Dynamical model of harmonic generation in centrosymmetric semiconductors at visible and UV wavelengths

M. Scalora,[1,*] M. A. Vincenti,[2] D. de Ceglia,[2] N. Akozbek,[2] V. Roppo,[3] M. J. Bloemer,[1] and J. W. Haus[4]

[1]*Charles M. Bowden Research Center AMSRD-AMR-WSS, RDECOM, Redstone Arsenal, Alabama 35898-5000, USA*

[2]*AEgis Technologies Group, 410 Jan Davis Dr., Huntsville, Alabama 35806, USA*

[3]*Universitat Politècnica de Catalunya, Departament de Física i Eng. Nuclear, Rambla Sant Nebridi, E-08222 Terrassa, Spain*

[4]*Electro-Optics Program, University of Dayton, 300 College Park, Dayton, Ohio 45469, USA*

(Received 18 September 2011; published 8 May 2012)

We study second and third harmonic generation in centrosymmetric semiconductors at visible and UV wavelengths in bulk and cavity environments. Second harmonic generation is due to a combination of spatial symmetry breaking, the magnetic portion of the Lorentz force, and quadrupolar contributions from inner core electrons. The material is assumed to have a nonzero, third-order nonlinearity that gives rise to most of the third harmonic signal. Using the parameters of bulk silicon we predict that cavity environments modify the dependence of second harmonic generation on incident angle, while improving third harmonic conversion efficiency by several orders of magnitude relative to bulk silicon. This occurs despite the fact that the harmonic signals may be tuned to a wavelength range where the dielectric function of the material is negative: A phase-locking mechanism binds the generated signals to the pump and inhibits their absorption. These results point the way to alternative uses and flexibility of materials such as silicon as nonlinear media in the visible and UV ranges.

DOI: 10.1103/PhysRevA.85.053809 PACS number(s): 42.65.Ky

I. INTRODUCTION

Results on second harmonic generation (SHG) in centrosymmetric semiconductors such as silicon were first reported in Ref. [1]. A comparison of SHG between bulk silicon and silver showed that the main features of the reflected signals with respect to incident angle were similar. These observations also suggested that SHG in silicon was independent of angle cut and crystal orientation, i.e., azimuthal axis. The subject remained latent until the early 1980s, when new experimental results [2,3] implied that under intense illumination silicon displayed properties consistent with a second-order nonlinear coefficient having bulk dipolar origins. The authors argued that neither bulk quadrupolar sources nor spatial symmetry breaking, superficial oxide layers, or impurities sufficed to explain their observations, and that their results were consistent with a bulklike $\chi^{(2)}$ similar to that of GaAs or GaP in symmetry and strength. Further experimental results [4] suggested that the dependence of SHG on the azimuthal angle arose even at low intensities, and that surface and bulk sources gave way to comparable contributions. In light of these conflicting results and interpretations it was postulated that the experimental results could be explained by a combination of dipolar surface and quadrupolar bulk sources [5].

In the intervening years since this debate first took place a consensus has developed among researchers that materials like Si are characterized by intrinsic nonlinear surface anisotropy with surface dipolar and bulk quadrupolar sources [6–23] divided in almost equal parts. In Ref. [23] it was noted that even though bulk contributions are slightly smaller, the interference between these two types of sources cannot generally be neglected, making it difficult to identify and separate surface and bulk sources unambiguously [24,25]. With recent advances in the understanding of linear and nonlinear optical phenomena at the nanoscale many researchers are taking a new look at well-known materials like silicon and exploring the possibility of using them in different regimes and in unique ways. An example is provided by the investigation of harmonic generation in silicon nanostructures at visible wavelengths [26].

As extensively reported in the literature the theoretical study of nonlinear interactions, and harmonic generation in particular, in materials like silicon is complicated by aspects related to nonlinear surface anisotropies [1–23]. Here we wish to build a self-consistent, classical model to describe frequency conversion and other nonlinear processes in a centrosymmetric, silicon-like material that lacks dipolar, $\chi^{(2)}$ bulk contributions, may have a nonzero $\chi^{(3)}$, and where a combination of yet-to-be-determined nonlinear surface and bulk sources gives rise to harmonic generation. As we will see below, the level of complexity of the system of equations that governs the dynamics renders impractical a purely analytical treatment of the problem. Therefore, using classical arguments we will derive equations of motion that will be integrated using spectral methods. Thanks to this approach we will highlight an interesting feature: At visible wavelengths silicon is highly absorptive, becoming metallic [$\mathrm{Re}(\varepsilon) < 0$] below 300 nm [27]. In Ref. [28] it was experimentally demonstrated that a GaP substrate supports the propagation of phase-locked third harmonic (TH) pulses at 223 nm, in the metallic region. The phase-locked component corresponds to the inhomogeneous solution of the wave equation. In this work we will show that a phase-locking mechanism [28,29] occurs also in centrosymmetric media, rendering the material transparent to the harmonic wavelengths despite material dispersion. In addition, an exponential increase of third harmonic generation (THG) occurs in the 80–300 nm range in a Fabry-Pérot étalon sandwiched between distributed Bragg reflectors (DBRs).

*michael.scalora@us.army.mil

A study of THG at infrared and visible wavelengths in absorbing materials including Si, Ge, and the noble metals [30] revealed the form of the $\chi^{(3)}$ tensors, along with dispersive properties that arise due to differences between conduction and valence electrons. A full band structure calculation of the dispersion of the $\chi^{(3)}$ tensor in Si, Ge, and GaAs was performed in Ref. [31]. Crystal symmetry affects the dependence of harmonic generation on the angle of incidence. For example, bulk cubic materials always have some degree of anisotropy [32], and for the purposes of THG it is important that this symmetry be reflected in the choice of the $\chi^{(3)}$ tensor. As another example, in isotropic, centrosymmetric media SH efficiency strongly depends on the polarization of the incident beam [33] and on the chosen crystallographic direction [34], and approaches zero when the pump field is incident normally on the surface [1].

II. THE MODEL

It is well known that in noble metals bound charges contribute to SHG [35]. The role that bound charges play was recently outlined in detail using a harmonic oscillator model [36,37] and later used to describe harmonic generation in GaAs-filled, metallic nanocavities [38] and a GaP substrate [28]. In addition to being a powerful pedagogical tool, the nonlinear oscillator model provides a detailed description of dynamical factors that contribute to harmonic generation. Effects due to bound charges may be seen explicitly in the case of SHG and THG from a GaP substrate at UV wavelengths [28]. In what follows we generalize the model to include contributions of bound quadrupoles because in the absence of a dipolar second-order nonlinearity they become primary nonlinear sources. We do not treat anisotropic surface sources and assume that harmonic generation is independent of the azimuthal angle, although the model may be extended to include arbitrary nonlinear surface distributions.

The nonlinear polarization of a medium may be written according to the well-established hierarchy that the electric dipole $\gg$ (electric quadrupole + magnetic dipole) $\gg$ (electric octupole + magnetic quadrupole) $\cdots$ [39]. Using this classification, the lowest-order contributions to the nonlinear polarization of a generic medium may be written as [1,3,4,40,41]

$$P_i^{(NL)} = \chi_{ijk}^{(2),ed} E_i E_j + \chi_{ijkl}^{(3),ed} E_j E_k E_l + \chi_{ijk}^{(2),md} E_j B_j + \chi_{ijkl}^{(2),eq} E_j \nabla_k E_l + \cdots , \quad (1)$$

where the subscripts i,j,k,l are Cartesian coordinates; $\chi_{ijk}^{(2),ed}$ and $\chi_{ijkl}^{(3),ed}$ are the tensor components of second- and third-order nonlinear coefficients; the superscripts ed, md, eq stand for electric dipole, magnetic dipole, and electric quadrupole, respectively; $E_{i,j,k}$ and $B_{i,j,k}$ are the Cartesian components of the electric and magnetic fields; and the usual summation convention has been assumed on the right-hand side. The description of a classical, local oscillator model is a natural starting point, and so one may simply begin with an equation of motion for a multipolar charge distribution assumed to be under the action of internal forces (damping, harmonic, and anharmonic restoring forces) and external forces due to the applied electromagnetic fields. A possible, quite simplistic way to describe nonlinear optical processes is then to modify the Lorentz model of the atom by introducing appropriate nonlinear terms. For example, in Ref. [36] the authors used a microscopic oscillator approach to describe analytically harmonic generation from amorphous silicon by introducing a linear restoring force and a bulk, dipolar $\chi_{ijk}^{(2),ed}$. With reference to Fig. 1, external electric and magnetic forces on the generic charge distribution e may be written as [1,42]

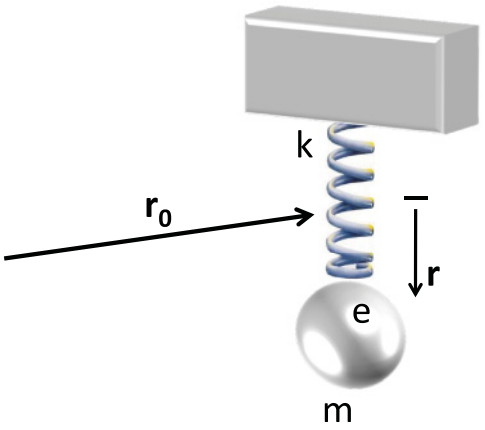

FIG. 1. (Color online) Simple harmonic oscillator with generic charge e on a spring of constant k, subject to external electric and magnetic forces, and to internal linear and nonlinear restoring forces. The origin is at $\mathbf{r}_0$; $\mathbf{r}$ is the displacement from equilibrium. Higher poles are excited by distortions of the charge distribution due to nonuniform fields.

