Layer-by-layer analysis of second harmonic generation at a simple surface

J.E. Mejía, C. Salazar\textsuperscript{a}, and B.S. Mendoza\textsuperscript{b}

Centro de Investigaciones en Optica, A.C.,
León Guanajuato, México,
\textsuperscript{b}e-mail: bms@cio.mx.

Recibido el 19 de noviembre de 2002; aceptado el 18 de agosto de 2003

We present a general scheme, based on a microscopic formulation, to obtain the second harmonic signal produced by each atomic layer of a semi-infinite crystal. Using the simple Si(111):H (1 × 1) surface as an example, we obtain that the nonlinear polarization in the bulk does not decay to zero due to the lack of centrosymmetry of the individual layers. However, the sum of this polarization follows the physically correct picture that the second harmonic signal is zero at the centrosymmetric bulk and finite in the selvedge as the surface is approached and the centrosymmetry is broken. The results show that the selvedge region includes the surface and just a few layers below it.

Keywords: Non-linear optics; second harmonic generation; surfaces.

Presentamos un procedimiento general, basado en una formulación microscópica, para obtener la señal del segundo armónico producida por cada capa atómica de un cristal semi-infinito. Usando la superficie sencilla Si(111):H (1 × 1) como un ejemplo, obtenemos que la polarización no lineal en el bulbo no decrece a cero debido a la falta de centrosimetría de las capas individuales. Sin embargo, la suma de esta polarización sigue el comportamiento físicamente correcto, mediante el cual la señal del segundo armónico es cero en el bulbo centrosimétrico y finita en la región superficial conforme se aproxima la superficie y la centrosimetría se pierde. Los resultados muestran que esta región superficial incluye la superficie misma y tan sólo unas cuantas capas por debajo.

Descriptores: Optica No-lineal; generacion de segundo armónico; superficies.

PACS: 42.65.Ky; 42.65.An; 78.66.-w

1. Introduction

In recent years, surface nonlinear optical spectroscopies and in particular, second harmonic generation (SHG) have evolved as useful non-destructive and non-invasive tool to study properties of surfaces and interfaces, like atomic structure, phase transitions and adsorption of atoms [1–10]. The high sensitivity of SHG spectroscopy is due to the fact that, within the dipole approximation, the bulk SHG signal of centrosymmetric materials is identically zero, and thus only the surface, where the inversion symmetry is broken, can radiate [11]. On the experimental side, new tunable laser systems have made SHG spectroscopy applicable to a wide range of systems [1–10]. The SHG advances on the experimental side are being followed by the theoretical development of the field. In particular, the semi empirical tight-binding (SETB) approach, with a sp\textsuperscript{3}s\textsuperscript{*} basis set [12–16], and \textit{ab-initio} methods (within the density-functional theory (DFT) in the local-density approximation (LDA)) [15, 17–19], have been developed in the last few years, to calculate the nonlinear optical properties of semiconductor surfaces, yielding results in qualitative agreement with experiments and clarifying the atomic and electronic structure of different surfaces and the adsorption of foreign atoms.

In this article we explore, for the first time, a theoretical scheme through which one can calculate the second harmonic (SH) contribution of each atomic layer of a semi-infinite system that goes from the surface to the bulk of the material. The scheme permits to understand how the SH is generated when the centrosymmetry of the bulk is broken as we approach the surface. For the example chosen, we show that the individual contribution of each noncentrosymmetric layer to the SH signal or to the nonlinear polarization (susceptibility), does not decay as one moves into the bulk of the system. However, the destructive interference of the nonlinear polarization among the different planes, is the one responsible for rendering a null SH signal from the centrosymmetric bulk, and producing a finite SH signal that comes from the surface layer and a small region below it known as the selvedge.

