Review of hyper–Rayleigh and second-harmonic scattering in minerals and other inorganic solids

DAVID B. HOLLIS
School of Materials, Division of Ceramics, Glasses and Polymers, Sheffield University, Northumberland Road, Sheffield S10 2TZ, England

ABSTRACT

Recent studies of light scattering at the second harmonic in solids have shown erroneous conclusions and confusion over the mechanism of scattering of radiation. This article presents a simple outline of the theory and the experiments necessary to observe particular mechanisms. Consequences of the presence or absence of particular scattering mechanisms for structural determinations of solids are then considered. Specific examples of scattering in inorganic crystals are reviewed, including perovskites, oxides, fluorides, ammonium halides, micas, feldspars, silicate minerals, and films on surfaces.

INTRODUCTION

Recently, there has been confusion in the literature between hyper-Rayleigh scattering (HRS) and second-harmonic generation (SHG). Further, several workers have used these effects to study inorganic solids, including minerals and have drawn erroneous conclusions from their results.

Essentially there are two main processes for scattering photons at twice the excitation frequency. The first is a spontaneous, isotropic, dephased re-emission of one photon after two excitation photons have been absorbed. This process of absorption and re-emission depends upon the parts of the dipole induced in the medium by a nonlinear interaction with the incident field. The second process arises through an interaction between the electric field and its spatial derivative. Emission by this process is highly anisotropic, coherent, and phase-coupled to the excitation and is therefore subject to phase-matching effects between the electric waves associated with the process.

The former process is called hyper–Rayleigh scattering. The latter process is called second-harmonic generation. Some confusion seems to exist, however, concerning the differences between these effects, and their relevance to the study of structures of materials. This short review is intended to clarify the differences between the two processes and their mechanisms and to draw attention to experimentally observable phenomena that help distinguish between them.

HYPER–RAYLEIGH SCATTERING (HRS)

Hyper–Rayleigh scattering is a spontaneous scattering process analogous to Rayleigh scattering (Long, 1971). To understand its origin, consider an incident light field producing induced dipoles in a material, such that

\[ P = P(0) + \alpha x E + \beta / 2 x EE + \cdots, \]

where \( P(0) \) is the static dipole present, \( \alpha \) is the first polarizability term, \( \beta \) is the second polarizability term, \( E \) is the strength of the incident electric field, and the dots between the polarizability and field terms indicate tensor products.

If the light field is sinusoidal (as is usually assumed), then

\[ E = E_0 \sin \omega_0 t. \]

Further, the polarizability terms of the material are modulated by the atomic motions in the material:

\[ \alpha = \alpha_0 + \alpha \sin \omega_m t + \cdots, \]

where higher-order terms are normally much smaller. Here, \( \omega_m \) is a normal vibrational or rotational frequency of the system \( \omega_0 \) is the frequency of the incident radiation. \( E_0 \) is the amplitude of the electric vector associated with the incident light field. The terms \( \alpha_0, \alpha_1, \alpha_2, \text{etc.} \) are the components of the total polarizability \( \alpha \) contributed by each of the harmonics of \( \omega_m \), the normal-mode frequency of the \( n \)th atomic motion. The effect of the atomic motions and the exciting field on the total dipole in the solid will thus be to create terms of the type

\[ P(1) = \alpha_0 E_0 \sin \omega_m t \sin \omega_0 t, \]

which gives rise to Rayleigh scattering terms of the form

\[ \alpha_0 E \sin \omega_0 t, \]

and Raman scattering terms of the form

\[ \alpha_0 E \sin(\omega_L \pm \omega_m)t. \]

There will also be terms of the type

\[ P(2) = \beta_0 EE \sin^2 \omega_0 t \sin \omega_m t. \]

This can be written as

\[ \beta_0 E \sin(2\omega_0 t + \omega_m t), \]
which gives rise to the hyper-Rayleigh scattering term,
\[ \frac{1}{2} \beta_{0}: EE \sin 2\omega_m t, \tag{8} \]
and the hyper-Raman scattering term,
\[ \frac{1}{2} \beta_{0}: EE \sin (2\omega_1 \pm \omega_m) t. \tag{9} \]

