Harmonic Generation in Thin Films and Multilayers

William S. Kolthammer, Dustin Barnard, Nicole Carlson, Aaron D. Edens, Nathan A. Miller, and Peter N. Saeta*

Physics Department, Harvey Mudd College, Claremont, CA 91711

(Dated: September 6, 2004)

A general method for computing harmonic generation in reflection and transmission from planar nonmagnetic multilayer structures is described. The method assumes plane waves and treats harmonic generation in the parametric approximation. The method is applied in studying the second- and third-harmonic generation properties of thin crystal silicon layers surrounded by thermal oxide. Most independent components of the nonlinear susceptibility tensor have unique signatures with silicon layer thickness d, allowing their strength to be determined in principle by measuring harmonic generation as a function of d. Surface and bulk contributions to third-harmonic generation are cleanly distinguished, with the bulk signal dominating. Four of six nonvanishing components of $\chi^{(2)}$ are independent. An approximate value for the bulk susceptibility component δ', which is accessible only in multibeam experiments and has not previously been measured, is obtained.

PACS numbers: 42.65.An, 42.65.Ky, 42.70.Nq

I. INTRODUCTION

Optical harmonic generation launched the field of nonlinear optics in 1961,[1] soon after the invention of the laser, and has remained an active area of fundamental research and applications in the over forty years since. Harmonic generation with femtosecond laser pulses produces strong signal levels with only modest energy deposition in absorbing samples,[2] and can provide valuable information on crystal structure and orientation,[3,4] as well as structure and bonding at surfaces and interfaces.[5–7] Furthermore, the technique is nondestructive and compatible with a variety of experimental conditions.[2,8–10]

Early papers by Bloembergen et al. presented the basic parametric theory of harmonic generation within a homogeneous medium[11] and at the plane interface between a linear and nonlinear medium,[12] including a treatment of the reflected and transmitted second-harmonic light from a thin nonlinear dielectric slab. They showed that for layer thickness d satisfying $d \ll \lambda$, the generated harmonic field depends linearly on d. Hence, the second-harmonic intensity grows quadratically with thickness in this range, since all the atoms in the layer radiate coherently.

In the simplest formulation, one neglects reflections at the interfaces and considers a single beam within the layer. For semiconducting or metallic layers surrounded by transparent media, the dielectric contrast at interfaces can be large, and failure to include the reflections distorts the calculation significantly. For layer thicknesses that are comparable to, or smaller than, the absorption depth at the fundamental and at the harmonic frequency, interference between the forward and backward waves is significant and appreciably modifies the observed harmonic light in reflection and transmission. In the case of crystal silicon layers surrounded by SiO_2, the (field) reflection coefficient for a fundamental wave with $\lambda = 800$ nm is $\sim 40\%$, rising to $\sim 55\%$ at the second harmonic, and $\sim 80\%$ at the third harmonic. These reflections cause deep modulations in the strength of the incident wave

inside the silicon layer, and hence in the strength of the nonlinear source polarization generated within it. In systems comprising layers with large reflection coefficients, such as Si/SiO_2, therefore, it is essential to consider the nonlinear sources arising not only from the incident wave, but from all combinations of the incident and reflected fundamental waves inside the layer.

A matrix method has long been employed to compute the fundamental waves in multilayer structures.[13,14] We describe here a generalization of the standard matrix method which computes the reflected and transmitted waves of the nth harmonic that are generated both within a set of nonlinear layers in a multilayer structure, and from the interfaces between the layers. It is used to describe the reflected second-harmonic light, and transmitted third-harmonic light, from Si(001) layers surrounded by thick thermal oxide layers as a function of the thickness of the upper SiO_2 and the silicon layer. In general, the harmonic waves arising from different source polarizations, whether bulk or surface, depend differently on the silicon layer thickness, d. Consequently, the thickness dependence of the reflected or transmitted harmonic wave can yield valuable information on the relative strength of the various contributions to harmonic generation, and can discriminate between surface and bulk sources.

In the case of third-harmonic generation (THG) from silicon, surface and bulk sources have significantly different dependence for $d < 20$ nm and this can be used to assess suggestions of a significant surface enhancement to THG.[15,16] For second-harmonic generation (SHG) from a (001) surface, distinguishing the surface and bulk contributions is more involved. Many authors have argued strenuously that the second-harmonic signal from Si(001) arises from the surface, not the bulk.[8,17–19] The arguments have been based on symmetry considerations, as well as the influence of surface modifications, such as reconstruction in UHV and exposure to oxidation. Heinz et al. observed a pronounced decrease in second-harmonic generation from 7×7 reconstructed Si(111) surfaces that were exposed to sufficient oxygen to produce roughly one

atomic layer.[8] They attributed this drop to a disordering of the surface electronic states caused by the oxidation, although some portion of the decrease undoubtedly arises from the shifting of dangling electronic states by bonding with oxygen atoms. Regardless of the mechanism, however, they argued that the strong influence of a single atomic layer of oxygen on the signal demonstrated that the second-harmonic signal was dominated by the surface.

As demonstrated by Sipe et al., however, there is a fundamental difficulty in separating the SHG arising from the bulk isotropic source (parameterized by γ in Eq. (4) below) and from one of the surface sources for reflected SHG from a semi-infinite slab.[20,21] Unfortunately, this remains true for a thin layer of varying thickness, as well. Lüpke et al. made clever use of oxidized vicinal (miscut) Si surfaces to effect a separation, finding that the isotropic bulk contribution to the second-harmonic signal to be an order of magnitude greater than the surface contribution,[22] in contradiction to many previous findings. However, the results for the isotropic response were not unambiguous on a Si(001) surface.[23] In the present case of SHG from thin layers, all terms but the pair discussed by Sipe have unique signatures with layer thickness d; hence, their relative importance can be investigated by measuring the thickness dependence of the reflected second-harmonic light.

In the following section, the method we use to compute the nth harmonic generated by a multilayer structure are outlined, with the details left to an appendix. We then describe the fabrication and characterization of $SiO_2/Si/SiO_2$ thin-film structures, which we produce by oxidation of silicon-on-insulator substrates. By using a spatial temperature gradient during a portion of the oxidation process, the silicon layers are gently tapered to facilitate measurements of the harmonic generation properties as a function of silicon layer thickness.[16,24,25] We report the third-harmonic-generation properties of the multilayers in rotation and in translation across the thickness gradient. The results are found to be well described by the dipole-allowed bulk source, with negligible surface contribution. We then report measurements, in rotation and as a function of layer thickness, of the SHG from these samples, and discuss their significance in light of literature values for the second-harmonic susceptibility components. We conclude with a discussion of the potential of the approach, and the appendix which discusses in detail the matrix method of computing harmonic generation from multilayers.

II. THEORY

As light passes through a multilayer structure it refracts and reflects at each interface, leading to waves inside the structure that propagate in both the forward and backward directions. When the strength of the reflected wave is appreciable, interference between the in-
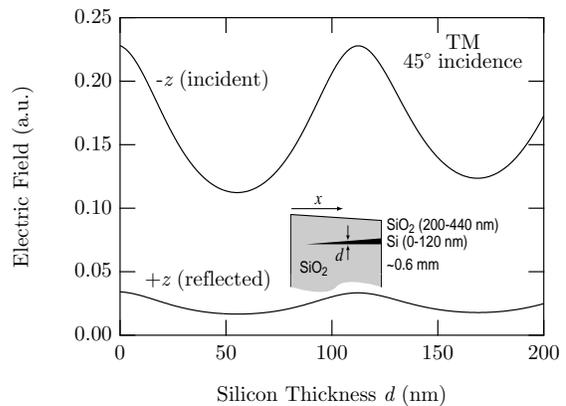
FIG. 1: The electric field strength just inside the front surface of a silicon layer on a SiO_2 substrate as a function of silicon layer thickness. The calculation assumes an upper layer of SiO_2 produced by partial oxidation of a 200-nm Si(001) layer, as illustrated in the inset.

cident and reflected waves leads to a significant modulation of the fundamental field strength within the layers. In the Si/SiO_2 system, for example, the field strength for a p-polarized fundamental wave at 800 nm, incident at $45°$, varies by more than 50% with silicon layer thickness d, as shown in Fig. 1. The variation for an s-polarized wave is even greater. Since the field in the layers generates an nth order nonlinear polarization proportional to the nth power of the field strength, the nonlinear polarization exhibits an even stronger modulation than the fundamental field. It is thus necessary to include both the incident and the reflected waves in the computation of the nonlinear polarization.

The reflected fundamental wave in the multilayer gives rise to another significant difference in the problem. The nth-order polarization at frequency $\Omega = n\omega$, $P^{(n)}(\Omega)$, arises from n factors of the fundamental field, which now has both forward- and backward-going terms. Hence, the nonlinear polarization comes from all combinations of forward and backward waves. The backward wave being weaker, the terms tend to diminish with the number of factors of the backward field. However, the phase mismatch for a source polarization that combines fields from both the forward and backward waves—as well as combinatorics—may compensate this reduction, so that terms combining both forward and backward waves may dominate the nonlinear polarization. Such is indeed the case for third-harmonic generation with a fundamental beam at $\lambda = 800$ nm in thin silicon layers surrounded with oxide. Furthermore, the presence of forward and backward waves in a cubic material relaxes a symmetry constraint on the second-harmonic susceptibility, leading to an additional bulk contribution to the nonlinear polarization that is not present in single-beam experiments.

We have developed an extension to the standard matrix method for computing (linear) transmission and reflection from multilayers to calculate the reflected and
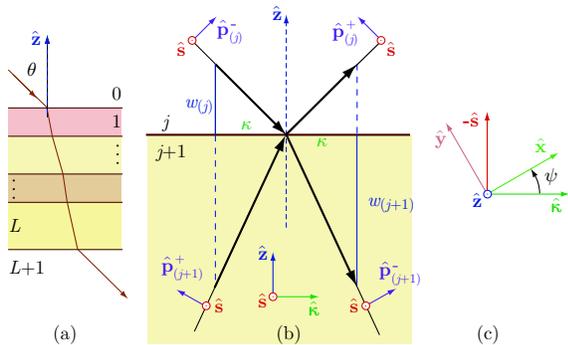
FIG. 2: (Color online) (a) The beam is incident at θ on layer 1, whose upper surface is at $z = 0$. Note that the outward normal of layer 1 is in the positive z direction. For convenience in certain expressions, the incident medium is numbered 0 and the substrate is numbered $L+1$. (b) Reflection and refraction at the boundary between layers j and $j + 1$. All waves have the same component κ parallel to the interface in the plane of incidence (POI); the z component of the wave vector in each layer j is denoted w_j and is found from Snel's law. (c) Relationship between POI axes $(\hat{\boldsymbol{\kappa}}, -\hat{\mathbf{s}}, \hat{\mathbf{z}})$ and crystallographic axes $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$. The azimuthal angle ψ measures the rotation of crystallographic [100] from the $\hat{\boldsymbol{\kappa}}$ direction.

transmitted harmonic waves from multilayer systems composed of both linear and nonlinear layers. We assume nonmagnetic materials ($\mu = 1$), and hence waves with their electric vector in the plane of incidence (p-polarized) are decoupled from those with their electric vector perpendicular to the plane (s-polarized). The principal steps of the calculation are outlined here, with details left to the appendix.

