INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/453-12

TRAINING COLLEGE ON PHYSICS AND CHARACTERIZATION OF LASERS AND OPTICAL FIBRES

(5 February - 2 March 1990)

BASIC LASER PHYSICS

PROPERTIES OF LASER BEAMS

O. Svelto

Centro di Elettronica Quantistica
e Strumentazione Elettronica
Istituto di Fisica del Politecnico
Milano, Italy

pump level is empty, the rate at which the upper laser level 2 becomes populated by the pumping, $(dN_2/dt)_p$, can in general be written as

$$\left(\frac{dN_2}{dt}\right)_p = W_p N_g \tag{1.10}$$

Here N_g is the population of the ground level (i.e., level 1 or 0 in Figs. 1.4a and 1.4b, respectively) and W_p is a coefficient that will be called the *pump rate*. To achieve the threshold condition, the pump rate must reach a threshold or critical value that we shall indicate by W_{cp}. Specific expressions for W_{cp} will be obtained in Chapter 5.

1.4. PROPERTIES OF LASER BEAMS

Laser radiation is characterized by an extremely high degree of (1) monochromaticity, (2) coherence, (3) directionality, and (4) brightness. To these properties a fifth can be added, viz., (5) short time duration. This refers to the capability for producing very short light pulses, a property that, although perhaps less fundamental, is nevertheless very important. We shall now consider these properties in some detail.

1.4.1. Monochromaticity

Briefly, we can say that this property is due to the following two circumstances: (1) Only an e.m. wave of frequency ν given by (1.1) can be amplified. (2) Since the two-mirror arrangement forms a resonant cavity, oscillation can occur only at the resonant frequencies of this cavity. The latter circumstance leads to the laser linewidth being often much narrower (by as much as six orders of magnitude!) than the usual linewidth of the transition $2 \rightarrow 1$ as observed in spontaneous emission.

1.4.2. Coherence

To first order, for any e.m. wave, one can introduce two concepts of coherence, namely, spatial and temporal coherence.

To define spatial coherence, let us consider two points P_1 and P_2 that, at time $t = 0$, lie on the same wave front of some given e.m. wave and let $E_1(t)$ and $E_2(t)$ be the corresponding electric fields at these points. By definition, the difference between the phases of the two fields at time $t = 0$ is zero. Now, if this difference remains zero at any time $t > 0$, we will say that there is a perfect coherence between the two points. If this occurs for any two points of the e.m. wave front, we will say that the wave has *perfect spatial coherence.* In practice, for any point P_1, the point P_2 must lie within some finite area

around P_1 if we want to have a good phase correlation. In this case we will say that the wave has a *partial spatial coherence* and, for any point P, we can introduce a suitably defined coherence area $S_c(P)$.

To define temporal coherence, we now consider the electric field of the e.m. wave at a given point P, at times t and $t + \tau$. If, for a given time delay τ, the phase difference between the two field values remains the same for any time t, we will say that there is temporal coherence over a time τ. If this occurs for any value of τ, the e.m. wave will be said to have perfect time coherence. If this occurs for a time delay τ such that $0 < \tau < \tau_0$, the wave will be said to have partial temporal coherence, with a coherence time equal to τ_0. An example of an e.m. wave with a coherence time equal to τ_0 is shown in Fig. 1.5. This shows a sinusoidal electric field undergoing phase jumps at time intervals equal to τ_0. We see that the concept of temporal coherence is, in this case, directly connected with that of monochromaticity. We will in fact show in Chapter 7—as, indeed, is apparent from the example shown in Fig. 1.5—that a stationary e.m. wave with a coherence time τ_0 has a bandwidth $\Delta\nu \cong 1/\tau_0$. In the same chapter, however, it will be shown that, for a nonstationary beam (e.g., a Q-switched or a mode-locked laser beam), the coherence time is not related to the inverse of the oscillation bandwidth $\Delta\nu$ and may actually be much longer than $1/\Delta\nu$.