The Raman-type terms appear in optical spectra as features to either side of, and with much weaker scattering than, the corresponding Rayleigh terms. For the present purpose, the concern is with the Rayleigh-type terms and in particular with those that cause scattering at twice the incident frequency—that is, the hyper-Rayleigh term:
\[ P(2) = \frac{1}{2} \beta_{0}: EE \sin 2\omega_1 t. \tag{10} \]

One important consequence of the tensor form of the interaction between the incident electrical field and the polarizabilities is that for centrosymmetric systems, hyper-Rayleigh scattering vanishes (Long, 1971; Chemla, 1980). Vogt and Neumann (1977) have verified this experimentally. However, disordering will cause the hyper-Rayleigh peak to appear, as demonstrated by Vogt and Neumann (1978). A further important consequence arises from the spontaneous form of the scattering. In this process, an electron is excited to a (possibly “virtual”) level and then falls back to its ground state spontaneously. All coherence with the incident radiation is lost, and re-radiated energy will be isotropic and dephased. In hyper-Rayleigh scattering, therefore, it is not possible to get the phase-matching effects associated with second-harmonic generation. The total final intensity is thus proportional to the sum of the squares of all the induced hyper-Rayleigh dipole terms (rather than the square of the sum, as in the coherent second-harmonic-generation process). Hyper-Rayleigh scattering is therefore likely to be weaker than second-harmonic scattering.

To sum up, hyper-Rayleigh scattering is a weak, non-coherent, spontaneous process, which originates from dipoles induced by nonlinear interaction between the medium and the exciting field. It cannot arise in centrosymmetric materials.

A weak octupole effect arising from the nonlinear polarizations also contributes to the hyper-Rayleigh intensity (Chemla, 1980, p. 1199). This effect also disappears in centrosymmetric systems.

**SECOND-HARMONIC GENERATION (SHG)**

Second-harmonic generation arises from two fields that are coupled to each other via a nonlinear term. The fields are associated with the exciting wave and its spatial derivative and with the re-radiated (second-harmonic) wave. In this case, the phases of the several fields are locked, and under the correct circumstances, when the waves are all in phase as they leave the medium, the electric fields of all the induced dipoles may be directly added to give the total re-radiated field. The radiation at the second harmonic is then much more intense than for hyper-Rayleigh scattering. The direct addition of the electric fields of all the induced dipoles is known as the phase-matching condition and is usually only fulfilled when the angles between the excitation beam, second-harmonic beam, and the axes of the induced dipoles are within a few degrees of specified angles for any given compound (Chemla, 1980, p. 1203–1205). To obtain an explanation of second-harmonic generation, one needs to link the time derivatives of the field and the induced nonlinear dipoles and then convert, for the given frequencies and propagation velocities, to spatial distributions. Chemla (1980, p. 1204–1205) obtained an explanation of SHG by considering Maxwell’s equations and writing the electric induction in terms of the usual linear polarization \( P(\omega) = \mu_0 \chi^{(1)} E(\omega) \), plus higher (nonlinear) terms involving the coupling of fields at the incident and second-harmonic frequencies. He obtained a form of the Maxwell equations containing on its left-hand side the usual equation for propagation of a free wave at excitation frequency \( \delta \) and on its right-hand side a driving term,
\[ -\mu_0 \partial^2 P_{NL}/\partial t^2_{NL}. \tag{11} \]

This term accounts for the nonlinear effects.

The second-harmonic generation arises from
\[ P_{NL}(2\omega) = \chi^{(2)}:E(\omega)E(\omega) + \cdots. \tag{12} \]
Here, \( \chi^{(2)} \) is the second susceptibility term of the medium. Since \( \chi^{(2)} \) is usually much larger than \( \chi^{(1)} \), etc., only the \( \chi^{(2)} \) part is considered in Equation 12 in most analyses of \( P_{NL} \).

Since, for a coherent plane-wave source,
\[ E(r,t) = E_0(\omega) \exp[i(\omega t - kr)], \tag{13} \]
Maxwell’s equation can be reduced from a temporal form to a spatial form:
\[ \partial E/\partial r + \alpha E/2 = i\mu_0 \omega \rho_0 P_{NL}/2n \tag{14} \]
where \( \mu_0 \) is the susceptibility of free space, and \( n \) is the refractive index. Here, \( \alpha \), the linear absorption coefficient, is given by
\[ \alpha = \mu_0 \sigma/2n, \tag{15} \]
where \( \sigma \) is the absorption cross section at \( \omega \).