The incident wave has the form

$$\mathbf{e}_{(0)}(\mathbf{r}, t) = \left[e_{(0)s}^{-}\hat{\mathbf{s}} + e_{(0)p}^{-}\hat{\mathbf{p}} \right] e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} + \text{c.c.} \, ,$$

where the wave vector is $\mathbf{k} = \kappa\hat{\boldsymbol{\kappa}} - w_{(0)}\hat{\mathbf{z}}$, and $\kappa = \tilde{\omega}\sin\theta$, $w_{(0)} = \tilde{\omega}\cos\theta$, and $\tilde{\omega} = \omega/c$. The unit vectors $\hat{\mathbf{s}}$ and $\hat{\mathbf{p}}$ are indicated in Fig. 2, and we take $\hat{\mathbf{z}}$ to be the surface normal with positive projection along the direction of the reflected beam. (In most respects, the notation follows Ref 20.) We use lowercase symbols for the fundamental fields, and uppercase symbols for polarizations and fields at the nth harmonic. Subscripts in parentheses indicate the layer number.

The first step is to compute the fundamental fields throughout the multilayer using the usual matrix method,[13,14] which represents the forward and backward waves at each depth in the multilayer structure as a two-dimensional column vector. A diagonal 2×2 complex matrix describes the evolution of the waves as they propagate through each layer, and a symmetric 2×2 matrix describes reflection and transmission at each interface. Multiplying in order the succession of interface and layer matrices yields a single matrix which describes the coupling of forward and backward waves through the entire

structure:

$$\begin{pmatrix} e_{(0)s}^{+}(0) \\ e_{(0)s}^{-}(0) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} e_{(L+1)s}^{+}(-D) \\ e_{(L+1)s}^{-}(-D) \end{pmatrix} . \quad (1)$$

Since there is no wave incident from the back of the multilayer structure at $z = -D$, $e_{(L+1)s}^{+}(-D) = 0$, and we can solve for the reflected and transmitted fields, $e_{(0)s}^{+}(0)$ and $e_{(L+1)s}^{-}(-D)$, respectively. An expression identical in form exists for the p-polarized component of the incident wave, although the coefficients of the transfer matrix are generally different.

The next step is to compute the nonlinear polarization at the harmonic frequency $\Omega = n\omega$ induced in the bulk of a nonlinear layer. The form of this polarization depends on the order of the nonlinear process and the symmetry of the material. Details aside, the matrix method allows one to compute the reflected and transmitted harmonic light. For specificity in the case of layers with cubic symmetry, such as silicon, the bulk polarization for both second-harmonic and third-harmonic generation takes the form

$$P_i^{(2\omega)} = \chi_{ijkl}^{(2)} \tilde{e}_j \nabla_k \tilde{e}_l \, , \quad (2)$$

$$P_i^{(3\omega)} = \chi_{ijkl}^{(3)} \tilde{e}_j \tilde{e}_k \tilde{e}_l \, , \quad (3)$$

where $\chi_{ijkl}^{(n)}$ is a fourth-rank tensor, repeated indices are summed over, and the tilde indicates the sum of forward and backward waves. It is customary in these expressions to use the field inside the layer, thereby avoiding ambiguity at the discontinuity at the interface;[26] the gradients are with respect to the field coordinates. Cubic symmetry implies that components of $\chi^{(n)}$ having an odd number of indices along any Cartesian direction (in the conventional cubic crystal basis) must vanish. There are thus only two independent nonzero components of $\chi_{ijkl}^{(3)}$, which we denote χ_{1111} and χ_{1122}. The six distinct permutations of the indices in the second term are all equal.

The second-harmonic case is slightly more complicated, since one of the indices corresponds to a derivative and can thus be distinguished from the other two. Following Bloembergen, we will express the second-harmonic polarization in the form

$$P_i^{(2\omega)}(\mathbf{r}) = \beta e_i(\nabla \cdot \mathbf{e}) + \gamma \nabla_i(\mathbf{e} \cdot \mathbf{e}) + \zeta e_i \nabla_i e_i + \delta'(\mathbf{e} \cdot \nabla)e_i \, , \quad (4)$$

where β, γ, ζ, and δ' are phenomenological constants and the axes are assumed to coincide with the standard cubic crystal axes.[3,27,28] The first term vanishes for plane waves; the second (bulk isotropic) term produces a polarization independent of crystal orientation; the third (bulk anisotropic) term produces an anisotropic polarization, which leads to a modulation in the generated second-harmonic light as the crystal is rotated about its surface normal; and the fourth (mixed source) term vanishes for isolated plane waves, but in multilayer structures produces a polarization parallel to the layers and in the

plane of incidence. Measurements on multilayer structures can thus provide information on a phenomenological susceptibility component, δ', that is inaccessible to single-beam experiments on thick samples.

Irrespective of the detailed form of the nonlinear source polarization, it generates bound and free harmonic waves in the bulk of the jth layer that together solve the (inhomogeneous) wave equation[26]

$$\nabla(\nabla \cdot \mathbf{E}_{(j)}^m) - \nabla^2 \mathbf{E}_{(j)}^m - \tilde{\Omega}^2 N^2 \mathbf{E}_{(j)}^m = 4\pi \tilde{\Omega}^2 \mathbf{P}_{(j)}^m. \quad (5)$$

In this expression, the nonlinear polarization $\mathbf{P}_{(j)}^m$ includes n factors of the fundamental field, of which m come from the backward (reflected) fundamental wave and $(n - m)$ from the forward (transmitted) fundamental wave. Boundary conditions at the interface require that all waves have the same in-plane component of the wave vector, $K\hat{\boldsymbol{\kappa}} = K(\hat{\mathbf{z}} \times \hat{\mathbf{s}})$. The bound wave has the form

$$\mathbf{E}_{(j)}^m = \mathbf{A}_{(j)}^m e^{iKx + iq_{(j)}^m z - i\Omega t} + \text{c.c.}, \quad (6)$$

where $\mathbf{A}_{(j)}^m$ is a constant vector whose magnitude and direction depend on the nonlinear source polarization $\mathbf{P}_{(j)}^m$, as well as the linear optical response at ω and Ω, and the wave vector of the bound wave, and

$$q_{(j)}^m = (2m - n)w_{(j)} \quad (7)$$

is the z-component of the bound wave vector. The free waves traverse the layer in the forward and backward directions consistent with the interfacial boundary conditions; namely, with wave vectors

$$\mathbf{K}_{(j)}^{\pm} = K\hat{\boldsymbol{\kappa}} \pm W_{(j)}\hat{\mathbf{z}}. \quad (8)$$

Mathematically, they "interact" with the bound waves purely at the interfaces, where boundary conditions obtained from the Maxwell equations cause discontinuities in the freely propagating waves at Ω (see Eqs. (38)-(41)).

In addition to the bulk nonlinear polarization, there may be surface nonlinear sources, which also contribute to the discontinuities of the free harmonic waves. Crystal symmetry that prohibits even-order harmonic generation in the dipole approximation is broken at an interface, leading to an enhanced role for surfaces and interfaces. Furthermore, the strong gradient in electric field across an interface between dissimilar dielectric materials may cause a significant enhancement in harmonic generation.[21,28] Assuming that the interfacial region is at most a few atoms thick, the phase difference across the thickness can be neglected and the local and nonlocal nonlinear response of the interface can be represented[28] by a surface nonlinear polarization and susceptibility of the form

$$
\begin{aligned}
P_i^{\mathrm{S}}(2\omega) &= \chi_{ijk}^{\mathrm{S}} e_j e_k \, \delta(z - z_0) = \Pi_i \, \delta(z - z_0). \\
P_i^{\mathrm{S}}(3\omega) &= \chi_{ijkl}^{\mathrm{S}} e_j e_k e_l \, \delta(z - z_0) = \Pi_i \, \delta(z - z_0).
\end{aligned}
\quad (9)
$$

With no loss of generality, the nonlinear dipole sheet is assumed to lie just within the layer.[21] For an interface at z_0 between (upper) layer j and (lower) layer $j + 1$, layer j's dipole sheet is at $z_0 + \delta$ and layer $(j + 1)$'s is at $z_0 - \delta$.

Once the nonlinear sources are known, we can relate the freely propagating harmonic fields at each interface via

$$
\begin{pmatrix} E_{(j)}^+ \\ E_{(j)}^- \end{pmatrix} = t^{-1} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} E_{(j+1)}^+ \\ E_{(j+1)}^- \end{pmatrix} + \begin{pmatrix} S_{(j)}^+ \\ S_{(j)}^- \end{pmatrix}, \quad (10)
$$

where $r = r_{(j,j+1)}$ and $t = t_{(j,j+1)}$ are the reflection and transmission coefficients at Ω, and the terms $S_{(j)}^{\pm}$ combine the surface and bulk field discontinuities at the interface (see Eqs. (46) and (56)). The remainder of the solution follows the linear case. The key point here is that the bulk and surface polarization terms in $S_{(j)}^{\pm}$ are series of the form

$$\sum_{m=0}^n c_m \binom{n}{m} [e_{(j)}^-]^{n-m} [e_{(j)}^+]^m e^{iq_{(j)}^m d} + \text{c.c.} \quad (11)$$

in the thickness d of layer j. The coefficients c_m depend on the incident angle, polarization, linear dielectric properties of the layer at both ω and Ω, as well as the normal component of the source polarization (for the bulk terms). In general, different susceptibility components give rise to different coefficients c_m, and so have different dependence on the layer thickness d. This is illustrated in Fig. 3, which shows the thickness dependence of the one surface and four bulk sources of third-harmonic generation in the cubic material silicon. (Here and in the following, we assume negligible THG from SiO_2.) Clearly, a measurement of the thickness dependence of THG can distinguish between the surface and bulk mechanisms.