It is worth noting that the two concepts of temporal and spatial coherence are indeed independent of each other. In fact, examples can be given of a wave having perfect spatial coherence but only a limited temporal coherence (or vice versa). If, for instance, the wave shown in Fig. 1.5 were to represent the electric fields at points P_1 and P_2 mentioned earlier, the spatial coherence between these points would be complete while the wave would have a limited temporal coherence.

We conclude this section by emphasizing that the concepts of spatial and temporal coherence provide only a first-order description of the laser's coherence. Higher-order coherence properties will be discussed in Chapter 7. Such

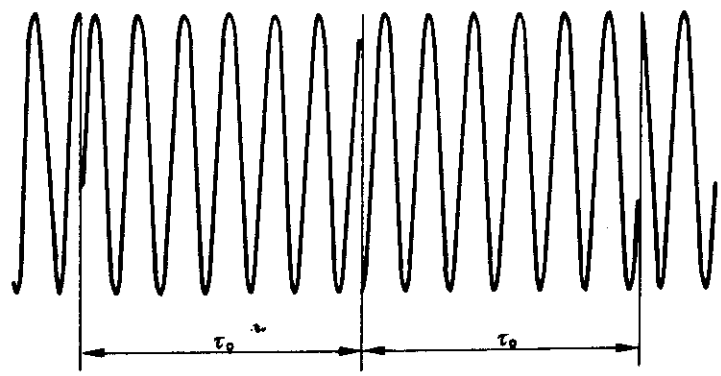

FIG. 1.5. Example of an e.m. wave with a coherence time of approximately τ_0.

a discussion is essential for a full appreciation of the difference between an ordinary light source and a laser. It will be shown in fact that, by virtue of the differences between the corresponding higher-order coherence properties, a laser beam is fundamentally different from an ordinary light source.

1.4.3. *Directionality*

This property is a direct consequence of the fact that the active material is placed in a resonant cavity such as the plane parallel one of Fig. 1.3. In fact, only a wave propagating along the cavity direction (or in a direction very near to it) can be sustained in the cavity. To gain a deeper understanding of the directional properties of laser beams (or, in general, of any e.m. wave), it is convenient to consider, separately, the case of a beam with perfect spatial coherence and the case of partial spatial coherence.

We first consider the case of perfect spatial coherence. Even for this case, a beam of finite aperture has an unavoidable divergence due to diffraction. This can be understood with the help of Fig. 1.6, where a beam of uniform intensity and plane wave front is assumed to be incident on a screen S containing an aperture D. According to Huygens' principle the wave front at some plane P behind the screen can be obtained from the superposition of the elementary waves emitted by each point of the aperture. We see that, on account of the finite size D of the aperture, the beam has a finite divergence θ_d. Its value can be obtained from diffraction theory. For an arbitrary amplitude distribution we get

$$\theta_d = \beta\lambda / D \tag{1.11}$$

where λ and D are the wavelength and the diameter of the beam. In equation (1.11) β is a numerical coefficient of the order of unity whose value depends on the shape of the amplitude distribution and on the way in which both the divergence and the beam diameter are defined. A beam whose divergence is given by equation (1.11) is described as being *diffraction limited.*

If the wave has partial spatial coherence, its divergence will be larger than the minimum value set by diffraction. Indeed, for any point P' of the wave front, the Huygens argument of Fig. 1.6 can only be applied for points

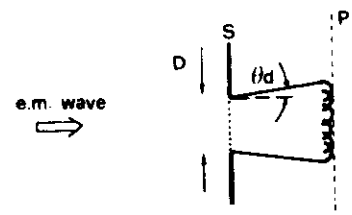

FIG. 1.6. **Divergence of a plane e.m. wave due to diffraction.**

lying within the coherence area S_c around point P'. The coherence area thus acts as a limiting aperture for the coherent superposition of the elementary wavelets. The beam divergence will now be given by

$$\theta_c = \beta\lambda/[S_c]^{1/2} \tag{1.12}$$

where again β is a numerical coefficient of the order of unity whose exact value depends on the way in which both the divergence θ_c and the coherence area S_c are defined.

