Cambridge University Press
978-0-521-87701-5 - Introduction to Nonlinear Optics
Geoffrey New
Excerpt
More information

1 Introduction

1.1 Nonlinearity in physics

The response of real physical systems is never exactly proportional to stimulus, which is a way of saying that we live in an inherently nonlinear world. The deviation from linearity may be very slight, especially if the stimulus is weak; the assumption of linearity will then be an excellent approximation and probably the only route to an analytical solution. But that does not change the fact that linearity is an idealisation.

A simple example of the linear approximation occurs in elementary mechanics where one assumes that restoring force is proportional to displacement from equilibrium (Hooke's law). This leads to the equation of motion

$$m\ddot{x} = -s_1 x \tag{1.1}$$

where m is the mass and s_1 is the Hooke's law constant (the stiffness of the system). Equation (1.1) has the simple harmonic motion solution

$$x = A\cos\{\omega_0 t + \phi\} \tag{1.2}$$

where $\omega_0 = \sqrt{s_1/m}$ is the angular frequency of oscillation, and A and ϕ are fixed by the initial conditions. Equation (1.1) is usually an excellent approximation when the amplitude A of oscillations is small. But, since real systems are never perfectly linear, their oscillations will never be precisely (co)sinusoidal, and will contain harmonics of ω_0.

Nonlinearity can be incorporated into Eq. (1.1) by including additional terms on the right-hand side to represent the fact that the restoring force is no longer linear in the displacement. For example, it might be appropriate to write

$$m\ddot{x} = -s_1 x + s_2 x^2 = -(s_1 - s_2 x)x \tag{1.3}$$

where the second form highlights the fact that the nonlinear term makes the stiffness dependent on amplitude. Notice that the inclusion of a term in x^2 makes the stiffness asymmetrical in the displacement; if $s_2 > 0$, the net stiffness is lower for positive x and higher for negative x.

Since Eq. (1.3) has no analytical solution, a numerical solution is generally called for. However, the role of the extra term in introducing harmonics can be appreciated by regarding it as a perturbation, with Eq. (1.2) as the zeroth-order solution. To the first order of approximation, the system will be subject to the additional force term

$$s_2 x^2 = s_2 A^2 \cos^2\{\omega_0 t + \phi\} = \tfrac{1}{2} s_2 A^2 [1 + \cos 2\{\omega_0 t + \phi\}], \tag{1.4}$$

Cambridge University Press
978-0-521-87701-5 - Introduction to Nonlinear Optics
Geoffrey New
Excerpt
More information

and this will lead (through Eq. 1.3) to a second harmonic component in the motion. In principle, the process can be repeated to higher orders of approximation, although the terms quickly become negligible if the nonlinearity is weak.

In many situations in life, one seeks to minimise the effects of nonlinearity. Manufacturers of audio equipment are for example keen to advertise the lowest possible figures for ‘harmonic distortion’. But in other circumstances, nonlinearity can be put to good use, and this book is about how it can be exploited to spectacular effect in optical physics.

1.2 The early history of nonlinear optics

In optics, one is interested in the response of atoms and molecules to applied electromagnetic (EM) fields. The interaction of light and matter is of course governed by the Schrödinger equation, which is linear in the wave function but nonlinear in the response of the wave function to perturbations. Despite this, optics proceeded quite successfully for many years on the assumption that the response of optical materials was linear in the applied electric field E. If P is the polarisation of the medium (i.e. the dipole moment per unit volume),[1] one writes

$$P = \varepsilon_0 \chi^{(1)} E \tag{1.5}$$

where $\chi^{(1)}$ is the linear susceptibility. So if $E = A\cos\omega t$, the consequence is that $P = \varepsilon_0 \chi^{(1)} A \cos\omega t$. It also follows from Eq. (1.5) that the electric displacement is

$$D = \varepsilon_0 E + P = \varepsilon_0 (1 + \chi^{(1)}) E = \varepsilon_0 \varepsilon E \tag{1.6}$$

where $\varepsilon = 1 + \chi^{(1)}$ is the relative dielectric constant. As we will see later, ε is the square of the refractive index, $n = \sqrt{1 + \chi^{(1)}}$.

