

# The anisotropic bond model of nonlinear optics

E. J. Adles and D. E. Aspnes\*

Department of Physics, North Carolina State University, Raleigh, NC 27695-8202, USA

Received 15 June 2007, revised 16 October 2007, accepted 17 December 2007

Published online 26 March 2008

PACS 42.65.An

\* Corresponding author: e-mail aspnes@ncsu.edu, Phone: +00 1 919 515 4261, Fax: +00 1 919 515 1333

The anisotropic-bond model (ABM) of nonlinear optics (NLO) provides a simple means of calculating NLO properties of materials by factoring the problem into four parts: first, determination of the local field at a bond-charge site; second, solution of the anharmonic force equation of the bond

charge; third, calculation of the radiation from the charge; and fourth, superposition of the radiation from all charges. Because this factorization is impossible in linear optics, this is one of the few cases where a nonlinear problem is simpler than its linear equivalent.

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

**1 Introduction** Physics takes place in real time and on the atomic scale. This is often forgotten in descriptions of optical properties of materials, where phenomena are usually expressed in macroscopic terms as dielectric functions  $\epsilon$ , which are themselves defined in terms of averages of atomic-scale fields and polarizations. However, physics is lost in the averaging process, as can be seen by considering the largely forgotten Ewald–Oseen extinction theorem of linear optics [1]. Here, reflection is described in terms of the coherent superposition of radiation from induced dipoles created when point charges (Ewald, 1912 [2]) or a continuum (Oseen, 1915 [3]) are accelerated by the field of an incoming plane wave. In this atomic-level picture, the Brewster angle is nothing more than the condition where the observer is viewing the induced dipoles end-on, rather than a consequence of particular values of macroscopic quantities. Since an oscillating dipole does not emit radiation along its axis, the absence of a reflected beam follows immediately from first principles.

The derivation of the Ewald–Oseen extinction theorem is complicated because the intensity of the emitted radiation (reflected and transmitted beams) is of the same order of magnitude as the plane wave that created it, so the problem must be solved self-consistently. We recently introduced a nonlinear-optic (NLO) analog, the anisotropic bond model (ABM), where observed NLO intensities are described as the superposition of radiation from bond

charges accelerated by an incoming plane wave. While the bond-charge picture was used in the early days of NLO, its use was restricted to calculation of static susceptibilities [4].

Our contributions are twofold. First, we not only formulate the problem in terms of bonds but define the bonds to be anisotropic, so crystal symmetry is incorporated at the atomic scale. This distinguishes our work from the Ewald–Oseen extinction theorem, and yields the macroscopic tensor properties automatically.

Second, we recognize that the NLO problem is actually much simpler than the linear version because the emerging radiation of interest is orders of magnitude weaker and occurs at energies different from that of the pump beam. This allows us to factor the NLO problem into 4 *independent* steps: (1) evaluation of the local field at a bond site; (2) solution of the mechanical equation  $F = ma$  to obtain the acceleration of the bond charge by the local field; (3) calculation of the radiation from the accelerating charge; and (4) superposition of the radiation from all charges to obtain the observed intensity. In contrast, in the linear problem steps (1), (3), and (4) are coupled. This factoring not only results in an enormous simplification, but provides the physical insight that is lacking in standard macroscopic approaches. Further, the independence allows considerable flexibility in setting up a calculation, because different levels of approximation appropriate to different situations can be introduced at each step. This is one of the

few cases where a nonlinear problem is simpler than the linear equivalent.

The point of submitting a NLO paper to a conference on (linear) spectroscopic ellipsometry (SE) is that the ABM provides previously lacking insights on optical properties at the atomic level and anticipates the time when NLO-SE becomes a reality. Because of their higher-rank tensorial nature, NLO data in principle offer considerably more diagnostic information than their linear equivalents. However, this promise has not been achieved, in part because of the difficulty of performing NLO spectroscopy, but also because NLO data have been difficult to analyze. But the ABM shows that there is a third reason: most of the information contained in the tensorial properties of NLO data is simply a result of crystal structure, and hence provides little new information about a material. Spectroscopic data are required. With further advances in technology we expect SE to move into NLO areas, so a discussion of NLO in a SE conference is relevant.

**2 Theory** We have applied a simpler version of the ABM to a wide range of NLO phenomena, including dipole-allowed second- [5], third- [6], and fourth- [7] harmonic generation (SHG, THG, FHG); dipole-forbidden SHG [8]; and the generation of terahertz (THz) radiation [9]. We are currently investigating SHG from nanoscopically inhomogeneous materials, for example the configuration recently investigated by Figliozzi et al. [10]. Our discussion below focuses on this application, because it allows us to demonstrate a NLO calculation that requires the use of all 4 steps.