The H-terminated Si(111) (1 × 1) surface, is chosen as an example, since it is the most simple of all semiconductor surfaces, in the sense that its equilibrium relaxed geometry is almost identical to an ideally terminated (111) surface. Also, as H saturates every Si dangling bond, there are no surface-related electronic states in the forbidden gap. Therefore, this surface presents a very good reference system for SHG studies, since neither the atomic structure nor electronic surface states in the gap region should be of concern, when comparing theoretical and experimental SHG spectra. To obtain explicit results we use the semi-empirical tight binding (SETB) formalism, with a sp\textsuperscript{3}s\textsuperscript{*} basis set, successfully used in other SH calculations [12–16]. We show that the SETB calculation qualitatively reproduces the experimental spectrum, that covers the E\textsubscript{1} and E\textsubscript{2} critical points of Si. Also, and more importantly, the SETB approach allows us to apply the scheme for the layer-by-layer analysis in very simple terms. Thus, we can explore with detail the physical origin of the surface SH process.

The article is organized as follows. Section 2 briefly describes the theoretical procedure to obtain the layer-by-layer contribution to SHG. In Sec. 3 the results are presented and conclusions are given in Sec. 4.
2. Theory

The SHG efficiency is defined as the nonlinear reflection coefficient, \( R = I(2\omega)/I^2(\omega) \), with \( I \) the intensity of the corresponding incoming linear (\( \omega \)) field or outgoing non-linear (2\( \omega \)) field. We restrict to a \( p \)-polarized incoming (fundamental) beam and to a \( p \)-polarized outgoing (SH) beam, since this combination of polarization gives the strongest signal and it also suffices for the purpose of this article. Then, we have that [19]

\[
R_{pp} = \frac{32\pi^3}{(n\epsilon^0)^2c^2} \frac{\omega^2}{\cos^2 \theta} \left| T_P^2(\omega)T_p^2(\omega)r_{pp} \right|^2, \quad (1)
\]

with

\[
r_{pp} = \sin \theta \left( \sin^2 \theta \chi_{\perp \perp \perp}^s + \frac{(c/\omega)^2 k_2^2(\omega)\chi_{\parallel \parallel \parallel}^s}{-} \right)
\]

\[
- \frac{(c/\omega)^2 k_2^2(\omega)\chi_{\parallel \perp \perp}^s}{\cos(3\phi)} + \frac{1}{2} \frac{(c/\omega)^2 k_2^2(\omega)\chi_{\parallel \parallel \parallel}^s}{}, \quad (2)
\]

where we have used the fact that for the symmetry group \( C_{3v} \) of an ideal (111) surface, only the following components of the surface nonlinear susceptibility \( \chi_{ijk} \) are non-zero

\[
\chi_{zzz}^s \equiv \chi_{\perp \perp \perp}^s, \quad \chi_{xzx}^s \equiv \chi_{\perp \parallel \parallel}^s, \quad \chi_{zxx}^s \equiv \chi_{\parallel \perp \perp}^s, \quad \chi_{xxy}^s \equiv \chi_{\parallel \parallel \perp}^s, \quad \text{and} \quad \chi_{xxz}^s \equiv -\chi_{zyy}^s \equiv \chi_{\parallel \parallel \parallel}^s.
\]

We have chosen the \( x \) and \( y \) axes along the [11\( \overline{2} \)] and [1\( \overline{1} \)T0] directions, respectively (see Fig. 1). Also, the intrinsic permutation symmetry \( \chi_{ijk}^s = \chi_{\alpha ij}^s \) is satisfied. Note that the \( \chi_{\parallel \parallel \parallel}^s \) component induces a SH response which is anisotropic with respect to the azimuthal angle \( \phi \) of the plane of incidence makes with the \( x \) axis (see Fig. 1). Also, in Eqs. (1) and (2), \( \theta \) is the angle of incidence, \( c \) the speed of light, \( \epsilon \) is the charge of the electron, \( n_0 \epsilon \) the electron density of the system, \( T_p \) the transmission Fresnel factor for \( p \)-polarization and \( k_2(\omega) = (\omega/c)(\epsilon(\omega) - \sin^2 \theta)^{1/2} \), with \( \epsilon(\omega) \) being the bulk dielectric function. These expressions are strictly valid within the dipole approximation. Nevertheless, even if quadrupolar corrections are considered, the isotropic and anisotropic bulk quadrupole terms in \( R \), have shown to yield negligible contributions as compared to the surface dipole terms [20]. Notice that we have chosen \( \chi_{ijk}^s \) as dimensionless functions, in such a way that the prefactor of Eq. (1) gives the correct units of \( R_{pp} \), i.e. area/power. To obtain \( \chi_{ijk}^s \) in the appropriate units we multiply it by \( 1/(n \epsilon^0) \) [21].