Equation (1.5) served as a good approximation for so long because the electric field strengths that scientists were able to deploy in those early years were far weaker than the fields inside atoms and molecules; the perturbations were therefore very small. It was not until the 1870s that the Rev. John Kerr, a lecturer at the Free Church Training College in Glasgow, UK, demonstrated that the refractive index of a number of solids and liquids is slightly changed by the application of a strong DC field [2]. This phenomenon, now known as the DC Kerr effect,[2] was the first nonlinear optical effect to be observed.

Two decades later, in the 1890s, Friedrich Pockels at the University of Göttingen studied a related process known today as the Pockels effect [3]. The Kerr effect and the Pockels effect differ in two respects. In the Kerr effect, the refractive index change is proportional to the *square* of the applied DC field, whereas in the Pockels effect, the change is directly

---

[1] The word ‘polarisation’ has two meanings in optics. It refers (as here) to the dipole moment per unit volume (in coulombs m^{-2}), and would be represented by the vector **P** in a more rigorous treatment. But the word is also used to describe the *polarisation of light* where it refers to the direction of the fields in the transverse EM wave. The direction (or plane) of polarisation is normally taken to be that of the electric field in the wave.

[2] Sometimes it is just called the Kerr effect, but it must be distinguished from the optical (or AC) Kerr effect.

proportional to the field.[3] Secondly, whereas the Kerr effect is observable in liquids and amorphous solids, the Pockels effect occurs only in crystalline materials that lack a centre of symmetry. This vital distinction and the reason behind it will be discussed in detail later.

1.3 Optical second harmonic generation

There now followed a long gap in the history of nonlinear optics. Further progress had to wait for a source of strong *optical frequency* fields to become available, in other words for the invention of the laser in 1960. With the arrival in the laboratory of the ruby laser, nonlinear optics underwent a second birth, and it has been flourishing ever since. Indeed, since lasers themselves are inherently nonlinear devices, one could even argue that laser physics is itself a compartment within the wider field of nonlinear optics. Traditionally, however, the field of nonlinear optics is taken to exclude lasers themselves, which is fortunate insofar as this book would otherwise be much longer.

The first nonlinear optics experiment of the laser era was performed in 1961 by a team led by the late Peter Franken at the University of Michigan in Ann Arbor [4]. As shown in Fig. 1.1, a ruby laser was focused into a slab of crystalline quartz to discover if the nonlinear response of the medium to the intense optical frequency radiation at 694.3 nm was strong enough to create a detectable second harmonic component at a wavelength of 347.15 nm.[4]

The way to think about this experiment is diagrammed in Fig. 1.2. Consider the response of the electrons in the quartz to the stimulus of the optical frequency electric field of the laser beam written as $E = A\cos\omega t$. The displacement of the electrons creates a dipole moment p per atom,[5] or $P = Np$ per unit volume where N is the atomic number density.

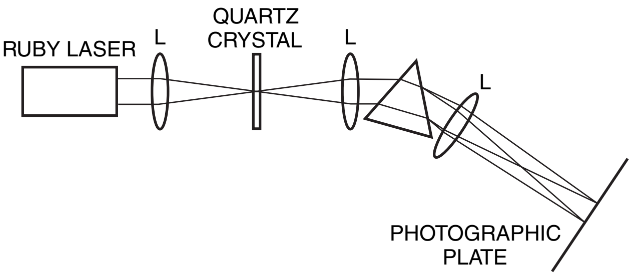

Fig. 1.1 Schematic diagram of the first second harmonic generation experiment by Peter Franken's group at the University of Michigan in 1961 [4].

---

[3] The Pockels effect is also known as the electro-optic effect, or sometimes (confusingly) as the *linear* electro-optic effect because the index change is linearly proportional to the DC field.

[4] The vacuum wavelength of the second harmonic component is of course half that of the fundamental.