We begin with the 4-potential of a point charge in free space:

$$[\phi(\mathbf{r}, t), \mathbf{A}(\mathbf{r}, t)] = \frac{1}{c} \int d^3 r' dt' [c\rho(\mathbf{r}', t'), \mathbf{J}(\mathbf{r}', t')] \times G(\mathbf{r}, \mathbf{r}', t, t'), \quad (1)$$

where

$$G(\mathbf{r}, \mathbf{r}', t, t') = \frac{\delta\left(t - t' - \frac{1}{c}|\mathbf{r} - \mathbf{r}'|\right)}{|\mathbf{r} - \mathbf{r}'|} \quad (2)$$

is the Green function for free space, and where  $\rho$  and  $\mathbf{J} = \rho\mathbf{v}$  are the charge and current densities, respectively. We suppose the  $\rho$  and  $\mathbf{J}$  are associated with a point charge  $q$  located at  $(\mathbf{r}_0 + \Delta\mathbf{r}(t))$ , where  $\Delta\mathbf{r}(t)$  is the displacement of  $q$  that results from an applied field  $\mathbf{E}$  and  $\mathbf{r}_0$  is the location of  $q$  with respect to an origin defined in the laboratory frame. Then

$$\rho(\mathbf{r}', t') = q\delta(\mathbf{r}' - \mathbf{r}_0 - \Delta\mathbf{r}(t')); \quad (3a)$$

$$\mathbf{J}(\mathbf{r}', t') = \left(q \frac{d\Delta\mathbf{r}(t')}{dt'}\right) \delta(\mathbf{r}' - \mathbf{r}_0 - \Delta\mathbf{r}(t')), \quad (3b)$$

where for second-harmonic generation  $\Delta\mathbf{r}(t)$  is obtained by solving the force equation

$$\mathbf{F} = m\mathbf{a} = m \frac{d^2\Delta\mathbf{r}(t)}{dt^2} = q\mathbf{E}(\mathbf{r}, t) - \tilde{\kappa}_1 \cdot \Delta\mathbf{r}(t) - \tilde{\kappa}_2 \cdot \Delta\mathbf{r}(t) \Delta\mathbf{r}(t), \quad (4)$$

where  $m$  is the mass of  $q$  and  $\tilde{\kappa}_1$  and  $\tilde{\kappa}_2$  are second- and third-rank tensors describing the linear (Hooke's Law) and first-order nonlinear restoring forces, respectively. For higher-harmonic generation other terms may be added as necessary. We now suppose that the driving field is a plane wave, so step (1) is trivial. We write  $\mathbf{E}(\mathbf{r}, t)$  as

$$\mathbf{E}(\mathbf{r}, t) = (1 + \Delta\mathbf{r} \cdot \nabla_{\mathbf{r}_0}) \mathbf{E}(\mathbf{r}_0) e^{-i\omega t}, \quad (5)$$

to bring out field-gradient effects explicitly. The function  $\mathbf{E}(\mathbf{r}_0)$  includes not only the standard plane-wave phase term  $e^{i\mathbf{k}_0 \cdot \mathbf{r}_0}$  but possibly more slowly varying longitudinal and transverse variations as well. Here,  $\mathbf{k}_0$  is the wave vector of the incoming plane wave, and  $\omega$  is its frequency.

We now evaluate  $\Delta\mathbf{r}(t)$ . From the form of  $\mathbf{E}(\mathbf{r}, t)$ ,  $\Delta\mathbf{r}(t)$  will have the general representation

$$\Delta\mathbf{r}(t) = \Delta\mathbf{r}_1 e^{-i\omega t} + \Delta\mathbf{r}_2 e^{-i2\omega t}, \quad (6)$$

which, when substituted into Eq. (4) and the solution obtained, will complete step 2. For simplicity, and because it is adequate to treat the spherical-inclusions problem, we assume that the bonds are rotationally symmetric so that the force equation can be divided into independent longitudinal and transverse parts. Letting  $\hat{b}$  and

$$\hat{t} = (\mathbf{E} - \hat{b}(\hat{b} \cdot \mathbf{E})) / \sqrt{\mathbf{E}^2 - (\hat{b} \cdot \mathbf{E})^2} \quad (7)$$

be the unit vectors along and transverse to the bond, with  $\hat{t}$  also lying in the  $\hat{b} - \mathbf{E}$  plane, we then divide  $\tilde{\kappa}_1$  into longitudinal  $\tilde{\kappa}_{1l}$  and transverse  $\tilde{\kappa}_{1t}$  parts, with  $\tilde{\kappa}_{2l}$  having only a longitudinal component, by symmetry. Substituting everything back into the force equation and isolating linear and second-order terms we have