The incident field, \( E \), and its spatial derivative have now been linked by the nonlinear field-coupling term \( P_{NL} \). For most second-harmonic-mixing crystals, the absorption loss, \( \alpha \), is small. Hence, approximately, one can write
\[ \partial E_1/\partial r = ikE_2 E_2 \exp(i\Delta kr) \tag{16} \]
\[ \partial E_2/\partial r = ikE_2^2 \exp(-i\Delta kr). \tag{17} \]
This phase mismatch is often expressed in terms of a coherence length, \( l_c \), where
\[ \Delta kr = \pi r/l_c. \tag{18} \]
Even if the phase mismatch is not eliminated, the second-harmonic generation is still present.

For this situation, Chemla (1980, p. 1206) has given an approximate solution:
\[ I_2(r) = \frac{2\omega_2^2 d_2^2 r^2}{\varepsilon_0 m_2 n_2^2 c^3} F_2 \left( \frac{\sin(\Delta kr/2)}{\Delta kr/2} \right)^2. \tag{19} \]
In this expression, the symbols have their usual meanings: \( I_0 \) = intensity of the incident light; \( I_{2\omega} \) = intensity at the second harmonic; \( \omega \) = frequency of the incident radiation; \( \epsilon_0 \) = permittivity of free space; \( n_r \) = refractive index at frequency \( \omega \); \( c \) = velocity of light; \( d \) = nonlinear coupling coefficient (\( d \) is equal to one-half of the second-order susceptibility \( x^{(2)} \)); \( \Delta K \) = momentum mismatch between incident and second-harmonic photons; \( r \) = distance over which the waves have propagated through the dispersive medium. Equation 19 is a function that oscillates with a characteristic pattern as the crystal is moved, as seen by Terhune et al. (1962). It is a function of the thickness \( r \) of the material and is proportional to the squared product of the nonlinear susceptibility, the fundamental intensity, and the coherence length.

A strong maximum at \( \Delta K = 0 \) represents a situation where a large proportion of radiation at frequency \( \omega \) is converted to radiation at \( 2\omega \). In this case, a more rigorous solution applies:

\[
I_0 = L_{\omega c(-0)} = \frac{1}{2} \mu_0 n_\omega c |E_0|^2, \\
I_{2\omega}(r) = I_0 \tanh^2 (KEc r), \\
I(r) = I_0 \operatorname{sech}^2 (KEc r).
\]

Beam focusing and double-refraction walk off, as well as nonzero absorption by the crystal, prevent the theoretical total conversion from being reached. Change of refractive indices, \( n_r \), because of heating by absorption of radiation, limits the total power throughput of the system. However, in excess of 50% conversion for certain phase-matched crystals and configurations is now common.

To obtain phase matching,

\[
\Delta K = 0.
\]

That is, the two refractive indices cancel:

\[
n_{2\omega} = n_\omega.
\]

In practice, the most convenient technique to fulfill this condition is to use the birefringence of the crystal to overcome the dispersion. Phase-matching by this method in uniaxial and biaxial crystals is now used in the electro-optical industry. Chemla (1980) has discussed phase-matching by this and other techniques in more detail.

From the foregoing discussion, several points arise: first, there cannot be any second-harmonic conversion by this mechanism in centrosymmetric crystals. Second, phase-matching is necessary for strong second-harmonic signals to be obtained, but it is not observed in most materials because of incorrect matching of dispersion to the birefringence.

WEAK SCATTERING EFFECTS AT THE SECOND HARMONIC

Apparent second-harmonic generation may occur in stressed or deformed systems, and a weak second-harmonic signal can appear even in centrosymmetric systems. These effects are due to four mechanisms:

1. Quadrupole scattering, the intensity of which depends on the magnitude of the exciting field and its spatial gradient. Kiellich (1968a, 1968b) and Meisner (1979) have given a theoretical analysis, as have Denisov et al. (1980a, 1980b), in their study of rutile. Bjorkholm and Siegman (1967) and Terhune et al. (1962) have studied this mechanism in calcite. Orman and Vogt (1976) have shown that, since the nonlinear susceptibility involved is a tensor of fourth rank, this form of SHG may give a small signal in centrosymmetric crystals. The quadrupole and octupole electronic contributions can both cause weak SHG (and HRS) signals in centrosymmetric crystals (Kiellich, 1968a, 1968b; Meisner, 1979; Califano et al., 1981).