[5] One could also write $p = \varepsilon_0 \alpha E$ where α is the atomic polarisibility, in which case $\chi^{(1)} = N\alpha$.

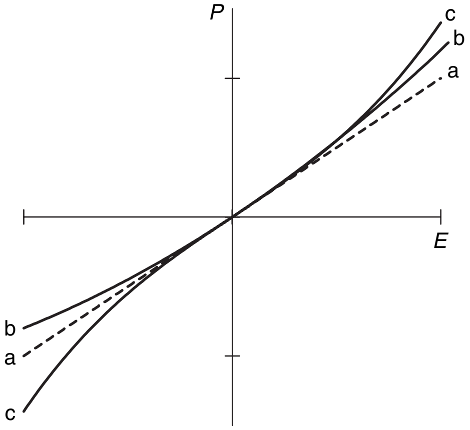

Fig. 1.2 Linear and nonlinear response of polarisation *P* to applied electric field *E*: (a) linear case (dotted), (b) quadratic nonlinearity, (c) cubic nonlinearity.

We allow for the possibility of a nonlinear response by writing

$$P = \varepsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \cdots). \tag{1.7}$$

For $\chi^{(2)} > 0$, the dependence of P on E is represented by curve b in Fig. 1.2, while in the absence of the nonlinear term, Eq. (1.7) reverts to Eq. (1.5), and the linear relationship between E and P represented by the dotted straight line a in the figure is recovered.

If $E = A\cos\omega t$ is substituted into Eq. (1.7), the polarisation now reads

$$P = \varepsilon_0(\underbrace{\chi^{(1)}A\cos\omega t}_{\text{linear term}} + \underbrace{\tfrac{1}{2}\chi^{(2)}A^2(1+\cos 2\omega t)}_{\text{nonlinear term}}). \tag{1.8}$$

The analogy between Eqs. (1.7)–(1.8) and Eqs. (1.3)–(1.4) is obvious. Figure 1.3 shows graphs of the total polarisation (in bold) and the linear term on its own (bold dotted); notice how the waveform is stretched on positive half-cycles and flattened on negative half-cycles, in accordance with curve b of Fig. 1.2. The solid grey line in Fig. 1.3 representing the nonlinear term in Eq. (1.8) clearly contains a DC offset (dotted grey), the significance of which is discussed in Section 1.6, and a second harmonic component, which is what Franken and his team were looking for in 1961. A useful picture is to regard the second harmonic polarisation of the nonlinear medium as an optical frequency antenna. Just as radio and TV signals are broadcast by accelerating charges in the transmitter aerial, so the oscillating nonlinear polarisation in Franken's quartz crystal should radiate energy at twice the frequency of the incident laser field. The key question in 1961 was simply: will the radiation from the second harmonic antenna be strong enough for photons at 347.15 nm to be observed experimentally?

The University of Michigan experiment was a huge success to the extent that *second harmonic generation* (SHG) was detected. The down side was that the harmonic intensity was extremely weak, so weak in fact that the photographic plate reproduced in the 15 August 1961 issue of *Physical Review Letters* [4] appears to be totally blank! In fact, the

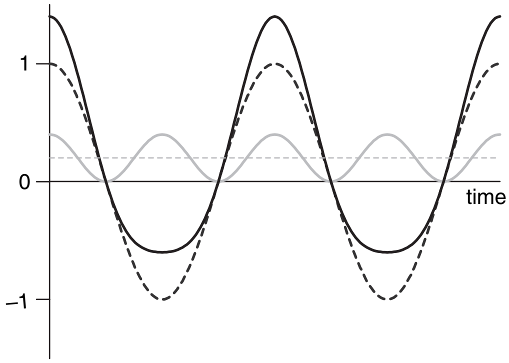

Fig. 1.3 Response (bold line) of a system exhibiting a quadratic nonlinearity to a (co)sinusoidal stimulus (dotted line). The difference between the two (grey line) contains both second harmonic and DC frequency components.

energy conversion efficiency from fundamental to harmonic was about 1 part in 10^8, a clear demonstration that SHG was a real effect although, at this minuscule efficiency, it was clearly a curiosity rather than a practical means of generating ultraviolet (UV) light.

