Chapter 9

Nonlinear Optics

9.1 Overview

Communication technology is undergoing a revolution, and computer technology may do so soon, in which the key devices used (e.g., switches and communication lines) are changing from radio and microwave frequency devices to optical frequencies. This revolution has been made possible by the invention and development of lasers (most especially semiconductor diode lasers) and the discovery and development of dielectric crystals whose polarization P_i is a nonlinear function of the applied electric field, $P_i = \epsilon_0(\chi_{ij}E^j + \chi_{ijk}E^jE^k + \chi_{ijkl}E^jE^kE^l + \cdots)$. In this chapter we shall study lasers, nonlinear crystals, and various *nonlinear optics* applications that are based on them.

Most courses in elementary physics idealize the world as linear. From the simple harmonic oscillator to Maxwell's equations to the Schrödinger equation, most all the elementary physical laws one studies are linear, and most all the applications one studies make use of this linearity. In the real world, however, nonlinearities abound, creating such phenomena as avalanches, breaking ocean waves, holograms, optical switches, and neural networks; and in recent years nonlinearities and their applications have become major themes in physics research, both basic and applied. This chapter, with its exploration of nonlinear effects in optics, serves as a first introduction to some fundamental nonlinear phenomena and their present and future applications. In later chapters we shall revisit some of these phenomena and shall meet others, in the context of fluids (Chaps. 14 and 15), plasmas (Chap. 22), and spacetime curvature (Chaps. 24, 25, 26).

Since highly coherent and monochromatic laser light is one of the key foundations on which modern nonlinear optics has been built, we shall begin in Sec. 9.2 with a review of the basic physics principles that underlie the laser: the pumping of an active medium to produce a molecular population inversion, and the stimulated emission of radiation from the inverted population of molecules. Then we shall describe the details of how a number of different lasers are pumped and the characteristics of the light they emit. Most important among these characteristics are high frequency stability and high power. In Sec. 9.3 we shall meet our first example of an application of nonlinear optics: holography. In holography a three-dimensional, monochromatic image of an object is produced by a two step process: recording a hologram, and then passing coherent light through the hologram.

Holography differs from more modern nonlinear optics applications in not being a real-time process. Real-time processes have been made possible by nonlinear crystals. In Sec. 9.4 we study an example of a real-time, nonlinear-optics process: phase conjugation of light by a phase-conjugating mirror; and we see how such phase conjugation can be used to prevent distortion of images and signals carried in an optical fiber. In Sec. 9.5 we study the wave-wave mixing in nonlinear crystals that makes possible phase conjugation and other nonlinear optical processes; we examine two different ways to construct a phase conjugating mirror from such crystals; and we discuss several other applications of nonlinear crystals: frequency doubling and the production of squeezed light in Sec. 9.6.

9.2 Lasers

9.2.1 Basic Principles of the Laser

In quantum mechanics one identifies three different types of interaction of light with material systems (atoms, molecules, atomic nuclei, electrons, ...): (i) *Spontaneous emission*, in which a material system in an excited state spontaneously drops into a state of lesser excitation and emits a photon in the process. (ii) *Absorption*, in which an incoming photon is absorbed by a material system, exciting it. (iii) *Stimulated emission*, in which a material system, initially in some excited state, is "tickled" by passing photons and this tickling stimulates it to emit a photon of the same sort (in the same state) as the photons that tickled it.

As peculiar as stimulated emission may seem at first sight, it in fact is easily understood and analyzed classically. It is nothing but "negative absorption:" In classical physics, when a light beam $\Psi = Ae^{i(kz-\omega t+\phi)}$ (with A real) travels through an absorbing medium, its amplitude A decays exponentially with the distance propagated, $A \propto e^{-\mu z/2}$ (corresponding to an intensity decay $I \propto e^{-\mu z}$), while its frequency ω, wave number k, and phase ϕ remain very nearly constant. For normal materials, the absorption rate $\mu = I^{-1}dI/dz$ is positive and the energy lost goes ultimately into heat. However, one can imagine a material with an internally stored energy that amplifies a passing light beam. Such a material would have a negative absorption rate, $\mu < 0$, and correspondingly the amplitude of the passing light would grow with the distance traveled, $A \propto e^{+|\mu|z/2}$, while its frequency, wave number, and phase would remain constant. Such materials do exist; they are called "active media" and their amplification of passing waves is called "stimulated emission."

This elementary, classical description of stimulated emission is equivalent to the quantum mechanical description in the domain where the stimulated emission is strong: the domain of large photon occupation numbers (which, as we learned in Chaps. 2 and 3, is the domain of classical waves).

The classical description of stimulated emission takes for granted the existence of an active medium. To understand the nature of such a medium, we must turn to quantum mechanics:

As a first step toward such understanding, consider a beam of monochromatic light that impinges on a collection of molecules (or atoms or charged particles) which are all in the same quantum mechanical state $|1\rangle$. Suppose that the light has angular frequency ω and the molecules have a second state $|2\rangle$ with energy higher than that of $|1\rangle$ by $E_2 - E_1 = \hbar\omega$.

Then the light will resonantly excite the molecules from their initial state $|1\rangle$ to the higher state $|2\rangle$, and in the process photons will be absorbed. The strength of the interaction is proportional to the intensity of the beam. Stated more precisely, the rate of absorption of photons is proportional to the number of photons in the beam, i.e., to its intensity, in accord with the classical description of absorption.

Suppose, now, that when the light beam first arrives, the atoms are all in the higher state $|2\rangle$ rather than the lower state $|1\rangle$. There will still be a resonant interaction, but this time the interaction will deexcite the atoms, with an accompanying emission of photons. As in the absorption case, the strength of the interaction is proportional to the intensity of the beam, i.e., the rate of emission of new photons is proportional to the number of photons that the beam already has, i.e., to its intensity. A quantum mechanical analysis shows that the photons from this stimulated emission come out in the same quantum state as is occupied by the photons of the incoming beam. Correspondingly, when viewed classically, the beam will be amplified with no change of its frequency, wave number, or phase.

In Nature molecules typically have their energy levels populated in accord with the laws of statistical (thermodynamic) equilibrium. Such thermalized populations, as we saw in Chap. 3, entail a ratio $N_2/N_1 = \exp[-(E_2 - E_1)/kT] < 1$ for the number N_2 of molecules in state $|2\rangle$ to the number N_1 in state $|1\rangle$. Here T is the molecular temperature, and for simplicity it is assumed that the states are nondegenerate. Since there are more molecules in the lower state $|1\rangle$ than the higher one $|2\rangle$, an incoming light beam will experience more absorption than stimulated emission.

On the other hand, occasionally in Nature and often in the laboratory a collection of molecules develops a "population inversion" in which $N_2 > N_1$. The two states can then be thought of as having a negative temperature with respect to each other. Light propagating through population-inverted molecules will experience more stimulated emission than absorption; i.e., it will be amplified. The result is "light amplification by stimulated emission," or "laser" action.

This basic principle underlying the laser has been known since the early years of quantum mechanics, but only in the 1950s did physicists succeed in designing, constructing, and operating real lasers. The first proposals for practical devices were made, independently, by Weber (1953), by Gordon, Zeiger, and Townes (1954), and by Basov and Prokhorov (1954, 1955). The first successful construction and operation of a laser was by Gordon, Zeiger, and Townes (1954, 1955)—though their first laser actually used radiation not at optical frequencies but rather at microwave frequencies (based on a population inversion of ammonia molecules)[1] and thus was called a *maser*. The first optical frequency laser, one based on a population inversion of chromium ions in a ruby crystal, was constructed and operated by Maiman (1960).

The key to laser action is the population inversion. Population inversions are incompatible with thermodynamic equilibrium; thus, to achieve them, one must manipulate the molecules in a nonequilibrium way. This is usually done by some concrete variant of the process shown in the energy level diagram of Fig. 9.1 Some sort of pump mechanism (to be discussed in the next section) rapidly excites molecules from the ground state into some group of "absorption"

[1]For the basic principles of the ammonia maser, see, e.g., Chap. 9 of Feynman, Leighton, and Sands (1965).

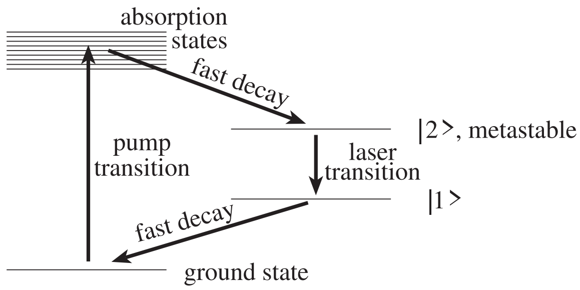

Fig. 9.1: The mechanism for creating the population inversion that underlies laser action. The horizontal lines and band represent energy levels of a molecule, and the arrows represent transitions in which the molecules are excited by pumping or decay by emission of photons.

states. The molecules then decay rapidly from the absorption states into the state $|2\rangle$, which is metastable (i.e., has a long lifetime against spontaneous decay). The laser transition is from state $|2\rangle$ into state $1\rangle$. Once a molecule has decayed into state $|1\rangle$, it quickly decays on down to the ground state and then may be quickly pumped back up into the absorption states.

If the pump acts suddenly and briefly, this process will produce a temporary population inversion of states $|2\rangle$ and $|1\rangle$, with which an incoming, weak burst of "seed" light can interact to produce a burst of amplification. The result is a pulsed laser. If the pump acts continually, the result may be a permanently maintained population inversion with which continuous seed light can interact to produce continuous-wave laser light.

As the laser beam travels through the active medium (the population-inverted molecules), its intensity I builds up with distance z as $dI/dz = I/l_o$, so $I(z) = I_o e^{z/l_o}$. Here I_o is the initial intensity, and $l_o \equiv 1/\mu$, the e-folding length, depends on the strength of the population inversion and the strength of the coupling between the light and the active medium. Typically l_o is so long that strong lasing action cannot be achieved by a single pass through the active medium. In this case, the lasing action is enhanced by placing the active medium inside a Fabry-Perot cavity (Fig. 9.2). The length L of the cavity is adjusted so the lasing transition frequency $\omega = (E_2 - E_1)/\hbar$ is an eigenfrequency of the cavity. The lasing action then excites a standing wave mode of the cavity, from which the light leaks out through one or both cavity mirrors. If F is the cavity's finesse [the average number of times a photon bounces back and forth inside the cavity before escaping through a mirror; cf. Eq. (8.46)], then the cavity increases the distance that typical photons travel through the active medium by a factor F, thereby increasing the intensity of the light output by a factor e^{FL/l_o}.

For an ideal laser (one, e.g., with a perfectly steady pump maintaining a perfectly steady population inversion that in turn maintains perfectly steady lasing), the light comes out in the most perfectly classical state that quantum mechanics allows. This state, called a "quantum mechanical coherent state," has a perfectly sinusoidally oscillating electric field on which is superimposed the smallest amount of noise (the smallest wandering of phase and amplitude) allowed by quantum mechanics: the noise of quantum electrodynamical vacuum fluctuations. The value of the oscillations' well defined phase is determined by the phase

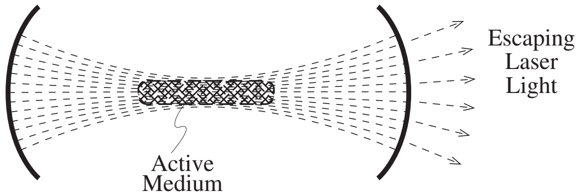

Fig. 9.2: The use of a Fabry-Perot cavity to enhance the interaction of the light in a laser with its active medium.

of the seed field from which the coherent state was built up by lasing. Real lasers have additional noise due to a variety of practical factors, but nevertheless, their outputs are usually highly coherent, with long coherence times.

9.2.2 Types of Pumping and Types of Lasers

Lasers can be pumped radiatively, collisionally, chemically, electrically, and even by nuclear explosions; and each method of pumping produces a laser with special properties that have special uses. In this section we shall explore a few examples.

Radiative pumping: In radiative pumping a burst of "pump" photons excites the active medium's molecules from their ground state to a group of absorption states. The pump photons are typically produced by a flash tube which surrounds the active medium, or whose light is focussed onto the active medium by mirrors. This was the type of pumping used by Maiman in his first ruby maser. The strongest pulsed lasers now available (Neodymium glass lasers) use a variant of this called *Q-switching*. In Q-switching, the resonant interaction between the laser light and the active medium is temporarily turned off (e.g., by removing the Fabry-Perot cavity from around the active medium) while the medium is radiatively pumped. Thereby a very strong population inversion is built up, and when the resonance is turned back on, an enormous but very brief pulse of laser light is produced—as much as 100 J in a picosecond.