$$\Delta\mathbf{r}_{1l} = \Delta r_{1l} \hat{b} = \frac{q(\hat{b} \cdot \mathbf{E})}{\kappa_{1l} - m\omega^2} \hat{b}; \quad (8a)$$

$$\Delta\mathbf{r}_{1t} = \Delta r_{1t} \hat{t} = \frac{q(\hat{t} \cdot \mathbf{E})}{\kappa_{1t} - m\omega^2} \hat{t}; \quad (8b)$$

$$\Delta\mathbf{r}_{2l} = \Delta r_{2l} \hat{b} = \frac{q(\Delta\mathbf{r}_1 \cdot \nabla_{\mathbf{r}_0})(\hat{b} \cdot \mathbf{E}) - \kappa_{2l} \Delta r_{1l} \Delta r_{1l}}{\kappa_{1l} - 4m\omega^2} \hat{b}; \quad (8c)$$

$$\Delta\mathbf{r}_{2t} = \Delta r_{2t} \hat{t} = \frac{q(\Delta\mathbf{r}_1 \cdot \nabla_{\mathbf{r}_0})(\hat{t} \cdot \mathbf{E})}{\kappa_{1t} - 4m\omega^2} \hat{t}. \quad (8d)$$

For SHG step 2 is now completed. If we neglect bond anharmonicity, Eq. (8c) and (8d) reduce to the  $(\mathbf{E} \cdot \nabla) \mathbf{E}$  result reported in macroscopic treatments [11, 12].

We now perform step 3, calculating the far-field limit  $E_{\text{ff}}$  of the field radiated by the motion of  $q$ . Combining Eqs. (1)–(3) we can evaluate the integral over  $\mathbf{r}'$  explicitly, obtaining

$$A(\mathbf{r}, t) = \frac{q}{c} \int dt' \frac{d\Delta \mathbf{r}'(t')}{dt'} \frac{\delta\left(t - t' - \frac{1}{c} |\mathbf{r} - \mathbf{r}_0 - \Delta \mathbf{r}'(t')|\right)}{|\mathbf{r} - \mathbf{r}_0 - \Delta \mathbf{r}'(t')|}. \quad (9)$$

The integral over  $t'$  is nontrivial because  $\Delta \mathbf{r}'(t')$  is itself a function of  $t'$ , which must be taken into account in evaluating the  $\delta$ -function. We do this by expanding the absolute-value function about the laboratory origin to obtain

$$t' = t_{\text{ret}} \approx t - \frac{r}{c} + \frac{1}{c} \mathbf{k} \cdot \mathbf{r}_0 + \frac{1}{c} \mathbf{k} \cdot \Delta \mathbf{r}'(t'), \quad (10)$$

where  $\mathbf{k}$  is the wavevector in the observation direction. Equation (10) is still self-consistent, but we can accommodate this to first order in  $v/c$  by substituting  $t_{\text{ret}}$  for  $t'$  in  $\Delta \mathbf{r}'(t')$ . We then obtain

$$A(\mathbf{r}, t) = \frac{q}{rc} \left( \frac{d\Delta \mathbf{r}'(t')}{dt'} \right)_{t=t_{\text{ret}}}, \quad (11)$$

where  $r = |\mathbf{r} - \mathbf{r}_0| \approx |\mathbf{r}|$ . In the radiation zone  $E_{\text{ff}}$  is given by

$$E_{\text{ff}} = -\frac{1}{c} \frac{\partial A(\mathbf{r}, t)}{\partial t} - \nabla \phi(\mathbf{r}, t) \quad (12a)$$

$$= -\frac{1}{c} \frac{\partial A_{\perp}(\mathbf{r}, t)}{\partial t}, \quad (12b)$$

since the longitudinal radiation originating from the scalar potential cancels the longitudinal component originating from  $A$ .