1.4 Phase matching

Why was the efficiency so small? The reason is easy to understand as soon as one includes the spatial dependence of the fields in the equations. Assuming that the ruby laser field is a plane wave propagating in the z-direction, we write

$$E_1 = A_1 \cos\{\omega t - k_1 z\} \tag{1.9}$$

where the angular wave number of the fundamental beam is $k_1 = n_1\omega/c$, and n_1 is the refractive index.[6] The second harmonic term in the polarisation is therefore

$$P_2 = \tfrac{1}{2}\varepsilon_0 \chi^{(2)} A_1^2 \cos\{2\omega t - 2k_1 z\} \tag{1.10}$$

which seems fine at first sight. But now compare the space-time dependence in Eq. (1.10) with that of a freely propagating field at 2ω namely

$$E_2 = A_2 \cos\{2\omega t - k_2 z\} \tag{1.11}$$

where $k_2 = n_2 2\omega/c$ by analogy with k_1. Notice that the arguments of the cosines in Eqs (1.10) and (1.11) are different unless $2k_1 = k_2$, which is only true if the refractive indices n_1 and n_2 are the same. But dispersion ensures that the indices are usually *not* the

---

[6] Here and throughout the book, the default space-time dependence of an optical field is $(\omega t - kz)$, where the parameter k is called the *angular wave number*, by analogy with angular frequency $\omega = 2\pi\nu$. However, many authors use $(kz - \omega t)$ instead of $(\omega t - kz)$. The choice is immaterial under a cosine but, with complex exponentials, it leads to a difference in sign that permeates into many subsequent equations. See Appendix A for more discussion.

Cambridge University Press
978-0-521-87701-5 - Introduction to Nonlinear Optics
Geoffrey New
Excerpt
More information

same. The severity of the mismatch is easily quantified by asking over what distance the cosine terms in Eqs (1.10) and (1.11) get π radians out of step. This distance, known as the *coherence length* for the second harmonic process, is easily shown to be[7]

$$L_{\mathrm{coh}} = \frac{\pi}{|\Delta k|} = \frac{\lambda}{4\,|n_2 - n_1|} \tag{1.12}$$

where $\Delta k = k_2 - 2k_1$ is called the mismatch parameter. For typical optical materials, L_{coh} is a few tens of microns[8] (μm) after which, as we will show in the next chapter, the SHG process goes into reverse, and energy is converted from the harmonic back into the fundamental wave. Hence, only a few microns of the 1 mm quartz sample contributed to the signal in the first SHG experiment, and the rest was redundant.

The phase-matching problem was soon solved. Within a few months of the University of Michigan experiment, researchers at the Ford Motor Company's laboratories, around 30 miles from Ann Arbor in Dearborn, Michigan, had exploited birefringence in a KDP crystal[9] to keep the ω and 2ω waves in step with each other [5]. They did this by finding a particular direction of propagation in KDP for which the refractive index of the ordinary wave at ω was the same as that of the extraordinary wave at 2ω. The process is called *phase matching* or sometimes *birefringent phase matching* (BPM) to distinguish it from other methods of achieving the same end; see Chapter 2. Under phase-matched conditions, the SHG conversion efficiency jumped to tens of percent with careful experimental adjustment, and from that moment on, nonlinear optics stopped being a curiosity and became a practical proposition.

Phase matching is of vital importance in many nonlinear processes. The key principle to grasp is that the direction of energy flow between fundamental and harmonic waves is determined by the relative phase between the nonlinear polarisation and the harmonic field. If the two can be held in step by making the fundamental and harmonic refractive indices the same, the energy will keep flowing in the same direction over a long distance. Otherwise it will cycle backwards and forwards between the fundamental and the harmonic, reversing direction more frequently the more severe the phase mismatch.