Collisional pumping: The continuous-wave helium-neon laser uses collisional pumping. A mixture of helium and neon gas (roughly 10 helium atoms for each neon atom) is subjected to a continuous electrical discharge. The electrons in the discharge collide with the many helium atoms, exciting them into absorption states that then decay rapidly into a long-lived metastable state. The resulting population inversion of the helium atoms, however, is not used directly for laser action. Rather, the many excited helium atoms collide with the fewer ground-state neon atoms, resonantly exciting them into a metastable neon state that has nearly the same energy as that of helium. The resulting population inversion of neon then acts as the laser's active medium. [There actually are several metastable states of neon that get population inverted in this way, and the helium-neon laser thereby can lase at several different wavelengths: 0.6328 microns (in the red), 1.15 and 3.39 microns (in the infrared), and others.]

Chemical pumping: In chemical pumping a nonequilibrium chemical reaction creates products in excited, metastable states that then lase. An example is the reaction $H + F \to HF$, which leaves the hydrogen flouride molecule in a metastable, lasing state.

Electrical pumping: In electrical pumping, electric fields and associated currents are used to produce population inversions. Two important examples are semiconductor diode lasers and free electron lasers. In *semiconductor diode lasers*, the flows of electrons and holes,[2] in response to an electrical bias, populate a portion of the conduction band and depopulate a portion of the valence band in a thin layer of a semiconductor (e.g., a 0.2 micron thick layer of Gallium Arsenide that is sandwiched between one material that injects conduction-band electrons into the Gallium Arsenide and another that injects holes). A weak beam of light passing along the thin layer stimulates electrons or holes to drop out of the conduction band into the valence band, thereby emitting photons that amplify the light. The resulting continuous-wave laser is easily modulated at frequencies as high as 10's of GHz by modulating the bias voltage. This and the diode's tiny size makes such lasers ideally suited for optical communication. In the *free electron laser*, a nearly monoenergetic beam of electrons, created by a particle accelerator, is sent through a static, transverse, spatially alternating magnetic field. The field is called an "undulator" because of its diffraction-grating-like undulations. The field's alternating Lorentz force causes the moving electrons to oscillate back and forth transversely and radiate. These electron oscillations resonate with the light they emit; the light moves forward, relative to each electron, by one optical wavelength while the electron undergoes one oscillation. In this device the electrons' population-inverted energy distribution (many high-speed electrons, fewer lower-speed electrons) is produced electrically, by the particle accelerator, and the photon emission drives the electrons from their initial, strongly populated states of high kinetic energy to more sparsely populated states of lower kinetic energy.

Nuclear-explosion pumping. A device much ballyhooed in America during the reign of Ronald Reagan, but never built, was a futuristic, super-powerful X-ray laser pumped by a nuclear explosion. As part of Reagan's Strategic Defense Initiative ("Star Wars"), this laser was supposed to shoot down Soviet missiles. In Ex. 9.1 the reader is invited to speculate about the design of such a laser.

Discussion: As the above examples show, lasers come in a wide variety of configurations, and the light they produce can have a wide variety of properties. Pulsed lasers can achieve very high instantaneous powers (100 J in a picosecond, corresponding to 10^{14} Watts). Continous-wave lasers can also achieve large powers; for example CO_2 lasers putting out as much as 10^9 Watts can have their light concentrated into regions with transverse dimensions as small as one wavelength (a micron, but no smaller because of diffraction), thereby yielding a local power density of 10^{21} W m^{-2}.

For many applications large power is irrelevant or undesireable, but high frequency stability (a long coherence time) is crucial. By locking the frequencies of lasers to optical cavities or to molecular transitions, one can suppress the wandering of the phase of the laser light and thereby achieve frequency stabilities as high as $\Delta f \sim 1$ mHz, corresponding to coherence times of ~ 1000 sec and coherence lengths of $\sim 3 \times 10^8$ km.

When first invented, lasers were called "a solution looking for a problem." Now they permeate everyday life as well as high technology. Examples are supermarket bar-code

[2] A "hole" is the absence of an electron in a "degenerate Fermi sea;" i.e., it is an empty single-particle state (mode) of the electron field, in a distribution function for which, up to some momentum, most all the other electron states are occupied.

readers, CD players, eye surgery, laser printers, laser gyroscopes (which are now standard on commercial aircraft), laser-based surveying, Raman spectroscopy, laser fusion, optical communication, optically based computers, holography, maser amplifiers, and hydrogen-maser clocks (the most stable of all atomic clocks on timescales of seconds to minutes).

****

EXERCISES

Exercise 9.1 *Challenge: Nuclear Powered X-Ray Laser*

How would *you* design a nuclear powered X-ray laser for use in Star Wars? The energy for the pump comes from a nuclear explosion that you set off in space above the earth. You want to use that energy to create a population inversion in an active medium that will lase at X-ray wavelengths; and you want to focus the resulting X-ray beam onto an intercontinental ballistic missile that is rising out of the earth's atmosphere. What would you use for the active medium? How would you guarantee that a population inversion is created in the active medium? How would you focus the resulting X-ray beam? (*Note*: This is a highly nontrivial exercise, intended more as a stimulus for thought than as a test of one's understanding of things taught in this book.)

****

9.3 Holography

Thus far in this book, our study of optics has focused on situations where waves propagate linearly, i.e., where they superpose linearly (additively). In the 1970's and 1980's the technology of lasers and of "nonlinear crystals" began to make possible processes in which light waves interact with each other nonlinearly (Sec. 9.5 below). The resulting nonlinear optics has promising applications in such diverse areas as computers, communication, optical astronomy, gravitational-wave detection, spectroscopy, holography, ...

Holography is an old and well-explored example of nonlinear optics—an example in which the nonlinear interaction of light with itself is produced not in real time, but rather by means of a recording followed by a later readout.[3]

Holography is to be contrasted with ordinary photography. Ordinary photography (Fig. 9.3) produces a colored, 2-dimensional image of 3-dimensional objects. Holography (Figs. 9.4, 9.6 below) produces a monochromatic 3-dimensional image of 3-dimensional objects. Note that, roughly speaking, the two processes contain the same amount of information: two items of information at each location in the image. The two items in an ordinary photograph are the intensity and the color; the two items in a holographic photograph (hologram) are the

[3] Holography is discussed and analyzed in most standard optics textbooks; e.g., Chapter 8 of Ghatak and Thyagarajan (1978). A number of practical applications of holography are discussed by Iizuka (1987) and by Cathey (1974).

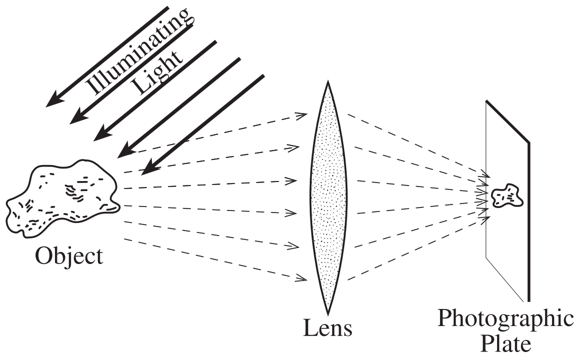

Fig. 9.3: Ordinary photography.

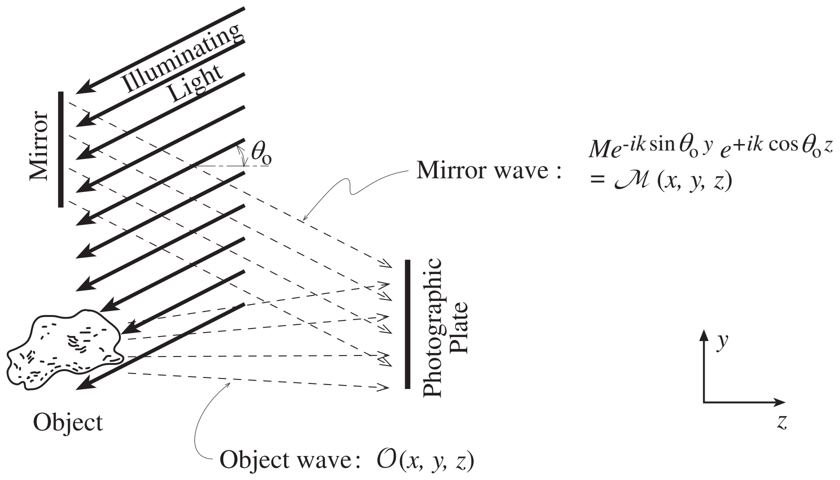

Fig. 9.4: Recording a hologram.

intensity and the phase of the monochromatic light. It is the phase of the light, lost from an ordinary photograph but preserved in a hologram, that contains the information about the third dimension: Our brain deduces the distance to a point on an object from the difference in the directions of propagation of the point's light as it arrives at our two eyes. Those propagation directions are encoded in the light as variations of the light's phase with transverse location [see, e.g., the point-spread function for a thin lens, Eq. (7.30)]. Thus, the transverse variations in phase contain the three-dimensional information. It is just those transverse phase variations that are preserved in a hologram.

In an ordinary photograph (Fig. 9.3), white light scatters off an object, with different colors scattering with different strengths. The resulting colored light is focused through a lens to form a colored image on a photographic plate. The plate records the color and intensity of the light at each point in the focal plane, thereby producing the ordinary photograph.

In holography one records a hologram (Fig. 9.4), and one then uses the hologram to reconstruct the holographic image (Fig. 9.6 below).

Consider, first, the recording of the hologram. Monochromatic, plane-wave light (angular frequency ω and wave number $k = \omega/c$) illuminates the object and also a mirror as shown in Fig. 9.4. The light must be spatially coherent over the entire region of mirror plus object. The illuminating light propagates in the y–z plane, at some angle θ_o to the z axis, and the mirror lies in an x–y plane. The mirror reflects the illuminating light, producing a so-called "reference beam," which we shall call the *mirror wave*:

$$\psi_{\rm mirror} = M e^{ik(z\cos\theta_o - y\sin\theta_o)} , \tag{9.1}$$

where M is a real constant. (Here and in the rest of this section, the fields are monochromatic with frequency ω and wave number k, and we suppress their trivial time dependences $e^{-i\omega t}$ from all formulas.) The object scatters the illuminating light, producing a wave propagating toward the photographic plate (z direction) that we shall call the *object wave* and shall denote

$$\psi_{\rm object} = \mathcal{O}(x,y,z)e^{ikz} . \tag{9.2}$$

It is the slowly varying complex amplitude $\mathcal{O}(x,y,z)$ of this object wave that carries the three-dimensional, but monochromatic, information about the object's appearance, and it thus is this $\mathcal{O}(x,y,z)$ that will be reconstructed in the second step of holography.

In the first step (Fig. 9.4 (a)) and Eq. (9.2), the object wave propagates along the z-direction to the photographic plate at $z = 0$, where it interferes with the mirror wave to produce a transverse intensity pattern

$$\begin{aligned} I(x,y) &= |\mathcal{O} + M e^{-iky\sin\theta_o}|^2 \\ &= M^2 + |\mathcal{O}(x,y,z=0)|^2 + \mathcal{O}(x,y,z=0)Me^{iky\sin\theta_o} + \mathcal{O}^*(x,y,z=0)Me^{-iky\sin\theta_o}. \end{aligned} \tag{9.3}$$

(Here and throughout this chapter a * denotes complex conjugation.) The plate is blackened at each point in proportion to this intensity. The plate is then developed and a positive or negative print (it doesn't matter which because of Babinet's principle) is made on a transparent sheet of plastic or glass. This print, the *hologram*, has a transmissivity as a function of x and y that is proportional to the intensity distribution (9.3):

$$T(x,y) \propto M^2 + |\mathcal{O}(x,y,z=0)|^2 + M\mathcal{O}(x,y,z=0)e^{iky\sin\theta_o} + M\mathcal{O}^*(x,y,z=0)e^{-iky\sin\theta_o} . \tag{9.4}$$

In this transmissivity we meet our first example of nonlinearity: $T(x,y)$ is a nonlinear superposition of the mirror wave and the object wave. Stated more precisely, the superposition is not a linear sum of wave fields, but instead is a sum of products of one wave field with the complex conjugate of another wave field. A further nonlinearity will arise in the reconstruction of the holographic image, Eq. (9.6) below.