We mention that in the calculation of \( \chi_{ijk}^s \) [22], one uses the long wavelength approximation, by which the fundamental electric field oscillating at \( \omega \), \( \hat{E}(\omega) \), which induces the non-linear response, is taken inside the surface. In particular, these fields are simply given by the external fields properly multiplied by the corresponding Fresnel factors. A more detailed description of the fields, which incorporates the spatial variation of the dielectric function near the surface within the three-layer model, shows for this surface, no change in the SHG peaks positions, and only a slight difference in their intensity [19]. This screening was thought to affect \( \chi_{\perp \perp \perp}^s \) more than any other component, however Ref. [19] shows that this is not the case. In this respect, the nonlocal nature of \( \chi_{ijk}^s \) is integrated out, leaving a susceptibility that depends only on the position where we evaluate the response. The full treatment of the surface screening is still lacking, and further improvement of the present formulation can be made along this line. Also, at the present state of the calculations of non-linear optical properties of semiconducting surfaces, like the one presented here, local-field and excitonic effects are beyond current capabilities and are thus neglected throughout. Again, these effects along with the surface screening may prove to be crucial for a quantitative comparison of theory and experiment.

Now, we present the most relevant points taken to calculate the SH of every atomic layer through \( \chi_{ijk} \), that is the main theoretical goal of the article. The semi-infinite crystal is simulated by a slab of \( N \) atomic layers, extending from \( z = 0 \) to \( z = N d \), \( d \) being the interlayer distance. Since the slab is intrinsically centrosymmetric, the non-linear second-order dipolar susceptibility, \( \chi_{ijk} = 0 \) [11]. This fact stems from the destructive interference between the first and second halves of the slab. Therefore, to circumvent this fact, one has to introduce a cut function associated with the quantum mechanical operator that gives the second-order non-linear polarization. Following Ref. [12] we write the modified momentum operator as

\[
\vec{p} = \frac{1}{2}(\hat{p}S(z) + S(z)\hat{p}), \quad (3)
\]

where \( \vec{p} = -i\hbar \vec{\nabla} \) is the standard momentum operator, and \( S(z) \) is the function that uncouples the slab to render a finite \( \chi_{ijk} \) representative of the surface in question. The typical
choice for \( S(z) \) is a step function that is one on one half of the slab and zero on the other [12]. However, we find that this function can be taken in such a way that one can isolate the contribution of every atomic layer to \( \chi_{ijk} \), and thus one can study the SH process layer by layer. Indeed, taking \( z_n \) as the \( z \)-position of the \( n \)-th layer we write
\[
S(z) = \Theta(z - z_n + d/2)\Theta(z_n + d/2 - z),
\]
where this function will include the contribution of the neighborhood that goes from above to below the \( n \)-th layer by half the interlayer spacing (\( d/2 \)). Using Eq. (4) in Eq. (3), and this in the corresponding relations for \( \chi_{ijk} \) [12], it can be easily shown that one can define the nonlinear susceptibility of \( n \)-th layer, that we simply denote as \( \chi_{ijk}(n) \). Through, \( \chi_{ijk}(n) \) we obtain the surface nonlinear susceptibility as
\[
\chi_{ijk}^s = \sum_{n=1}^{N/2} \chi_{ijk}(n),
\]
which is identical to that of Ref. [12], where \( S(z) \) is taken a step function centered in the middle of the slab. Other choices for \( S(z) \) have been explored in Ref. [22], where using smoother functions yields the same SHG line shape with only some changes in \( R_{pp} \).

To obtain explicit relations for the matrix elements of \( S(z) \) between any two quantum states \( r \) and \( s \), which are required for the evaluation of \( \chi_{ijk}(n) \) [12], the SETB method is particular suitable in the present context, because it allows for a natural assignment of the electronic states to specific atomic layers. Indeed, we can show easily that if the electronic wave functions are expanded into (or projected onto) atomic orbitals, then the matrix elements of \( S(z) \) are simply given by \( (S(z))_{r,s} = S(z_n)\delta_{r,s} \) where \( S(z_n) = 1 \) if the position \( z_n \) corresponds to the atom described by the quantum state \( r = s \), and is zero otherwise. We mention that our approach of separating the atomic layer contributions to \( \chi_{ijk} \) is not restricted to a tight-binding description of the slab electronic structure, but may also be combined with plane-wave or finite-difference schemes, as outlined in Ref. [18].