1.5 Symmetry considerations

Before moving on, it is worth considering some other key features of the SHG process. First of all, notice that the quadratic term in Eq. (1.6), represented by curve b in Fig. 1.2, implies that the medium responds differently according to the direction of the electric field. A positive field creates a slightly greater (positive) response than that of curve a, whereas a negative field produces a slightly smaller response than in the linear case. This implies that

---

[7] Sadly, there is no generally agreed definition of 'coherence length' in nonlinear optics, and definitions with 1, 2, π, and 2π in the numerator of Eq. (1.12) can be found in different textbooks. The definition in Eq. (1.12) is the one used almost universally by researchers in the 1960s, and I see no reason to change it.

[8] 'Micron' is still widely used for micrometre in colloquial usage.

[9] KDP stands for potassium dihydrogen phosphate.

Cambridge University Press
978-0-521-87701-5 - Introduction to Nonlinear Optics
Geoffrey New
Excerpt
More information

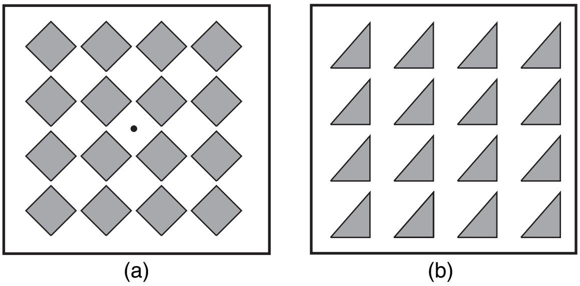

Fig. 1.4 (a) Centrosymmetric pattern and (b) non-centrosymmetric pattern.

SHG (and all other nonlinear processes that depend on $\chi^{(2)}$) can occur only in media that are structurally different in one direction from the opposite direction. Media of this kind are variously said to *lack inversion symmetry*, to *lack a centre of symmetry*, or to be *non-centrosymmetric*. In simple terms, one could say that they have an inherent 'one-wayness', a property that only crystalline materials can possess. Fortunately, many non-centrosymmetric media also exhibit double refraction (birefringence), and so are candidates for birefringent phase matching.

A more detailed discussion of inversion symmetry is given in Chapter 3, but the simple schematic pictures shown in Fig. 1.4 may be helpful at this stage. Imagine you are looking into a crystal along the line of the laser beam. The structure shown on the left possesses inversion symmetry. It can be flipped horizontally ($x \to -x$) or vertically ($y \to -y$) and the pattern is unchanged; the centre of symmetry at the mid-point is marked with a dot. It is therefore impossible in principle for this medium to exhibit the asymmetrical characteristics shown in Fig. 1.3. The structure on the right is different because, in this case, a flip in either direction results in a different pattern. So left is different from right, up is different from down, and the response in Fig. 1.3 is no longer forbidden.

A simple mathematical proof of the principle runs as follows [6]. In one dimension, consider the term $P_x = \varepsilon_0 \chi^{(n)} E_x^n$. Under inversion symmetry, P_x must change sign if E_x changes sign. But, when n is even, E_x^n is unchanged if $E_x \to -E_x$, and it follows that $\chi^{(2)}$, $\chi^{(4)}$, etc. must be zero in this case. No such restriction applies for odd n, so processes involving $\chi^{(3)}$ for example (see Section 1.10) can occur in centrosymmetric media.

1.6 Optical rectification

After the successful demonstration of second harmonic generation, the nonlinear optics bandwagon began to roll, and a multitude of other nonlinear processes were discovered in a 'gold rush' period in the mid-1960s. One such process was *optical rectification*, which has already appeared in the polarisation in Eq. (1.8) above [7]. This is probably the simplest nonlinear optical process both to visualise and to observe. To detect optical rectification, all one has to do is to pass a laser beam through a non-centrosymmetric crystal located

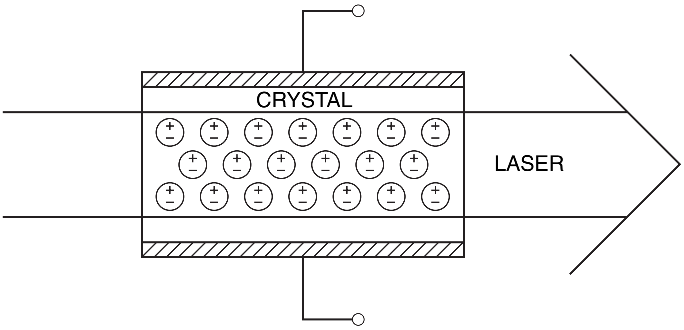

Fig. 1.5 Schematic diagram of optical rectification. Positive and negative charges in the nonlinear medium are displaced, causing a potential difference between the plates and current to flow in an external circuit.