III. EXPERIMENT

Thin silicon layers surrounded by SiO_2 were prepared from Unibond silicon-on-insulator substrates specially manufactured by Soitec, which had a 200-nm c-Si(001) layer press-bonded to a fused silica substrate.[29,30] Roughly rectangular pieces were cut from the 4-inch wafers, with their long axis aligned with the [110] direction. The pieces were oxidized in dry O_2 in a quartz tube furnace at temperatures between 950°C and 1050°C. A spatial temperature gradient was used to produce a silicon layer of gently tapered thickness[24,25] varying by approximately 30 nm over a lateral distance of ~ 30 mm. The structure is illustrated in the inset of Fig. 1. The gradual taper of the layer permits the thickness dependence of the harmonic light generated either in reflection or transmission to be investigated on a single sample with consistent orientation and oxide-interface conditions. To explore layer thicknesses d from 0 to 120 nm, several different samples were prepared having overlapping thickness ranges.
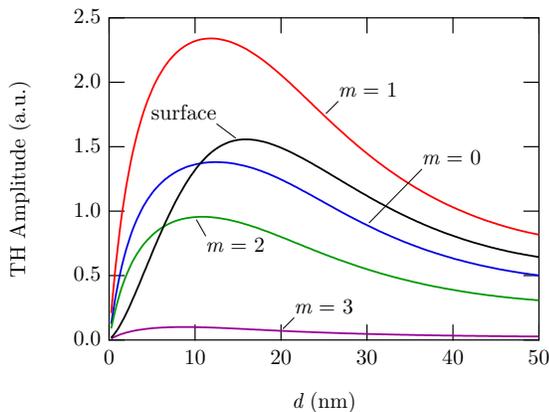
FIG. 3: (Color online) Silicon layer thickness dependence of each term in the third-order nonlinear polarization to the transmitted third-harmonic amplitude at normal incidence. Notice that among the bulk terms, the term arising from two factors of the field in the incident direction and one in the reflected direction ($m = 1$) produces the strongest signal. Note also that the rise in the surface term is appreciably slower for $d < 10$ nm than the bulk terms. A measurement of the thickness dependence of the transmitted THG thus permits one to distinguish surface and bulk contributions.

FIG. 4: (Color online) Normal incidence transmission spectra through a tapered $SiO_2/Si/SiO_2$ multilayer structure. The scans are made in a Cary 2400 spectrophotometer using a mask to limit the size of the 1-nm-bandwidth beam to $\approx$ 1 mm $\times$ 2 mm on the sample. Spectra were fitted (smooth curves) to deduce the thickness of the upper oxide layer and the silicon layer (indicated in the legend), with a resolution of ≈ 0.2 nm averaged over the rectangular beam. The inset shows the thickness profile obtained from the fits.

The roughness of the upper Si-SiO_2 interface was investigated with atomic force microscopy after removing the upper oxide layer with buffered HF. This treatment causes minimal distortion of the silicon surface. The RMS roughness of the exposed surface was typically ≈ 0.2 nm for a layer thickness of 15 nm, as determined from scans 230 nm on a side.

The thickness profile of the silicon layer and oxide overlayer were determined by measuring normal-incidence transmission spectra with a spectrophotometer in the range 200–800 nm, and fitting to the thin-film equations. Typical spectra and fits are shown in Fig. 4, from which a smooth mapping of position x along the length of the sample into silicon layer thickness d was created for each sample. An example is shown in the inset of the figure.

With decreasing d below 6 nm, departures from the dielectric function of bulk c-Si (Ref. 14) were observed, particularly in the neighborhood of the E_1 point at 3.39 eV.[31,32] In this thickness range quantum confinement effects significantly modify the dielectric function. However, good fits were obtained using Yamaguchi's thickness-dependent semi-empirical model of the dielectric function of c-Si layers,[32] as illustrated in Fig. 5.

The third-harmonic generation properties of the silicon layers were studied using a 100-MHz mode-locked Ti:sapphire laser (Clark-MXR NJA-5) producing 60-fs pulses centered at 820 nm, and the setup illustrated in Fig. 6.[16] The beam was polarized with a Glan-laser polarizer and focused with a 40-mm focal length lens to a spot radius of 7 μm; the peak intensity was below 20 GW/cm^2. The transmitted THG was analyzed with an identical Glan-laser polarizer, then dispersed with an

FIG. 5: (Color online) Illustration of a fit to the transmission spectrum of a $SiO_2/Si/SiO_2$ multilayer. The measured normal-incidence transmission data are shown in open circles. The best fit using the index of refraction of bulk silicon is shown in the dashed line, which exhibits appreciable disagreement with the data throughout the visible and ultraviolet range of the spectrum. The significantly improved fit using a thickness-dependent semiempirical index expression developed by Yamaguchi, shown in the heavy line, agrees much better with the data and yields a silicon layer thickness of 1.9 nm.
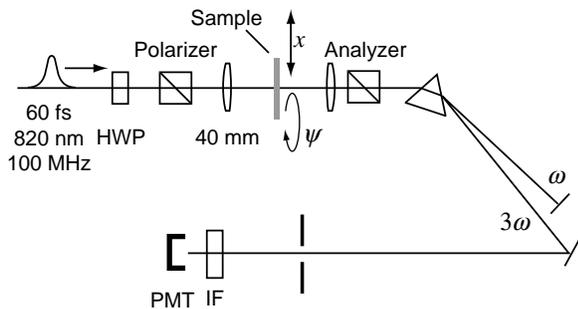
FIG. 6: Optical setup of the THG measurements. The signal from the PMT was recorded either via photon counting, for measurements as a function of film thickness, or as current using a picoammeter, for the rotational scans.

equilateral fused silica prism, and detected by a thermo-electrically cooled photomultiplier tube (PMT). A 20-nm bandpass interference filter centered at 266 nm rejected stray fundamental light. Care was taken to align the sample plane with the translation axis to within 2 mrad to avoid changes in focusing at the surface when measuring the thickness dependence of the THG. The TH signal generated in transmission for a fundamental beam incident along the sample normal was measured as a function of sample rotation ψ at fixed sample position, using a rotation stage. It was also measured as a function of position for fixed sample rotation, using a translation stage.

The same laser system was used to investigate second-harmonic generation from the same set of oxide-cladded silicon layers in reflection at 45° incidence. For these measurements, the final interference filter was replaced with a pair of blue glass filters to block stray fundamental light.

In addition, the SHG properties of the samples were studied using a 1-kHz amplified Ti:sapphire laser system (Spectra-Physics Millennia-pumped Tsunami seeding an Evolution-pumped Spitfire) producing 60-fs pulses. The setup is illustrated in Fig. 7. The beam was chopped at ~ 320 Hz and the output of the PMT was detected with a digital lock-in amplifier. A 780-nm long-pass filter was placed immediately prior to the focusing lens to remove any blue light generated at mirror surfaces prior to the sample, and to provide a reference reflection. This reflected beam was focused into a KDP crystal and the generated blue light was detected with an amplified silicon photodiode to monitor laser stability. Additional amplified photodiodes were used to monitor the reflected and transmitted fundamental beams.

To avoid excessive intensity on the sample, the beam was focused with a 10-cm focal length cylindrical lens to a width of ~ 20 μm along the direction of the silicon thickness gradient, as measured by replacing the sample with a razor blade and monitoring the transmitted beam. The average irradiance was less than 90 W/cm^2 and the peak intensity was approximately 600 GW/cm^2.

Although this peak intensity exceeds the reported
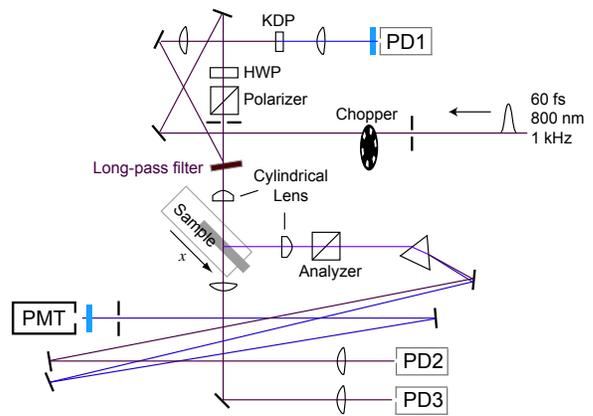
FIG. 7: (Color online) Setup for measuring SHG with the amplified laser system. The 1-kHz beam is mechanically chopped, and filtered with a long-pass glass filter to remove blue light generated at metal mirror surfaces prior to the sample. The beam is focused with a cylindrical lens onto the sample mounted parallel to the translation axis of a computerized stage. The reflected SHG is detected with a cooled PMT and measured with a digital lock-in amplifier. The reflected and transmitted fundamental beams are detected with amplified photodiodes PD2 and PD3, as is the second-harmonic produced by the portion of the input beam reflected from the long-pass filter (PD1).

damage threshold for silicon of 100 GW/cm^2,[7] we observed no permanent damage to the sample at these intensities and obtained results entirely consistent with those using the unamplified laser system and < 20 GW/cm^2 on all samples. Presumably, the thick upper oxide layer serves to stabilize the silicon surface.

IV. RESULTS

A. Third Harmonic

Rotational scans of the third-harmonic intensity polarized parallel and perpendicular to the incident field are shown in Fig. 8. The parallel-polarization data were obtained at a silicon thickness of 34 nm; they were fitted to the expression

$$I_\parallel(3\omega) \propto [3\chi_{1111} + 3\chi_{1212} - (3\chi_{1212} - \chi_{1111})\cos 4\psi]^2$$
$$\propto [(4+\sigma) - \sigma\cos 4\psi]^2 , \quad (12)$$

where $\sigma \equiv 3\chi_{1212}/\chi_{1111} - 1$ is the anisotropy parameter.[33] The resulting value of $\sigma = 0.71\pm0.01$ is consistent with previous work on bulk samples using 770-nm femtosecond pulses,[15,22] and differs slightly from a value at 819 nm using nanosecond pulses.[33]

The perpendicular THG has the expected eightfold dependence given by

$$I_\perp(3\omega) \propto \chi_{1111}^2 \sin^2 4\psi . \quad (13)$$
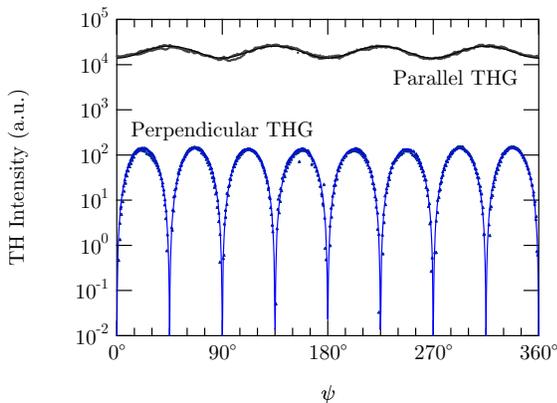
FIG. 8: (Color online) Normal-incidence transmitted THG from thin silicon layers as a function of azimuth ψ with respect to the (100) direction. The THG polarized parallel to the fundamental beam was taken where the layer thickness was 34 nm; the perpendicular THG was taken at 22-nm thickness. The smooth curves are fits to Eqs. (12) and (13), as discussed in the text.

The scan shown here was taken at a layer thickness of 22 nm; similar traces were obtained for both parallel and perpendicular polarization throughout the range of this sample, which was 0–35 nm. The rotational THG scans thus confirm nicely the cubic symmetry of the silicon samples, although they cannot distinguish surface and bulk contributions, which have identical symmetry.

