\delta_{ab}\delta_{nm} - \hbar L^{\mathrm{ab}}_{nm}. \tag{A7}$$

Equation (A6) generalizes the usual expression of $(r^a_{nm})_{;k^b}$ for a local Hamiltonian [28,29,38,39] to the case of a nonlocal Hamiltonian. Note that the layered term $\mathcal{T}^{\mathrm{ab}}_{nm}$ reduces to T^{ab}_{nm} for the full-slab/bulk case.

Again, we use Eqs. (9), (11), and (12), along with $[\mathbf{r}, F(\mathbf{r})] = 0$, valid for and any function $F(\mathbf{r})$, to obtain

$$(\mathcal{C}_{nm})_{;\mathbf{k}} = i\sum_q (\mathbf{r}_{nq}\mathcal{C}_{qm} - \mathcal{C}_{nq}\mathbf{r}_{qm}) + i\,\mathbf{r}_{nm}(\mathcal{C}_{mm} - \mathcal{C}_{nn}),$$

where we remind the reader that $\mathbf{r}_{nm}$ is calculated through Eq. (16) for LDA.

4. Expressions for $\mathcal{V}^S_{nm}$ and $\left(\mathcal{V}^S_{nm}\right)_{;\mathbf{k}}$

From Eq. (22),

$$\mathcal{V}^S_{nm} = \frac{1}{2}\sum_q \left(\mathbf{v}^S_{nq}\mathcal{C}_{qm} + \mathcal{C}_{nq}\mathbf{v}^S_{qm}\right), \tag{A8}$$

where $\sum_q |q\mathbf{k}\rangle\langle q\mathbf{k}| = 1$ was used and $\mathbf{v}^S_{nm}$ is given in Eq. (15). Taking the generalized derivative of Eq. (A8) and applying the chain rule, we obtain

$$\left(\mathcal{V}^S_{nm}\right)_{;\mathbf{k}} = \frac{1}{2}\sum_q \left(\left(\mathbf{v}^S_{nq}\right)_{;\mathbf{k}}\mathcal{C}_{qm} + \mathbf{v}^S_{nq}(\mathcal{C}_{qm})_{;\mathbf{k}} + (\mathcal{C}_{nq})_{;\mathbf{k}}\mathbf{v}^S_{qm} + \mathcal{C}_{nq}\left(\mathbf{v}^S_{qm}\right)_{;\mathbf{k}}\right).$$

Again, from Eq. (15),

$$\left(\mathbf{v}^S_{nm}\right)_{;\mathbf{k}} = i\,\Delta f_{mn}(\mathbf{r}_{nm})_{;\mathbf{k}},$$

a result that is in agreement with Eq. A(6) of Ref. [39].

[1] W. Daum, H. J. Krause, U. Reichel, and H. Ibach, Phys. Rev. Lett. **71**, 1234 (1993).

[2] J. F. McGilp, M. Cavanagh, J. R. Power, and J. D. O'Mahony, Opt. Eng. **33**, 3895 (1994).

[3] C. Meyer, G. Lüpke, U. Emmerichs, F. Wolter, H. Kurz, C. H. Bjorkman, and G. Lucovsky, Phys. Rev. Lett. **74**, 3001 (1995).

[4] J. R. Power, J. D. O'Mahony, S. Chandola, and J. F. McGilp, Phys. Rev. Lett. **75**, 1138 (1995).

[5] P. Godefroy, W. De Jong, C. W. Van Hasselt, M. A. C. Devillers, and T. Rasing, Appl. Phys. Lett. **68**, 1981 (1996).

[6] U. Höfer, Appl. Phys. A **63**, 533 (1996).