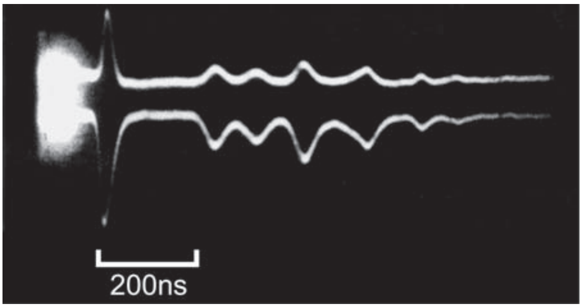

Fig. 1.6 Optical rectification signal (top trace) and laser monitor (lower trace).

between a pair of capacitor plates, as shown in Fig. 1.5. As suggested in the figure, the asymmetric response of the material to the field of the laser displaces the centre of gravity of the positive and negative charges in the medium, creating a DC polarisation in the medium. Dipoles oscillating at the *optical* frequency are of course created in all materials, and these are represented by the linear polarisation of Eq. (1.5). What is shown in Fig. 1.5 is rather the DC offset of Fig. 1.3, and its effect is to induce a potential difference between the plates, which allows a rectified component of the optical frequency field to be detected; see Problem 1.2.

An example of an optical rectification (OR) signal is shown in the dual-beam oscilloscope record of Fig. 1.6, where the upper and lower traces are the OR and the laser monitor signals, respectively. In fact, so accurately does the OR signal track the laser intensity that it is impossible to tell them apart from the record itself.[10] Since the OR process also leaves the laser field essentially unchanged, it was considered as a possible basis for laser power measurement in the early days of the subject. But other technologies prevailed in the end.

Notice that the phase-matching issue does not arise in the case of optical rectification, because the DC term in the polarisation (analogous to the second harmonic contribution in

---

[10] The fact that one trace is the negative of the other is a trivial result of the polarity of the electrical connections.

Cambridge University Press
978-0-521-87701-5 - Introduction to Nonlinear Optics
Geoffrey New
Excerpt
More information

Eq. 1.10) is

$$P_{dc} = \tfrac{1}{2}\varepsilon_0 \chi^{(2)} A_1^2 \tag{1.13}$$

which has no space-time dependence.

1.7 The Pockels effect

The *Pockels effect* [3] relates to the change in the refractive index of a non-centrosymmetric medium when a DC electric field is applied. The effect can be discovered within the framework of Eq. (1.7) if a DC term is included in Eq. (1.9) so that

$$E = A_0 + A_1 \cos\{\omega t - k_1 z\}. \tag{1.14}$$

Substituting this form into Eq. (1.6) produces a new nonlinear term in the polarisation with the same space-time dependence as the linear polarisation. When the two are combined, the net contribution at frequency ω is

$$P = \varepsilon_0(\chi^{(1)} + 2\chi^{(2)} A_0) A_1 \cos\{\omega t - k_1 z\} + \cdots \tag{1.15}$$

The new term is proportional to the DC field A_0, and is in phase with the optical frequency field provided $\chi^{(2)}$ is real. No phase-matching considerations arise since, as for optical rectification, phase matching is automatically ensured. It is clear from Eq. (1.15) that the linear susceptibility has been modified by $2\chi^{(2)} A_0$, and the refractive index has correspondingly changed from $n = \sqrt{1 + \chi^{(1)}}$ to $n = \sqrt{1 + \chi^{(1)} + 2\chi^{(2)} A_0}$.