As is clear from Fig. 3, the dependence of the transmitted THG on silicon layer thickness d can distinguish the two contributions. The data for the critical region satisfying $d < 30$ nm are shown in the upper panel of Fig. 9, along with computed curves that assume either pure bulk third-harmonic generation or pure surface third-harmonic generation. The polarizer and analyzer were parallel for these data, and aligned with the [110] axis of the silicon layers. The signal level at negative layer thickness (corresponding to positions with no remaining silicon, as indicated in the inset of Fig. 1) represents a background level of stray fundamental light leaking through the interference filter, as was confirmed by its linear dependence on laser intensity. We thus confirm that third-harmonic generation from a bare SiO_2 surface is weaker by at least a factor of 100 than that from a silicon layer a few nanometers thick. Note that the background light was more effectively suppressed in the rotational data of Fig. 8 due to the significantly greater distance between the prism and the PMT in that setup.

B. Second Harmonic

The second-harmonic case in silicon is significantly more involved than the third-harmonic case, since there are many more symmetry-allowed susceptibility components. In addition to the bulk terms γ, ζ, and
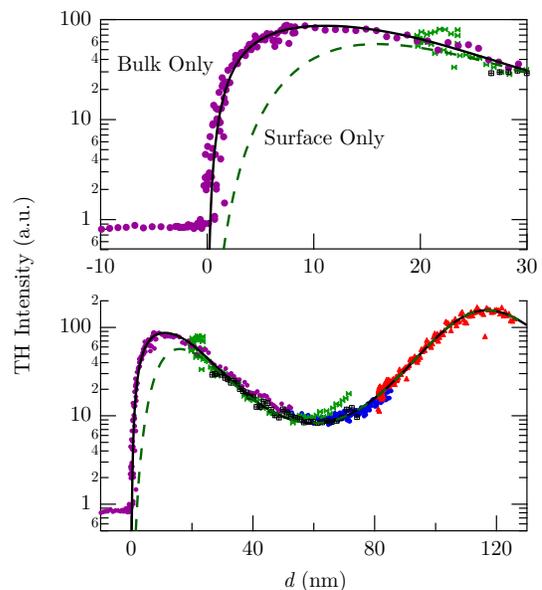
FIG. 9: (Color online) Transmitted third-harmonic intensity as a function of silicon layer thickness at normal incidence. Different symbols correspond to different samples. Data at negative silicon layer thickness correspond to regions on the sample in which the silicon layer has been entirely oxidized. These serve to show the background signal level, which is dominated by scattered fundamental light. The smooth curve shows a calculation of the expected THG assuming purely a bulk source, scaled to match the data at large d; similarly, the dashed curve assumes solely a surface source. The lower panel shows the data from five different samples; the upper panel shows the same data in the range $d < 30$ nm.

δ' of Eq. (4), for a Si(100) surface there is a surface polarization[26]

$$\begin{pmatrix} \Pi_x \\ \Pi_y \\ \Pi_z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & \partial_{15} & 0 \\ 0 & 0 & 0 & \partial_{15} & 0 & 0 \\ \partial_{31} & \partial_{31} & \partial_{33} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{e}_x^2 \\ \tilde{e}_y^2 \\ \tilde{e}_z^2 \\ 2\tilde{e}_y\tilde{e}_z \\ 2\tilde{e}_x\tilde{e}_z \\ 2\tilde{e}_x\tilde{e}_y \end{pmatrix} \quad (14)$$

that is parameterized by three independent susceptibility components, $\partial_{31} \equiv \partial_{zxx}$, $\partial_{33} \equiv \partial_{zzz}$, and $\partial_{15} \equiv \partial_{xxz}$. Note that the Cartesian indices here refer to the standard cubic basis. All three of these terms produce a nonlinear polarization that lies in the plane of incidence, and hence they contribute only to p-polarized SHG. Furthermore, they are independent of azimuth ψ for rotation about the surface normal. The same is true for the bulk terms proportional to γ and δ'. Only the ζ term gives rise to rotational anisotropy; it alone produces s-polarized SHG for an s- or p-polarized incident beam. For p-polarized incident light, the s-polarized SHG has the form

$$I_{ps}(\psi) \propto \sin^2(4\psi). \quad (15)$$

The bottom two traces in Fig. 10 show the measured
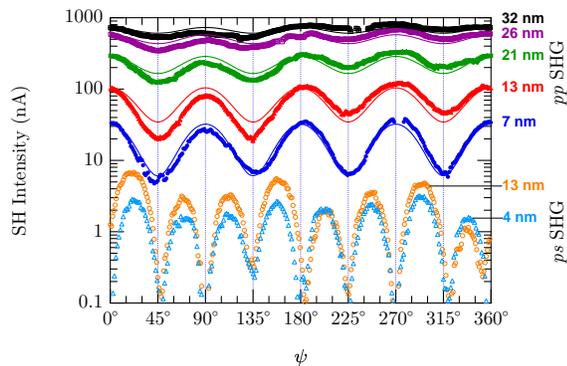
FIG. 10: (Color online) Intensity of reflected second harmonic from thin silicon layers as a function of rotation angle ψ about the sample normal for a p-polarized fundamental beam incident at $45°$. The lower two traces show s-polarized SHG for 4 and 13 nm layers. They show the expected zeroes every $45°$, but uneven maxima between them. The upper 5 traces show the p-polarized SHG from the silicon layer from thicknesses between 7 nm and 32 nm. These approximate the 4-fold symmetry expected for (001) surfaces, and a degree of modulation that clearly depends on the layer thickness.

ps-SHG as a function of sample rotation ψ for two thicknesses of silicon layer. The p-polarized laser beam was incident at $45°$; the data were recorded by rotating the sample at the rate of 1 revolution per minute and digitizing the output of a picoammeter monitoring the PMT current. Care was taken to ensure that the incident laser beam was focused on the center of rotation, and that the sample normal coincided with the rotation axis. Scans frequently ran for 2 revolutions, to ensure consistency and to check for any laser-induced modification of the sample. Although both traces exhibit the expected zeroes every $\pi/4$, the eight maxima at $\frac{\pi}{8}(2n+1)$ are not all equivalent, as predicted by Eq. (15). Care was taken to ensure that the polarizer and analyzer were properly oriented by adjusting for maximum extinction of the fundamental beam, with a resolution of $0.1°$. Traces were also taken with the polarizers slightly misaligned, resulting in increased asymmetry of the 8 peaks. The two traces shown in Fig. 10 are representative of the most symmetric $I_{ps}(\psi)$ scans obtained on this and other samples.

The upper 5 curves show $I_{pp}(\psi)$ for thicknesses from 7 nm to 32 nm. The expected form for these curves is

$$I_{pp}(\psi) \propto |1 + \rho\cos(4\psi)|^2, \qquad (16)$$

where the fourfold term arises from the bulk anisotropic source (ζ), whose relative strength ρ is compared to the combination of the surface sources and the two isotropic bulk sources (γ and δ'). As is the case for I_{ps}, the curves of $I_{pp}(\psi)$ depart from the expected fourfold symmetry of Eq. (16), shown in the smooth curves. Despite this disagreement, it is clear from the traces that the relative magnitude of the modulation (the relative strength of the ζ term to the sum of the other terms) depends
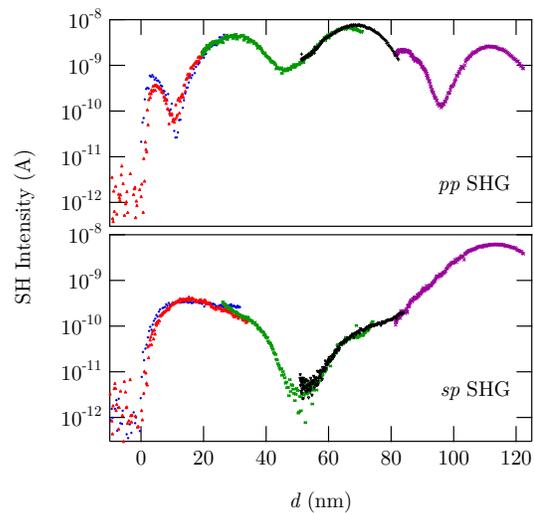
FIG. 11: (Color online) p-polarized SHG from a p-polarized fundamental (upper panel) and from an s-polarized fundamental (lower panel) as a function of silicon thickness d. These data were taken at $45°$ incidence using the amplified laser system with lock-in detection of the PMT current. Data from the five samples of Fig. 9 are superposed here. Data at negative silicon thickness are taken past the edge of the silicon film and show the background level from the SiO$_2$ substrate.

appreciably on the silicon layer thickness. This is to be expected from Eq. (11), and is consistent with the principal idea that the dependence of harmonic generation on d can provide information on the strength of the various sources and susceptibility components.

The thickness dependence of p-polarized SHG was studied using the setup of Fig. 7 and the five samples whose third-harmonic properties had been studied previously. They were translated parallel to the plane of incidence and in the direction of the silicon thickness gradient along [110]. In addition, the intensity of the reflected and transmitted fundamental beams were recorded, as was a reference SHG beam obtained by doubling a reflected portion of the incident beam in a KDP crystal. The results are shown in Fig. 11. The various scans were scaled modestly to match in the regions of overlap, consistent with day-to-day variations in the laser intensity.

Both $I_{pp}(2\omega)$ and $I_{sp}(2\omega)$ exhibit minima near $d = 50$ nm and maxima near $d = 110$ nm, consistent with the minimum and maximum in the laser field inside the silicon layer (see Fig. 1). However, $I_{pp}(2\omega)$ displays additional pronounced minima at ≈ 8 nm and ≈ 95 nm. The minimum at 8 nm is particularly remarkable, as this length is significantly shorter than any other length scale in the problem (see Table I). However, it is a robust feature that appears in multiple samples. Furthermore, the linear reflectivity is perfectly well behaved in this region, as was the third-harmonic signal from these samples (Fig. 9).

TABLE I: Length scales in silicon. λ_0 is the vacuum wavelength, λ is the wavelength in silicon, $2\alpha^{-1}$ is the electric field e-folding length, where α is the (intensity) absorption coefficient.

Wave	n	λ_0	λ	$2\alpha^{-1}$
ω	$3.69 + 0.0065\,i$	800 nm	217 nm	19.6 μm
2ω	$5.57 + 0.39\,i$	400 nm	72 nm	163 nm
3ω	$1.86 + 4.47\,i$	267 nm	144 nm	9.4 nm

V. DISCUSSION

We begin with the third-harmonic data, which is simpler to interpret. Figure 8 shows excellent agreement with the expected 4-fold symmetry of the cubic material silicon, demonstrating clean THG signals and bulk symmetry. Furthermore, the thickness-dependence data of Fig. 9 show excellent agreement with the bulk-only curve, and significant disagreement with the surface-only curve, for the range $0 \le d \le 20$ nm over which the curves are distinguishable. At minimum, 90% of the THG signal arises from the bulk. For thicknesses greater than 20 nm, the two curves coincide, showing a pronounced dip at $d = 50$ nm, and a maximum near $d = 120$ nm. These modulations track the intensity of the fundamental beam, which arise from thin-film interference at the laser wavelength (see Fig. 1). The absorption depth at 3ω is only $\alpha^{-1} = 4.8$ nm, so that light generated in the backward direction and reflected off the input face is strongly attenuated in propagating through the layer to the exit face. Significant interference between the forward and backward waves induced at 3ω only occurs for $d \lesssim 2\alpha^{-1}$, and it is this interference which gives rise to the difference between surface and bulk contributions to harmonic generation. For $d \lesssim 2\alpha^{-1}$ the thickness dependence of harmonic generation can thus distinguish between different sources of harmonic generation.