To calculate the contribution of any given layer \( n \) to \( R_{pp} \) we simply substitute \( \chi_{ijk}(n) \), instead of \( \chi_{ijk}^s \), into \( R_{pp} \) of Eq. (2) to obtain \( r_{pp}(n) \). Therefore we can define
\[
R_{pp}(\ell, \ell') = \frac{32\pi^3}{(n\ell e)^2} \frac{\omega^2}{\text{cal}^2 \theta^2} |T_p(2\omega)T_p^2(\omega) \sum_{n=\ell}^{\ell'} r_{pp}(n)|^2,
\]
as the contribution to SH coming from the \( \ell \) till the \( \ell' \) layer. Then, \( R_{pp}(\ell, \ell) \) is the contribution to SH coming from layer \( \ell \). Likewise,
\[
\chi_{ijk}(\ell, \ell') = \sum_{n=\ell}^{\ell'} \chi_{ijk}(n),
\]
gives the contribution to \( \chi_{ijk} \) from layer \( \ell \) up to layer \( \ell' \), and the non-linear polarization could be obtained for a given layer or a sum of them, by using the well known relationship
\[
p_i = (1/n_0 e) \sum_{jk} \chi_{ijk} E_j E_k.
\]
Notice that while \( \chi_{ijk}(\ell, \ell') \) is additive, \( R_{pp}(\ell, \ell) \) is not, i.e., \( \chi_{ijk}(1,2) = \chi_{ijk}(1,1) + \chi_{ijk}(2,2) \), etc., but \( R_{pp}(1,2) \neq R_{pp}(1,1) + R_{pp}(2,2) \). Then, in principle one can analyze \( \chi_{ijk}(\ell, \ell') \), however, its difficult to relate structures in \( R_{pp}(\ell, \ell') \) with those of \( \chi_{ijk}(\ell, \ell') \), since \( \chi_{ijk}(\ell, \ell') \) is complex, and its different components are added in a non-trivial way in \( r_{pp} \) (see Eq. (2)). Then, even though \( R_{pp}(\ell, \ell') \) is not additive, as we will see in the results, it still allow us to analyze the SH process.

### 3. Results

We take the equilibrium atomic positions of the Si(111):H(1 × 1) surface from Ref. [19], that were obtained from DFT-LDA. Then, we follow Ref. [12] to obtain, within the SETB scheme, the energies and the momentum matrix elements required to evaluate \( \chi_{ijk}(n) \). As in the experiment, we take \( \theta = 65^\circ \), and \( \phi = 30^\circ \), with which \( \chi_{\parallel\parallel}(n) \) drops from \( r_{pp}(n) \).

To asses the validity of the SETB electronic states, in Fig. 2 we compare \( R_{pp} \) for the SETB model along with the \textit{ab-initio} and experimental spectra from Ref. [19]. In both theoretical schemes \( N = 38 \) layers gives well converged results. The spectra show well defined \( E_1 \) and \( E_2 \) SH peaks in agreement with the experiment, that sows \( E_1 \sim 3.3 \) eV and \( E_2 \sim 4.3 \) eV. We notice that the \textit{ab-initio} spectrum is blue shifted, whereas the SETB spectrum is red shifted with respect to the experimental results. However, the energy separation between \( E_1 \) and \( E_2 \) is well reproduced by both theoret-