$$\mathbf{F}(\mathbf{r}_0+\mathbf{r},t) = e\mathbf{E}(\mathbf{r}_0+\mathbf{r},t) + \frac{e}{c}\dot{\mathbf{r}} \times \mathbf{B}(\mathbf{r}_0+\mathbf{r},t). \quad (2)$$

The fields at the electron position $\mathbf{r}_0+\mathbf{r}$ may be expanded about the origin $\mathbf{r}_0$, so that

$$\mathbf{E}(\mathbf{r}_0+\mathbf{r},t) = \mathbf{E}(\mathbf{r}_0,t) + (\mathbf{r}\cdot\nabla_{\mathbf{r}})\mathbf{E}(\mathbf{r}_0,t) + \tfrac{1}{2}\nabla_{\mathbf{r}}(\mathbf{rr} : \nabla_{\mathbf{r}}\mathbf{E}(\mathbf{r}_0,t)) + \cdots \quad (3)$$

and

$$\mathbf{B}(\mathbf{r}_0+\mathbf{r},t) = \mathbf{B}(\mathbf{r}_0,t) + (\mathbf{r}\cdot\nabla_{\mathbf{r}})\mathbf{B}(\mathbf{r}_0,t) + \tfrac{1}{2}\nabla_{\mathbf{r}}(\mathbf{rr} : \nabla_{\mathbf{r}}\mathbf{B}(\mathbf{r}_0,t)) + \cdots . \quad (4)$$

The colon symbol signifies double dot product. Collecting lowest-order terms, the force Eq. (2) becomes [36,37,42–46]

$$m^*\ddot{\mathbf{r}}(t) + \gamma m^*\dot{\mathbf{r}}(t) + k\mathbf{r}(t) = e\mathbf{E}(\mathbf{r}_0,t) + e\,(\mathbf{r}\cdot\nabla_{\mathbf{r}})\,\mathbf{E}(\mathbf{r}_0,t) + \frac{e}{4}\nabla_{\mathbf{r}}\,[\mathbf{rr} : \nabla_{\mathbf{r}}\mathbf{E}(\mathbf{r}_0,t)] + \frac{e}{c}\dot{\mathbf{r}} \times [\mathbf{B}(\mathbf{r}_0,t) + (\mathbf{r}\cdot\nabla_{\mathbf{r}})\,\mathbf{B}(\mathbf{r}_0,t) + \cdots], \quad (5)$$

where one may recognize $\mathbf{p} = e\mathbf{r}$ and $\mathbf{Q} = (e/2)\mathbf{rr}$ as the electric dipole and quadrupole contributions, respectively; m^* is the effective mass of the oscillator; k is the spring constant; e is the charge; c is the speed of light in vacuum; γ is the damping coefficient. Equation (5) is a modified version of Eq. (2) found in Ref. [36].

There is a more canonical way to proceed that allows one to also arrive at Eq. (5). Any realistic charge distribution may be thought of as a superposition of multipoles. For

example, the more loosely bound valence electrons of neutral atoms respond to the applied field and contribute most to the polarization of matter, a picture that can explain most linear and nonlinear optical phenomena. At the same time, inner core electrons that occupy lower orbitals and that interact more strongly with the nucleus form an electron cloud that is generally slightly distorted, giving way to small imbalances in the charge distribution which then interacts weakly with the externally applied fields in the form of quadrupolar transitions. By the same token, the total force on the charge distribution described above can be thought of as a vector sum of separate forces acting on dipolar and quadrupolar species, and higher poles if present. Therefore, one may distill simplified equations of motion by writing a Hamiltonian for a multipolar charge distribution [1], and by taking its negative gradient to find the total force acting on the system. Because the Hamiltonian approach is formal and perhaps more intuitive than the expansion we used above, we think it is instructive to also present the energetic route separately in the Appendix. Next we focus on the development of Eq. (5).

For simplicity, we begin examining Eq. (5) by dropping the subscript $\mathbf{r}$ from the ∇ operator, and by expanding the quadrupolar terms in Cartesian coordinates:

$$\nabla\mathbf{E} = \begin{pmatrix} \frac{\partial}{\partial x}\mathbf{i} \\ \frac{\partial}{\partial y}\mathbf{j} \\ \frac{\partial}{\partial z}\mathbf{k} \end{pmatrix} \left(E_x\mathbf{i} + E_y\mathbf{j} + E_z\mathbf{k}\right) = \begin{pmatrix} \frac{\partial E_x}{\partial x} & \frac{\partial E_y}{\partial x} & \frac{\partial E_z}{\partial x} \\ \frac{\partial E_x}{\partial y} & \frac{\partial E_y}{\partial y} & \frac{\partial E_z}{\partial y} \\ \frac{\partial E_x}{\partial z} & \frac{\partial E_y}{\partial z} & \frac{\partial E_z}{\partial z} \end{pmatrix}, \tag{6}$$

and

$$\mathbf{rr} = \begin{pmatrix} x\mathbf{i} \\ y\mathbf{j} \\ z\mathbf{k} \end{pmatrix} (x\mathbf{i} \;\; y\mathbf{j} \;\; z\mathbf{k}) = \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix}. \tag{7}$$

Combining Eqs. (6) and (7) the quadrupolar portion of the force is

$$\begin{aligned} \nabla(\mathbf{rr} : \nabla\mathbf{E}) = \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right)\Bigg(x^2\frac{\partial E_x}{\partial x} + xy\frac{\partial E_x}{\partial y} \\ + xz\frac{\partial E_x}{\partial z} + yx\frac{\partial E_y}{\partial x} + y^2\frac{\partial E_y}{\partial y} + yz\frac{\partial E_y}{\partial z} \\ + zx\frac{\partial E_z}{\partial x} + zy\frac{\partial E_z}{\partial y} + z^2\frac{\partial E_z}{\partial z}\Bigg). \end{aligned} \tag{8}$$

After a straightforward application of the spatial derivatives, Eq. (8) becomes

$$\begin{aligned} &\nabla(\mathbf{rr} : \nabla\mathbf{E}) \\ &\approx \mathbf{i}\left(2x\frac{\partial E_x}{\partial x} + y\frac{\partial E_x}{\partial y} + z\frac{\partial E_x}{\partial z} + y\frac{\partial E_y}{\partial x} + z\frac{\partial E_z}{\partial x}\right) \\ &\quad + \mathbf{j}\left(x\frac{\partial E_x}{\partial y} + x\frac{\partial E_y}{\partial x} + 2y\frac{\partial E_y}{\partial y} + z\frac{\partial E_y}{\partial z} + z\frac{\partial E_z}{\partial y}\right) \\ &\quad + \mathbf{k}\left(x\frac{\partial E_x}{\partial z} + y\frac{\partial E_y}{\partial z} + x\frac{\partial E_z}{\partial x} + y\frac{\partial E_z}{\partial y} + 2z\frac{\partial E_z}{\partial z}\right). \end{aligned} \tag{9}$$

We have neglected terms that contain products of two coordinates and higher-order spatial derivatives of the fields. Looking ahead, we may get another point of view if we write Eq. (9) in terms of the macroscopic polarizations:

$$\begin{aligned} \nabla(\mathbf{rr} : \nabla\mathbf{E}) \approx \frac{1}{n_0 e}\Bigg\{&\mathbf{i}\left(2P_x\frac{\partial E_x}{\partial x} + P_y\frac{\partial E_x}{\partial y} + P_z\frac{\partial E_x}{\partial z} + P_y\frac{\partial E_y}{\partial x} + P_z\frac{\partial E_z}{\partial x}\right) \\ &+ \mathbf{j}\left(P_x\frac{\partial E_x}{\partial y} + P_x\frac{\partial E_y}{\partial x} + 2P_y\frac{\partial E_y}{\partial y} + P_z\frac{\partial E_y}{\partial z} + P_z\frac{\partial E_z}{\partial y}\right) \\ &+ \mathbf{k}\left(P_x\frac{\partial E_x}{\partial z} + P_y\frac{\partial E_y}{\partial z} + P_x\frac{\partial E_z}{\partial x} + P_y\frac{\partial E_z}{\partial y} + 2P_z\frac{\partial E_z}{\partial z}\right)\Bigg\}, \end{aligned} \tag{10}$$

where n_0 is particle density, and $P_j \equiv n_0 e r_j$ are the Cartesian components of the macroscopic polarization. We now simplify the geometry by assuming that structures have two-dimensional symmetry, with no dependence of the fields on the azimuthal angle. With reference to Fig. 2, we assume that the structure is infinite along the x direction (into the page), that the wave propagates on the y-z plane, and refer to z as the longitudinal coordinate. Then, a transverse magnetic (TM)-polarized field has E-field components that point along y and z; a transverse electric (TE)-polarized field has corresponding H-field components that point in the directions y and z, and the fields are independent of x. Suppressing derivatives with respect to x, Eq. (10) simplifies to

$$\begin{aligned} &\nabla(\mathbf{rr} : \nabla\mathbf{E}) \\ &\approx \frac{1}{n_0 e}\Bigg\{\mathbf{i}\left(P_y\frac{\partial E_x}{\partial y} + P_z\frac{\partial E_x}{\partial z}\right) \\ &\quad + \mathbf{j}\left(P_x\frac{\partial E_x}{\partial y} + 2P_y\frac{\partial E_y}{\partial y} + P_z\frac{\partial E_y}{\partial z} + P_z\frac{\partial E_z}{\partial y}\right) \\ &\quad + \mathbf{k}\left(P_x\frac{\partial E_x}{\partial z} + P_y\frac{\partial E_y}{\partial z} + P_y\frac{\partial E_z}{\partial y} + 2P_z\frac{\partial E_z}{\partial z}\right)\Bigg\}. \end{aligned} \tag{11}$$