2. Deformations and stresses in materials. These give rise to an internal field that gives a nonlinear polarization.

3. Surface phenomena, of which only three examples are known to me, as described by Baranov and Bobovich (1982), Murphy et al. (1982), and Girling et al. (1986).

4. Lowering of the symmetry of the crystal by an externally applied electric field, as in d.c. field-induced Raman scattering.

It should be noted that all four mechanisms are coherent with the exciting radiation and that some of them exist in both centrosymmetric and noncentrosymmetric systems.

EXPERIMENTAL RATIONALE AND METHODS

To distinguish between HRS and SHG from any of the above effects in crystals, it is possible to use the coherence properties of SHG and look for the disappearance (or otherwise) of the HRS (isotropic, noncoherent) part of the scattered radiation. Strong SHG may be obtained when the beam propagates in the direction required for exact matching of the birefringence to the dispersion in the medium (Terhune et al., 1962; Hobden, 1967; Chemla, 1980). Terhune-Maker-Savage fringes (Terhune et al., 1962) can be produced by rotating the crystal such that the beams are scanned through several coherence lengths (this paper, Eq. 19).

For powdered samples, a different, less rigorous, but usually sufficiently accurate approach can be employed. Since HRS is totally forbidden in centrosymmetric crystals, but SHG is not, centrosymmetry can only be uniquely determined by the disappearance of the HRS (isotropic) component of the scattering. As will be described below, however, calibration of the experimental system with a known standard powder—usually \( \alpha \)-quartz—may be used to distinguish the centrosymmetric, weak SHG scatterers from the truly noncentrosymmetric HRS and/or strong SHG scatterers.

When only centrosymmetry is to be determined, experimental observations are carried out with powdered samples to overcome phase-matching problems encountered when SHG, rather than HRS, is the dominant mechanism. Otherwise, oriented crystals are used. The calibrator normally employed is powdered \( \alpha \)-quartz. For centrosymmetry to be confirmed, the ratio of scattering intensity from the powdered test sample to that of powdered \( \alpha \)-quartz should be less than \( 10^{-3} \). If the test spec-
imen is to be classed as noncentrosymmetric, then the ratio should be greater than $10^{-2}$. It should also be stressed that even a small quantity of noncentrosymmetric material in a nominally centrosymmetric sample can lead to erroneous conclusions, as shown in Appendix 1.

Since the scattering effects employed here are weak, sensitive techniques have to be employed. Typically, if a high-power pulsed laser is employed, then the Rayleigh scattering intensity is about $10^{-3}$ that of the excitation intensity at the sample, and the hyper-Rayleigh scattering is about $10^{-4}$ that of the excitation at the sample. The SHG intensity depends strongly on the phase effects discussed in the previous section, but will typically be somewhat stronger than the HRS in a strongly noncentrosymmetric sample. To observe the weak scattering, the scattered radiation is collected, monochromated, and then observed by a time-gated low-noise photomultiplier system. The advantage of pulsed-laser excitation and gating of the photomultiplier is that “noise counts” during the “off” period of the laser can be “gated” out, and the high excitation and scattering intensities during the “gated on” period greatly improve the signal to noise ratio at the detector.

**HRS and SHG in solids**

Reviews of SHG in solids have been given by Hulme (1973) and Chemla (1980). Reviews of HRS in solids have been given by Vogt (1981), Vogt (1982), and Hollis (1983).

Some more recent examples are considered here.

**Space-group assignment**

Zilczer and Loiacono (1982) studied a series of potassium feldspars. The signals obtained clearly suggested that one sample was noncentrosymmetric and that the other was centrosymmetric. However, the second sample was a type of feldspar that is known to exhibit pyroelectric activity. Taylor (1933) and Wooster (1973) have shown that pyroelectricity arises from microscopic domains of noncentrosymmetric structure within the specimen. One wonders if a weak SHG or HRS signal would have been observed had Zilczer and Loiacono been able to detect SHG or HRS an order of magnitude weaker than their stated experimental limit.