Fig. 9.5 shows an example. Fig. 9.5 (a) is an ordinary photograph of an object, (b) is a hologram of the same object, and (c) is a blow-up of a portion of that hologram. The object is not at all recognizable in the hologram because the object wave $\mathcal{O}$ was not focused to form an image at the plane of the photographic plate. Rather, light from each region of the object was scattered to and recorded by all regions of the photographic plate. Nevertheless, the plate contains the full details of the scattered light $\mathcal{O}(x,y,z=0)$, including its phase. That information is recorded in the piece $M(\mathcal{O}e^{iky\sin\theta_o} + \mathcal{O}^*e^{-iky\sin\theta_o}) = 2M\,\Re(\mathcal{O}e^{iky\sin\theta_o})$ of the hologram's transmissivity. This piece oscillates sinusoidally in the y direction with

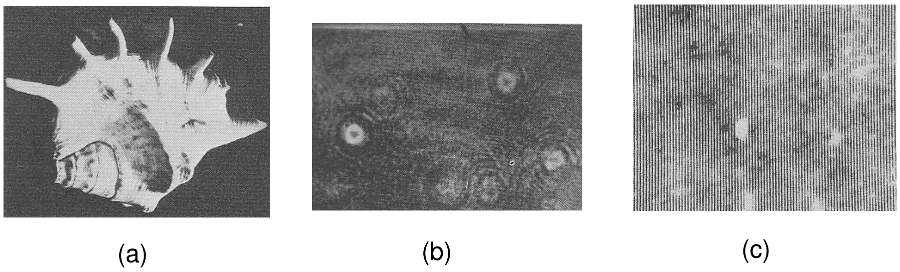

Fig. 9.5: (a) Ordinary photograph of an object. (b) Hologram of the same object. (c) Enlargement of the hologram. [From Fig. 8.3 of Ghatak and Thyagarajan, 1978.].

wavelength $2\pi/k\sin\theta_o$; and the amplitude and phase of its oscillations are modulated by the object wave $\mathcal{O}(x, y, z = 0)$. Those modulated oscillations show up clearly when one magnifies the hologram (Fig. 9.5 (c)); they make the hologram into a sort of diffraction grating, with the object wave $\mathcal{O}(x, y, z = 0)$ encoded as variations of the darkness and spacings of the grating lines.

What about the other pieces of the transmissivity (9.4), which superpose linearly on the diffraction grating? One piece, $T \propto M^2$, is spatially uniform and thus has no effect except to make the lightest parts of the hologram slightly grey rather than leaving it absolutely transparent (since this hologram is a negative rather than a positive). The other piece, $T \propto |\mathcal{O}|^2$, is the intensity of the object's unfocussed, scattered light. It produces a greying and whitening of the hologram (Fig. 9.5 (b)) that varies on lengthscales long compared to the grating's wavelength $2\pi/k\sin\theta_o$, and that thus blots out the diffraction grating a bit here and there, but does not change the amplitude or phase of the grating's modulation.

To reconstruct the object's 3-dimensional wave, $\mathcal{O}(x, y, z)e^{ikz}$, one sends through the hologram monochromatic, plane-wave light identical to the mirror light used in making the hologram; cf. Fig. 9.6. If, for pedagogical simplicity, we place the hologram at the same location $z = 0$ as was previously occupied by the photographic plate, then the incoming light has the same form (9.1) as the original mirror wave, but with an amplitude that we shall denote as R corresponding to the phrase "reference beam" that is used to describe this incoming light:

$$\psi_{\rm reference} = Re^{ik(z\cos\theta_o - y\sin\theta_o)} . \tag{9.5}$$

In passing through the hologram at $z = 0$, this reference beam is partially absorbed and partially transmitted. The result, immediately upon exiting from the hologram, is a "reconstructed" light-wave field whose value $\psi_{\rm reconstructed} = \mathcal{R}(x, y, z = 0)$ and normal derivative $\psi_{{\rm reconstructed}\ ,z} = \mathcal{Z}(x, y, z = 0)$ are given by [cf. Eq. (9.4)]

$$\begin{aligned}
\mathcal{R}(x, y, z = 0) &= T(x, y)Re^{-iky\sin\theta_o} \\
&= M^2 Re^{-iky\sin\theta_o} + |\mathcal{O}(x, y, z = 0)|^2 Re^{-iky\sin\theta_o} \\
&+ MR\mathcal{O}(x, y, z = 0) + MR\mathcal{O}^*(x, y, z = 0)e^{-i2ky\sin\theta_o} \; ; \\
\mathcal{Z}(x, y, z = 0) &= ik\cos\theta_o \mathcal{R}(x, y, z = 0).
\end{aligned} \tag{9.6}$$

This field and normal derivative act as initial data for the subsequent evolution of the reconstructed wave. Note that the field and derivative, and thus also the reconstructed

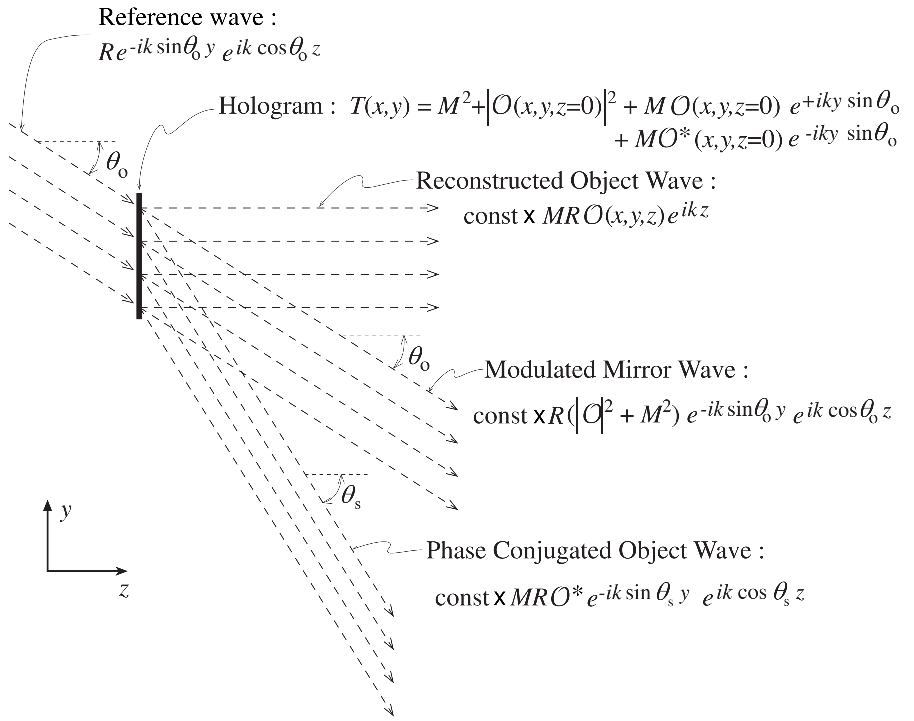

Fig. 9.6: Reconstructing the holographic image from the hologram. Note that $\sin\theta_1 = 2\sin\theta_o$.

wave, are triply nonlinear: each term in Eq. (9.6) is a product of the original mirror wave M used to construct the hologram or the original object wave $\mathcal{O}$, times $M^* = M$ or $\mathcal{O}^*$, times the reference wave R that is being used in the holographic reconstruction.

The evolution of the reconstructed wave beyond the hologram (at $z > 0$) can be computed by combining the initial data, Eq. (9.6) for $\psi_{\rm reconstructed}$ and $\psi_{{\rm reconstructed}\ ,z}$ at $z = 0$ with the Helmholtz-Kirchhoff formula (7.4); see Ex. 9.2. From the four terms in the initial data, Eq. (9.6) [which arise from the four terms in the hologram's transmissivity $T(x, y)$, Eq. (9.4)], the reconstruction produces four wave fields; see Fig. 9.6. The direction of propagation of each of these waves can easily be inferred from the vertical spacing of its phase fronts along the outgoing face of the hologram, or equivalently from the relation $\partial\psi_{\rm reconstructed}/\partial y = ik_y\psi_{\rm reconstructed} = -ik\sin\theta\psi$, where θ is the angle of propagation relative to the horizontal z direction. Since, immediately in front of the hologram, $\psi_{\rm reconstructed} = \mathcal{R}$, the propagation angle is

$$\sin\theta = \frac{\partial\mathcal{R}/\partial y}{-ik\mathcal{R}}\ . \tag{9.7}$$

Comparing with Eq. (9.4) and Eq. (9.6), we see that the first two, slowly spatially varying terms in the transmissivity, $T \propto M^2$ and $T \propto |\mathcal{O}|^2$, both produce waves that propagates in the same direction as the reference wave, $\theta = \theta_o$. This combined wave has an uninteresting, smoothly and slowly varying intensity pattern.

The two diffraction-grating terms in the hologram's transmissivity produce two interesting waves. One, arising from $T \propto \mathcal{O}(x, y, z = 0)Me^{iky\sin\theta_o}$ [and produced by the first term on the third line of the initial conditions (9.6)], is precisely the same object wave

$\psi_{\text{object}} = \mathcal{O}(x, y, z)e^{ikz}$ (aside from overall amplitude) as one would have seen while making the hologram if one had replaced the photographic plate by a window and looked through it. This object wave, carrying [encoded in $\mathcal{O}(x, y, z)$] the famous holographic image with full 3-dimensionality, propagates in the z-direction, $\theta = 0$.

The transmissivity's second diffraction-grating term, $T \propto \mathcal{O}^*(x, y, z = 0)Me^{-iky\sin\theta_o}$, acting via the second term on the third line of the initial conditions (9.6), gives rise to a *secondary* wave which [according to Eq. (9.7)] propagates at an angle θ_s to the z-axis, where $\sin\theta_s = 2\sin\theta_o$ This secondary wave carries an image that is encoded in the complex conjugate $\mathcal{O}^*(x, y, z = 0)$ of the transverse (i.e., x, y) part of the original object wave. Since complex conjugation of an oscillatory wave just reverses the sign of the wave's phase, this wave in some sense is a "phase conjugate" of the original object wave.

When one recalls that the electric and magnetic fields that make up an electromagnetic wave are actually real rather than complex, and that we are using complex wave fields to describe electromagnetic waves only for mathematical convenience, one then realizes that this phase conjugation of the object wave is actually a highly nonlinear process. There is no way, by linear manipulations of the real electric and magnetic fields, to produce the phase-conjugated wave from the original object wave.

In the next section we shall develop in detail the theory of *phase-conjugated waves*, and in Exercise 9.2 (b), we shall relate our holographically constructed secondary wave to that theory. As we shall see, our secondary wave is not quite the same as the "phase-conjugated object wave," but it *is* the same aside from some distortion along the y direction and a change in propagation direction. Using the theory in the next section, we shall deduce [Exercise 9.2 (b)] the following comparison between our holographic object wave and our secondary wave: If one looks into the object wave with one's eyes (i.e. if one focuses it onto one's retinas), one sees the original object in all its three-dimensional glory, though single colored, sitting behind the hologram at the object's original position. Because the image one sees is behind the hologram, it is called a "virtual image". If, instead, one looks into the secondary wave with one's eyes (i.e. if one focuses it onto one's retinas), one sees the original three-dimensional object, sitting in front of the hologram but turned inside out and distorted; for example, if the object is a human face, the secondary image looks like the interior of a mask made from that human face, with some distortion along the y direction. Because this secondary image appears to be in front of the hologram, it is called a "real image"—even though one can pass one's hands through it and feel nothing but thin air.

There are many variants on the basic holographic technique that we have described in Figs. 9.4–9.6. In one, called "volume holography," the hologram is a number of wavelengths deep rather than being just two-dimensional. For example, it could be made from a photographic emulsion a number of wavelengths thick, in which the absorption length for light (before developing) is longer than the thickness. Such a hologram has a three-dimensional grating structure (grating "surfaces" rather than grating "lines"), and when one reconstructs the holographic image from it in the manner of Fig. 9.6, the third dimension of the grating suppresses the phase-conjugated wave while enhancing the (desired) object wave. In another variant, one reflects light off the hologram instead of transmitting light through it; in such reflection, the hologram's diffraction grating produces a three-dimensional holographic image by the same process as in transmission.

Other variants are optimized for reconstructing the holographic image with white light (light that has a broad range of frequencies). Even for the simple two-dimensional hologram of Fig. 9.6, if one sends in white light at the angle θ_o, one will get a three-dimensional object wave: The hologram's grating will diffract various wavelengths in various directions. In the direction of the original object wave (the horizontal direction in Fig. 9.6), one will get a 3-dimensional reconstructed image of the same color as was used when constructing the hologram. When one moves vertically away from that direction (as shown in Fig. 9.6), one will see the color of the 3-dimensional image continuously change. A "white-light" hologram of this type (though one relying on reflection rather than transmission) is used on many credit cards as an impediment to counterfeiting, and has even been used on postage stamps.

Holograms are much used in everyday life and technology. Examples are credit cards and holographic lenses in supermarket checkouts. Other examples, that are still in the developmental stage but that may become widespread in a few years, are volume holograms used for three-dimensional movies, and volume holograms for storage of large amounts of data.

Just as one can draw two-dimensional pictures numerically, pixel-by-pixel, so one can also create and modify holograms numerically.