[7] J. I. Dadap, Z. Xu, X. F. Hu, M. C. Downer, N. M. Russell, J. G. Ekerdt, and O. A. Aktsipetrov, Phys. Rev. B **56**, 13367 (1997).
[8] N. Bloembergen, Appl. Phys. B **68**, 289 (1999).
[9] J. F. McGilp, Surf. Rev. Lett. **6**, 529 (1999).
[10] T. Suzuki, D. E. Milovzorov, S. Kogo, M. Tsukakoshi, and M. Aono, Appl. Phys. B **68**, 623 (1999).
[11] S. A. Mitchell, M. Mehendale, D. M. Villeneuve, and R. Boukherroub, Surf. Sci. **488**, 367 (2001).
[12] J. L. P. Hughes and J. E. Sipe, Phys. Rev. B **53**, 10751 (1996).
[13] P. Guyot-Sionnest and Y. R. Shen, Phys. Rev. B **38**, 7985 (1988).
[14] M. C. Downer, Y. Jiang, D. Lim, L. Mantese, P. T. Wilson, B. S. Mendoza, and V. Gavrilenko, Phys. Status Solidi A **188**, 1371 (2001).
[15] Y. R. Shen, Appl. Phys. B **68**, 295 (1999).
[16] Y. R. Shen, Nature (London) **337**, 519 (1989).
[17] C. K. Chen, A. R. B. de Castro, and Y. R. Shen, Phys. Rev. Lett. **46**, 145 (1981).
[18] B. S. Mendoza, A. Gaggiotti, and R. Del Sole, Phys. Rev. Lett. **81**, 3781 (1998).
[19] M. C. Downer, B. S. Mendoza, and V. I. Gavrilenko, Surf. Interface Anal. **31**, 966 (2001).
[20] A. M. Zheltikov, G. Ferrante, and M. Zarcone, Laser Phys. **10**, 600 (2000).
[21] G. Lüpke, Surf. Sci. Rep. **35**, 75 (1999).
[22] R. V. Salazar-Aparicio, R. A. Vázquez-Nava, N. Arzate, and B. S. Mendoza, Phys. Rev. B **90**, 155403 (2014).
[23] M. Cazzanelli, F. Bianco, E. Borga, G. Pucker, M. Ghulinyan, E. Degoli, E. Luppi, V. Véniard, S. Ossicini, D. Modotto *et al.*, Nat. Mater. **11**, 148 (2014).
[24] P. N. Butcher and T. P. McLean, Proc. Phys. Soc. **81**, 219 (1963).
[25] D. E. Aspnes, Phys. Rev. B **6**, 4648 (1972).
[26] J. E. Sipe and E. Ghahramani, Phys. Rev. B **48**, 11705 (1993).
[27] Z. H. Levine, Phys. Rev. B **49**, 4532 (1994).
[28] C. Aversa and J. E. Sipe, Phys. Rev. B **52**, 14636 (1995).
[29] S. N. Rashkeev, W. R. L. Lambrecht, and B. Segall, Phys. Rev. B **57**, 3905 (1998).
[30] For bulk calculation schemes of the SH susceptibility tensor beyond the independent particle approximation, see, for instance, Refs. [68–75].
[31] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[32] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[33] M. J. Lucero, T. M. Henderson, and G. E. Scuseria, J. Phys. Condens. Matter **24**, 145504 (2012).
[34] Z. H. Levine and D. C. Allan, Phys. Rev. Lett. **63**, 1719 (1989).
[35] Z. H. Levine and D. C. Allan, Phys. Rev. B **43**, 4187 (1991).
[36] R. D. Sole and R. Girlanda, Phys. Rev. B **48**, 11789 (1993).
[37] L. Kipp, D. K. Biegelsen, J. E. Northrup, L.-E. Swartz, and R. D. Bringans, Phys. Rev. Lett. **76**, 2810 (1996).
[38] F. Nastos, B. Olejnik, K. Schwarz, and J. E. Sipe, Phys. Rev. B **72**, 045223 (2005).
[39] J. L. Cabellos, B. S. Mendoza, M. A. Escobar, F. Nastos, and J. E. Sipe, Phys. Rev. B **80**, 155205 (2009).
[40] L. Reining, R. Del Sole, M. Cini, and J. G. Ping, Phys. Rev. B **50**, 8411 (1994).
[41] N. Arzate and B. S. Mendoza, Phys. Rev. B **63**, 125303 (2001).
[42] B. S. Mendoza, M. Palummo, G. Onida, and R. Del Sole, Phys. Rev. B **63**, 205406 (2001).