Bish et al. (1979) investigated micas. The ideal micas are centrosymmetric (Tak6uchi, 1965). As expected, Bish et al. were unable to detect SHG or HRS from these micas, but they did obtain a signal from the disordered (and therefore noncentrosymmetric) micas such as zinnwaldite. The situation is less clear-cut for the brittle micas including margarite and ephesite. These have either a centrosymmetric unit cell with disordered substitution (Takéchi, 1965), or a noncentrosymmetric unit cell with ordered substitution (Guggenheim and Bailey, 1975, 1977). In either case, one would expect SHG or HRS. The important detail—its strength relative to that of powdered α-quartz—is not given by Bish et al., so no conclusion can be drawn for these micas on the basis of their results.

Loiacono et al. (1982) have examined several minerals whose space groups were not certain. Some give strong signals and are therefore noncentrosymmetric. However, others give signals that are between $10^{-2}$ and $10^{-3}$ of the signal of α-quartz, so that their space groups remain uncertain.

**“SHG” from noncentrosymmetric domains**

The recent publications by Inoue and Akamoto (1983) on BaTiO$_3$, Uwe and Vogt (1984) on KTaO$_3$, Denisov et al. (1983), on SrTiO$_3$, and Hollis (1983) on KZnF$_3$ indicate strong “hyper-Rayleigh” scattering in nominally cubic perovskite compounds. A review by Bruce and Cowley (1980) indicated that the hyper-Rayleigh activity may be strong because of mixing between the hyper-Rayleigh peak and a soft mode. At low enough temperatures, this soft mode leads to a permanent displacive transition to another, noncubic phase. However, these crystals were all studied well above the temperature at which the soft mode would coalesce with the hyper-Rayleigh peak and give rise to the phase change.

The point that this phenomenon is a form of HRS, not SHG, is further emphasized in a study by Shin et al. (1987) of the central mode in the hydrogen-bonded ferroelectric CsH$_2$PO$_4$. Again, the conclusion was that domains with noncentrosymmetric symmetry were formed in the nominally centrosymmetric crystal, but that the domains had short life times.

Hastings et al. (1978) introduced defects into SrTiO$_3$ by reducing crystals with hydrogen. An increase of defect density by 200 times gave an increase of “hyper-Rayleigh” scattering of only three times. Defects cannot therefore be considered as a prime cause of the peak at the second harmonic. Similar results and conclusions were arrived at for cubic zirconia by Shin and Ishigame (1986).

A study by Lehner et al. (1982) of KZnF$_3$ shows that the positions of the F atoms have already become unstable at temperatures well above that of any possible phase transition. They commented that it is not possible to assign the instability to an incipient displacive transition of the Zn atoms or to an incipient disordering transition of the ZnF$_6$ octahedra. However, Hollis (1983) gave spectra that show the strong “hyper-Rayleigh” peak for KZnF$_3$. Furthermore, the incipient ferroelectric behavior of KTaO$_3$ is well known, as are related electrical phenomena in SrTiO$_3$ and BaTiO$_3$. Therefore, transient domains of noncubic symmetry must be present in these structures.

Since Ortmann and Vogt (1976) have shown that defects, rotational disordering, and noncentrosymmetric structures can produce weak SHG and Vogt and Neumann (1978) have also shown that disordering can produce HRS, and since Long (1971) has shown that noncentrosymmetric structures can produce strong HRS, it is not possible to say which is the dominant mechanism in these experiments. Information on relative scattering...
intensities and phase coherency between excitation radiation and scattered radiation is required.

Hall (1981) and Al-Rubaiey (1981) have studied ammonium halides. They showed, by a series of experiments carried out while crystals were being heated or cooled, that hysteresis in the intensity of the HRS (or SHG) peak exists in these compounds. They also found an exponential fall in intensity with temperature similar to that noted for KTaO, by Uwe and Vogt (1984; Vogt and Uwe, 1984). They concluded that the hysteresis arose from the slow rate of reordering at the phase transition, such that domains of the previous phase remained frozen into the new phase. That is, the thermal activation energies associated with reordering the structures are high compared to the average thermal energies of the compounds at the phase transitions. Hall and Al-Rubaiey were able to monitor the reordering processes until as few as one per thousand of the unit cells was noncubic.