****

EXERCISES

Exercise 9.2 *Derivation: The Holographically Reconstructed Wave*

(a) Use the Helmholtz-Kirchhoff law to compute all four pieces of the holographically reconstructed wave field. Show that the piece generated by $T \propto \mathcal{O}(x, y, z = 0)Me^{iky\sin\theta_o}$ is the same as the field $\psi_{\rm object} = \mathcal{O}(x, y, z)e^{-i\omega t}$ that would have resulted, when making the hologram (Fig. 9.4), had the mirror wave been absent and the photographic plate been replaced by a window. Show that the other pieces have the forms and propagation directions indicated heuristically in Fig. 9.6.

(b) Consider the secondary wave generated by $T \propto M\mathcal{O}^* e^{-iky\sin\theta_o}$. Assume, for simplicity, that the mirror and reference waves propagate nearly perpendicular to the hologram, so $\theta_o \ll 90$ degrees and $\theta_s \simeq 2\theta_o \ll 90$ degrees; but assume that θ_s is still large enough that fairly far from the hologram the object wave and secondary waves separate cleanly from each other. Then, taking account of the fact that the object wave field has the form $\mathcal{O}(x, y, z)e^{ikz}$, show that the secondary wave is the "phase conjugated object wave" as defined in Sec. 9.5, except that it is propagating in the $+z$ direction rather than $-z$, i.e. it has been reflected through the $z = 0$ plane. Then use this, and the discussion of phase conjugation in Sec. 9.4, to show that the secondary wave carries an image that resides in front of the hologram and is turned inside out, as discussed near the end of Sec. 9.4. Show, further, that if θ_o is not $\ll 90$ degrees (but is < 30 degrees, so θ_s is a real angle and the secondary image actually exists), then the secondary image is changed by a distortion along the y direction. What is the nature of the distortion, a squashing or a stretch?

(c) Suppose that plane-parallel white light is used in the holographic reconstruction of Fig. 9.6. Derive an expression for the direction in which one sees the object's three-dimensional image have a given color (or equivalently wave number). Assume that the original hologram was made with green light and $\theta_o = 45$ degrees. What are the angles at which one sees the image as green and as red?

Exercise 9.3 *Problem: Compact Disks*

Information on compact disks ("CDs") is recorded and read out holographically. Explain why one can expect to record, at the very most, one bit of information per square wavelength of the recording light on the disk's surface. Typical CDs achieve only a few per cent of this upper limit on information density. Estimate the number of volumes of the Encyclopedia Britannica that can be recorded on a CD.

****

9.4 Phase Conjugate Optics

Nonlinear optical techniques make it possible to phase conjugate an optical wave in real time, by contrast with holography where the phase conjugation requires recording a hologram and then reconstructing the wave later. In this section we shall explore the properties of phase conjugated waves, and in the next section we shall discuss the technology by which real-time phase conjugation is achieved.

The basic ideas and foundations for phase conjugation of waves were laid by Boris Yakovovich Zel'dovich and his colleagues (1972) and by Amnon Yariv (1977). (Zel'dovich is the famous son of a famous Russian/Jewish physicist, Yakov Borisovich Zel'dovich, who with Andrei Dmitreivich Sakharov fathered the Soviet hydrogen bomb and then went on to become a dominant figure in astrophysics and cosmology; Yariv is an Israeli physicist who works at Caltech and is a founder not only of the field of phase conjugate optics but also of nonlinear optics quite generally.)

Phase conjugation is the process of taking a wave $\Psi_{\rm O} = \psi(x, y, z)e^{-i\omega t}$ and from it constructing the wave $\Psi_{\rm PC} = \psi^*(x, y, z)e^{-i\omega t}$.

In discussing phase conjugation, and nonlinear optics more generally, we shall (as throughout the last few chapters) think of ψ as representing, in complex notation, the real electric field E of some specific polarization state of an electromagnetic wave:

$$E(t, x, y, z) = \Re(\Psi) \ . \tag{9.8}$$

Then the electric fields of the original wave $\Psi_{\rm O}$ and of its phase conjugate $\Psi_{\rm PC}$ are

$$\begin{aligned} E_{\rm O} &= \frac{1}{2}(\psi e^{-i\omega t} + \psi^* e^{+i\omega t}), \\ E_{\rm PC} &= \frac{1}{2}(\psi^* e^{-i\omega t} + \psi e^{+i\omega t}). \end{aligned} \tag{9.9}$$

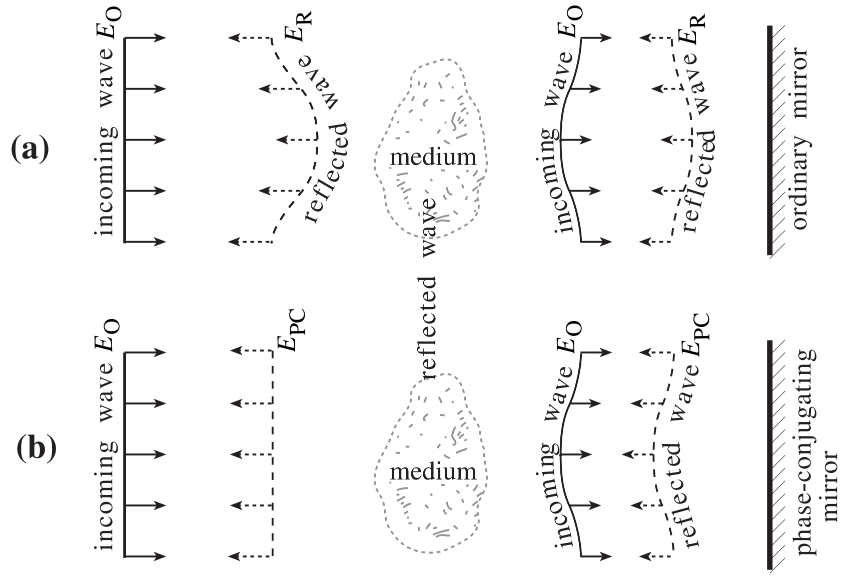

Fig. 9.7: A rightward propagating wave and the reflected wave produced by (a) an ordinary mirror and (b) a phase-conjugating mirror. In both cases the waves propagate through a medium of spatially variable index of refraction, which distorts their phase fronts. In case (a) the distortion is reinforced by the second passage through the variable medium; in case (b) the distortion is removed.

Notice that the phase conjugated wave $E_{\rm PC}$ is obtainable from the original wave $E_{\rm O}$ by *time reversal*, $t \to -t$. This has a number of important consequences. One is that $E_{\rm PC}$ propagates in the opposite direction to $E_{\rm O}$. Others are explained most clearly with the help of a *phase-conjugating mirror*:

Consider an electromagnetic wave $E_{\rm O}$ with spatial modulation (i.e., a wave that carries a picture or a signal of some sort). Let the wave propagate in the z-direction (rightward in Fig. 9.7), so $\psi = \mathcal{E} e^{ikz}$ and

$$E_{\rm O} = \Re[\mathcal{E}(x,y,z)e^{i(kz-\omega t)}] \,, \tag{9.10}$$

where $\mathcal{E}$ is a complex amplitude whose modulus and phase vary slowly in x, y, z (slowly compared to the wave's wavelength $\lambda = 2\pi/k$). Suppose that this wave propagates not through vacuum, but rather through a medium with an index of refraction $n(x,y,z) = n(\mathbf{x})$ that varies slowly compared to λ and that therefore distorts the wave as it propagates. The propagation equation $\nabla^2 E_{\rm O} + (n/c)^2\omega^2 E_{\rm O} = 0$ for the electric field $E_{\rm O}$ implies the following propagation equation for the slowly distorting, complex amplitude $\mathcal{E}$:

$$\nabla^2\mathcal{E} + \Big[\frac{n^2(\mathbf{x})\omega^2}{c^2} - k^2\Big]\mathcal{E} + 2ik\frac{\partial\mathcal{E}}{\partial z} = 0 \,. \tag{9.11}$$

Let the slowly distorting, rightward propagating wave $E_{\rm O}$ reflect off a mirror located at $z = 0$. If the mirror is a phase-conjugating one, then very near it (at z near zero) the reflected wave will have the form

$$E_{\rm PC} = \Re[\mathcal{E}^*(x,y,z=0)e^{i(-kz-\omega t)}] \,, \tag{9.12}$$

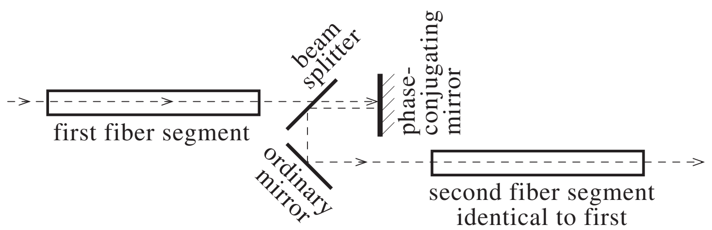

Fig. 9.8: The use of a phase-conjugating mirror in an optical transmission line to prevent the fiber from distorting an optical image. The distortions put onto the image as it propagates through the first segment of fiber are removed during propagation through the second segment.

while if it is an ordinary mirror, then the reflected wave will be

$$E_{\rm R} = \Re[-\mathcal{E}(x, y, z = 0)e^{i(-kz-\omega t)}] \,. \tag{9.13}$$

(The sign flip in $\mathcal{E}$ guarantees that $E_{\rm O} + E_{\rm R} = 0$ at the surface of the highly conducting ordinary mirror.)

These two waves, the phase-conjugated one and the ordinary reflected one, have very different surfaces of constant phase (*phase fronts*): Denote by $\phi(x, y, z)$ the slowly varying phase of $\mathcal{E}(x, y, z)$. Then the phase of the incoming wave $E_{\rm O}$ [Eq. (9.10)] as it nears the mirror ($z = 0$) is $\phi + kz$, so (taking account of the fact that ϕ is slowly varying), the surfaces of constant phase are $z = -\phi(x, y, z = 0)/k$. Similarly, the phase of the wave $E_{\rm R}$ [Eq. (9.13)] reflected from the ordinary mirror is $\phi - kz$, so its surfaces of constant phase near the mirror are $z = +\phi(x, y, z = 0)/k$, which are reversed from those of the incoming wave as shown in the upper right of Fig. 9.7. Finally, the phase of the wave $E_{\rm PC}$ [Eq. (9.12)] reflected from the phase-conjugating mirror is $-\phi - kz$, so its surfaces of constant phase near the mirror are $z = -\phi(x, y, z = 0)/k$, which are the same as those of the incoming wave (lower right of Fig. 9.7), even though the two waves are propagating in opposite directions.

The phase fronts of the original incoming wave and the phase conjugated wave are the same not only near the phase conjugating mirror; they are the same everywhere. More specifically, as the phase-conjugated wave $E_{\rm PC}$ propagates away from the mirror [near which it is described by Eq. (9.12)], the propagation equation forces it to evolve in such a way as to remain always the phase conjugate of the incoming wave:

$$\begin{aligned} E_{\rm PC} &= \Re[\mathcal{E}^*(x, y, z)e^{i(-kz-\omega t)}] \\ &= \Re[\mathcal{E}(x, y, z)e^{i(kz+\omega t)}]. \end{aligned} \tag{9.14}$$

This should be obvious from the fact that expression (9.14) for $E_{\rm PC}$ satisfies the same propagation equation $\nabla^2 E + (n/c)^2\omega^2 E = 0$ as does the $E_{\rm O}$ of Eq. (9.10).

This fact that the reflected wave $E_{\rm PC}$ remains always the phase conjugate of the incoming wave $E_{\rm O}$ means that *the distortions put onto the incoming wave, as it propagates rightward through the inhomogeneous medium, get removed from the phase conjugated wave as it propagates back leftward;* see Fig. 9.7.

This removal of distortions has a number of important applications. One is for image transmission in optical fibers. Normally when an optical fiber is used to transmit an optical image, the transverse spatial variations $n(x, y)$ of the fiber's index of refraction (which are required to hold the light in the fiber) distort the image somewhat. The distortions can be eliminated by using a sequence of identical segments of optical fibers separated by phase-conjugating mirrors (Fig. 9.8). A few other applications include (*i*) real time holography, (*ii*) removal of phase distortions in Fabry Perot cavities by making one of the mirrors a phase conjugating one, with a resulting improvement in the shape of the beam that emerges from the cavity, (*iii*) devices that can memorize an optical image and compare it to other images, (*iv*) the production of squeezed light (Ex. 9.10), and (*v*) focusing of laser light for laser fusion (Part V of this book).