[43] J. E. Mejía, B. S. Mendoza, M. Palummo, G. Onida, R. Del Sole, S. Bergfeld, and W. Daum, Phys. Rev. B **66**, 195329 (2002).
[44] H. Sano, G. Mizutani, W. Wolf, and R. Podloucky, Phys. Rev. B **66**, 195338 (2002).
[45] J. E. Mejía, C. Salazar, and B. S. Mendoza, Revista Mexicana de Física **50**, 134 (2004).
[46] S. Ismail-Beigi, E. K. Chang, and S. G. Louie, Phys. Rev. Lett. **87**, 087402 (2001).
[47] E. Luppi, H.-C. Weissker, S. Bottaro, F. Sottile, V. Véniard, L. Reining, and G. Onida, Phys. Rev. B **78**, 245124 (2008).
[48] J. E. Sipe and A. I. Shkrebtii, Phys. Rev. B **61**, 5337 (2000).
[49] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[50] V. Véniard, E. Luppi, and H. Hübener (unpublished).
[51] E. N. Adams, J. Chem. Phys. **21**, 2013 (1953).
[52] E. I. Blount, Solid State Phys. **13**, 305 (1962).
[53] V. Olevano (private communication).
[54] C. Motta, M. Giantomassi, M. Cazzaniga, K. Gaál-Nagy, and X. Gonze, Comput. Mater. Sci. **50**, 698 (2010).
[55] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
[56] B. Adolph, V. I. Gavrilenko, K. Tenelsen, F. Bechstedt, and R. Del Sole, Phys. Rev. B **53**, 9797 (1996).
[57] B. S. Mendoza, F. Nastos, N. Arzate, and J. E. Sipe, Phys. Rev. B **74**, 075318 (2006).
[58] In Ref. [76], a non-Hermitian operator was used in the context of linear optical response.
[59] N. Tancogne-Dejean, B. S. Mendoza, and V. Véniard, Phys. Rev. B **90**, 035212 (2014).
[60] The half-slab layer extends to the middle of the vacuum region between consecutive (front-back or back-front) surfaces of the repeated supercell scheme.
[61] The ABINIT code is a common project of the Universit Catholique de Louvain, Corning Incorporated, and other contributors (URL http://www.abinit.org). X. Gonze *et al.*, Comp. Mater. Sci. **25**, 478 (2002); Zeit. Crystallogr, **220**, 558 (2005).
[62] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[63] V. Olevano, L. Reining, and F. Sottile, http://www.dp-code.org/.
[64] L. Caramella, C. Hogan, G. Onida, and R. Del Sole, Phys. Rev. B **79**, 155447 (2009).
[65] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **52**, 1905 (1995).
[66] P. García-González and R. W. Godby, Comput. Phys. Commun. **137**, 108 (2001).
[67] R. Asahi, W. Mannstadt, and A. J. Freeman, Phys. Rev. B **62**, 2552 (2000).
[68] E. Luppi, H. Hübener, and V. Véniard, Phys. Rev. B **82**, 235201 (2010).
[69] R. Leitsmann, W. G. Schmidt, P. H. Hahn, and F. Bechstedt, Phys. Rev. B **71**, 195209 (2005).
[70] B. Adolph and F. Bechstedt, Phys. Rev. B **62**, 1706 (2000).

[71] E. Luppi, H. Hübener, and V. Véniard, J. Chem. Phys. **132**, 241104 (2010).
[72] H. Hübener, E. Luppi, and V. Véniard, Phys. Rev. B **83**, 115205 (2011).
[73] M. L. Trolle, G. Seifert, and T. G. Pedersen, Phys. Rev. B **89**, 235410 (2014).
[74] C. Attaccalite and M. Grüning, Phys. Rev. B **88**, 235113 (2013).
[75] M. Grüning and C. Attaccalite, Phys. Rev. B **89**, 081102 (2014).
[76] C. Hogan, R. Del Sole, and G. Onida, Phys. Rev. B **68**, 035405 (2003).