The Pockels effect allows the polarisation properties of light to be controlled electrically and, in combination with a polarising beam splitter, is routinely used to create an optical switch. A more detailed discussion of a Pockels cell can be found in Chapter 4.

1.8 Sum frequency generation

The catalogue of second-order nonlinear processes can be extended further by considering the process of sum frequency generation (SFG), a possibility that emerges if the incident electric field contains two different frequency components namely

$$E = A_1 \cos\{\omega_1 t - \mathbf{k_1}.\mathbf{r}\} + A_2 \cos\{\omega_2 t - \mathbf{k}_2.\mathbf{r}\}. \tag{1.16}$$

The angular wave numbers have now been written as vectors to allow for the possibility that the waves at ω_1 and ω_2 are non-collinear (i.e. travelling in different directions).[11] Inserting Eq. (1.16) into the polarisation expansion of Eq. (1.7) yields a term of the form

$$P = \varepsilon_0(\cdots + \chi^{(2)} A_1 A_2 \cos\{\omega_3 t - (\mathbf{k_1} + \mathbf{k_2}).\mathbf{r}\} + \cdots) \tag{1.17}$$

---

[11] When written as a vector, the angular wave number is variously called the wave vector, the propagation vector or simply the **k**-vector.

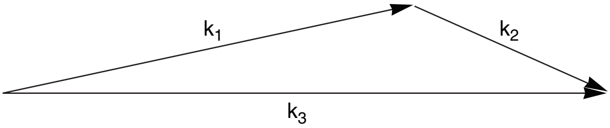

Fig. 1.7 Wave vector triangle for phase-matching non-collinear beams.

where $\omega_3 = \omega_1 + \omega_2$. It must be stressed that the subscripts 1–3 are now merely labels, their only numerical significance being that ω_3 is the highest frequency of the set. In this notation, second harmonic generation is the special case where $\omega_1 = \omega_2 = \omega_3/2$.

If the polarisation wave of Eq. (1.17) is to match the space-time dependence of a freely propagating field at the sum frequency ω_3 and space-time dependence $\cos\{\omega_3 t - \mathbf{k}_3.\mathbf{r}\}$, the phase-matching condition to be satisfied is

$$\mathbf{k_1} + \mathbf{k}_2 = \mathbf{k}_3 \tag{1.18}$$

which is represented by the vector triangle of Fig. 1.7. For collinear beams, this equation becomes

$$n_1\omega_1 + n_2\omega_2 = n_3\omega_3 \tag{1.19}$$

which is more complicated than for SHG where phase matching simply requires that the fundamental and harmonic refractive indices are the same. The term 'phase matching' is to be preferred to 'index matching' for this reason. In media that exhibit normal dispersion,[12] the right-hand side of Eq. (1.19) will always be larger than the left-hand side (see Problem 1.2). This means that adjusting the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$ is not sufficient on its own to fix the phase-mismatch problem, although it may still be useful for adjusting the phase-matching condition. An example of the use of non-collinear beams for increasing the bandwidth of a nonlinear interaction can be found in Chapter 7.

At the photon level, one can regard the SFG process as the annihilation of photons $\hbar\omega_1$ and $\hbar\omega_2$, and the creation of a photon $\hbar\omega_3$, so that

$$\hbar\omega_1 + \hbar\omega_2 \leftrightharpoons \hbar\omega_3 \tag{1.20}$$

which is a statement of conservation of energy. This equation has been written to suggest that the process may go into reverse, which will happen for certain phase relationships between the three fields.

Notice that multiplying all terms in Eq. (1.18) by $\hbar$ yields

$$\hbar\mathbf{k_1} + \hbar\mathbf{k_2} = \hbar\mathbf{k_3}. \tag{1.21}$$

Since momentum is $\mathbf{p} = \hbar\mathbf{k}$, this equation demonstrates that the phase-matching condition for the interacting waves is equivalent to the momentum-matching condition for the interacting photons.

---

[12] Normal dispersion means that the refractive index rises with frequency (or falls with wavelength).