As we shall see in the next section, phase conjugating mirrors rely crucially on the sinusoidal time evolution of the wave field; they integrate up that sinusoidal evolution coherently over some timescale $\hat{\tau}$ (typically microseconds to nanoseconds) in order to produce the phase conjugated wave. Correspondingly, if an incoming wave varies on timescales τ long compared to this $\hat{\tau}$ (e.g., if it carries a temporal modulation with bandwidth $\Delta\omega \sim 1/\tau$ small compared to $1/\hat{\tau}$, then the wave's temporal modulations will *not* get time reversed by the phase conjugating mirror. For example, if the wave impinging on a phase conjugating mirror has a frequency that is ω_a initially, and then gradually, over a time τ, increases to $\omega_b = \omega_a + 2\pi/\tau$, then the phase conjugated wave will *not* emerge from the mirror with frequency ω_b first and ω_a later. Rather, it will emerge with ω_a first and ω_b later (same order as for the original wave). When the incoming wave's temporal variations are fast compared to the mirror's integration time, $\tau \ll \hat{\tau}$, the mirror encounters a variety of frequencies during its integration time, and ceases to function properly. Thus, even though phase conjugation is equivalent to time reversal in a formal sense, a phase conjugating mirror cannot time reverse a temporal signal. It only time reverses monochromatic waves (which might carry a spatial signal).

9.5 Wave-Wave Mixing in Nonlinear Crystals

9.5.1 Nonlinear Dielectric Susceptibilities

In any medium that is electrically polarized, with polarization (electric dipole moment per unit volume) $\mathbf{P}$, but that has no free charges or currents, Maxwell's equations in SI units take the standard form

$$\boldsymbol{\nabla} \cdot \mathbf{E} = -\frac{\boldsymbol{\nabla} \cdot \mathbf{P}}{\epsilon_0}, \quad \boldsymbol{\nabla} \cdot \mathbf{B} = \mathbf{0}\ , \quad \boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial \mathbf{t}}\ , \quad \boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \Big(\frac{\partial \mathbf{P}}{\partial \mathbf{t}} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial \mathbf{t}} \Big)\ . \tag{9.15}$$

If the electric field is sufficiently weak and the medium is isotropic (the case treated in most textbooks on electromagnetic theory), the polarization $\mathbf{P}$ is proportional to the electric field: $\mathbf{P} = \epsilon_0 \chi_0 \mathbf{E}$, where χ_0 is the medium's electrical susceptibility. In this case the medium does not introduce any nonlinearities into Maxwell's equations. In many dielectric media, however, a strong electric field can produce a polarization that is nonlinear in the field. In such "nonlinear media," the general expression for the polarization $\mathbf{P}$ in terms of the electric field is

$$P_i = \epsilon_0 (\chi_{ij} E_j + \chi_{ijk} E_j E_k + \chi_{ijkl} E_j E_k E_l + \ldots)\ . \tag{9.16}$$

Here χ_{ij}, the linear susceptibility, is proportional to the 3-dimensional metric, $\chi_{ij} = \chi_0 g_{ij}$, if the medium is isotropic (i.e., if all directions in it are equivalent), but otherwise is more complicated; and the χ_{ijk}, χ_{ijkl}, ... are nonlinear susceptibilities. The normalizations used for these susceptibilities differ from one researcher to another; sometimes the factor ϵ_0 is omitted; sometimes factors of 2 or 4 or ... are inserted.

When the nonlinear susceptibilities are important, the nonlinearities lead to harmonic generation—i.e., the production of waves with frequencies 2ω, 3ω, ... from an input wave with frequency ω; see below. As a result, an electric field in the medium cannot oscillate at just one frequency, and each of the electric fields in expression (9.16) for the polarization must be a sum of pieces with different frequencies. Because the susceptibilities can depend on frequency, this means that when using expression (9.16) one sometimes must break P_i and each E_i up into its frequency components and use different values of the susceptibility to couple the different frequencies together. For example, one of the terms in Eq. (9.16) will become

$$P_i^{(1)} = \epsilon_0 \chi_{ijkl}^{(1234)} E_j^{(2)} E_k^{(3)} E_l^{(4)} , \tag{9.17}$$

where $P_i^{(1)}$ oscillates at frequency ω_1, $E_j^{(A)}$ oscillates at frequency ω_A, and $\chi_{ijkl}^{(1234)}$ depends on the four frequencies $\omega_1, \ldots, \omega_4$. Although this is complicated in the general case, in most practical applications resonant couplings (or equivalently energy and momentum conservation for photons) guarantee that only a single set of frequencies is important, and the resulting analysis simplifies substantially.

Because all the tensor indices on the susceptibilities except the first index get contracted into the electric field in expression (9.16), we are free to (and it is conventional to) define the susceptibilities as symmetric under interchange of any pair of indices that does not include the first. When [as has been tacitly assumed in Eq. (9.16)] there is no hysteresis in the medium's response to the electric field, the energy density of interaction between the polarization and the electric field is

$$\varepsilon = -\epsilon_0 \left(\frac{\chi_{ij} E_i E_j}{2} + \frac{\chi_{ijk} E_i E_j E_k}{3} + \frac{\chi_{ijkl} E_i E_j E_k E_l}{4} + \cdots \right) , \tag{9.18}$$

and the polarization is related to this energy of interaction, in Cartesian coordinates, by

$$P_i = -\frac{\partial \varepsilon}{\partial E_i} , \tag{9.19}$$

which agrees with Eq. (9.16) providing the susceptibilities are symmetric under interchange of *all* pairs of indices, including the first. We shall assume such symmetry.[4] If the crystal is isotropic, then each of its tensorial susceptibilities is constructable from the metric $g_{ij} = \delta_{ij}$ and a single scalar susceptibility; see Ex. 9.4:

$$\chi_{ij} = \chi_0 g_{ij} , \quad \chi_{ijk} = 0 , \quad \chi_{ijkl} = \frac{1}{3}\chi_4 (g_{ij} g_{kl} + g_{ik} g_{jl} + g_{il} g_{jk}) , \quad \cdots . \tag{9.20}$$

A simple model of a crystal which explains how nonlinear susceptibilities can arise is the following. Imagine each ion in the crystal as having a valence electron that can oscillate

[4]For further details see, e.g., Secs. 16.2–16.4 and 16.7 of Yariv (1989).

in response to a sinusoidal electric field. The electron can be regarded as residing in a potential well which, for low-amplitude oscillations, is very nearly harmonic (potential energy quadratic in displacement; restoring force proportional to displacement; "spring constant" independent of displacement). However, if the electron's displacement from equilibrium becomes a significant fraction of the interionic distance, it will begin to feel the electrostatic attraction of the neighboring ions, and its spring constant will weaken. This means that the potential the electron sees is really not that of a harmonic oscillator, but rather that of an *anharmonic oscillator*, $V(x) = \alpha x^2 - \beta x^3 + \cdots$, where x is the electron's displacement from equilibrium. The nonlinearities in this potential cause the electron's amplitude of oscillation, when driven by a sinusoidal electric field, to be nonlinear in the field strength, and that nonlinear displacement causes the crystal's polarization to be nonlinear.[5] For most crystals, the spatial arrangement of the ions causes the electron's potential energy V to be different for displacements in different directions, and this causes the susceptibilities to be anisotropic.

Because the total energy required to liberate the electron from its lattice site is roughly one eV, and the separation between lattice sites is $\sim 10^{-10}$ m, the characteristic electric field for strong instantaneous nonlinearities is $\sim 1\mathrm{V}/10^{-10}\mathrm{m} = 10^{10}\mathrm{V\ m}^{-1}$. Correspondingly, since χ_{ijk} has dimensions 1/(electric field) and χ_{ijkl} has dimensions 1/(electric field)^2, the largest that we can expect their Cartesian components to be is $\chi_{ijk} \sim 10^{-10}\mathrm{m\ V}^{-1}$, and $\chi_4 \sim \chi_{ijkl} \sim 10^{-20}\mathrm{m}^2\ \mathrm{V}^{-2}$. These numbers, together with the fact that the electric fields in light waves typically are not anywhere near as intense as $10^{10}\mathrm{V\ m}^{-1}$, dictates that, unless the third-order χ_{ijk} are suppressed by isotropy, they will produce much larger effects than the fourth-order χ_{ijkl}, which in turn will dominate over all higher orders.

Among the dielectric crystals with especially strong nonlinear susceptibilities are barium titonate ($\mathrm{BaTiO_R}$) and lithium niobate ($\mathrm{LiNbO_R}$); they have $\chi_{ijk} \sim (1 \text{ to } 10) \times 10^{-11}$ (Volt/meter)$^{-1}$ at optical frequencies; i.e., they get as large as our rough estimate for the upper limit.

9.5.2 Wave-Wave Mixing

The nonlinear susceptibilities produce wave-wave mixing when a beam of light is sent through a crystal. The mixing produced by χ_{ijk} is called three-wave mixing because three electric fields appear in the polarization-induced interaction energy, Eq. (9.18); and the mixing produced by χ_{ijkl} is similarly called four-wave mixing. Three-wave mixing dominates in an anisotropic medium, but is suppressed when the medium is isotropic, leaving four-wave mixing as the leading-order nonlinearity.

Let us examine three-wave mixing in a general anisotropic crystal. Because the nonlinear susceptibilities are so small (i.e., because the input waves will always be far weaker than $10^{10}\mathrm{V\ m}^{-1}$, the nonlinearities can be regarded as small perturbations. Suppose that two waves, labeled $A = 1$ and $A = 2$, are injected into the anisotropic crystal, and let their wave vectors be $\mathbf{k}_A$ when one ignores the (perturbative) nonlinear susceptibilities but keeps the large linear χ_{ij}. Because χ_{ij} is an anisotropic function of frequency, the dispersion relation for these waves (ignoring the nonlinearities), $\Omega(\mathbf{k})$, will typically be anisotropic. The frequencies

[5] Quantitative details are worked out, e.g., in Sec. 16.3 of Yariv (1989).

of the two input waves satisfy this dispersion relation, $\omega_A = \Omega(\mathbf{k}_A)$, and their forms are

$$E_j^{(A)} = \Re\left(\mathcal{E}_j^{(A)} e^{i(\mathbf{k}_A\cdot\mathbf{x}-\omega_A t)}\right) = \frac{1}{2}\left(\mathcal{E}_j^{(A)} e^{i(\mathbf{k}_A\cdot\mathbf{x}-\omega_A t)} + \mathcal{E}_j^{(A)*} e^{i(-\mathbf{k}_A\cdot\mathbf{x}+\omega_A t)}\right) , \tag{9.21}$$

where we have denoted their complex amplitudes by $\mathcal{E}_j^{(A)}$. We shall adopt the convention that wave 1 is the one with the larger frequency, so $\omega_1 - \omega_2 \geq 0$.

These two input waves couple, via the third-order nonlinear susceptibility χ_{ijk}, to produce the following contribution to the medium's polarization vector:

$$\begin{aligned} P_i &= 2\epsilon_0 \chi_{ijk} E_j^{(1)} E_k^{(2)} \\ &= \epsilon_0 \chi_{ijk} \Re\left(\mathcal{E}_j^{(1)}\mathcal{E}_k^{(2)} e^{i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{x}} e^{i(\boldsymbol{\omega}_1+\omega_2)t} \;+\; \mathcal{E}_j^{(1)}\mathcal{E}_k^{(2)*} e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{x}} e^{i(\boldsymbol{\omega}_1-\omega_2)t}\right) . \end{aligned} \tag{9.22}$$

This sinusoidally oscillating polarization produces source terms in Maxwell's equations (9.15): an oscillating, polarization-induced charge density $\rho_P = -\boldsymbol{\nabla}\cdot P$ and current density $\mathbf{j}_P = \partial\mathbf{P}/\partial t$. This polarization charge and current, like $\mathbf{P}$ itself [Eq. (9.22)], consist of two traveling waves, one with frequency and wave vector

$$\omega_3 = \omega_1 + \omega_2 , \quad \mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2 ; \tag{9.23}$$

the other with frequency and wave vector

$$\omega_3 = \omega_1 - \omega_2 , \quad \mathbf{k}_3 = \mathbf{k}_1 - \mathbf{k}_2 . \tag{9.24}$$

If either of these $(\omega_3, \mathbf{k}_3)$ satisfies the medium's dispersion relation $\omega = \Omega(\mathbf{k})$, then the charge-current wave will generate electromagnetic waves that propagate along in resonance with itself. These new electromagnetic waves, with frequency ω_3 and wave vector $\mathbf{k}_3$ will grow, as they propagate, with their growth being along the direction of the group velocity $V_g^j = (\partial\Omega/\partial k_j)_{\mathbf{k}=\mathbf{k}_3}$. The waves will be weakest at the "back" of the crystal (the side where $\mathbf{V}_g\cdot\mathbf{x}$ is smallest), and strongest at the "front" (the side where $\mathbf{V}_g\cdot\mathbf{x}$ is largest).