Interpreting the second-harmonic results presents significant challenges compared to the third-harmonic data, both because there are many more unknown susceptibility terms and because of the increased prominence of the surface. Third-harmonic generation is not symmetry forbidden in the dipole approximation in silicon, and so one expects the bulk to dominate by virtue of the greater number of atoms able to participate. This is indeed confirmed here in Fig. 9.

Second-harmonic generation is dipole-forbidden in the bulk, and so we expect surface nonlinear sources to play a significant, if not dominant, role. For an ideal (001) surface, the additional 3 terms—parameterized by 3 complex susceptibility components, or 6 real coefficients—present no conceptual complication beyond the well-known fact that the bulk isotropic term γ is indistinguishable from the surface term ∂_{31}.[20]

However, an ideal (001) surface should produce four-fold symmetry in p-polarized SHG and eightfold symmetry in s-polarized SHG on rotation about the normal.
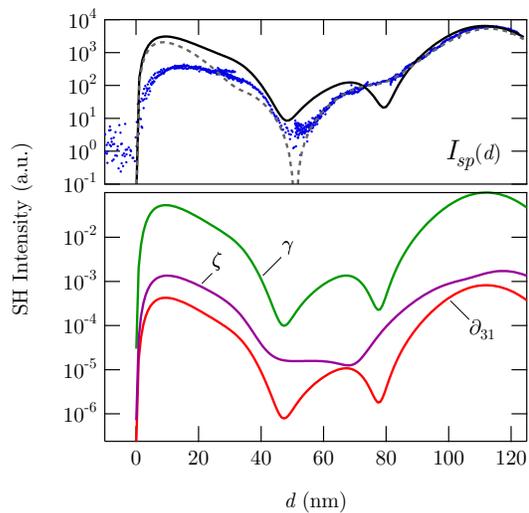
FIG. 12: (Color online) p-polarized SHG as a function of layer thickness for s-polarized excitation. The upper panel shows the data and two calculated curves. The solid curve uses the susceptibility components of Table II; the broken curve assumes $\zeta = -120 + 170\,i$ (holding γ and ∂_{31} unchanged). The lower panel shows the intensity produced by each susceptibility component in isolation. Note that γ and ∂_{31} are degenerate.

The third-harmonic rotational scans (Fig. 8) do exhibit analogous symmetries because of the dominant role of the bulk, but the second harmonic results of Fig. 10 show significant departures. Regarding our samples, we incline to Pauli's view that "God made solids, but surfaces were the work of the Devil." Ours are unintentionally miscut away from the (001) surface in an unknown direction and magnitude.

Lüpke et al. used surfaces intentionally miscut along directions of high symmetry to deduce from rotational SHG scans the magnitudes and phases of the various bulk and surface susceptibility components at $\lambda = 765$ nm.[22] Their results appear in Table II. Assuming we may neglect dispersion in $\chi^{(2)}$ between 765 nm and 800 nm, these coefficients should provide a reasonable approximation to our $I_{sp}(d)$ data, which depend only on the bulk anisotropic term ζ and the sum of the bulk isotropic term γ and the zxx surface term ∂_{31}, which are degenerate. The curve computed with values of Table II is shown superimposed on the data in the solid curve of the upper panel of Fig. 12. The calculation has been scaled to match the data approximately in the region of strongest signal near $d = 70$ nm. Clearly, it disagrees significantly from the data in a number of respects, most notably in an order-of-magnitude overshoot in the initial rise with d.

The lower panel of the figure shows curves of reflected second-harmonic intensity computed for each susceptibility component in isolation; i.e., the predicted reflected second-harmonic intensity obtained from a silicon layer whose second-order susceptibility components all vanish except the single component indicated on each curve,

TABLE II: Silicon SHG susceptibility components at $\lambda = 765$ nm reported by Lüpke *et al.*[22] for oxide layers grown at $550°$ C in steam to a thickness of 1–1.5 nm. Note that the value of the bulk isotropic term γ is arbitrarily defined to be 100, and the others are scaled with respect to this value.

Bulk	Surface
$\gamma = 100$	$\partial_{31} = -4.3 + 1.2\,i$
$\zeta = -66 - 5\,i$	$\partial_{15} = -29$
$\delta' = $ (not accessible)	$\partial_{33} = 35 - 9\,i$
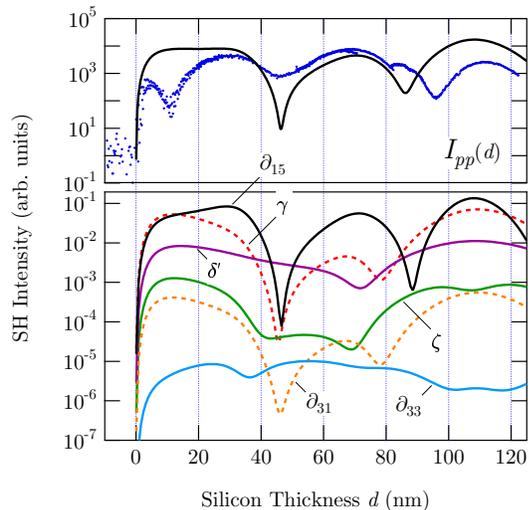
FIG. 13: (Color online) *pp*-SHG as a function of silicon thickness. The upper panel shows the data and a curve calculated from the susceptibility components of Table II. The lower panel shows the intensity produced by each susceptibility component in isolation. Note that γ and ∂_{31} (shown dashed) are degenerate. Lacking any previous measurements, a value of $\delta' = 100$ has been assumed for the mixed-wave term.

which takes the value of Table II. In the vicinity of $d = 80$ nm, where the γ curve has a pronounced minimum in disagreement with the data, the ζ curve does not (see the lower panel of Fig. 12). By increasing the magnitude of ζ by slightly more than a factor of 3 it is possible to track the data quite well in this region, as shown in the dashed curve. This suggests that the value of the bulk anisotropic term is probably greater than found in Ref. 22. The disagreement with the data near the minimum at 51 nm is probably inconsequential, as the signal is not far from the noise level there. However, the disagreement for $d < 20$ nm is still considerable and we hesitate to draw firm conclusions from the computed curve. Nonetheless, the strong modulation of the *pp*-SHG data for $d = 7$ nm shown in Fig. 10 demonstrates the significant contribution of the bulk anisotropic term ζ and suggests a value of $|\zeta/\gamma|$ greater than in Table II.

As seen in the upper panel of Fig. 13, the *pp*-SHG data also show significant disagreement with the susceptibilities of Table II. The solid curve was calculated

for $45°$ incidence along a (110) direction, and assumes $\delta' = 0$. Again, the data at small silicon layer thickness ($d < 30$ nm) differ appreciably from the predicted curve, which shows no hint of a dip near $d = 8$ nm. In fact, each component rises in the same quadratic way for small d, as they must.[12] By 8 nm there begins to be a modest separation among the possible contributions, and hence, with appropriate weighting, it is possible to arrange a minimum in this region. Holding all other susceptibilities at their values in Table II, and increasing the magnitude and phase of δ' until a minimum near 8 nm is obtained yields curve (b) of Fig. 14 for $\delta' = -370 + 190i$. However, this curve clearly disagrees with the data for $d > 35$ nm. Curves (c) and (d) illustrate the sensitivity of the reflected second harmonic to variations in the magnitude of δ', showing a 15% reduction and a 20% increase.

Alternatively, we can take the value $\zeta = -120 + 170\,i$ from the curve of Fig. 12 and seek the value of δ' that produces the best agreement. Varying only δ' generates curve (e) and a value $\delta' = -325 + 100\,i$, but this curve misses the minimum at 100 nm. Curve (f) yields a smaller value of δ' by effectively halving the value of ∂_{15} and introducing a $90°$ phase shift, which captures the deep minimum near 100 nm but appears shifted by about 10 nm in the range from 50–80 nm. Curves (g) and (h) show attempts to match both minima and maxima in the data by allowing ζ even greater latitude of variation. Curve (h) follows the data for $d < 80$ nm best, but fails to capture the depth of the minimum at 100 nm.

In short, we have found no set of parameters to describe the data quantitatively, but find qualitatively better agreement in the *sp* data for $|\zeta| = 210$, a factor of 3 greater than in Ref. 22, and that $|\delta'/\gamma|$ must be at least 0.8 to provide qualitative agreement with the minimum observed near 8 nm in the *pp* data. Some of this disagreement is almost certainly due to surface terms arising from misorientation of the silicon-oxide interfaces, which we do not model. However, the rotational curves of Fig. 10 suggest that the extent of the extraneous surface contributions is not so large as to vitiate a qualitative assessment of the importance of these susceptibility terms.

The $I_{sp}(d)$ data, particularly in the neighborhood of $d = 80$ nm (Fig. 12), suggest that ζ is significantly larger than determined in Ref. 22. In that work, the authors neglect the depletion field at an oxidized or bare silicon surface, which arises from trapping of majority carriers at midgap surface/interface states.[34] However, the 60 Ω cm, P-doped silicon samples of that study would have depletion fields of order volts per micron. Our samples are essentially undoped and should have negligible static fields along the normal, since dopants present in the original silicon source wafer diffuse out of the 200-nm Si layer during the initial flip-bonding and annealing process that produced the Unibond wafers, and certainly during the lengthy high-temperature oxidation we perform to thin the silicon layer. Static fields can significantly perturb the *p*-polarized SH signal[9,19] and would tend to enhance the isotropic response. Since our samples lack that en-
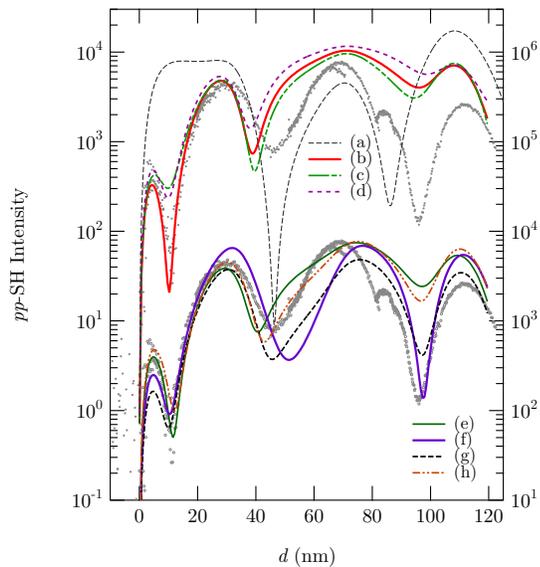
Curve	ζ	δ'	∂_{15}
(a)	$-66-5i$	0	-29
(b)		$-370+190i$	
(c)		$-320+160i$	
(d)		$-450+230i$	
(e)	$-120+170i$	$-325+100i$	
(f)	$-120+170i$	$-92+22i$	$0.7+14i$
(g)	$-175+52i$	$-65+44i$	$3.2+8.2i$
(h)	$-170+240i$	$-220+50i$	

FIG. 14: (Color online) Attempts to fit the $I_{pp}(d)$ data. The upper panel shows the curves calculated with the values of Table II, and different values of δ'. Values not shown in the table are identical to those of curve (a), and for all curves $\gamma = 100$, $\partial_{31} = -4.3 + 1.2\,i$, and $\partial_{33} = 35 - 9\,i$. To show additional curves, the data are shown shifted down in the lower portion of the figure. Curves (e) and (f) take the value of ζ estimated in Fig. 12 and adjust either δ' or additionally ∂_{15}. Curves (g) and (h) are attempts to fit with even greater departures for ζ.

hancement, it is reasonable for us to see a comparatively greater contribution from the bulk anisotropic term.