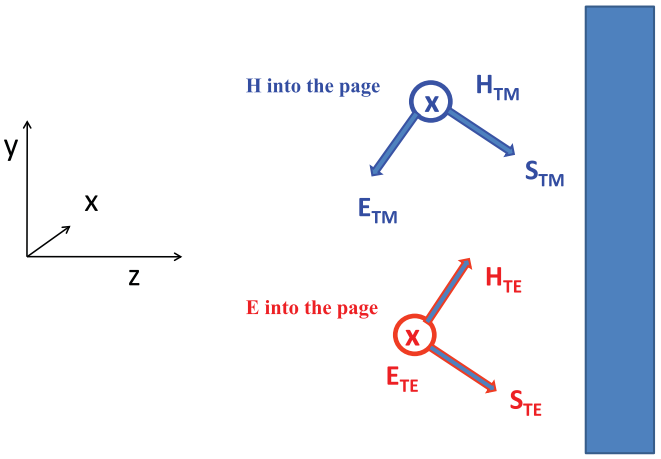

FIG. 2. (Color online) Typical geometry of a structure with two-dimensional symmetry, i.e., propagation occurs on the y-z plane, and derivatives with respect to x are suppressed. $\mathbf{S}_{\rm TM}$ and $\mathbf{S}_{\rm TE}$ are the Poynting vectors relative to TM and TE polarizations, and indicate direction of propagation. The rectangular box represents the sample.

For a coherent pump field of frequency ω, the general solution for the position vector may be written as a superposition of harmonics as usual:

$$\mathbf{r}(t)=\mathbf{r}_{\omega}(t)e^{-i\omega t}+\mathbf{r}_{2\omega}(t)e^{-2i\omega t}+\mathbf{r}_{3\omega}(t)e^{-3i\omega t}+\text{c.c.} \quad (12)$$

Switching over to a macroscopic representation, we define the polarizations:

$$\mathbf{P}_{\omega}\equiv n_0 e\mathbf{r}_{\omega},\quad \mathbf{P}_{2\omega}\equiv n_0 e\mathbf{r}_{2\omega},\quad \mathbf{P}_{3\omega}\equiv n_0 e\mathbf{r}_{3\omega}. \quad (13)$$

We now write the fields in terms of generic space- and time-dependent envelope functions and arbitrarily extract the carrier frequency ω [36,37]:

$$\begin{aligned}\mathbf{E}(\mathbf{r}_0,t)&=\mathbf{E}_{\omega}(\mathbf{r}_0,t)e^{-i\omega t}+\mathbf{E}_{2\omega}(\mathbf{r}_0,t)e^{-2i\omega t}\\&\quad+\mathbf{E}_{3\omega}(\mathbf{r}_0,t)e^{-3i\omega t}+\text{c.c.},\\\mathbf{H}(\mathbf{r}_0,t)&=\mathbf{H}_{\omega}(\mathbf{r}_0,t)e^{-i\omega t}+\mathbf{H}_{2\omega}(\mathbf{r}_0,t)e^{-2i\omega t}\\&\quad+\mathbf{H}_{3\omega}(\mathbf{r}_0,t)e^{-3i\omega t}+\text{c.c.}\end{aligned} \quad (14)$$

The extraction of a carrier frequency does not preempt the possibility of treating a dispersive medium, because the envelope functions are allowed to vary rapidly in both space and time, and are substituted back into Maxwell's equation without approximations. Using Eqs. (12) and (14), the term $(\mathbf{r}\cdot\nabla_{\mathbf{r}})\mathbf{E}(\mathbf{r}_0,t)$ in Eq. (5) may be expanded into its harmonic contributions if we replace $\nabla_{\mathbf{r}}\to i\mathbf{K}$. For example, $(\mathbf{r}_{\omega}\cdot\nabla_{\mathbf{r}})\mathbf{E}_{\omega}(\mathbf{r}_0,t)e^{-2i\omega t}\sim i(\mathbf{K}\cdot\mathbf{r}_{\omega})\mathbf{E}_{\omega}(\mathbf{r}_0,t)e^{-2i\omega t}$. Then, without any claim of rigor and with some hand waving, the Coulomb terms may be *approximated* as follows:

$$\begin{aligned}&i\mathbf{K}\cdot n_0 e\mathbf{r}_{\omega}\approx\nabla\cdot\mathbf{P}_{\omega},\quad 2i\mathbf{K}\cdot n_0 e\mathbf{r}_{2\omega}\approx\nabla\cdot\mathbf{P}_{2\omega},\\&3i\mathbf{K}\cdot n_0 e\mathbf{r}_{3\omega}\approx\nabla\cdot\mathbf{P}_{3\omega}.\end{aligned} \quad (15)$$

The issue of how to best represent the Coulomb terms is still the subject of debate [44]. For example, writing the Coulomb term proportional to $(\nabla\cdot\mathbf{P})\mathbf{E}$ best describes a dielectric medium composed of individual point charges. On the other hand, if the Coulomb term is proportional to $(\mathbf{P}\cdot\nabla)\mathbf{E}$ [43] then the description is that of a dielectric medium composed of individual dipoles. Apparent differences notwithstanding, these diverse representations yield the same net Coulomb force in both microscopic and macroscopic regimes [44]. Substitution of Eqs. (11)–(15) into Eq. (5) leads to three coupled equations for the field and polarization envelope functions [37]:

$$\begin{aligned}\ddot{\mathbf{P}}_{\omega}+\tilde{\gamma}_{\omega}\dot{\mathbf{P}}_{\omega}+\tilde{\omega}^2_{0,\omega}\mathbf{P}_{\omega}&=\frac{e\lambda_0}{m^*c^2}\left\{n_0 e\lambda_0\mathbf{E}_{\omega}-\frac{1}{2}(\nabla\cdot\mathbf{P}_{2\omega})\mathbf{E}^*_{\omega}+2(\nabla\cdot\mathbf{P}^*_{\omega})\mathbf{E}_{2\omega}+(\dot{\mathbf{P}}^*_{\omega}+i\omega\mathbf{P}^*_{\omega})\times\mathbf{H}_{2\omega}\right.\\&\quad\left.+(\dot{\mathbf{P}}_{2\omega}-2i\omega\mathbf{P}_{2\omega})\times\mathbf{H}^*_{\omega}+\frac{1}{4}\mathbf{F}_{\omega}\right\},\\\ddot{\mathbf{P}}_{2\omega}+\tilde{\gamma}_{2\omega}\dot{\mathbf{P}}_{2\omega}+\tilde{\omega}^2_{0,2\omega}\mathbf{P}_{2\omega}&=\frac{e\lambda_0}{m^*c^2}\left\{n_0 e\lambda_0\mathbf{E}_{2\omega}+(\nabla\cdot\mathbf{P}_{\omega})\,\mathbf{E}_{\omega}-\frac{1}{3}(\nabla\cdot\mathbf{P}_{3\omega})\,\mathbf{E}^*_{\omega}-3(\nabla\cdot\mathbf{P}^*_{\omega})\mathbf{E}_{3\omega}\right.\\&\quad\left.+(\dot{\mathbf{P}}_{\omega}-i\omega\mathbf{P}_{\omega})\times\mathbf{H}_{\omega}+\frac{1}{4}\mathbf{F}_{2\omega}\right\},\\\ddot{\mathbf{P}}_{3\omega}+\tilde{\gamma}_{3\omega}\dot{\mathbf{P}}_{3\omega}+\tilde{\omega}^2_{0,3\omega}\mathbf{P}_{3\omega}&=\frac{e\lambda_0}{m^*c^2}\left\{n_0 e\lambda_0\mathbf{E}_{3\omega}+\frac{1}{2}(\nabla\cdot\mathbf{P}_{2\omega})\,\mathbf{E}_{\omega}+2(\nabla\cdot\mathbf{P}_{\omega})\,\mathbf{E}_{2\omega}+(\dot{\mathbf{P}}_{2\omega}-i\omega\mathbf{P}_{2\omega})\times\mathbf{H}_{\omega}\right.\\&\quad\left.+(\dot{\mathbf{P}}_{\omega}-i\omega\mathbf{P}_{\omega})\times\mathbf{H}_{2\omega}+\frac{1}{4}\mathbf{F}_{3\omega}\right\},\end{aligned} \quad (16)$$

where we have retained lowest-order magnetic contributions. The coefficients $\tilde{\gamma}_{l\omega}=(\gamma-li\omega)$ and $\tilde{\omega}^2_{0,l\omega}=[\omega_0^2-(l\omega)^2+i\gamma l\omega]$ are scaled damping and resonance frequency, respectively, ω_0 is the natural frequency of the oscillator, and $l=1,2,3$ denotes the given harmonic. Differentiation is now with respect to the dimensionless coordinates, $\xi=z/\lambda_0$, $\tilde{y}=y/\lambda_0$, $\tilde{x}=x/\lambda_0$, and $\tau=ct/\lambda_0$, and we have chosen a convenient reference wavelength $\lambda_0=1\ \mu$m. The lowest-order quadrupolar contributions to each harmonic are