For most choices of the input waves, i.e. most choices of $\mathbf{k}_1$, $\omega_1 = \Omega(\mathbf{k}_1)$, $\mathbf{k}_2$, and $\omega_2 = \Omega(\mathbf{k}_2)$, neither of the charge-density waves, ($\mathbf{k}_3 = \mathbf{k}_1 \pm \mathbf{k}_2$, $\omega_3 = \omega_1 \pm \omega_2$) will satisfy the medium's dispersion relation, and thus neither will be able to create a third electromagnetic wave resonantly; the wave-wave coupling is ineffective. However, for certain special choices of the input waves, resonant coupling *will* be achieved, and a strong third wave will be produced.

The resonance conditions have simple quantum mechanical interpretations—a fact that is not at all accidental: quantum mechanics underlies the classical theory that we are developing. Each classical wave is carried by photons which have discrete energies $E_A = \hbar\omega_A$ and discrete momenta $\mathbf{p}_A = \hbar\mathbf{k}_A$. The input waves are able to produce, resonantly, waves with $\omega_3 = \omega_1 \pm \omega_2$ and $\mathbf{k}_3 = \mathbf{k}_1 \pm \mathbf{k}_2$, if those waves satisfy the dispersion relation. Restated in quantum mechanical terms, the condition of resonance with the "+" sign rather than the "−" is

$$E_3 = E_1 + E_2 , \quad \mathbf{p}_3 = \mathbf{p}_1 + \mathbf{p}_2 . \tag{9.25}$$

This has the quantum mechanical meaning that one photon of energy E_1 and momentum $\mathbf{p}_1$, and another of energy E_2 and momentum $\mathbf{p}_2$ combine together, via the medium's nonlinearities, and are annihilated (in the language of quantum field theory), and by their annihilation they create a new photon with energy $E_3 = E_1 + E_2$ and momentum $\mathbf{p}_3 = \mathbf{p}_1 + \mathbf{p}_2$. Thus, the classical condition of resonance is the quantum mechanical condition of energy-momentum conservation for the sets of photons involved in a quantum annihilation and creation process. For this process to proceed, not only must energy-momentum conservation be satisfied, but all three photons must have energies and momenta that obey the photons' semiclassical Hamiltonian relation $E = H(\mathbf{p})$ (i.e., the dispersion relation $\omega = \Omega(\mathbf{k})$ with $H = \hbar\Omega$, $E = \hbar\omega$, and $\mathbf{p} = \hbar\mathbf{k}$).

Similarly, the classical conditions of resonance with the "−" sign rather than the "+" can be written (after bringing photon 2 to the left-hand side) as

$$E_3 + E_2 = E_1 \ , \quad \mathbf{p}_3 + \mathbf{p}_2 = \mathbf{p}_1 \ . \tag{9.26}$$

This has the quantum mechanical meaning that one photon of energy E_1 and momentum $\mathbf{p}_1$ gets annihilated, via the medium's nonlinearities, and from its energy and momentum there are created two photons, with energies E_2, E_3 and momenta $\mathbf{p}_2$, $\mathbf{p}_3$ that satisfy energy-momentum conservation. This process has a rate (number of such photon annihilation/creation events per second) that is proportional to the number of photons (the total energy) in wave 1 because those photons are being absorbed, and also proportional to the number (the total energy) in wave 2 because those photons are being created via stimulated emission. Since the energies are proportional to $|\mathcal{E}_1|^2$ and $|\mathcal{E}_2|^2$, and since the rate of reactions is proportional to the rate of growth of $|\mathcal{E}_3|^2$ from its initial zero value, this quantum rate dependence, rate $\propto |\mathcal{E}_1|^2|\mathcal{E}_2|^2$, means that $\mathcal{E}_3$ must grow at a rate proportional to $|\mathcal{E}_1||\mathcal{E}_2|$, which is indeed the case as we shall see in the following example:

9.5.3 Resonance Conditions and Growth Equation for an Idealized, Dispersion-Free, Isotropic Medium

Consider, as an example, the simple and idealized case where the linear part of the susceptibility χ_{ij} is isotropic and frequency-independent, $\chi_{jk} = \chi_0 g_{jk}$. Then the dispersion relation, ignoring the nonlinearities, takes the simple, nondispersive form (Ex. 9.5)

$$\omega = \frac{c}{n}k, \quad \text{where } k = |\mathbf{k}|, \quad n = \sqrt{1+\chi_0} \tag{9.27}$$

with n a constant. Let the input waves $\mathbf{E}_1$ and $\mathbf{E}_2$ both propagate in the z direction. Then the resonance condition (energy-momentum conservation for photons) requires that the new, third wave also propagate in the z direction, with frequency and wave number $\omega_3 = \omega_1 \pm \omega_2$, $\mathbf{k}_3 = \mathbf{k}_1 \pm \mathbf{k}_2$. This is a highly unusual situation: because the input waves satisfy the dispersion relation (9.27), so also do *both* of the possible output waves. Making use of the fact that the lengthscale on which the new waves grow is long compared to a wavelength (which is always the case because the fields are always much weaker than $10^{10}\mathrm{Vm}^{-1}$), Maxwell's equations (9.15) imply the following growth equation for the new waves' amplitudes (Ex. 9.5):

$$\begin{aligned}
\frac{d\mathcal{E}_i^{(3)}}{dz} &= -2i\frac{k_3}{n^2}\chi_{ijk}^{(123)}\mathcal{E}_j^{(1)}\mathcal{E}_k^{(2)} \quad \text{at } \omega_3 = \omega_1 + \omega_2 \ , \quad k_3 = k_1 + k_2 \\
\frac{d\mathcal{E}_i^{(3)}}{dz} &= -2i\frac{k_3}{n^2}\chi_{ijk}^{(123)}\mathcal{E}_j^{(1)}\mathcal{E}_k^{(2)*} \quad \text{at } \omega_3 = \omega_1 - \omega_2 \ , \quad k_3 = k_1 - k_2.
\end{aligned} \tag{9.28}$$

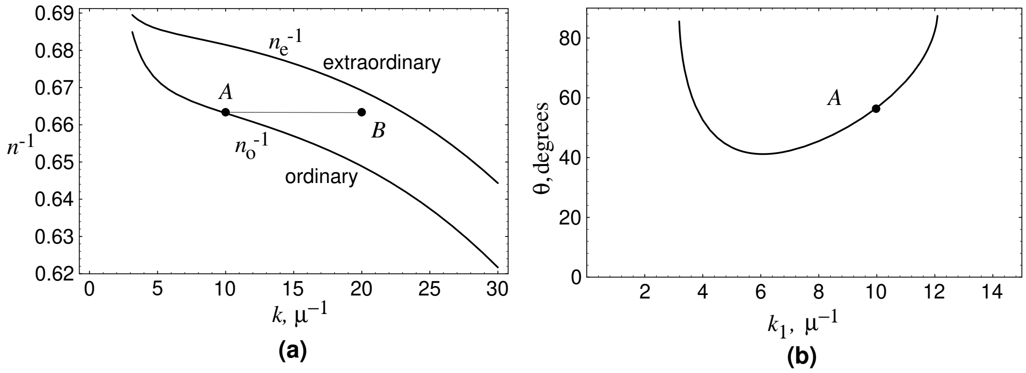

Fig. 9.9: (a) The inverse of the index of refraction (equal to the phase speed in units of the speed of light) for electromagnetic waves propagating along or orthogonal to the symmetry axis of a KDP crystal, as a function of wave number k in reciprocal microns. (b) The angle to the symmetry axis at which waves with wave number k_1 must propagate in order to be frequency doubled or phase conjugated via 3-wave mixing.

9.5.4 Resonance Conditions and Growth Equation for an Anisotropic, Axisymmetric Medium

In reality, all nonlinear media have frequency-dependent dispersion relations and many are anisotropic. An example is the crystal KH_2PO_4, also called "KDP", which is axisymmetric. If we orient its symmetry axis along the z direction, then its linear susceptibility χ_{ij} has as its only nonzero components $\chi_{11} = \chi_{22}$ and χ_{33}, which we embody in two indices of refraction,

$$n_o = \sqrt{1 + \chi_{11}} = \sqrt{1 + \chi_{22}}\ , \quad n_e = \sqrt{1 + \chi_{33}}\ , \tag{9.29}$$

that depend on frequency as shown in Fig. 9.9 (a).

Maxwell's equations imply (Ex. 9.6) that for plane, monochromatic waves propagating in the $x - z$ plane at an angle θ to the symmetry axis [$\mathbf{k} = k(\sin\theta \mathbf{e}_x + \cos\theta \mathbf{e}_z)$], there are two dispersion relations corresponding to the two polarizations of the electric field: (i) If $\mathbf{E}$ is orthogonal to the symmetry axis, then it must also be orthogonal to the propagation direction (i.e., must point in the $\mathbf{e}_y$ direction), the waves are called *ordinary*, and their dispersion relation is

$$\frac{\omega/k}{c} = (\text{phase speed in units of speed of light}) = \frac{1}{n_o}\ . \tag{9.30}$$

This phase speed is the lower curve in Fig. 9.9 (a); at $k = 10\mu^{-1}$, the phase speed is $0.663c$, while at $k = 20\mu^{-1}$, it is $0.649c$. (ii) If $\mathbf{E}$ is not orthogonal to the symmetry axis, then it must lie in the plane formed by $\mathbf{k}$ and the symmetry axis (the $x - z$) plane, with $E_x/E_z = -(n_e/n_o)^2 \cot\theta$ [which means that $\mathbf{E}$ is *not* orthogonal to the propagation direction unless the crystal is isotropic, $n_e = n_o$]; in this case the waves are called *extraordinary* and their dispersion relation is

$$\frac{\omega/k}{c} = \frac{1}{n} = \sqrt{\frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2}}\ . \tag{9.31}$$

As the propagation direction varies from parallel to the symmetry axis ($\cos\theta = 1$) to perpendicular ($\sin\theta = 1$), this extraordinary phase speed varies from c/n_o ($0.663c$ at $k = 10\mu^{-1}$) to c/n_e ($0.681c$ at $k = 10\mu^{-1}$).

Now, consider the resonance conditions for a *frequency-doubling device* (discussed in greater detail in the next section): one in which the two input waves are identical, so $\vec{k}_1 = \vec{k}_2$ and $\vec{k}_3 = 2\vec{k}_1$ point in the same direction. Let this common propagation direction be at an angle θ to the symmetry axis. Then the resonance conditions reduce to the demands that the output wave number be twice the input wave number, $k_3 = 2k_1$, and the output phase speed be the same as the input phase speed, $\omega_3/k_3 = \omega_1/k_1$. Now, for waves of the same type (both ordinary or both extraordinary), the phase speed is a monotonic decreasing function of wave number [Fig. 9.9 and Eq. (9.30), (9.31)], so there is no choice of propagation angle θ that enables these resonance conditions to be satisfied. The only way to satisfy them is by using ordinary input waves and extraordinary output waves, and then only for a special, frequency-dependent propagation direction.

For example, if the input waves are ordinary, with $k = 10\mu^{-1}$ [point A in Fig. 9.9 (a)], then the output waves must be extraordinary and must have the same phase speed as the input waves [same height in Fig. 9.9 (a); i.e., point B]. This phase speed is between c/n_e and c/n_o, and thus can be achieved for a special choice of propagation angle: $\theta = 56.7^{\rm o}$ [point A in Fig. 9.9 (b)]. In general, Eq. (9.30) , (9.31) imply that the unique propagation direction θ at which the resonance conditions can be satisfied is the following function of the input wave number k_1:

$$\sin^2\theta = \frac{1/n_o^2(k_1) \ - \ 1/n_o^2(2k_1)}{1/n_e^2(2k_1) - 1/n_o^2(2k_1)} \ . \tag{9.32}$$

This resonance angle is plotted as a function of frequency for KDP in Fig. 9.9 (b).

This special case of identical input waves illustrates the very general phenomenon, that at fixed input frequencies, the resonance conditions can be satisfied only for special, discrete input and output directions.

Also quite generally, once one has found wave vectors and frequencies that satisfy the resonance conditions, the growth rate of the new waves is governed by equations similar to Eq. (9.28), but with the growth along the direction of the group velocity of the new, growing wave; see Ex. 9.7. For example, in any 3-wave process with $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$ and $\omega_3 = \omega_1 + \omega_2$, the rate of growth of the new wave with distance s along the group-velocity direction is (Ex. 9.7)

$$\frac{d\mathcal{E}_a^{(3)}}{ds} = \left(-2i\alpha\frac{k_3}{{n_3}^2}\chi_{ijk}^{(123)}\mathcal{E}_i^{(1)}\mathcal{E}_j^{(2)}f_k^{(3)}\right)f_a^{(3)} \ , \tag{9.33}$$

where α is a coefficient of order unity that depends on the dispersion relation, and $f_a^{(3)}$ is a unit vector pointing along the electric field direction of the new wave (3).