![Figure 2. \( R_{pp} \) vs the two photon energy for \( \phi = 30^\circ \) and \( \theta = 65^\circ \). The solid and dotted lines are the SETB and \textit{ab-initio} results of Ref. [19], respectively (left vertical scale). The experimental spectrum, [19] is shown with pluses (right vertical scale). The inset shows the SETB results (solid line) divided by 5 and blue shifted by 0.3 eV, to have the \( E_1 \) and \( E_2 \) resonances coinciding with the experiment (dotted line).](image-url)
ical approaches. On one hand, the quasi particle corrections required in LDA to get the experimental energy gap, overestimate the correct positions of each critical point $E_1$ and $E_2$ by $\sim 0.5$ eV, and on the other hand, the deficiency of SETB to account for the surface effects on the electron self-energies [23], underestimates them by $\sim 0.4$ eV. Below $E_1$ the spectra is zero as there are no surface related states. The overall SETB $\mathcal{R}_{pp}$ is a factor of $\sim 5$ larger than the experiment (see inset in Fig. 2), that may be related to the approximation involved in the method. However, according to Ref. [19], the use of a three-layer model for surface screening will close this difference. We also note that $\mathcal{R}_{pp}(E_1) > \mathcal{R}_{pp}(E_2)$, as opposed to the experimental result that reveals a larger $\mathcal{R}_{pp}$ for $E_2$ (see inset in Fig. 2), which is a SETB limitation for producing the correct weights of the electronic transitions involved in the $E_2$ resonance. Even though these deficiencies, the SETB approach still gives a good qualitative description of SH for this surface.

We now analyze $\mathcal{R}_{pp}$ by using the atomic layer decomposition given in Eq. (6). In Fig. 3, we show $\mathcal{R}_{pp}(\ell, \ell')$ for the following cases of $\ell, \ell'$. In the top panel, $\ell = \ell' = 1, 2, \cdots, 7$, that represents the SHG coming from each individual layer. In the middle panel, we have taken $(\ell, \ell') = (2, 4), (3, 7)$ till $(17, 19)$ (i.e., middle of the slab), which is equivalent to the radiation of every three consecutive Si layers. On the other hand, in the bottom panel, we have taken $(\ell, \ell') = (2, 3), (4, 5)$ till $(12, 13)$, which is now equivalent to the radiation of every two consecutive Si layers. From the top panel is clear that the radiation coming from the H layer (=1) is zero, due to the fact that H saturates the Si dangling bond, and the response of it is quenched. Therefore, in the next panels the different $\mathcal{R}_{pp}(\ell, \ell')$ do not need to consider the first layer, and we can start with the first Si layer (=2, see Fig. 1). At first sight it may seem from the top panel that the radiation coming from the inner Si layers is larger than from the layers closer to the surface, in contrast with the argument that SH must be null due to the centrosymmetry of the bulk. However, the SH process is not the addition of the different SH contributions of every layer, but rather, as in Eq. (6), is the addition of the susceptibilities (or polarization) of every layer, and thus interference effects play a dominant role. Indeed, in the middle and bottom panels of Fig. 3, we can see how this interference works. In the middle panel we have added every 3 consecutive Si layers, and although the signal is smaller for the inner layers, still no conclusive null signal is seen as we approach the bulk (or the middle of the slab). On the other hand, in the bottom panel, we add every two consecutive Si layers, and indeed a very clear picture is now apparent. The first four Si layers or surface layers dominate the signal, whereas the rest of them, give a diminishing signal that is already negligible for $(\ell, \ell') = (12, 13)$, that is, before we even reach the middle of the slab.

The results reached by analyzing the bottom panel of Fig. 3 can be understood from the intrinsic symmetry of the atomic structure seen in Fig. 1. Adding every two consecutive layers, starting with the top-most Si layer, gives a basic unit from which the slab can be constructed just by piling up these units bellow each other, in such a way that the bottom-most Si can be H saturated with the same surface symmetry as the top-most Si. In other words, the irreducible slab along the $z$ direction consists exactly of 2 Si layers, and the correct centrosymmetry derived by the repetition of this unit is evident in the SH signal. One or three Si layers are just not the correct $z$-unit cell. To complement on this idea, we show in Fig. 4 $\mathcal{R}_{pp}(\ell, \ell')$ for $(\ell, \ell') = (2, 3), (2, 5), (2, 15)$ which consists of adding the SH signal from the surface till the $\ell' = 15$ layer in steps of two layers, and for $(\ell, \ell') = (2, 4), (2, 7), (2, 16)$ which consists of adding the SH signal from the surface till the $\ell' = 16$ layer but in steps of three layers. Although $\mathcal{R}_{pp}(1, 1)$ (for the case of summing every two layers), and $\mathcal{R}_{pp}(1, 16)$ (for the case of summing every three layers), are now very similar to the full spectrum (Fig. 2), is very clear that the SH signal builds up systematically for the first case and erratically for the second. Therefore, once more is clear that the centrosymmetry is constructed by $z$-units of two Si layers. As we move toward the surface the centrosymmetry is broken thus allowing the SH generation in the sel-