$$\begin{aligned}\mathbf{F}_{\omega}=\mathbf{i}&\left[\frac{\partial E_{\tilde{x},2\omega}}{\partial\tilde{y}}P^*_{\tilde{y},\omega}+\left(\frac{\partial E_{\tilde{x},\omega}}{\partial\tilde{y}}\right)^*P_{y,2\omega}+\frac{\partial E_{\tilde{x},2\omega}}{\partial\xi}P^*_{\xi,\omega}+\left(\frac{\partial E_{\tilde{x},\omega}}{\partial\xi}\right)^*P_{\xi,2\omega}\right]\\&+\mathbf{j}\left\{\frac{\partial E_{\tilde{x},2\omega}}{\partial\tilde{y}}P^*_{\tilde{x},\omega}+\left(\frac{\partial E_{\tilde{x},\omega}}{\partial\tilde{y}}\right)^*P_{\tilde{x},2\omega}+2\frac{\partial E_{\tilde{y},2\omega}}{\partial\tilde{y}}P^*_{\tilde{y},\omega}+2\left(\frac{\partial E_{\tilde{y},\omega}}{\partial\tilde{y}}\right)^*P_{\tilde{y},2\omega}+\frac{\partial E_{\tilde{y},2\omega}}{\partial\xi}P^*_{\xi,\omega}\right.\end{aligned}$$

$$+\frac{\partial E_{\xi,2\omega}}{\partial\tilde{y}}P^{*}_{\xi,\omega}+\left(\frac{\partial E_{\tilde{y},\omega}}{\partial\xi}\right)^{*}P_{\xi,2\omega}+\left(\frac{\partial E_{\xi,\omega}}{\partial\tilde{y}}\right)^{*}P_{\xi,2\omega}\Bigg\}$$

$$+\mathbf{k}\left\{2\frac{\partial E_{\xi,2\omega}}{\partial\xi}P^{*}_{\xi,\omega}+2\left(\frac{\partial E_{\xi,\omega}}{\partial\xi}\right)^{*}P_{\xi,2\omega}+\frac{\partial E_{\tilde{y},2\omega}}{\partial\xi}P^{*}_{\tilde{y},\omega}+\frac{\partial E_{\xi,2\omega}}{\partial\tilde{y}}P^{*}_{\tilde{y},\omega}+\left(\frac{\partial E_{\tilde{y},\omega}}{\partial\xi}\right)^{*}P_{\tilde{y},2\omega}\right.$$

$$\left.+\left(\frac{\partial E_{\xi,\omega}}{\partial\tilde{y}}\right)^{*}P_{\tilde{y},2\omega}+\frac{\partial E_{\tilde{x},2\omega}}{\partial\xi}P^{*}_{\tilde{x},\omega}+\left(\frac{\partial E_{\tilde{x},\omega}}{\partial\xi}\right)^{*}P_{\tilde{x},2\omega}\right\}, \tag{17}$$

$$\mathbf{F}_{2\omega}=\mathbf{i}\left[\frac{\partial E_{x,\omega}}{\partial\tilde{y}}P_{\tilde{y},\omega}+\frac{\partial E_{\tilde{x},\omega}}{\partial\xi}P_{\xi,\omega}\right]+\mathbf{j}\left[\frac{\partial E_{\tilde{x},\omega}}{\partial\tilde{y}}P_{\tilde{x},\omega}+2\frac{\partial E_{\tilde{y},\omega}}{\partial\tilde{y}}P_{\tilde{y},\omega}+\left(\frac{\partial E_{\tilde{y},\omega}}{\partial\xi}+\frac{\partial E_{\xi,\omega}}{\partial\tilde{y}}\right)P_{\xi,\omega}\right]$$

$$+\mathbf{k}\left[\frac{\partial E_{\tilde{x},\omega}}{\partial\xi}P_{\tilde{x},\omega}+2\frac{\partial E_{\xi,\omega}}{\partial\xi}P_{\xi,\omega}+\left(\frac{\partial E_{\tilde{y},\omega}}{\partial\xi}+\frac{\partial E_{\xi,\omega}}{\partial\tilde{y}}\right)P_{\tilde{y},\omega}\right], \tag{18}$$

$$\mathbf{F}_{3\omega}=\mathbf{i}\left[\frac{\partial E_{\tilde{x},2\omega}}{\partial\tilde{y}}P_{\tilde{y},\omega}+\frac{\partial E_{\tilde{x},\omega}}{\partial\tilde{y}}P_{\tilde{y},2\omega}+\frac{\partial E_{\tilde{x},2\omega}}{\partial\xi}P_{\xi,\omega}+\frac{\partial E_{\tilde{x},\omega}}{\partial\xi}P_{\xi,2\omega}\right]$$

$$+\mathbf{j}\left[\frac{\partial E_{\tilde{x},2\omega}}{\partial\tilde{y}}P_{\tilde{x},\omega}+\frac{\partial E_{\tilde{x},\omega}}{\partial\tilde{y}}P_{\tilde{x},2\omega}+2\frac{\partial E_{\tilde{y},2\omega}}{\partial\tilde{y}}P_{\tilde{y},\omega}+2\frac{\partial E_{\tilde{y},\omega}}{\partial\tilde{y}}P_{\tilde{y},2\omega}+\left(\frac{\partial E_{\xi,2\omega}}{\partial\tilde{y}}+\frac{\partial E_{\tilde{y},2\omega}}{\partial\xi}\right)P_{\xi,\omega}+\left(\frac{\partial E_{\xi,\omega}}{\partial\tilde{y}}+\frac{\partial E_{\tilde{y},\omega}}{\partial\xi}\right)P_{\xi,2\omega}\right]$$

$$+\mathbf{k}\left[\frac{\partial E_{\tilde{x},2\omega}}{\partial\xi}P_{\tilde{x},\omega}+\frac{\partial E_{\tilde{x},\omega}}{\partial\xi}P_{\tilde{x},2\omega}+\left(\frac{\partial E_{\tilde{y},2\omega}}{\partial\xi}+\frac{\partial E_{\xi,2\omega}}{\partial\tilde{y}}\right)P_{\tilde{y},\omega}+\left(\frac{\partial E_{\tilde{y},\omega}}{\partial\xi}+\frac{\partial E_{\xi,\omega}}{\partial\tilde{y}}\right)P_{\tilde{y},2\omega}+2\frac{\partial E_{\xi,2\omega}}{\partial\xi}P_{\xi,\omega}+2\frac{\partial E_{\xi,\omega}}{\partial\xi}P_{\xi,2\omega}\right]. \tag{19}$$

Equations (16) are solved in the time domain together with Maxwell's equations using a fast Fourier transform pulse propagation method [37].

The validity of the Taylor expansion [42] in Eq. (5) may be called into question if the boundary between two media were infinitely sharp. In tackling numerical problems there are at least two important questions one must always ask: Does the calculation converge and how fast? Is the result accurate? One obvious and ordinary aspect of working in the time domain is the requirement of finite spatial and temporal time steps, and signals that have a finite turn-on time. The polarization thus grows from zero as a function of time as the wave front approaches, and for different spatial step sizes at any instant the polarization boundary may look like that depicted in Fig. 3. The junction points at the top and bottom of the oblique sections of each curve may be rounded out to insure that the functions are well behaved and differentiable everywhere, but this operation is usually unnecessary. In the absence of intrinsic instabilities the calculations converge rapidly to a given trajectory. Convergence is checked by decreasing spatial and temporal step sizes simultaneously. While decreasing spatial step size leads to the functions shown in Fig. 3, a smaller time step causes pulses to advance and to sample the boundary at slower rates. The derivatives of typical variables may be large but they are always finite, with the result that "hard" interfaces are in fact quite manageable provided enough terms are retained in the Taylor expansion for a given harmonic, in neighborhoods with steep slopes. Finally, in Eqs. (16) one finds three types of sources: (i) Terms proportional to $\nabla\cdot\mathbf{P}_{N\omega}$ that are zero everywhere except at the surface have quadrupolar quality but dipolar origin; (ii) magnetic terms $(\dot{\mathbf{P}}_{\omega}-i\omega\mathbf{P}_{\omega})\times\mathbf{H}_{\omega}$ and $(\dot{\mathbf{P}}_{2\omega}-i\omega\mathbf{P}_{2\omega})\times\mathbf{H}_{\omega}$; (iii) quadrupolar contributions that contain spatial derivatives of polarizations and fields that exhibit surface and volume character simultaneously.

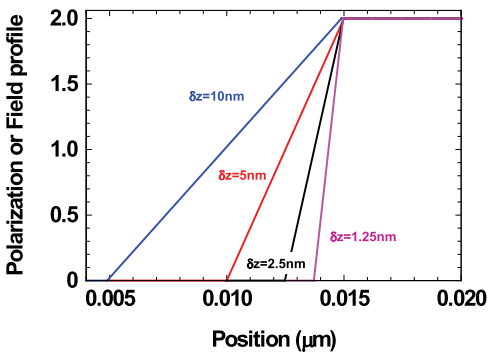

FIG. 3. (Color online) Typical polarization profile as a function of position for different spatial integration steps, as indicated in the figure.