For our frequency doubling example, the extraordinary dispersion relation Eq. (9.31) for the output wave can be rewritten as

$$\omega = \frac{ck}{n} = \Omega_e(\mathbf{k}) = c\sqrt{\frac{k_z^2}{n_o(k)^2} + \frac{k_x^2}{n_e(k)^2}} \ , \quad \text{where} \quad k = \sqrt{k_x^2 + k_z^2} \ . \tag{9.34}$$

Correspondingly, the group velocity $v_g^j = \partial\Omega/\partial k_j$ for the output waves has components

$$\begin{aligned} v_g^x &= v_{\rm ph} \sin\theta \left(\frac{n^2}{n_e^2} - \frac{n^2 \cos^2\theta}{n_o^2} \frac{d\ln n_o}{d\ln k} - \frac{n^2 \sin^2\theta}{n_e^2} \frac{d\ln n_o}{d\ln k} \right) , \\ v_g^z &= v_{\rm ph} \cos\theta \left(\frac{n^2}{n_o^2} - \frac{n^2 \cos^2\theta}{n_o^2} \frac{d\ln n_o}{d\ln k} - \frac{n^2 \sin^2\theta}{n_e^2} \frac{d\ln n_o}{d\ln k} \right) , \end{aligned} \tag{9.35}$$

where $v_{\rm ph} = \omega/k = c/n$ is the phase velocity. For ordinary input waves with $k_1 = 10\mu^{-1}$ and extraordinary output waves with $k_3 = 20\mu^{-1}$, these formulae give for the direction of the output group velocity (direction along which the output waves grow) $\theta_g = \arctan(v_g^x/v_g^z) = 58.4^{\rm o}$, compared to the direction of the common input-output phase velocity $\theta = 56.7^{\rm o}$; and they give for the magnitude of the group velocity $v_g = 0.628c$, compared to the common phase velocity $v_{\rm ph} = 0.663c$. Thus, the differences between group velocity and phase velocity are small.

****

EXERCISES

Exercise 9.4 *Derivation and Example: Nonlinear Susceptibilities for an Isotropic Medium*

Explain why the nonlinear susceptibilities for an isotropic medium have the forms given in Eq. (9.20). [Hint: Use the facts that the χ's must be symmetric in all their indices, and that, because the medium is isotropic, the χ's must be constructable from the only isotropic tensors available to us, the (symmetric) metric tensor g_{ij} and the (antisymmetric) Levi-Civita tensor ϵ_{ijk}.] What are the corresponding forms, in an isotropic medium, of χ_{ijklm} and χ_{ijklmn}?

Exercise 9.5 *Derivation: Growth Equation in Idealized Three-Wave Mixing*

Use Maxwell's equations to derive the propagation equations (9.28) for the new wave produced by three-wave mixing under the idealized dispersion-free conditions described in the text.

Exercise 9.6 *Example: Dispersion Relation for an Anisotropic Medium*

Consider a wave propagating through a dielectric medium that is anisotropic, but not necessarily—for the moment—axisymmetric. Let the wave be sufficiently weak that nonlinear effects are unimportant. Define the wave's displacement vector in the usual way, $D_i = \epsilon_0 E_i + P_i$.

(a) Show that

$$D_i = \epsilon_o \epsilon_{ij} E_i \ , \quad \text{where } \epsilon_{ij} \equiv \delta_{ij} + \chi_{ij} \equiv \text{“dielectric tensor”}; \tag{9.36}$$

ϵ_o is often absorbed into the dielectric tensor, but we find it more convenient to normalize ϵ_{ij} so that in vacuum $\epsilon_{ij} = \delta_{ij}$.

(b) Show that, in terms of this displacement vector, Maxwell's equations (9.15) reduce to the familiar form for a dielectric medium (Jackson 1998)

$$\boldsymbol{\nabla}\cdot\mathbf{D}=\mathbf{0}\,,\quad \boldsymbol{\nabla}\cdot\mathbf{B}=\mathbf{0}\,,\quad \boldsymbol{\nabla}\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial\mathbf{t}}\,,\quad \boldsymbol{\nabla}\times\mathbf{B}=\mu_{\mathbf{o}}\frac{\partial\mathbf{D}}{\partial\mathbf{t}}\,. \tag{9.37}$$

(c) By taking the curl of the third of these equations and the time derivative of the fourth and then combining them, show that

$$-\nabla^2\mathbf{E}+\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{E})=-\boldsymbol{\epsilon}\cdot\frac{\partial^2\mathbf{E}}{\partial t^2}\,. \tag{9.38}$$

(d) Now specialize to a monochromatic plane wave with angular frequency ω and wave vector $\mathbf{k}$. Show that the wave equation (9.38) reduces to

$$L_{ij}E_j=0\,,\quad \text{where}\quad L_{ij}=k_ik_j-k^2\delta_{ij}+\frac{\omega^2}{c^2}\epsilon_{ij}\,. \tag{9.39}$$

This equation says that $\mathbf{E}$ is an eigenvector of L with vanishing eigenvalue, which is possible if and only if

$$\det||L_{ij}||=0\,. \tag{9.40}$$

This vanishing determinant is the waves' dispersion relation.

(e) Now specialize to an axisymmetric medium and orient the symmetry axis along the z direction, so the only nonvanishing components of ϵ_{ij} are $\epsilon_{11}=\epsilon_{22}$ and ϵ_{33}. Show that in this case L_{ij} has the form

$$||L_{ij}||=\left\|\begin{array}{ccc}(n_o/n)^2-\cos^2\theta & 0 & \sin\theta\cos\theta\\ 0 & (n_o/n)^2-1 & 0\\ \sin\theta\cos\theta & 0 & (n_e/n)^2-\sin^2\theta\end{array}\right\|\,, \tag{9.41}$$

and the dispersion relation (9.40) reduces to

$$\left(\frac{1}{n^2}-\frac{1}{n_o^2}\right)\left(\frac{1}{n^2}-\frac{\cos^2\theta}{n_o^2}-\frac{\sin^2\theta}{n_e^2}\right)=0\,, \tag{9.42}$$

where $1/n=\omega/kc$, $n_o=\sqrt{\epsilon_{11}}=\sqrt{\epsilon_{22}}$, and $n_e=\sqrt{\epsilon_{33}}$, in accord with Eq. (9.29).

(f) Show that this dispersion relation has the two solutions (ordinary and extrarodinary) discussed in the text, Eq. (9.30), (9.31), and show that the electric fields associated with these two solutions point in the directions described in the text.

Exercise 9.7 *Growth Equation in Realistic Wave-Wave Mixing*

Consider a wave-wave mixing process in which the new wave satisfies a linearized wave equation of the form

$$L_{ij}(-i\boldsymbol{\nabla},i\partial/\partial t)E_j=0\,, \tag{9.43}$$

where L_{ij} is some function of its indicated arguments. For an anisotropic dielectic medium, L_{ij} will have a form that can be read off Eqs. (9.38) and (9.39). We use the more general form (9.43) for the wave equation so that our analysis will be valid for waves in a plasma (Chap. 20) as well as in a dielectric. Let $\mathbf{k}$, ω, and $\mathbf{f}$ be a wave vector, angular frequency, and unit polarization vector such that $E_j = e^{-i\omega t + \mathbf{k}\cdot\mathbf{x}} f_j$ satisfies this wave equation.

(a) Suppose that several other waves interact nonlinearly to produce a piece $P_j^{\rm NL}$ of the polarization vector that propagates as a plane wave with wave vector and angular frequency $\mathbf{k}$ and ω, i.e., propagates in resonance with the new wave. Show that the this polarization vector along f_j will resonantly generate the new wave in a manner described by the equation

$$\left(\frac{\partial L_{ab}}{\partial \omega}\frac{\partial}{\partial t} - \frac{\partial L_{ab}}{\partial k_j}\frac{\partial}{\partial x_j}\right) E_b = i\mu_o \frac{\partial^2}{\partial t^2} P_c^{\rm NL} \ . \tag{9.44}$$

Here in the functional form of L_{ab}, $-i\boldsymbol{\nabla}$ has been replaced by $\mathbf{k}$ and $i\partial/\partial t$ by ω.

(b) Orient the axes of a (primed) coordinate system along the eigendirections of L_{ab}, so its only nonzero components (when evaluated for the resonant wave vector and frequency) are $L_{1'1'} = \lambda_1 = 0$, $L_{2'2'} = \lambda_2 \neq 0$, and $L_{3'3'} = \lambda_3 \neq 0$. (The vanishing of one of the eigenvalues is demanded by the dispersion relation $\det||L_{ab}|| = 0$.) Show that the group velocity of the resonant waves is given by

$$v_g^j = -\frac{\partial\lambda_1/\partial\omega}{\partial\lambda_1/\partial k_j} \ . \tag{9.45}$$

(c) By a computation in the primed coordinate system, show that the growth equation (9.44) for the new wave has the form

$$\frac{d}{dt}\mathcal{E}_{1'} = \frac{-i\mu_o\omega^2 P_{1'}^{\rm NL}}{\partial\lambda_1/\partial\omega} \ , \tag{9.46}$$

where d/dt is the time derivative moving with the group velocity,

$$\frac{d}{dt} = \frac{\partial}{\partial t} + v_g^J \frac{\partial}{\partial x^j} \ . \tag{9.47}$$

(d) For 3-wave mixing (e.g., frequency-doubling), show that this growth equation takes the form, Eq. (9.33), and evaluate the coefficient α.

****

9.6 Applications of Wave-Wave Mixing: Frequency Doubling, Phase Conjugation, and Squeezing

9.6.1 Frequency Doubling

Frequency doubling (also called *second harmonic generation*) is one of the most important applications of wave-wave mixing. As we have seen in the previous section, it is achieved by passing a single wave (which plays the role of both wave $A = 1$ and wave $A = 2$) through a nonlinear crystal, with the propagation direction chosen to satisfy the resonance conditions [Eq. (9.32) and Fig. 9.9]. Three-wave coupling produces an output wave $A = 3$ with $\omega_3 = 2\omega_1$ that grows with distance inside the crystal at a rate given by a variant of Eq. (9.28). By doing a sufficiently good job of satisfying the resonance conditions and choosing the thickness of the crystal appropriately, one can achieve close to 100% conversion of the input-wave energy into frequency-doubled energy.

As an example, the Neodymium:YAG (Nd^{3+}:YAG) laser, which is based on an Yttrium Aluminum Garnet crystal with trivalent Neodymium impurities, is the most attractive of all lasers for a combination of high frequency stability, moderately high power, and high efficiency. However, this laser operates in the infrared, at a wavelength of 1.0641 microns. For some purposes one wants optical light. This can be achieved by frequency doubling the output of the Nd^{3+}:YAG laser. Thereby one can convert nearly all of the laser's output power into 0.532 micron (green) light; cf. Ex. 9.7.

Frequency doubling also plays a key role in laser fusion, where intense, pulsed laser beams, focused on a pellet of fusion fuel, compress and heat the pellet to high densities and temperatures. Because the beam's energy flux is inversely proportional to the area of its focussed cross section, and because the larger the wavelength, the more seriously diffraction impedes making the cross section small, it is important to give the beam a very short wavelength. This is achieved by multiple frequency doublings.

9.6.2 Phase Conjugation

A second example of three-wave mixing is phase conjugation. Phase conjugation is achieved by pumping the crystal with a very strong plane wave $A = 1$ at a frequency $\omega_1 = 2\omega_2$, and using as wave $A = 2$ the much weaker one that is to be phase conjugated. In this case the propagation equation (9.28) produces the new wave, $A = 3$, with complex amplitude $\mathcal{E}_i^{(3)}$ proportional to $\mathcal{E}_k^{(2)*}$. An ordinary mirror can then be used to turn the new wave around (Fig. 9.10); the result is a fully phase-conjugated wave. Thus, the three-wave-coupling crystal and the ordinary mirror together function as a phase-conjugating mirror.

Four-wave mixing in an isotropic crystal can be used for phase conjugation, without the help of an ordinary mirror, and without the difficulties in maintaining temporal and spatial resonance simultaneously that plague three-wave mixing—but at the price of working with a much weaker nonlinearity.