If the observed scattering is SHG, then it should persist weakly in the centrosymmetric phase. Hall and Al-Rubaiey were able to show its complete disappearance, however, and hence, for these crystals, the mechanism of scattering was definitely HRS, not SHG. Furthermore, since coupling between acoustic phonons, or entropy fluctuations (tensor of even rank and parity), and the hyperpolarizability (tensor of third rank and odd parity) is forbidden (Al-Rubaiey, 1981; Hall, 1981), the scattering observed was true HRS and not a kind of hyper-Brillouin central peak of the type described by Hayes and Loudon (1978). It arose from static “domains” of noncubic symmetry.

Since the temperature dependence of the scattering intensity was single exponential, there appeared to be only one mechanism involved, rather than several. Hall and Al-Rubaiey believed that the mechanism was disordering caused by the development of vibrational motions into reorientational motions of ammonium ions, by overcoming an activation energy barrier. Such a system would display a temperature-sensitive Arrhenius-type single exponential behavior. Also, at higher temperatures, permanent displacements in all the unit cells of the structure would lead to the formation of a new phase. Both of these effects were indeed observed by Hall and Al-Rubaiey.

CONCLUSION

In conclusion, it can be said that observation of scattering at the second harmonic is a powerful, though as yet little explored, tool for the study of phase transitions, phase-change kinetics, and centrosymmetry in solids. Care has to be taken to avoid erroneous conclusions about weak scatterers. Weak SHG may be confused with strong HRS from domains of noncentrosymmetry. Calibration of the scattering strength at the second harmonic of the incident beam, using powdered α-quartz, should therefore be considered an absolute necessity. All test samples should then be compared, in a powdered form, to the results from the standard.

If sizable crystals are available, the coherent, anisotropic nature of SHG, as compared to the isotropic, noncoherent behavior of HRS, can be used to determine whether the system is completely centrosymmetric and free of small domains of strained, distorted, defector vacancy-containing material or other noncentrosymmetric phases.

REFERENCES CITED


Kiely, S. (1968a) Molecular interactions in optically induced non-linear-
HOLLIS: HYPER-RAYLEIGH AND SECOND-HARMONIC SCATTERING IN MINERALS


Manuscript received July 28, 1986
Manuscript accepted March 4, 1988

APPENDIX I. EFFECT OF STRONGLY SCATTERING IMPURITIES UPON THE SHG SCATTERING INTENSITY VALUE OBTAINED FOR A WEAK SHG SCATTERER

Let \( \eta_1, \eta_2, \eta_3, \cdots \) be the ratio of the scattering strengths of components no. 1, no. 2, \cdots of the mixture being tested to the scattering strength of powdered \( \alpha \)-quartz. Let \( x_1, x_2, x_3, \cdots \) be the percentages by weight of the components in the mixture.

The total scattering strength relative to the same weight of \( \alpha \)-quartz should be \( R \). In actual fact it will be

\[
\eta_1 x_1 + \eta_2 x_2 + \eta_3 x_3 + \cdots
\]

Hence, the ratio of the value obtained to the correct value will be

\[
R = x_1 + (\eta_1/\eta_2)x_2 + (\eta_1/\eta_3)x_3 + \cdots
\]

If \( x_2 \) or any other component is a strong scatterer and \( x_1 \) is a weak scatterer, then typically \( \eta_2 \approx 100 \eta_1 \). Hence, the ratio of the observed value to the correct one will be approximately

\[
R = x_1 + 100x_2 + \cdots
\]

Even if \( x_2 \) is only \( \approx 10\% \) and \( x_1 = 99\% \), the ratio will be

\[
R = 0.9 + 100x(10/100) = 11,
\]

which may be sufficiently large to cause the wrong choice of space group. If \( x_2 = 1\% \) and \( x_1 = 99\% \), the ratio will be

\[
R = 0.99 + 100x(1/100) = 1.99 \approx 2.
\]

Although this ratio is still appreciably too large, it should not give rise to erroneous assignments of noncentrosymmetric space groups.