We now provide some perspective on the quadrupolar terms in Eqs. (10) and (11) by comparing them with quadrupolar terms extracted from energy arguments that are used to define nonlinear polarization sources. For SHG, we have [1,3,4,22,24,25,40,41,47,48]

$$P_i^{2\omega}=\sum_{jkl}\chi^{(2),eq}_{ijkl}E_j\nabla_k E_l, \tag{20}$$

where i,j,k,l are Cartesian coordinates, and $\chi^{(2),eq}_{ijkl}$ are effective second-order components of a fourth-rank tensor that reflects the crystallographic properties of the medium. For cubic materials, $\chi^{(2),eq}_{ijkl}$ has 21 nonzero elements, four of which are independent [49]. Expanding Eq. (20) and neglecting

derivatives with respect to x,

$$\begin{pmatrix} P_x^{2\omega} \\ P_y^{2\omega} \\ P_z^{2\omega} \end{pmatrix} = \begin{pmatrix} \chi^{(2),eq}_{xyyx} E_y \frac{\partial E_x}{\partial y} + \chi^{(2),eq}_{xzzx} E_z \frac{\partial E_x}{\partial z} + \chi^{(2),eq}_{xxyy} E_x \frac{\partial E_y}{\partial y} + \chi^{(2),eq}_{xxzz} E_x \frac{\partial E_z}{\partial z} \\ \chi^{(2),eq}_{yxyx} E_x \frac{\partial E_x}{\partial y} + \chi^{(2),eq}_{yyyy} E_y \frac{\partial E_y}{\partial y} + \chi^{(2),eq}_{yzzy} E_z \frac{\partial E_y}{\partial z} + \chi^{(2),eq}_{yzyz} E_z \frac{\partial E_z}{\partial y} + \chi^{(2),eq}_{yyzz} E_y \frac{\partial E_z}{\partial z} \\ \chi^{(2),eq}_{zxzx} E_x \frac{\partial E_x}{\partial z} + \chi^{(2),eq}_{zzyy} E_z \frac{\partial E_y}{\partial y} + \chi^{(2),eq}_{zyzy} E_y \frac{\partial E_y}{\partial z} + \chi^{(2),eq}_{zyyz} E_y \frac{\partial E_z}{\partial y} + \chi^{(2),eq}_{zzzz} E_z \frac{\partial E_z}{\partial z} \end{pmatrix}. \tag{21}$$

Although the polarization components P_j are rather complicated solutions of Eqs. (16), for discussion and comparison purposes only we may arbitrarily extract the leading terms and write them in the form $P_j = \chi_L E_j$, where χ_L is the linear susceptibility. Equation (11) becomes

$$\begin{aligned} \nabla(\mathbf{rr} : \nabla\mathbf{E}) \approx \frac{\chi_L}{n_0 e} \Bigg[& \mathbf{i} \left(E_y \frac{\partial E_x}{\partial y} + E_z \frac{\partial E_x}{\partial z} \right) \\ & + \mathbf{j} \left(E_x \frac{\partial E_x}{\partial y} + 2E_y \frac{\partial E_y}{\partial y} + E_z \frac{\partial E_y}{\partial z} + E_z \frac{\partial E_z}{\partial y} \right) \\ & + \mathbf{k} \left(E_x \frac{\partial E_x}{\partial z} + E_y \frac{\partial E_y}{\partial z} + E_y \frac{\partial E_z}{\partial y} + 2E_z \frac{\partial E_z}{\partial z} \right) \Bigg]. \end{aligned} \tag{22}$$

A comparison between Eqs. (21) and (22) reveals that aside from the proportionality constants all terms present in Eq. (22) are also present in Eq. (21). However, there is a disparity in the number of terms present because we have made simple assumptions about the nature of the medium. While on one hand the model clearly captures the basic dynamics of the system, despite its simplicity, on the other it clearly cannot fully account for the crystalline nature of materials. This shortcoming may be cured by either going back to Eq. (5) and introducing anisotropies in the material model, or more simply by adding the missing terms to our Eq. (11). For our purposes it suffices to choose the latter course and simply add the missing terms to Eq. (11).

We now show an example that demonstrates stability, convergence, and accuracy when Eqs. (16) are integrated together with Maxwell's equations using a quadrupole moment apt for cubic materials. We calculate reflected SH efficiencies from a thick silicon substrate as a function of time by progressively decreasing spatial and temporal integration steps, and plot the results in Fig. 4. A TM-polarized, 0.5 GW/cm^2, 20 fs pulse tuned to $\lambda = 780$ nm is incident at 30° on a slab of silicon. The figure shows that the calculations are stable even for relatively large integration steps, with converging solutions that become progressively more accurate with decreasing step size. The peak amplitudes of the generated harmonic pulses are insensitive to further reductions of the integration steps beyond those shown in the figure. Figure 4 shows that all spatial derivatives in Eqs. (16) are manageable, that one is not required to introduce *ad hoc* surface nonlinearities, or guess signs and amplitudes because they are determined dynamically as the pulse crosses the interface. This makes the addition of anisotropic surface sources in a three-dimensional model relatively easy to implement.

Although part of the TH signal originates from terms in the last of Eqs. (16), the bulk $\chi^{(3),ed}_{ijkl}$ is responsible for most THG. $\chi^{(3),ed}_{ijkl}$ effects are introduced as a third-order nonlinear polarization whose Cartesian components are given by

$$P_i^{NL} = \sum_{j=1,3,} \sum_{k=1,3,} \sum_{l=1,3,} \chi^{(3),ed}_{ijkl} E_j E_k E_l, \tag{23}$$

where i,j,k,l are Cartesian coordinates. Equation (23) yields 81 components [49], with nonzero terms that track those of Eq. (21). The result is

$$\begin{aligned} P_x^{NL} &= \chi^{(3),ed}_{xxxx} E_x^3 + \left(\chi^{(3),ed}_{xyyx} + \chi^{(3),ed}_{xyxy} + \chi^{(3),ed}_{xxyy}\right) E_y^2 E_x + \left(\chi^{(3),ed}_{xzzx} + \chi^{(3),ed}_{xzxz} + \chi^{(3),ed}_{xxzz}\right) E_z^2 E_x, \\ P_y^{NL} &= \chi^{(3),ed}_{yyyy} E_y^3 + \left(\chi^{(3),ed}_{yyxx} + \chi^{(3),ed}_{yxyx} + \chi^{(3),ed}_{yxxy}\right) E_x^2 E_y + \left(\chi^{(3),ed}_{yzzy} + \chi^{(3),ed}_{yzyz} + \chi^{(3),ed}_{yyzz}\right) E_z^2 E_y, \\ P_z^{NL} &= \chi^{(3),ed}_{zzzz} E_z^3 + \left(\chi^{(3),ed}_{zzxx} + \chi^{(3),ed}_{zxzx} + \chi^{(3),ed}_{zxxz}\right) E_x^2 E_z + \left(\chi^{(3),ed}_{zzyy} + \chi^{(3),ed}_{zyzy} + \chi^{(3),ed}_{zyyz}\right) E_y^2 E_z. \end{aligned} \tag{24}$$

For simplicity we may set all coefficients equal, so that Eq. (24) simplifies to

$$\begin{aligned} P_x^{NL} &= \chi^{(3),ed}\left(E_x^2 + E_y^2 + E_z^2\right) E_x, \\ P_z^{NL} &= \chi^{(3),ed}\left(E_x^2 + E_y^2 + E_z^2\right) E_y, \\ P_z^{NL} &= \chi^{(3),ed}\left(E_x^2 + E_y^2 + E_z^2\right) E_z. \end{aligned} \tag{25}$$

Substitution of Eq. (12) into Eq. (25) generates nonlinear source terms for the pump and all its harmonics, that are then inserted back into Eqs. (16). In what follows we choose $\chi^{(3),ed}_{cgs} = 0.75 \times 10^{-10}$ esu, equivalent to $\chi^{(3),ed}_{MKS} \sim 10^{-18}$ (m/V)^2.