Assembling all equations and performing the indicated differentiations leads to the final result

$$E_{\text{ff}} = (\tilde{I} - \hat{k}\hat{k}) \cdot \left[ \frac{k^2 q}{r} \Delta \mathbf{r}_1 e^{i\mathbf{k}\mathbf{r} - i\mathbf{k}\cdot\mathbf{r}_0 - i\omega t} - 2i \frac{k^2 q}{r} \Delta \mathbf{r}_1 (\mathbf{k} \cdot \Delta \mathbf{r}_1) e^{i2\mathbf{k}\mathbf{r} - i2\mathbf{k}\cdot\mathbf{r}_0 - i2\omega t} + 4 \frac{k^2 q}{r} \Delta \mathbf{r}_2 e^{i2\mathbf{k}\mathbf{r} - i2\mathbf{k}\cdot\mathbf{r}_0 - i2\omega t} \right], \quad (13)$$

where  $\tilde{I}$  is the unit tensor and the prefactor  $(\tilde{I} - \hat{k}\hat{k})$  eliminates the longitudinal component of the following expression. The first term in brackets in Eq. (13) is the linear response, the second the magnetic dipole/electric quadrupole (MD/EQ) contribution that arises from the retarded time, and the final term describes the direct second-harmonic radiation from  $q$ .

Before considering the application, we note the following points about Eqs. (8) and (13) that the above derivation demonstrates. First, the MD/EQ term is equivalent to fre-

quency- or phase-modulation. The physics behind this nonlinearity is that it takes slightly longer for the radiation from  $q$  to arrive at the observer when  $q$  is farther away than when it is nearer, and given the fact that  $c$  is finite, the nonlinearity follows. Second, both the MD/EQ and gradient terms in Eq. (8) are written entirely in terms of the linear response, so they provide no new information about a sample. That the field-gradient provides no new information is not surprising, because if the field has a gradient along the direction of motion, then it will be slightly stronger at one limit of the motion than another. This leads to an anharmonicity indistinguishable from that of an anharmonic restoring force. This can be seen directly in Eq. (8c).

We now consider step 4, combining the radiation from the different charges  $q$ . As a result of the phase dependence of the driving field, the linear and nonlinear terms contain the phase factors  $e^{i(\mathbf{k}_0 - \mathbf{k})\cdot\mathbf{r}_0}$  and  $e^{i(2\mathbf{k}_0 - 2\mathbf{k})\cdot\mathbf{r}_0}$ , respectively. These are the only places that the location  $\mathbf{r}_0$  appears in  $E_{\text{ff}}$ . In the linear case, where  $|\mathbf{k}| = |\mathbf{k}_0|$ , the sum over bonds causes destructive interference except in specific directions, for example in the specular-reflectance direction in the Ewald–Oseen extinction theorem. This is the familiar result obtained from diffraction theory.

In the nonlinear case generally  $|2\mathbf{k}| \neq 2|\mathbf{k}_0|$  as a result of the frequency dependence  $n(\omega)$  of the refractive index  $n$ , so additional considerations apply. This is particularly important for disordered transparent materials. First, in this case the only possible finite contribution occurs in the forward direction, which simplifies the remaining calculations because we need consider only this direction. Second, it means that the contributing region has a finite depth before destructive interference dominates. If in fact the two indices were closely matched, as for example in a green laser pointer, then this region would be so large that we would need an additional equation to describe energy removal as a function of depth from the beam. However, for most materials this region is relatively short, which allows us to make a second simplification. If we suppose that  $|2\mathbf{k}| = 2|\mathbf{k}_0|$  over this finite region then  $\mathbf{r}_0$  drops out completely, which means that  $q$  gives the same contribution no matter where it is located in the region. In this case we can put all bonds at a single site, calculating the response by averaging the single bond over all directions and multiplying the result by the number of bonds in the volume.

**3 Application** We consider the configuration of Figliozzi et al., which consists of a region of crystalline Si spheres (*nc*-Si) of dimension of the order of 5 nm embedded in an amorphous SiO<sub>2</sub> matrix and illuminated with a Gaussian beam of radius  $r$ . A detailed treatment of this configuration has been given by Brudny et al. [11] and Mochán et al. [12] in the macroscopic “dipolium” approximation, where the *nc*-Si spheres and the amorphous host are described by isotropic dielectric functions, and where SHG is obtained by calculating the overall dipole of

the sphere using macroscopic equations. In our case we suppose that the beam propagates in the positive  $\hat{z}$  direction and is polarized along  $\hat{x}$ . From the result of step 4 discussed above we therefore need consider only the observation direction  $\hat{z}$ . We do the averaging over directions by assuming that  $\hat{b} = \hat{x} \sin \theta \cos \phi + \hat{y} \sin \theta \sin \phi + \hat{z} \cos \theta$ , then integrating over the entire solid angle. It is straightforward to show that the MD/EQ and anharmonic component in the direction of propagation are identically zero.