Many workers have reported time-dependent SH signals from silicon surfaces covered with thin oxide layers in air.[2,35–37] Heating effects can alter the efficiency with which a surface generates pp-SHG,[2] but this does not account for all observations. The mechanism proposed by Mihaychuk and coworkers involves electron transport through the thin oxide overlayer, combined with trapping at the free surface mediated by oxygen in the ambient atmosphere. The magnitude of the effect diminishes rather rapidly with the thickness l of the oxide layer, becoming negligible for $l \gtrsim 10$ nm.[36] We do not observe this effect in our samples, which have oxide layers thicker than 170 nm.

VI. CONCLUSIONS

We have developed a method for computing the reflected and transmitted harmonic light from a multilayer structure consisting of parallel layers of nonmagnetic materials. The calculation handles properly the significant variations in the amplitude of the fundamental field within the nonlinear layer(s) that arises from thin-film interference, and includes all combinations of fundamental fields in the generation of the harmonic waves. The calculation solves for the forward and backward going fundamental waves in each layer using the standard matrix method using only the linear dielectric constant at ω. These are used parametrically to compute the induced bulk and surface nonlinear polarizations at the harmonic frequency $\Omega = n\omega$. Augmented 3×3 matrices describe the coupling of the freely propagating waves at Ω across interfaces, accounting for the surface and bulk nonlinear sources, permitting the computation for the inhomogeneous wave equation to proceed formally in much the same way as the familiar matrix approach to solving the homogeneous problem. Because, in general, different susceptibility components have different layer-thickness (d) dependence, measuring reflected or transmitted harmonic light as a function of d can allow different contributions to be distinguished.

The method was applied to second- and third-harmonic generation in thin layers of c-Si surrounded by SiO$_2$ using $\sim$60 fs pulses at 800-nm from a Ti:sapphire laser. The nonlinear optical response of the silicon layers was described phenomenologically, using the symmetry-allowed susceptibilities for a (001) surface. At 3ω, bulk nonlinear polarization is dipole-allowed and the agreement between theory and experiment is excellent (see Fig. 9). The results demonstrate clearly that the dominant signal arises in the bulk. The second-harmonic case is complicated by the greater number of nonvanishing terms that must be considered. Dipole SHG is forbidden in the bulk, but electric quadrupole and magnetic dipole terms may be parameterized by three susceptibility components, one of which (δ') is studied here for the first time. For an ideal (001) surface there are three surface susceptibility components that contribute to the isotropic SH response, but for a miscut surface additional terms with rotational anisotropy emerge, complicating the analysis.

Although quantitative agreement of $I_{sp}(d)$ and $I_{pp}(d)$ was not obtained, certain features of these curves provide insights into the relative strength of the various susceptibility components. The lack of a deep minimum at $d = 80$ nm suggests that the signal from the isotropic and anisotropic terms are comparable and yields a value of $|\zeta/\gamma|$ that is significantly larger than Ref. 22. The $I_{pp}(d)$ data exhibit an unexpected, puzzling, and yet consistent and reproducible minimum at 8 nm, which is on a length scale appreciably shorter than any other natural length scale in the problem. This feature suggests an accidental cancelation among the various terms. Attempts

to fit the $I_{pp}(d)$ data were unsuccessful, but computed curves suggest that the value of the nonlinear susceptibility δ', which mixes the forward and backward fundamental waves, is comparable to the bulk isotropic susceptibility γ.

Appendix

We use a coordinate system in which the outward normal of the first layer is along the positive z direction, with the upper surface of the first layer at $z = 0$. As illustrated in Fig. 2, the incident fundamental beam at frequency ω propagates towards the interface from positive z and is incident at angle θ with respect to the normal.

We treat separately the case of s- and p-polarized fundamental waves. Consider first an s-polarized (TE) plane wave incident at angle θ in vacuum on the multilayer structure, as shown in Fig. 2. Inside layer j of (complex) index of refraction $n_{(j)} = \sqrt{\epsilon_{(j)}(\omega)}$, the electric and magnetic fields of this forward-going wave may be expressed as

$$
\begin{aligned}
\mathbf{e}^-_{(j)s} &= e^-_{(j)s} e^{i\mathbf{k}^-_{(j)}\cdot\mathbf{r}-i\omega t}\,\hat{\mathbf{s}} + \text{c.c.} \\
\mathbf{h}^-_{(j)s} &= -n_{(j)}e^-_{(j)s} e^{i\mathbf{k}^-_{(j)}\cdot\mathbf{r}-i\omega t}\,\hat{\mathbf{p}}^-_{(j)} + \text{c.c.},
\end{aligned}
\tag{17}
$$

where the wave vector is given by

$$
\begin{aligned}
\mathbf{k}^-_{(j)} &= \kappa\hat{\boldsymbol{\kappa}} - w_{(j)}\hat{\mathbf{z}}, \\
\kappa &= \tilde{\omega}\sin\theta, \\
\tilde{\omega} &= \omega/c, \\
w_{(j)} &= \sqrt{\epsilon_{(j)}\tilde{\omega}^2 - \kappa^2},
\end{aligned}
\tag{18}
$$

and the unit vector in the direction of the magnetic field $\mathbf{h}^-_{(j)s}$ is

$$
\hat{\mathbf{p}}^-_{(j)} = \frac{\kappa\hat{\mathbf{z}} + w_{(j)}\hat{\boldsymbol{\kappa}}}{n\tilde{\omega}}.
\tag{19}
$$

Lowercase symbols are used for waves at the fundamental frequency ω, and uppercase symbols will be used for waves at the harmonic frequency $\Omega = n\omega$. The in-plane component of the wave vector, κ, and the vacuum wave vector, $\tilde{\omega}$, are both real. When $n_{(j)}$ is complex, the normal component of the wave vector, $w_{(j)}$, is complex as well, and we take the root with positive real and imaginary components. Recall that we assume nonmagnetic layers ($\mu = 1$).

Reflections at the boundaries between layers produce a backward-going wave of the form

$$
\begin{aligned}
\mathbf{e}^+_{(j)s} &= e^+_{(j)s} e^{i\mathbf{k}^+_{(j)}\cdot\mathbf{r}-i\omega t}\,\hat{\mathbf{s}} + \text{c.c.}, \\
\mathbf{h}^+_{(j)s} &= -n_{(j)}e^+_{(j)s} e^{i\mathbf{k}^+_{(j)}\cdot\mathbf{r}-i\omega t}\,\hat{\mathbf{p}}^+_{(j)} + \text{c.c.}, \\
\mathbf{k}^+_{(j)} &= \kappa\hat{\boldsymbol{\kappa}} + w_{(j)}\hat{\mathbf{z}}, \\
\hat{\mathbf{p}}^+_{(j)} &= \frac{\kappa\hat{\mathbf{z}} - w_{(j)}\hat{\boldsymbol{\kappa}}}{n\tilde{\omega}}.
\end{aligned}
\tag{20}
$$

At the interface between layers j and $j + 1$, boundary conditions for the parallel component of the electric field and the magnetic field produce two equations linking the waves in the two layers. In the absence of sources, the parallel components of both electric and magnetic fields are continuous. From continuity of the electric field we obtain

$$
e^+_{(j)s} + e^-_{(j)s} = e^+_{(j+1),s} + e^-_{(j+1),s},
\tag{21}
$$

while continuity of the parallel component of the magnetic field gives

$$
\frac{w_{(j)}}{\tilde{\omega}}e^-_{(j)s} - \frac{w_{(j)}}{\tilde{\omega}}e^+_{(j)s} = \frac{w_{(j+1)}}{\tilde{\omega}}e^-_{(j+1),s} - \frac{w_{(j+1)}}{\tilde{\omega}}e^+_{(j+1),s}.
\tag{22}
$$

Equations (21) and (22) give the forward- and backward-going fields in the jth layer in terms of the fields in the $(j + 1)$st layer, the solution of which can be conveniently expressed in matrix form. If we define the transmission $t_{(j)}$ and reflectivity $r_{(j)}$ of the interface between layers j and $j + 1$ by

$$
\begin{aligned}
t_{(j),s} &\equiv \frac{2w_{(j)}}{w_{(j)} + w_{(j+1)}}, \\
r_{(j),s} &\equiv \frac{w_{(j)} - w_{(j+1)}}{w_{(j)} + w_{(j+1)}},
\end{aligned}
\tag{23}
$$

then the solution is

$$
\begin{pmatrix} e^+_{(j)s} \\ e^-_{(j)s} \end{pmatrix} = \frac{1}{t_{(j),s}} \begin{pmatrix} 1 & r_{(j),s} \\ r_{(j),s} & 1 \end{pmatrix} \begin{pmatrix} e^+_{(j+1),s} \\ e^-_{(j+1),s} \end{pmatrix}.
\tag{24}
$$

This equation defines the interface matrix $m_{(j,j+1)s}$ between the two layers.

Within layer j the fields accumulate phase proportional to the normal component of the wave vector, $\pm w_{(j)}$. The fields at the front surface ($z + d$) and back surface (z) of layer j of thickness d are related by

$$
\begin{pmatrix} e^+_{(j)s}(z+d) \\ e^-_{(j)s}(z+d) \end{pmatrix} = \begin{pmatrix} e^{iw_{(j)}d} & 0 \\ 0 & e^{-iw_{(j)}d} \end{pmatrix} \begin{pmatrix} e^+_{(j+1),s}(z) \\ e^-_{(j+1),s}(z) \end{pmatrix},
\tag{25}
$$

which defines the layer matrix $m_{(j)}$.