III. SHG AND THG FROM BULK SILICON

Now that we have tested numerical stability and convergence we examine harmonic generation from a thick silicon substrate. We use the same quadrupole moment used to compute Fig. 4, one that contains the terms found in Eq. (21) all placed on equal footing. Then, the largest contributions come

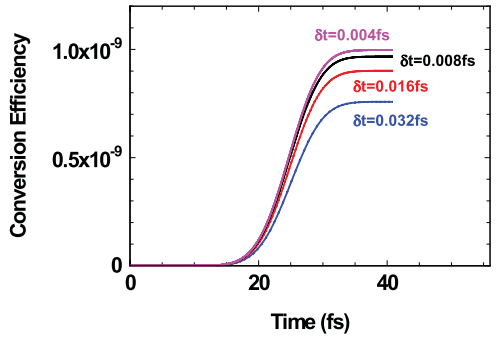

FIG. 4. (Color online) Reflected efficiency vs time for decreasing spatial and temporal step sizes. The calculations provide a fairly accurate answer even for the largest integration steps shown. In scaled units, $\delta\xi = z/\lambda_0 = \delta\tau = ct/\lambda_0$.

from just a few terms proportional to $\partial E_\xi/\partial\xi$. The complex dielectric function of silicon [27] is plotted in Fig. 5. Si has two absorption resonances in the UV range, and $\mathrm{Re}(\varepsilon) < 0$ for $\lambda < 300$ nm. The linear data are fitted using two detuned Lorentz oscillators:

$$\varepsilon(\omega) = 1 - \frac{\omega_{p,1}^2}{\omega^2 - \omega_{0,1}^2 + i\gamma_1\omega} - \frac{\omega_{p,2}^2}{\omega^2 - \omega_{0,2}^2 + i\gamma_2\omega}, \qquad (26)$$

where ω is in units of $\mu\mathrm{m}^{-1}$, γ_1, $\gamma_2 = 0.75$, 0.18 are damping constants, and $\omega_{0,1}$, $\omega_{0,2} = 3.3$, 2.8 and $\omega_{p,1}$, $\omega_{p,2} = 10$, 4.5 are resonance and plasma frequencies, respectively. We then integrate two distinct sets of Eqs. (16) simultaneously, one for each oscillator type, each having its own quadrupolar species. Incident pulses with peak intensities of 0.5 GW/cm^2 are tuned to 780 nm, where the absorption length is $\sim$25 μm. SH and TH signals are tuned to an absorption resonance (390 nm) and to the metallic range (260 nm), respectively. In Fig. 6 we plot the reflected SHG (a) and THG (b) conversion efficiencies with and without quadrupolar contributions. The SHG maxima are separated by just a few degrees, and with all else being equal quadrupoles enhance the intensity of the emitted signal by two orders of magnitude [right axis scale in Fig. 6(a)] without changing the basic shape

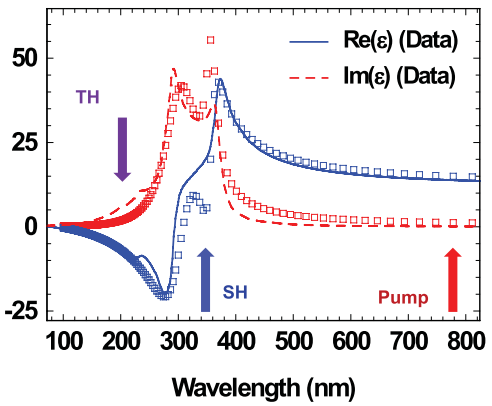

FIG. 5. (Color online) Dielectric response of silicon at visible and near-IR wavelengths. The thin solid curves correspond to the data, as indicated. Empty, color-coded squares that retrace the complex data set correspond to a plot of Eq. (26). The data are reproduced reasonably well across the entire range.

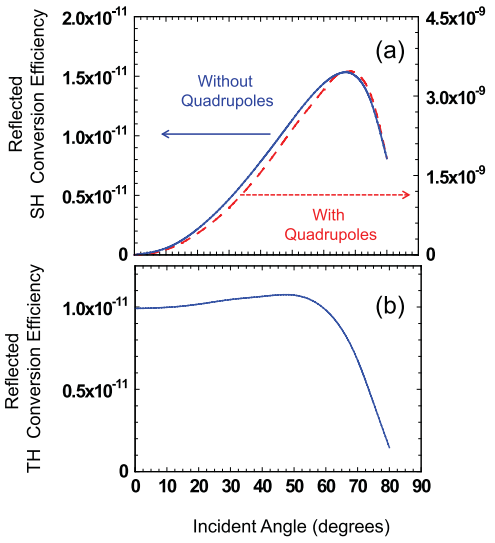

FIG. 6. (Color online) SHG (a) and THG (b) as a function of incident angle for a bulk silicon substrate. The comparison in (a) with and without quadrupoles suggests that most of the SH signal may have quadrupolar origin (right axis). For THG (b), the signal originates mostly with the bulk $\chi^{(3),ed}$, and closely resembles the predictions in Ref. [30].

of the curve. If all matrix elements of the quadrupole moment are taken to have equal weight then most of the signal originates with the quadrupolar term $P_{\tilde{y}}^{2\omega} = \chi_{yyzz}^{(2),eq} E_{\tilde{y}}(\partial E_\xi/\partial\xi)$ in Eq. (21), or equivalently, with the more generic term $P_{\tilde{y}}^{2\omega} = P_{\tilde{y}}(\partial E_\xi/\partial\xi)/n_0 e$ added to Eq. (11). In contrast, reflected THG is largely unaffected by sources unrelated to the $\chi_{ijkl}^{(3),ed}$, and the curve retains the same shape and amplitude regardless of the circumstances. In particular, we note that its shape is nearly identical to the Fresnel factor used to calculate TH conversion efficiencies from bulk silicon in Ref. [30].

IV. SHG AND THG FROM A FABRY-PÉROT ÉTALON AND MULTILAYER STACK

While the calculations for a thick substrate suggest that the shape of the emitted SH signal remains the same with or without quadrupole contributions (Fig. 6), this is not the case for a Fabry-Pérot étalon. In Fig. 7 we show the linear

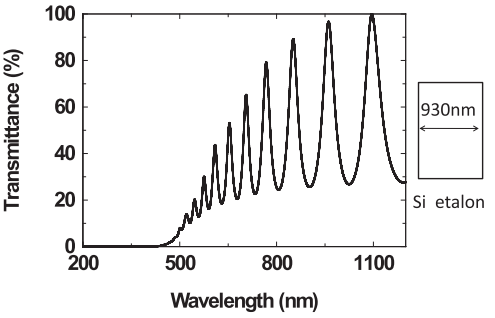

FIG. 7. Linear transmittance from a silicon étalon 930 nm thick at normal incidence. At 780 nm transmittance is $\sim$80%; wavelengths below 450 nm are absorbed, with no expectation to see harmonic signals pass through the étalon.

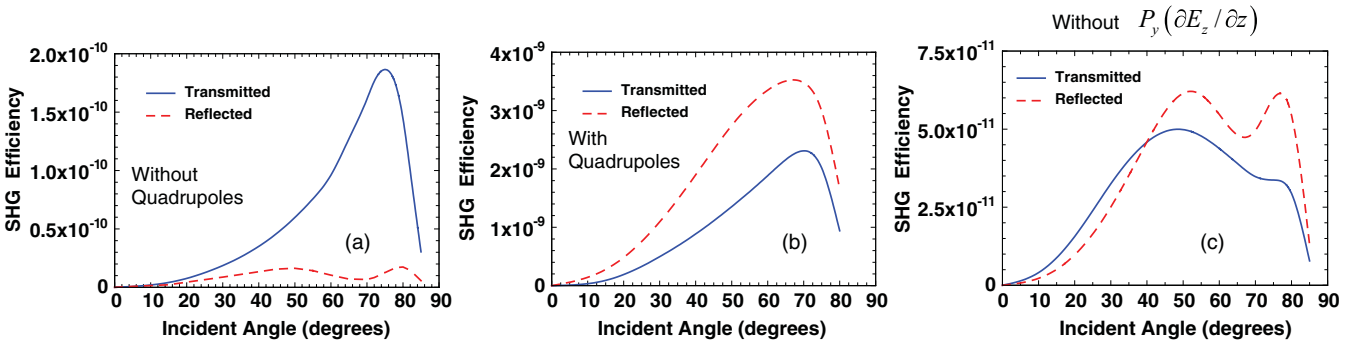

FIG. 8. (Color online) SHG without (a) and with (b) quadrupolar contributions, and (c) without the most dominant quadrupolar term. The shapes of the curves in (b) depend on the relative amplitudes of the matrix elements of the quadrupole tensor. Panel (c) shows that the curves can take on configurations intermediate between (a) and (b), with efficiencies similar to (a), so that the types of nonlinear sources discussed in the text contribute to similar degrees.

transmission function vs incident wavelength for a Si layer ∼930 nm thick, in the UV–near-IR range. Although a pump field tuned near 780 nm is still resonantly transmitted at a rate of ∼80%, the figure shows that below 450 nm transmittance shuts down due to linear absorption. Two mechanisms allow the observation of transmitted SH and/or TH signals: (i) phase locking, so that part of the signal that exits the étalon is forcibly carried by the pump, just as it occurs for a GaP substrate [28], and (ii) the interaction of the pump with the exit surface. In this case the transmitted TH signal is due to $\chi^{(3),ed}$. In Fig. 8 we plot SH efficiencies vs incident angle with and without quadrupolar contributions for the étalon depicted in Fig. 7. Figure 8(b) shows that the quadrupoles can significantly modify the angular dependence and the amplitude of the signal. A comparison between the conversion efficiencies of bulk [Fig. 6(a)] and étalon [Fig. 8(b)] shows that the reflected signal is practically unchanged, while the transmitted signal in Fig. 8(b) is now quite substantial. The actual shapes of transmitted and reflected curves shown will depend on the relative magnitudes of the components in Eqs. (21). The ultimate angular dependence may be intermediate between Figs. 8(a) and 8(b). An example of how the curves reshape is shown in Fig. 8(c), where SHG efficiency is calculated without the most dominant quadrupolar term. The results of THG for the étalon are far less ambiguous and are shown in Fig. 9. The total THG efficiency for the étalon is enhanced by a factor of 30 relative to bulk [Fig. 6(b)], most of it coming from the transmitted portion of the signal at normal incidence, notwithstanding the fact that the medium should be opaque in the TH range. As we mentioned above, one of the mechanisms that allows SH and TH signals to be transmitted in environments that are either strongly absorptive or metallic is phase locking: The pump traps the harmonics and imparts to them its own dispersive properties. It then becomes possible for the harmonics to resonate, even though the cavity is designed to resonate only at the fundamental wavelength [50]. THG has already been experimentally verified for bulk GaP in the metallic range [28], but there is still no experimental evidence that cavities can be used to enhance THG if the dielectric constant of the active layer is negative.