**Figure 3.** $\mathcal{R}_{pp}(\ell, \ell')$ for $(\ell, \ell') = (\text{Top panel}) (1, 1), (2, 2), (7, 7)$, equivalent to layer 1 (H) through layer 7 (Si); (Middle panel) $(2, 4), (5, 7)$, till $(17, 19)$, equivalent to add every three Si layers starting with the top-most Si (layer 2 of Fig. 1); (Bottom panel) $(2, 3), (4, 5)$, till $(12, 13)$, equivalent to add every two Si layers starting again with the top-most Si. The SETB red shifted SH $E_1$ and $E_2$ are shown for reference.
behavior also seen in R last, the analysis carried for χ, added, and in the bottom panel χ, (top panel 1 responding value of part of sin R the spectra. Second, the interference of χ contribution of χ, i.e. (that appear in r ter to plot them properly multiplied by the prefactors χ, sequence, the other two components of χ, never, instead of simply analyzing χ, N layers, this surface though, is the fact that it takes a large number of surface region of 3 ∼ 16, which includes the surface Si layer and a small sub- surface region of 3~4 atomic layers. A striking feature of this surface though, is the fact that it takes a large number of layers, N ∼ 16, to get the full SH signal.

Finally, we study the non-linear susceptibility, however, instead of simply analyzing χijk(l, l′), it is better to plot them properly multiplied by the prefactors that appear in rχ, since this sets the correct scale with which each component of χijk(l, l′) contributes to the Rχ,pp(l, l′). Then, in Fig. 5 we show the imaginary part of sin2 θχ,l⊥∥∥∥, (c/ω)2k⊥∥∥∥(ω)χ,l⊥∥∥∥(l, l′), and (c/ω)2k⊥∥∥∥(ω)χ,l⊥∥∥∥(2ω)χ,l⊥∥∥∥(l, l′) [24], multiplied by the corresponding value of 1/ε0ε and expressed in esu-cm. In the top panel (l, l′) = (2, 2) and (13,13), in the middle panel (l, l′) = (2, 4) and (11, 13) (i.e., three consecutive layers added), and in the bottom panel (l, l′) = (2, 3) and (12, 13) (i.e., two consecutive layers added). First, we notice that the contribution of χ,l⊥∥∥∥(l, l′) is negligible, and that in conse- quence, the other two components of χijk(l, l′) dominate the spectra. Second, the interference of χ,l⊥∥∥∥(l, l′), and χ,l⊥∥∥∥(l, l′), gives the corresponding Rχ,pp(l, l′). Third, and last, the analysis carried for Rχ,pp(l, l′), just follows the behavior also seen in χijk(l, l′), but is after we add their com-ponents as prescribed in Eq. (2) and take the absolute value squared of Eq. (6), that we can see the E1 and E2 resonances emerging from their sources.

4. Conclusions

We present for the first time, a method for analyzing the layer-by-layer contribution to the SH signal of a semi-infinite system. As an example, we studied the SHG of the Si(111):H(1 × 1) surface. To obtain the electronic structure of this surface, we used the SETB method, that has the advantage of being computationally less intensive than the ab-initio methods, and gives just as good qualitative results. Using our method for the layer-by-layer analysis, we concluded that the SH takes place in a few atomic layers in the selvedge region and that for this particular surface the structural cen- trosymmetry lies well bellow the surface. The contribution of the individual layers to the SH signal do not decay as one moves into the bulk of the system, however the susceptibility

of the different layers interferes destructively giving the surface originated SH signal and the cancellation of it in the centrosymmetric environment of the bulk. The atomic geometry along the perpendicular direction to the surface is responsible for the behavior found in the SH signal, thus showing the sensitivity of this non-linear optical probe to study surfaces and interfaces.

Acknowledgments

We acknowledge partial support from CONACyT-México under grant 36033-E. CS thanks the AMC for the scholarship while at CIO.