Since there is no fundamental beam incident from the back of the multilayer structure, the reflectivity and transmission of the structure, as well as the amplitude of both forward and backward waves in each layer may be obtained by requiring that only a forward-going wave emerge from the final layer. Multiplying in order the matrices for each successive interface and layer produces a matrix characterizing the entire structure of L layers, which gives the incident and reflected fields in terms of the transmitted field. Dividing through by the incident field produces expressions for the reflected and transmitted amplitudes,

$$
\begin{pmatrix} r_s \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 0 \\ t_s \end{pmatrix},
\tag{26}
$$

where the matrix for the entire multilayer structure is given by

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = m_{0,1} \, m_1 \, m_{1,2} \cdots m_L \, m_{L,L+1} \, . \qquad (27)$$

Solving Eq. (26) for the s-polarized reflection and transmission gives

$$\begin{aligned} t_s &= 1/a_{22} \, , \\ r_s &= a_{12}/a_{22} \, . \end{aligned} \qquad (28)$$

The strength of forward and backward waves at any point in the multilayer can then be found from the transmitted field by multiplication by the appropriate combination of matrices.

The p-polarized (TM) case is closely analogous. The electric and magnetic fields of the fundamental waves are

$$\begin{aligned} \mathbf{e}_{(j)}^{\pm} &= e_{(j)p}^{\pm} \, e^{i\mathbf{k}_{(j)}^{\pm} \cdot \mathbf{r} - i\omega t} \, \hat{\mathbf{p}}_{(j)}^{\pm} + \text{c.c.} \, , \\ \mathbf{h}_{(j)}^{\pm} &= n_{(j)} e_{(j)p}^{\pm} \, e^{i\mathbf{k}_{(j)}^{\pm} \cdot \mathbf{r} - i\omega t} \, \hat{\mathbf{s}} + \text{c.c.} \, . \end{aligned} \qquad (29)$$

The interface matrix defined in Eq. (24) is unchanged, provided that the expressions for the reflectivity and transmission in Eq. (23) are replaced by

$$\begin{aligned} t_{(j)p} &\equiv \frac{2n_{(j)} n_{j+1} w_{(j)}}{\epsilon_{j+1} w_{(j)} + \epsilon_{(j)} w_{j+1}} \, , \\ r_{(j)p} &\equiv \frac{\epsilon_{j+1} w_{(j)} - \epsilon_{(j)} w_{j+1}}{\epsilon_{j+1} w_{(j)} + \epsilon_{(j)} w_{j+1}} \, . \end{aligned} \qquad (30)$$

A. Nonlinear Sources

Within layer j the forward and backward waves propagating with wave vectors given by Eq. (18) and Eq. (20) generate source polarizations with wave vectors $(n-m)\mathbf{k}_{(j)}^{-} + m\mathbf{k}_{(j)}^{+}$ for $m \in \{0, 1, \ldots, n\}$. Here m represents the number of factors of the (weaker) backward-going wave. All these have identical components in the plane of the surface, which we define to be

$$\mathbf{K} = n\kappa\hat{\boldsymbol{\kappa}} \, , \qquad (31)$$

but differ in the z-component of the wave vector, $q_{(j)}^{m}$, defined in Eq. (7).

In parallel with the discussion of the solution for the fundamental waves, we define

$$\begin{aligned} \tilde{\Omega} &= \frac{\Omega}{c} = \frac{n\omega}{c} \, , \\ W_{(j)} &= \sqrt{\tilde{\Omega}^2 \epsilon_{(j)}(\Omega) - K^2} \, , \\ \hat{\mathbf{P}}_{(j)}^{\pm} &= \frac{K\hat{\mathbf{z}} \mp W_{(j)}\hat{\boldsymbol{\kappa}}}{N_{(j)}\tilde{\Omega}} \, , \end{aligned} \qquad (32)$$

where $\tilde{\Omega}$ is the magnitude of the vacuum wave vector at the harmonic frequency and $N_{(j)} = \sqrt{\epsilon_{(j)}(\Omega)}$ is the

(complex) index of refraction at the harmonic frequency. Then the source polarization arising from any one of these $(n+1)$ combinations of fields in the jth layer takes the form

$$\mathcal{P}_{(j)}^{\Omega,m}(\mathbf{r}, t) = \mathbf{P}_{(j)}^{m} \, e^{iq_{(j)}^{m} z + iKx - i\Omega t} + \text{c.c.} \, . \qquad (33)$$

This source polarization generates the harmonic wave, which solves the (inhomogeneous) wave equation

$$\nabla(\nabla \cdot \mathbf{E}_{(j)}^{m}) - \nabla^2 \mathbf{E}_{(j)}^{m} - \tilde{\Omega}^2 N^2 \mathbf{E}_{(j)}^{m} = 4\pi \tilde{\Omega}^2 \mathbf{P}_{(j)}^{m} e^{iq_{(j)}^{m} z + iKx} \, . \qquad (34)$$

Its solution consists of two parts: the freely propagating wave at frequency Ω, which solves the homogeneous equation with the right hand side set to zero, and the particular solution to the inhomogeneous equation. Boundary conditions at the interfaces require that all waves have the same in-plane component of the wave vector, $K\hat{\boldsymbol{\kappa}}$. The freely propagating waves therefore take the form

$$\begin{aligned} \mathbf{E}_{(j)s}^{\pm} &= E_{(j)s}^{\pm} \, e^{i\mathbf{K}_{(j)}^{\pm} \cdot \mathbf{r} - i\Omega t} \, \hat{\mathbf{s}} + \text{c.c.} \, , \\ \mathbf{E}_{(j)p}^{\pm} &= E_{(j)p}^{\pm} \, e^{i\mathbf{K}_{(j)}^{\pm} \cdot \mathbf{r} - i\Omega t} \, \hat{\mathbf{P}}^{\pm} + \text{c.c.} \, , \\ \mathbf{K}_{(j)}^{\pm} &= K\hat{\boldsymbol{\kappa}} \pm W_{(j)}\hat{\mathbf{z}} \, . \end{aligned} \qquad (35)$$

We look for a source wave solution of the form

$$\mathbf{E}_{(j)}^{m} = \mathbf{A}_{(j)}^{m} \, e^{iKx + iq_{(j)}^{m} z - i\Omega t} + \text{c.c.} \, . \qquad (36)$$

Substituting this expression into Eq. (34) gives

$$\begin{aligned} A_{(j)s}^{m} &= C_{(j)} N_{(j)}^2 \tilde{\Omega}^2 P_{(j)s}^{m} \, , \\ A_{(j)\kappa}^{m} &= C_{(j)} \left[W_{(j)}^2 P_{(j)\kappa}^{m} - q_{(j)}^{m} K P_{(j)z}^{m} \right] \, , \\ A_{(j)z}^{m} &= C_{(j)} \left\{ \left[N_{(j)}^2 \tilde{\Omega}^2 - (q_{(j)}^{m})^2 \right] P_{(j)z}^{m} - q_{(j)}^{m} K P_{(j)\kappa}^{m} \right\} \, , \\ C_{(j)} &\equiv -\frac{4\pi}{N_{(j)}^2 \left[W_{(j)}^2 - (q_{(j)}^{m})^2 \right]} \, . \end{aligned} \qquad (37)$$

Equations (35)–(37) specify the source waves arising from the nonlinear polarization induced by the fundamental waves travelling both forward and backward inside nonlinear layer j. Boundary conditions at the interfaces between layers for the parallel components of the electric and magnetic fields at Ω determine the freely propagating waves at Ω that are seen in reflection and transmission.

B. Nonlinear Interface Matrices

The boundary conditions at the interface between layer j and layer $j + 1$, both of which may have surface non-

linear source terms, are then[38]

$$H_{(j)\kappa} - H_{(j+1)\kappa} = +4\pi i\,\tilde{\Omega}\left[\Pi_{(j)s} + \Pi_{(j+1)s}\right], \qquad (38)$$

$$H_{(j)s} - H_{(j+1)s} = -4\pi i\,\tilde{\Omega}\left[\Pi_{(j)\kappa} + \Pi_{(j+1)\kappa}\right], \qquad (39)$$

$$E_{(j)\kappa} - E_{(j+1)\kappa} = -4\pi\left[\frac{1}{\epsilon_{(j)}}(\hat{\boldsymbol{\kappa}}\cdot\nabla)\Pi_{(j)z} + \right.$$
$$\left. \frac{1}{\epsilon_{(j+1)}}(\hat{\boldsymbol{\kappa}}\cdot\nabla)\Pi_{(j+1)z}\right], \qquad (40)$$

$$E_{(j)s} - E_{(j+1)s} = -4\pi\left[\frac{1}{\epsilon_{(j)}}(\hat{\mathbf{s}}\cdot\nabla)\Pi_{(j)z} + \right.$$
$$\left. \frac{1}{\epsilon_{(j+1)}}(\hat{\mathbf{s}}\cdot\nabla)\Pi_{(j+1)z}\right], \qquad (41)$$

where the dielectric function is to be evaluated at the harmonic frequency Ω.

These expressions allow the harmonic fields in the jth layer to be found from the fields in the $(j+1)$st layer, the source waves $\mathbf{A}_{(j)}^m$ in both layers, and the surface polarizations $\mathbf{\Pi}$ on either side of the interface. For s-polarized nth harmonic Eq. (41) gives

$$E_{(j)s}^+ + E_{(j)s}^- = E_{(j+1),s}^+ + E_{(j+1),s}^- + \sum_{m=0}^{n}(A_{(j+1),s}^m - A_{(j)s}^m), \qquad (42)$$

since the surface polarization terms do not vary in the $\hat{\mathbf{s}}$ direction, while Eq. (38) yields

$$\frac{W_{(j)}}{\tilde{\Omega}}(E_{(j)s}^- - E_{(j)s}^+) = \frac{W_{(j+1)}}{\tilde{\Omega}}(E_{(j+1)s}^- - E_{(j+1)s}^+)$$
$$+ \frac{1}{\tilde{\Omega}}\sum_{m=0}^{n}(A_{(j)s}^m q_{(j)}^m - A_{(j+1)s}^m q_{(j+1)}^m) \qquad (43)$$
$$+ 4\pi i\tilde{\Omega}(\Pi_{(j)s} + \Pi_{(j+1)s}).$$

Solving for the free waves in layer j in terms of the waves in layer $j+1$ and the source terms gives

$$\begin{pmatrix} E_{(j)s}^+ \\ E_{(j)s}^- \end{pmatrix} = \frac{1}{t_{(j)}}\begin{pmatrix} 1 & r_{(j)} \\ r_{(j)} & 1 \end{pmatrix}\begin{pmatrix} E_{(j+1)s}^+ \\ E_{(j+1)s}^- \end{pmatrix} + \begin{pmatrix} S_{(j)s}^+ \\ S_{(j)s}^- \end{pmatrix}, \qquad (44)$$

where the reflection and transmission coefficients are defined by

$$t_{(j)s} = \frac{2W_{(j)}}{W_{(j)} + W_{(j+1)}},$$
$$r_{(j)s} = \frac{W_{(j)} - W_{(j+1)}}{W_{(j)} + W_{(j+1)}}, \qquad (45)$$

and the source terms are given by

$$S_{(j)s}^{\pm} = \frac{1}{2}\sum_{m=0}^{n}\left[A_{(j+1)s}^m\left(1 \pm \frac{q_{(j+1)}^m}{W_{(j)}}\right) - A_{(j)s}^m\left(1 \pm \frac{q_{(j)}^m}{W_{(j)}}\right)\right]$$
$$\mp \frac{2\pi i\tilde{\Omega}^2}{W_{(j)}}(\Pi_{(j)s} + \Pi_{(j+1)s}). \qquad (46)$$