Here we predict that the enhancement of THG should be significant if we consider a Si étalon sandwiched between two DBRs. The multilayer stack–Fig. 10–consists of a Si layer 330 nm thick sandwiched between two, three-period SiO_2(100 nm)/Si_3N_4(140 nm) DBRs. Linear transmittance and

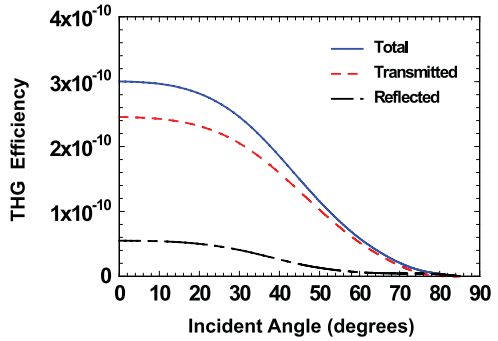

FIG. 9. (Color online) THG vs incident angle for the Fabry-Pérot étalon depicted in Fig. 7. When this curve is compared with Fig. 6, which depicts THG from bulk silicon, one can deduce that total efficiency is amplified by a factor of ∼30. Most of the enhancement comes from the transmitted signal, which is phase locked and resonates with the pump.

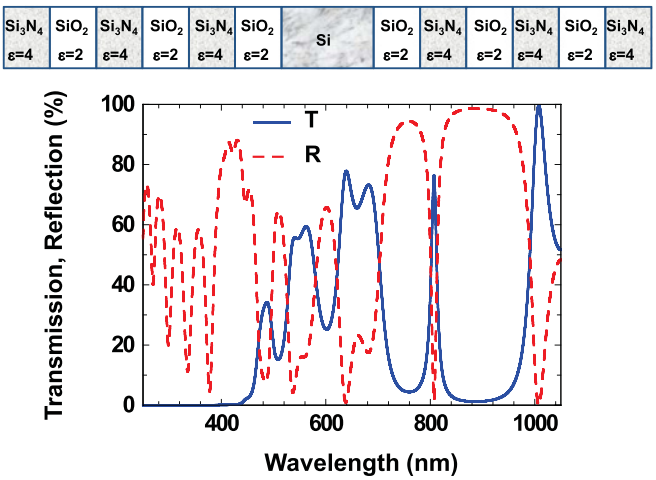

FIG. 10. (Color online) Linear transmittance and/or reflectance of a multilayer stack that contains a Si defect layer for right-to-left or left-to-right propagation. The resonance near 800 nm is ∼20 nm wide, and is resolved using subpicosecond, incident pulses [50].

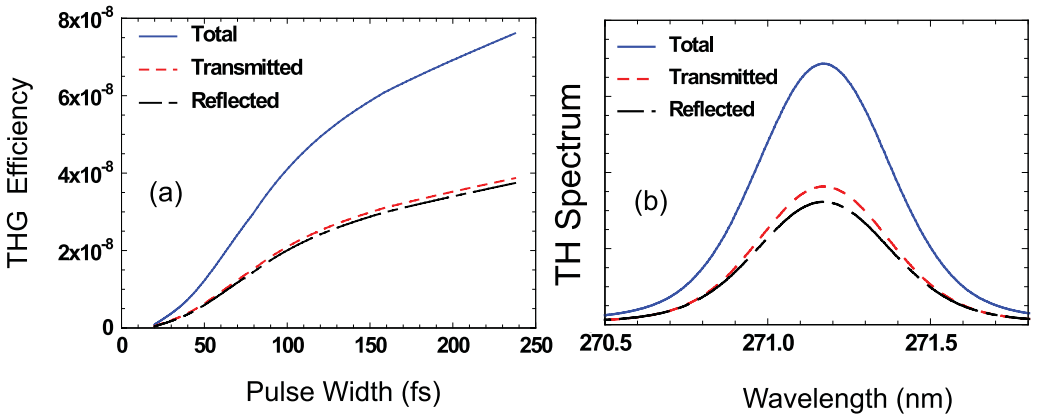

FIG. 11. (Color online) (a) Peak transmitted and reflected conversion efficiencies as a function of incident pulse width. The pulses are tuned at resonance near 813 nm. THG saturates as longer pulses resolve the resonance. (b) Transmitted and reflected THG spectra when the incident pulse is ∼250 fs in duration. Even though the cavity $Q \approx 40$, the total conversion efficiency increases by more than two orders of magnitude compared to the Si étalon. The far left axis scale is shared by both (a) and (b).

reflectance are shown in Fig. 10, with a Fabry-Pérot resonance at ∼813 nm. The estimated cavity $Q = \lambda/\delta\lambda \approx 40$. In Fig. 11 we plot transmitted and reflected TH efficiencies vs incident pulse duration at normal incidence [Fig. 11(a)], as well as transmitted and reflected spectra when the stack is illuminated with a pulse approximately 250 fs in duration [Fig. 11(b)]. We neglect SHG because at normal incidence it is minimized, although anisotropic surface sources might change that. First, we note that transmitted and reflected amplitudes are similar and saturate as a function of pulse duration, as expected, as the signal resolves the resonance. Second, a comparison between Figs. 9 and 11 shows that a cavity with a modest Q improves THG efficiency by more than two orders of magnitude, despite the fact that the TH is tuned in the metallic range. The widths of the spectral features shown in Fig. 11(b) depend on the pump spectrum, i.e., shorter pulses produce correspondingly narrower TH spectra. This behavior also occurs for a GaAs defect layer surrounded by DBRs when the SH signal is tuned below the absorption edge, and higher efficiencies are possible by adding more periods to the DBRs [50]. These results confirm theoretically that if the pump is tuned to a region of relative transparency phase locking triggers a form of electromagnetic transparency at the harmonic wavelengths regardless of material dispersion, with relatively large enhancement of conversion efficiency in cavity environments.

V. CONCLUSIONS

In summary, we have derived a new dynamical model of harmonic generation from centrosymmetric materials that includes surface, magnetic, and electric quadrupole sources. In silicon, the presence of quadrupoles can remodulate both amplitude and shape of the conversion efficiency curve of SHG, while THG can be enhanced significantly in cavity environments, even in ranges where the dielectric constant displays metallic behavior. These findings make it possible to rethink the role and functionality of semiconductors like silicon or germanium, perhaps by extending their usefulness as nonlinear media well into the UV range.

APPENDIX

It is well known that an arbitrary charge distribution may be expressed as a superposition of multipoles. An alternative way to arrive at Eq. (5) of our manuscript is to follow the Hamiltonian approach outlined in our Ref. [1], where instead of equations of motion the stage is set for an energetic route combined with perturbation theory. The Hamiltonian describes the interaction energy between externally applied fields and the charge distribution in the form of electric dipoles, $\mathbf{p} = e\mathbf{r}$, and electric quadrupoles, $\mathbf{Q} = (e/2)\mathbf{rr}$, and takes the following form (neglecting magnetization effects):

$$H = -\mathbf{p}\cdot\mathbf{E}(\mathbf{r}_0+\mathbf{r},t) - \tfrac{1}{2}\mathbf{Q}:\nabla\mathbf{E}(\mathbf{r}_0+\mathbf{r},t). \quad \text{(A1)}$$

$\mathbf{Q}$ is defined as a nonlinear function of position whose effects cannot be reproduced by a linear superposition of electric dipoles, i.e., the first term in Eq. (A1). The approach followed in Ref. [1] is to insert the quadrupole operator in a perturbation theory to show that the nonlinear polarization that arises from quadrupolar terms is proportional to $E_i\nabla_j E_k$. Instead, using the Hamiltonian in Eq. (A1) we examine a simpler, classical oscillator model capable of yielding practical equations of motion that are more accessible and can easily be tested. Taking the negative gradient of H in Eq. (A1) to calculate the total force on the multipolar charge distribution we obtain two terms in the form

$$\mathbf{F} = \mathbf{F}_{\text{dipoles}} + \mathbf{F}_{\text{quadrupoles}}. \quad \text{(A2)}$$

The theoretical aspects are complicated by the fact that dipolar and quadrupolar species may have different effective masses, damping, and spring constants. Putting that aside temporarily, using an identity valid for any pair of well-behaved vectors, and dropping any explicit time dependence, the dipolar force term becomes

$$\nabla[\mathbf{p}\cdot\mathbf{E}(\mathbf{r}_0)] = (\mathbf{p}\cdot\nabla)\mathbf{E}(\mathbf{r}_0) + [\mathbf{E}(\mathbf{r}_0)\cdot\nabla]\mathbf{p} + \mathbf{p}\times\nabla \times \mathbf{E}(\mathbf{r}_0) + \mathbf{E}(\mathbf{r}_0)\times\nabla\times\mathbf{p}. \quad \text{(A3)}$$