However, the gradient component perpendicular to the propagation direction is not zero, and gives rise to a finite contribution

$$\mathbf{E}_{\text{ff}} = \frac{4k^2 q}{15r} \left[ \hat{x}(3\alpha^2 + 4\alpha\beta + 8\beta^2) \frac{\partial \mathbf{E}}{\partial x} + \hat{y}(\alpha - \beta)^2 \frac{\partial \mathbf{E}}{\partial y} \right] e^{i2kr - i2\omega t}, \quad (14a)$$

where

$$\alpha = \frac{qE}{\kappa_{\text{vl}} - m\omega^2}; \quad \beta = \frac{qE}{\kappa_{\text{vl}} - m\omega^2}. \quad (14b, c)$$

From Eq. (14a) three insights are available. First, the longitudinal gradient contributes nothing, in contrast to the result of bulk response of planar Si [9]. Second, transverse field gradients are required to produce SHG from amorphous media. Third, SHG is described entirely in terms of the linear-optical properties of the material and beam geometry, so we learn nothing new about the material from this measurement.

The beam radius affects the overall intensity as the driving field goes down as  $1/r$ , where  $r$  is the beam radius. Given that  $\alpha$ ,  $\beta$ , and the gradient operation are all inversely proportional to  $r$ , it follows that the intensity from the amorphous host decreases as  $1/r^8$ .

We next consider the spherical inclusions. Here, we need to invoke step 1, because the field is nonuniform in the vicinity of the sphere. Since the diameter of the spheres is much smaller than the wavelength  $\lambda$ , we can treat the problem by electrostatics. Because the dielectric function of the sphere is much larger than that of the glass host, the field is essentially screened out of the interior of the sphere. Therefore, the interior makes a negligible contribution to the overall SHG signal.

We now consider the bonding arrangement. Because the sphere is a single crystal and the host surrounds the sphere, the Si–O bonds form an ordered array at the interface. However, because the SiO<sub>2</sub> host is outside the sphere, the Si–O bonds point in opposite directions on opposite sides of the sphere. If the driving field were uniform the radiation from a given bond would therefore be 180° out of phase with its twin on the other side, and their signals will cancel. Hence to lowest order the interface contributes nothing.

However, with the transverse gradient that results from focusing, the field is not uniform, but has different values on opposite sides of the sphere. We can express this as

$$\mathbf{E} \approx \mathbf{E}_0 + a\hat{\rho} \cdot \nabla \mathbf{E}, \quad (15)$$

where  $\mathbf{E}_0$  is the mean value,  $a$  is the radius of the sphere, and  $\hat{\rho}$  is the radial unit vector. Thus the signal no longer cancels, but a response is generated according to the local value of the field gradient.

However, we need to consider the field gradient more carefully. Because the gradient points in opposite directions on opposite sides of the beam, we still get a cancellation in the forward direction. However, upon returning to step 4 we realize that we can pick a propagation direction where the 180° phase difference on the two sides of the beam can be cancelled on average by a 180° difference in path length. Since this can be done two ways, there exist two directions where the intensity does not cancel, with the directions depending on the size of the beam. This gives rise to a two-lobed pattern seen experimentally [10].

Finally, the replacement of the gradient operation on the field with the radius of the sphere reduces the dimensional dependence of the SHG from the spheres to  $1/r^6 a^2$ . Thus by defocusing the drive beam, the signal from the amorphous host decreases faster than that of the nc-Si inclusions, reaching a point where only the nc-Si signal is detected, also in agreement with experiment.

## References

- [1] M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1975).
- [2] P. P. Ewald, Ph.D. thesis, Munich (1912); *Ann. Phys. (Leipzig)* **49**, 1 (1916).
- [3] C. W. Oseen, *Ann. Phys. (Leipzig)* **48**, 1 (1915).
- [4] See, for example: Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984) and references therein.
- [5] G. D. Powell, J.-F. T. Wang, and D. E. Aspnes, *Phys. Rev. B* **65**, 205320 (2002).
- [6] J.-F. T. Wang, G. D. Powell, R. S. Johnson, G. Lucovsky, and D. E. Aspnes, *J. Vac. Sci. Technol. B* **20**, 1699 (2002).
- [7] J.-K. Hansen, H.-J. Peng, and D. E. Aspnes, *J. Vac. Sci. Technol. B* **21**, 1798 (2003).
- [8] H.-J. Peng, E. J. Adles, J.-F. T. Wang, and D. E. Aspnes, *Phys. Rev. B* **72**, 205203 (2005).
- [9] H.-J. Peng and D. E. Aspnes, *Phys. Rev. B* **70**, 165312 (2004).
- [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] V. L. Brudny, B. S. Mendoza, and W. L. Mochán, *Phys. Rev. B* **62**, 11152 (2000).
- [12] W. L. Mochán, J. A. Maytorena, B. S. Mendoza, and V. L. Brudny, *Phys. Rev. B* **68**, 085318 (2003).