Suppose that the goal is to phase conjugate a wave $A = 1$ that enters an isotropic crystal moving in the $+z$ direction. Wave $A = 4$ is to be the fully phase conjugated wave, and thus must leave the crystal moving in the $-z$ direction. Waves $A = 2$ and 3, used

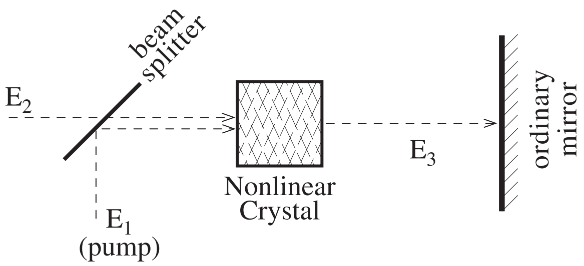

Fig. 9.10: A phase conjugating mirror based on 3-wave mixing.

to pump the crystal, are chosen to have the same frequency as the wave to be conjugated: $\omega_2 = \omega_3 = \omega_1 \ \equiv \ \omega$. The pump waves are injected into the crystal moving perpendicular to the incoming wave $A = 1$ and in opposite directions, so their wave vectors are $\mathbf{k}_2 = -\mathbf{k}_3 \perp \mathbf{k}_1$. Then the four-wave-mixing polarization produced by the two pump waves and the incoming wave will include a term

$$P_i = \epsilon_0 \Re\Big[\frac{1}{3}\chi_4 \mathcal{E}_j^{(2)} e^{i(\mathbf{k}_2\cdot\mathbf{x}-\omega t)} \mathcal{E}_j^{(3)} e^{i(\mathbf{k}_3\cdot\mathbf{x}-\omega t)} \mathcal{E}_i^{(1)*} e^{-i(kz-\omega t)}\Big] , \tag{9.48}$$

which generates via Maxwell's equations the fully phase-conjugated wave

$$E_i^{(4)} = \Re\Big[\text{const} \times \mathcal{E}_i^{(1)*} e^{-i(kz-\omega t)}\Big] . \tag{9.49}$$

See Exercise 9.9 for details.

****

EXERCISES

Exercise 9.8 *Problem: Efficiency of Frequency Doubling*

A Nd^{+3}:YAG laser puts out 10 Watts of linearly polarized light in a Gaussian beam at its lasing wavelength of 1.0641 microns. It is desired to frequency double a large fraction of this light using a $Ba_2NaNb_5O_{15}$ crystal for which the relevant component of the quadratic susceptibility has magnitude $|\chi_{ijk}| \sim 4 \times 10^{-11}$ (Volts/m)$^{-1}$. The crystal has a thickness of 1 centimeter. To what diameter d_o should the laser's light beam be focused before sending it through the crystal, in order to guarantee that a large fraction of its power will be frequency doubled? Show that diffraction effects are small enough that the beam diameter can remain $\simeq d_o$ all the way through the crystal. [Hint: Recall the spreading of a Gaussian-shaped beam, as described by Eqs. (7.38)–(7.41).]

Exercise 9.9 *Derivation and Example: Phase Conjugation by Four-Wave Mixing*

(a) Consider an idealized crystal which is isotropic and has scalar susceptibilities χ_0 and χ_4 that are independent of frequency. Show that Maxwell's equations in the crystal imply

$$-\left(\frac{n}{c}\right)^2 \frac{\partial^2 \mathbf{E}}{\partial t^2} + \nabla^2 \mathbf{E} = \text{const}\left[\frac{n^2}{c^2}\frac{\partial^2}{\partial t^2}(\mathbf{E}^2\mathbf{E}) - \boldsymbol{\nabla}\boldsymbol{\nabla}\cdot(\mathbf{E}^2\mathbf{E})\right] , \tag{9.50}$$

where n is the index of refraction and $\mathbf{E}^2 = \mathbf{E} \cdot \mathbf{E}$. What is the constant in terms of χ_4?

(b) Assume (as is always the case) that $\chi_4 \mathbf{E}^2 \ll 1$. Then Eq. (9.50) can be solved using perturbation theory. We shall do so, in this exercise, for the physical setup described at the end of Sec. 9.5.2, which produces a phase-conjugated output. As a concrete realization of that setup, assume that the incoming ("zero-order") waves inside the medium are (i) a signal wave which propagates in the z direction and has a slowly varying complex amplitude $\mathcal{E}^{(1)}(x, y, z)$ containing some sort of picture

$$\mathbf{E}^{(1)} = \Re[\mathcal{E}^{(1)}(x, y, z) e^{i(kz-\omega t)} \mathbf{e}_x] \; ; \tag{9.51}$$

and (ii) two pump waves which propagate in the x and $-x$ directions:

$$\mathbf{E}^{(2)} = \Re[\mathcal{E}^{(2)} e^{i(kx-\omega t)} \mathbf{e}_z] \; , \quad \mathbf{E}^{(3)} = \Re[\mathcal{E}^{(3)} e^{i(-kx-\omega t)} \mathbf{e}_z] \; , \tag{9.52}$$

Here $|\mathcal{E}^{(1)}| \ll |\mathcal{E}^{(2)}|$, $|\mathcal{E}^{(1)}| \ll |\mathcal{E}^{(3)}|$, and all the frequencies ω and wave numbers $k \equiv (n/c)\omega$ are identical. Give a list of the frequencies of all the new waves $\mathbf{E}^{(4)}$ that are generated from these three waves by four-wave mixing. Among those frequencies is that of the original three waves, ω. A narrow-band filter is placed on the output of the crystal to assure that only this frequency, ω, emerges.

(c) When one computes the details of the four-wave mixing using the propagation equation (9.50), one finds that the only new waves with frequency ω that can be generated have wave numbers $k = (n/c)\omega$. Explain why. Give a list of all such waves including (i) their propagation directions, (ii) their polarization directions, and (iii) their dependences on $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$.

(d) Show that the only wave which emerges from the crystal propagating in the $-z$ direction with frequency ω is the phase-conjugated signal wave. Suppose that the crystal extends along the z direction from $z = 0$ to $z = L$. What is the amplitude of this phase conjugated wave throughout the crystal and emerging from its front face, $z = 0$?

Exercise 9.10 *Important Example: Squeezed Light Produced by Phase Conjugation*

Consider a plane electromagnetic wave in which we ignore polarization, and for which we write the complex amplitude as $X_1 + iX_2$ so

$$E = \Re[(X_1 + iX_2) e^{i(kz-\omega t)}] \; . \tag{9.53}$$

This wave is slightly noisy: X_1 and X_2 are randomly varying functions of $t - zn/c$, with means $\bar{X}_1$, $\bar{X}_2$, variances $(\Delta X_1)^2 = (\Delta X_2)^2 \equiv \sigma^2$, and correlation time τ_*. One can describe such waves by an error box in the complex amplitude plane (Fig. 9.11 (a)). Because $(\Delta X_1)^2 = (\Delta X_2)^2$, the error box is round. This plane wave is split into two parts by a beam splitter, one part is reflected off a phase-conjugating mirror, the other part is reflected off an ordinary mirror, and the two parts are then recombined at the beam splitter.

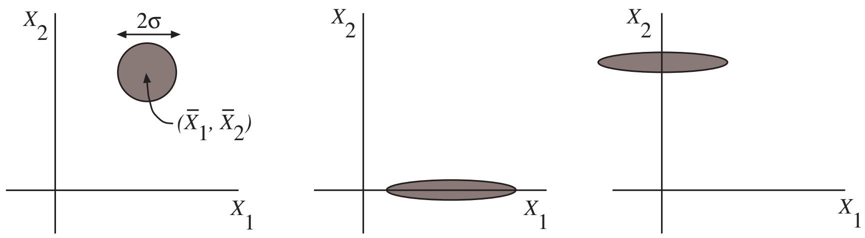

Fig. 9.11: Error boxes in the complex amplitude plane for several different waves: (a) The original wave discussed in Ex. 9.8. (b) A wave in a squeezed state with reduced phase noise. (c) A wave in a squeezed state with reduced amplitude noise.

(a) Suppose that the phase-conjugating mirror is a pumped, four-wave-mixing crystal of the type analyzed in Ex. 9.8, and that its length is L. Suppose, further, that the incoming wave's correlation time τ_* is long compared to the time, $2Ln/c$, required for light to propagate from one end of the crystal to the other and back. Explain why the phase-conjugating mirror will not time reverse the variations of the wave's complex amplitude.

(b) Suppose that the two mirrors (one phase-conjugating, the other ordinary) reflect their waves with slightly different efficiencies, so the beams that recombine at the beam splitter have complex amplitudes whose moduli differ by a fractional amount $\epsilon \ll 1$. Show that by appropriately adjusting the relative phase delay of the recombining beams, one can make the recombined light have the form

$$E = \alpha \Re[(2X_1 + i\epsilon X_2)e^{i(kz-\omega t)}] \, . \tag{9.54}$$

Here α is a constant and z is distance along the optic axis. In this recombined light the mean and the variance of X_2 are both reduced drastically (by a factor $\epsilon/2$) relative to those of X_1. The corresponding error box, shown in Fig. 9.11 (b), is squeezed along the X_2 direction. Correspondingly, the light itself is said to be in a *squeezed state*.

(c) By appropriately choosing the initial $\bar{X}_1$, $\bar{X}_2$, and squeeze factor ϵ, one can produce light with an error box of the form shown in Fig. 9.11 (c) rather than (b). How should they be chosen? Explain why the light in Fig. 9.11 (b) is said to have "reduced phase noise" and that in (c) to have "reduced amplitude noise."

(d) For the light with reduced phase noise, the electric field $E(t)$ measured at some fixed location in space lies inside the stippled band shown in Fig. 9.12, with 90 per cent confidence. Draw similar stippled bands depicting the electric field $E(t)$ for the original, unsqueezed light and for squeezed light with reduced amplitude noise.

Discussion: A quantum mechanical analysis shows that, not only is classical noise squeezed by the above process; so is quantum noise. However, in quantum mechanics X_1 and X_2 are Hermitean operators which do not commute, and correspondingly there is an uncertainty relation of the form $\Delta X_1 \Delta X_2 \geq$ (a quantum limit). As a result,

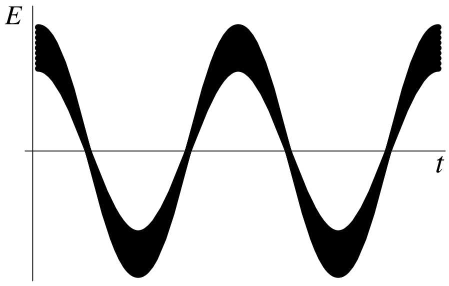

Fig. 9.12: The error band for the electric field $E(t)$, as measured at a fixed location in space when squeezed light with reduced phase noise passes by.

when one begins with minimum uncertainty light, $\Delta X_1 = \Delta X_2 =$(the minimum, corresponding to an energy $\frac{1}{2}\hbar\omega$ in the chosen wave mode), for example the vacuum state of quantum electrodynamics, squeezing that light will reduce ΔX_1 only at the price of increasing ΔX_2. Such squeezed light, including the "squeezed vacuum," has great promise for fundamental physics experiments and technology. For example, it can be used to reduce the photon shot noise of an interferometer or a laser below the "standard quantum limit" of $\Delta N = \sqrt{N}$, thereby improving the signal to noise ratio in certain communications devices, and in laser interferometer gravitational-wave detectors.[6]

****

Bibliography

Agrawal, G. P. & Boyd, R. W. eds. 1992 *Contemporary Nonlinear Optics* New York: Academic Press

Basov, N. G. & Prokhorov, A.M. 1954 JETP 27 431

Basov, N. G. & Prokhorov, A.M. 1955 JETP 28 249

Boyd, R. W. 1992 *Nonlinear Optics* New York: Academic Press

Cathey, W. T. 1974 *Optical Information Processing and Holography* New York: Wiley

Feynman, R. P., Leighton, R. B., & Sands, M. 1965 *The Feynman Lectures on Physics* New York: Addison Wesley

Ghatak, A. K. & Thyagarajan, K. 1978 *Contemporary Optics* New YorK: Plenum

Gordon, J. P., Zeiger, H. J., & Townes, C. H. 1954 Phys. Rev. 95 282

Gordon, J. P., Zeiger, H. J., & Townes, C. H. 1955 Phys. Rev. 99 1264

[6] For detailed discussions see, e.g., Walls (1983), Wu *et al.* (1986), and LaPorta *et al.* (1989).

Iizuka, K. 1987 *Engineering Optics* Berlin: Springer-Verlag

Jackson, J. D. 1999 *Classical Electrodynamics*New York: Wiley

La Porta, A., Slusher, R. E., & Yurke, B. 1989 Phys. Rev. Lett. 62 28

Maiman, T.H. 1960 *Nature* 187 493

Shen, Y.R. 1984 The Principles of Nonlinear Optics New York: Wiley

Walls, D.F. 1983 Nature 306 141

Weber, J. 1953 *IRE Trans. Prof. Group on Electron Devices* 3 1

Wu, L.-A., Kimble, H. J., Hall, J .L., & Wu, H. 1986 Phys. Rev. Lett. 57 2520

Yariv, A. 1977 J. Opt. Soc. Amer. 67 1

Yariv, A. 1989 Quantum Electronics New York: Wiley

Zel'dovich, B.Y., Popovichev, V.I., Ragulskii, V.V., & Faisullov, F.S. 1972 JETP 15 109