Aided by another identity, $\nabla \times \mathbf{p} = \nabla \times \mathbf{r} = 0$, and one of Maxwell's equations,

$$\nabla \times \mathbf{E}(\mathbf{r}_0) = -\frac{1}{c}\frac{\partial \mathbf{B}(\mathbf{r}_0)}{\partial t}, \tag{A4}$$

it follows that the negative gradient of the dipolar term in Eq. (A1) becomes

$$\begin{aligned}\nabla[\mathbf{p}\cdot\mathbf{E}(\mathbf{r}_0)] &= e\mathbf{E}(\mathbf{r}_0) + (\mathbf{p}\cdot\nabla)\mathbf{E}(\mathbf{r}_0) + \mathbf{p}\times\nabla\times\mathbf{E}(\mathbf{r}_0)\\ &= e\mathbf{E}(\mathbf{r}_0) + (\mathbf{p}\cdot\nabla)\mathbf{E}(\mathbf{r}_0) - \frac{1}{c}\mathbf{p}\times\frac{\partial\mathbf{B}(\mathbf{r}_0)}{\partial t}.\end{aligned} \tag{A5}$$

Comparing Eqs. (A4) to (5) in the text, we find that all first-order, dipolar terms match, except for the magnetic term, where time derivatives appear inverted. However, it can be shown that at least for monochromatic waves the magnetic force can take on two apparently different but equivalent forms, namely,

$$\frac{e}{c}\dot{\mathbf{r}}\times\mathbf{B} = -\frac{e}{c}\mathbf{r}\times\frac{\partial\mathbf{B}}{\partial t}. \tag{A6}$$

Assuming that Eq. (A6) holds generally, we continue on with the development of the Hamiltonian Eq. (A1), and evaluate the gradient of the quadrupolar interaction Hamiltonian, i.e.,

$$\mathbf{F}_{\text{quadrupoles}} = \tfrac{1}{2}\nabla[\mathbf{Q}:\nabla\mathbf{E}(\mathbf{r}_0+\mathbf{r},t)], \tag{A7}$$

which yields the remaining quadrupolar terms in Eq. (5). While in reality dipolar and quadrupolar species have different degrees of freedom, and may display different physical parameters (mass, damping constant, resonance frequency, displacement from equilibrium), in order to simplify the problem one may combine the forces into a single equation of motion by assuming that dipoles and quadrupoles are colocated, share the same coordinate system, and have the same density n_0. Then, if the medium is isotropic and both charge distributions are driven by the same field, one can reasonably expect that the position vector that describes the second, less influential, quadrupolar oscillator is roughly collinear and amounts to a small fraction of the vector that describes the more dominant, dipolar oscillator species. The differences in oscillator species are then absorbed by the effective mass, damping coefficient, resonance, and plasma frequency. As a result, it is possible to develop a simplistic model that overlooks unnecessary complications by combining Eq. (A5)–(A7) into a single equation of motion, Eq. (5) in the text.

[1] N. Bloembergen, R. K. Chang, S. S. Jha, and C. H. Lee, Phys. Rev. **174**, 813 (1968).
[2] T. A. Driscoll and D. Guidotti, Phys. Rev. B **28**, 1171 (1983).
[3] D. Guidotti and T. A. Driscoll, Il Nuovo Cimento **8**, 385 (1986).
[4] H. W. K. Tom, T. F. Heinz, and Y. R. Shen, Phys. Rev. Lett. **51**, 1983 (1983).
[5] J. A. Litwin, J. E. Sipe, and H. M. van Driel, Phys. Rev. B **31**, 5543 (1985).
[6] B. S. Mendoza, J. Phys.: Condens. Matter **5**, A181 (1993).
[7] B. S. Mendoza and W. L. Mochán, Phys. Rev. B **53**, R10473 (1996).
[8] T. Suzuki, D. E. Milovzorov, S. Kogo, M. Tsukakoshi, and M. Aono, Appl. Phys. B **68**, 623 (1999).
[9] Y. Q. An, R. Carriles, and M. C. Downer, Phys. Rev. B **75**, 241307 (2007).
[10] P. Figliozzi, L. Sun, Y. Jiang, N. Matlis, B. Mattern, M. C. Downer, S. P. Withrow, C. W. White, W. L. Mochán, and B. S. Mendoza, Phys. Rev. Lett. **94**, 047401 (2005).
[11] T. Ishihara, K. Koshino, and H. Nakashima, Phys. Rev. Lett. **91**, 253901 (2003).
[12] J. J. H. Gielis, Ph.D. thesis, Technische Universiteit Eindhoven, 2008.
[13] V. I. Gavrilenko, Phys. Status Solidi A **1884**, 1267 (2001).
[14] M. C. Downer, B. S. Mendoza, and V. I. Gavrilenko, Surf. Interface Anal. **31**, 966 (2001).
[15] D. Bodlaki, E. Freysz, and E. Borguet, J. Chem. Phys. **119**, 3958 (2003).
[16] T. Chu, M. Yamada, J. Donecker, M. Rossberg, V. Alex, and H. Riemann, Mater. Sci. Eng. B **91-92**, 174 (2002).
[17] S. R. Lederhandler, J. Appl. Phys. **30**, 1631 (1959).
[18] C. Schriever, C. Bohley, and R. B. Wehrspohn, Opt. Lett. **35**, 273 (2010).
[19] M. Kauranen, T. Verbiest, and A. Persoons, J. Mod. Opt. **45**, 403 (1998).
[20] V. Fomenko, D. Bodlaki, C. Faler, and E. Borgueta, J. Chem. Phys. **116**, 6745 (2002).
[21] I. M. Baranova and K. N. Evtyukhov, Quantum Electron. **25**, 1198 (1995).
[22] O. A. Aktsipetrov, I. M. Baranova, and Yu. A. Il'inskii, Sov. Phys. JETP **64**, 167 (1986).
[23] H. J. Peng, E. J. Adles, J. F. T. Wang, and D. E. Aspnes, Phys. Rev. B **72**, 205203 (2005).
[24] P. Guyot-Sionnest and Y. R. Shen, Phys. Rev. B **35**, 4420 (1987).
[25] P. Guyot-Sionnest and Y. R. Shen, Phys. Rev. B **38**, 7985 (1988).
[26] M. Galli, D. Gerace, K. Welna, T. F. Krauss, L. O'Faolain, G. Guizzetti, and L. C. Andreani, Opt. Exp. **18**, 26613 (2010).
[27] *Handbook of Optical Constants of Solids*, edited by E. D. Palik (Academic, San Diego, CA, 1985).
[28] V. Roppo, J. V. Foreman, N. Akozbek, M. A. Vincenti, and M. Scalora, Appl. Phys. Lett. **98**, 11105 (2011).
[29] V. Roppo, M. Centini, C. Sibilia, M. Bertolotti, D. de Ceglia, M. Scalora, N. Akozbek, M. J. Bloemer, J. W. Haus, O. G. Kosareva, and V. P. Kandidov, Phys. Rev. A **76**, 033829 (2007).
[30] W. K. Burns and N. Bloembergen, Phys. Rev. B **4**, 3437 (1971).
[31] D. J. Moss, E. Ghahramani, J. E. Sipe, and H. M. van Driel, Phys. Rev. B **41**, 1542 (1990).
[32] D. E. Aspnes and A. A. Studna, Phys. Rev. Lett. **54**, 1956 (1985).
[33] R. W. J. Hollering, J. Opt. Soc. Am. B **8**, 374 (1991).
[34] D. A. Koos, V. L. Shannon, and G. L. Richmond, Phys. Rev. B **47**, 4730 (1993).

[35] N. Bloembergen and Y. R. Shen, Phys. Rev. **141**, 298 (1966).
[36] E. J. Adles and D. E. Aspnes, Phys. Rev. B **77**, 165102 (2008).
[37] M. Scalora, M. A. Vincenti, D. de Ceglia, V. Roppo, M. Centini, N. Akozbek, and M. J. Bloemer, Phys. Rev. A **82**, 043828 (2010).
[38] M. A. Vincenti, D. de Ceglia, V. Roppo, and M. Scalora, Opt. Express **19**, 2064 (2011).
[39] O. L. de Lange and R. E. Raab, Am. J. Phys. **74**, 301 (2006).
[40] P. S. Pershan, Phys. Rev. **130**, 919 (1963).
[41] T. Verbiest, K. Clays, and V. Rodriguez, *Second-order Nonlinear Optical Characterization Techniques: An Introduction* (CRC Press–Taylor and Francis Group, Boca Raton, FL, 2009).
[42] J. D. Jackson, *The Classical Electromagnetic Field* (Wiley, New York, 1999).
[43] J. P. Gordon, Phys. Rev. A **8**, 14 (1973).
[44] S. M. Barnett and R. Loudon, J. Phys. B: At. Mol. Opt. Phys. **39**, S671 (2006).
[45] H. S. Freedhoff, J. Chem. Phys. **54**, 1618 (1971).
[46] H. S. Freedhoff, Phys. Rev. A **5**, 126 (1972).
[47] D. Epperlein, B. Dick, G. Marowsky, and G. A. Reider, Appl. Phys. B **44**, 5 (1987).
[48] F. X. Wang, F. J. Rodríguez, W. M. Albers, R. Ahorinta, J. E. Sipe, and M. Kauranen, Phys. Rev. B **80**, 233402 (2009).
[49] R. W. Boyd, *Nonlinear Optics* (Academic, Boston, 2003).
[50] V. Roppo, F. Raineri, R. Raj, I. Sagnes, J. Trull, R. Vilaseca, M. Scalora, and C. Cojocaru, Opt. Lett. **36**, 1809 (2011).