Equation (44) completely specifies the fields in layer j, as was desired, but the inhomogeneous term makes it awkward to express compactly the progression of forward- and backward-going waves from deepest to frontmost interface. This problem may be overcome by defining 3×3 matrices for the interfaces, $M_{(j)j+1}$, by

$$\begin{pmatrix} E_{(j)s}^+ \\ E_{(j)s}^- \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{t_{(j)s}} & \frac{r_{(j)s}}{t_{(j)s}} & S_{(j)s}^+ \\ \frac{r_{(j)s}}{t_{(j)s}} & \frac{1}{t_{(j)s}} & S_{(j)s}^- \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} E_{(j+1)}^+ \\ E_{(j+1)}^- \\ 1 \end{pmatrix}, \qquad (47)$$

and for the layers, $M_{(j)}$, by

$$\begin{pmatrix} E_{(j)s}^+(z+d) \\ E_{(j)s}^-(z+d) \\ 1 \end{pmatrix} = \begin{pmatrix} e^{iW_{(j)}d} & 0 & 0 \\ 0 & e^{-iW_{(j)}d} & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} E_{(j+1)s}^+(z) \\ E_{(j+1)s}^-(z) \\ 1 \end{pmatrix}, \qquad (48)$$

so that the entire structure can be described by the product matrix

$$M = M_{0,1}\,M_1\,M_{1,2}\,\cdots\,M_L\,M_{L,L+1} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 \end{pmatrix}. \qquad (49)$$

The amplitude of the reflected and transmitted second-harmonic waves may now be determined from the matrix M. Assuming that there are no waves incident at the harmonic frequency Ω, we have

$$\begin{pmatrix} E^R \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ E^T \\ 1 \end{pmatrix}, \qquad (50)$$

from which we deduce the transmitted and reflected harmonic amplitudes,

$$E^T = -\frac{b_{23}}{b_{22}},$$
$$E^R = b_{13} + b_{12}E_s^T = b_{13} - \frac{b_{23}}{b_{22}}. \qquad (51)$$

The p-polarized case may be handled in a similar way. The boundary conditions of Eqs. (40) and (39) yield the equations

$$\frac{W_{(j)}}{N_{(j)}\tilde{\Omega}}(E_{(j)p}^- - E_{(j)p}^+) = \frac{W_{(j+1)}}{N_{(j+1)}\tilde{\Omega}}(E_{(j+1)p}^- - E_{(j+1)p}^+)$$
$$+ \sum_{m=0}^{n}(A_{(j+1),\kappa}^m - A_{(j)\kappa}^m)$$
$$- 4\pi i K\left(\frac{\Pi_{(j)z}}{\epsilon_{(j)}} + \frac{\Pi_{(j+1),z}}{\epsilon_{(j+1)}}\right) \qquad (52)$$

and

$$N_{(j)}(E_{(j)p}^- + E_{(j)p}^+) = N_{j+1}(E_{(j+1)p}^- + E_{(j+1)p}^+)$$
$$+ \frac{1}{\tilde{\Omega}} \sum_{m=0}^{n} \left\{ K(A_{(j+1)z}^m - A_{(j)z}^m) - [A_{(j+1)\kappa}^m q_{(j+1)}^m - \right.$$
$$\left. A_{(j)\kappa}^m q_{(j)}^m] \right\} - 4\pi i \tilde{\Omega}(\Pi_{(j)\kappa} + \Pi_{(j+1)\kappa}),$$

(53)

for the parallel components of the electric field and magnetic field, respectively. These equations may be summarized by the matrix equation

$$\begin{pmatrix} E_{(j)p}^+ \\ E_{(j)p}^- \\ 1 \end{pmatrix} = \begin{pmatrix} 1/t_{(j)p} & r_{(j)}/t_{(j)p} & S_{(j)p}^+ \\ r_{(j)p}/t_{(j)p} & 1/t_{(j)p} & S_{(j)p}^- \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E_{(j+1)p}^+ \\ E_{(j+1)p}^- \\ 1 \end{pmatrix},$$

(54)

with the reflection and transmission coefficients

$$t_{(j)p} \equiv \frac{2N_{(j)}N_{j+1}W_{(j)}}{\epsilon_{j+1}W_{(j)} + \epsilon_{(j)}W_{j+1}}$$
$$r_{(j)p} \equiv \frac{\epsilon_{j+1}W_{(j)} - \epsilon_{(j)}W_{j+1}}{\epsilon_{j+1}W_{(j)} + \epsilon_{(j)}W_{j+1}},$$

(55)

and the nonlinear sources given by

$$S_{(j)p}^{\pm} = \frac{1}{2N_{(j)}\tilde{\Omega}} \sum_{m=0}^{n} \left[K(A_{(j+1),z}^m - A_{(j)z}^m) \right.$$
$$- A_{(j+1),\kappa}^m \left(q_{j+1}^m \pm \frac{N_{(j)}^2 \tilde{\Omega}^2}{W_{(j)}} \right)$$
$$+ A_{(j)\kappa}^m \left(q_{(j)}^m \pm \frac{N_{(j)}^2 \tilde{\Omega}^2}{W_{(j)}} \right) \right]$$
$$+ 2\pi i \left[\pm \frac{N_{(j)}\tilde{\Omega}K}{W_{(j)}} \left(\frac{\Pi_{(j)z}}{\epsilon_{(j)}} + \frac{\Pi_{(j+1)z}}{\epsilon_{j+1}} \right) \right.$$
$$\left. - \frac{\tilde{\Omega}}{N_{(j)}} (\Pi_{(j)\kappa} + \Pi_{(j+1)\kappa}) \right].$$

(56)

As before, the reflected and transmitted fields are given by Eqs. (49) and (51).

* saeta@hmc.edu; http://saeta.physics.hmc.edu; This work was supported by Harvey Mudd College and the Research Corporation (Cottrell College Science Award CC4262). We gratefully acknowledge assistance from P. U. Berry, K. Hultman, A. C. Johnson, A. W. Rollins, P. D. Sparks, J. C. Eckert, and T. D. Donnelly.

[1] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich, Phys. Rev. Lett. **7**, 118 (1961).

[2] J. I. Dadap, X. F. Hu, N. M. Russell, J. G. Ekerdt, J. K. Lowell, and M. C. Downer, IEEE J. Sel. Top. Quantum Electron. **1**, 1145 (1995).

[3] H. W. K. Tom, T. F. Heinz, and Y. R. Shen, Phys. Rev. Lett. **51**, 1983 (1983).

[4] D. J. Bottomley, G. Lüpke, J. G. Mihaychuk, and H. M. van Driel, J. Appl. Phys. **74**, 6072 (1993).

[5] Y. R. Shen, Nature **337**, 519 (1989).

[6] T. F. Heinz, in *Nonlinear Surface Electromagnetic Phenomena*, edited by H.-E. Ponath and G. Stegeman (North-Holland, Amsterdam, 1991), vol. 29, pp. 353–416.

[7] G. Lüpke, Surf. Sci. Reports **35**, 75 (1999).

[8] T. F. Heinz, M. M. T. Loy, and W. A. Thompson, J. Vac. Sci. Technol. B **3**, 1467 (1985).

[9] J. I. Dadap, X. F. Hu, M. H. Anderson, M. C. Downer, J. K. Lowell, and O. A. Aktsipetrov, Phys. Rev. B **53**, R7607 (1996).

[10] 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).

[11] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Phys. Rev. **127**, 1918 (1962).

[12] N. Bloembergen and P. S. Pershan, Phys. Rev. **128**, 606 (1962).

[13] O. S. Heavens, *Optical Properties of Thin Solid Films* (Dover, New York, 1965).

[14] E. D. Palik, ed., *Handbook of Optical Constants of Solids* (Academic Press, Orlando, 1985).

[15] T. Y. F. Tsang, Phys. Rev. A **52**, 4116 (1995).

[16] P. N. Saeta and N. A. Miller, Appl. Phys. Lett. **79**, 2704 (2001).

[17] C. V. Shank, R. Yen, and C. Hirlimann, Phys. Rev. Lett. **50**, 454 (1983).

[18] H. W. K. Tom, G. D. Aumiller, and C. H. Brito-Cruz, Phys. Rev. Lett. **60**, 1438 (1988).

[19] L. I. Kulyuk, D. A. Shutov, E. E. Strumban, and O. A. Aktsipetrov, J. Opt. Soc. Am. B **8**, 1766 (1991).

[20] J. E. Sipe, V. Mizrahi, and G. I. Stegeman, Phys. Rev. B **35**, 9091 (1987).

[21] P. Guyot-Sionnest and Y. R. Shen, Phys. Rev. B **38**, 7985 (1988).

[22] G. Lüpke, D. J. Bottomley, and H. M. van Driel, J. Opt. Soc. Am. B **11**, 33 (1994).

[23] D. J. Bottomley, G. Lüpke, C. Meyer, and Y. Makita, Opt. Lett. **20**, 453 (1995).

[24] P. N. Saeta and A. C. Gallagher, J. Appl. Phys. **77**, 4639 (1995).

[25] P. N. Saeta and A. C. Gallagher, Phys. Rev. B **55**, 4563 (1997).

[26] J. E. Sipe, D. J. Moss, and H. M. van Driel, Phys. Rev. B **35**, 1129 (1987).

[27] N. Bloembergen, R. K. Chang, S. S. Jha, and C. H. Lee, Phys. Rev. **174**, 813 (1968).

[28] P. Guyot-Sionnest, W. Chen, and Y. R. Shen, Phys. Rev. B **33**, 8254 (1986).

[29] M. Bruel, MRS Bull. December, 35 (1998).

[30] Q.-Y. Tong and R. W. Bower, MRS Bull. December, 40 (1998).

[31] N. V. Nguyen, D. Chandler-Horowitz, P. M. Amirtharaj, and J. G. Pellegrino, Apl **64**, 2688 (1994).

[32] T. Yamaguchi, M. Nasu, Z.-T. Jiang, M. Tabe, and Y. Kanda, Thin Solid Films **313-314**, 264 (1998).

[33] D. J. Moss, H. M. van Driel, and J. E. Sipe, Opt. Lett. **14**, 57 (1989).

[34] S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981).

[35] J. G. Mihaychuk, J. Bloch, Y. Liu, and H. M. van Driel, Opt. Lett. **20**, 2063 (1995).

[36] J. Bloch, J. G. Mihaychuk, and H. M. van Driel, Phys. Rev. Lett. **77**, 920 (1996).

[37] N. Shamir, J. G. Mihaychuk, H. M. van Driel, and H. J. Kreuzer, Phys. Rev. Lett. **82**, 359 (1999).

[38] Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984), p. 496.