We conclude this general discussion of the directional properties of e.m. waves by pointing out that, given suitable operating conditions, the output beam of a laser can be made diffraction limited.

1.4.4. *Brightness*

We define the brightness of a given source of e.m. waves as the power emitted per unit surface area per unit solid angle. To be more precise, let dS be the elemental surface area at point 0 of the source (Fig. 1.7). The power dP emitted by dS into a solid angle $d\Omega$ around the direction OO' can be written as

$$dP = B\cos\theta\, dS\, d\Omega \tag{1.13}$$

where θ is the angle between OO' and the normal $\mathbf{n}$ to the surface. The quantity B will generally depend on the polar coordinates θ and ϕ of the direction OO' and on the point O. This quantity B is called the source brightness at the point O in the direction of OO'. In equation (1.13) the factor $\cos\theta$ arises simply from the fact that the physically important quantity is the projection of dS onto a plane orthogonal to the OO' direction. When B is independent of θ and ϕ, the source is said to be isotropic (a Lambert source). A laser of even moderate power (e.g., a few milliwatts) has a brightness that is orders of magnitude greater than that of the brightest conventional sources. This is mainly due to the highly directional properties of the laser beam.

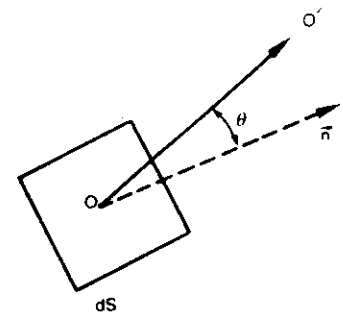

FIG. 1.7. Surface brightness at the point O for a source of e.m. waves.

1.4.5. Short Time Duration

Without going into any detail at this stage, we simply mention that by means of a special technique called mode locking, it is possible to produce light pulses whose duration is roughly equal to the inverse of the linewidth of the $2 \rightarrow 1$ transition. Thus, with gas lasers, whose linewidth is relatively narrow, the pulsewidth way be of ~0.1–1 ns. Such pulse durations are not regarded as particularly short because even some flashlamps can emit light pulses with a duration of somewhat less than 1 ns. On the other hand, the linewidth of solid state or liquid lasers can be 10^3–10^5 times larger than that of a gas laser, and, in this case, much shorter pulses may be generated (from 1 ps down to ~30 fs). This opens up exciting new possibilities for laser research and applications.

Notice that the property of short time duration, which implies energy concentration in time, can, in a sense, be considered to be the dual property of monochromaticity, which implies energy concentration in wavelength. Short time duration is, however, perhaps a less fundamental property than monochromaticity. While in fact all lasers can, in principle, be made extremely monochromatic, only lasers with a broad linewidth, and this in practice means solid state and liquid lasers, may produce pulses of very short time duration. On the other hand, gas lasers with their narrower lines lend themselves more readily to very monochromatic operation.

1.5. ORGANIZATION OF THE BOOK

The organization of the book is based on the fact that, as explained earlier in this chapter, a laser can be thought of as consisting of three elements: (1) an active material, (2) a pumping system, and (3) a suitable resonator. Accordingly, the next three chapters deal, respectively, with the interaction of radiation with matter, pumping processes, and the theory of passive optical resonators. The concepts introduced in this way are then used in Chapter 5 to develop a theory of the cw and transient behavior of lasers. The theory is developed within the lowest-order approximation, i.e., using the rate-equation approach. This treatment is, in fact, capable of describing most laser characteristics. Obviously, lasers based upon different types of active media have significant differences in their characteristics. It is therefore natural that, next, Chapter 6 should discuss the characteristic properties of each type of laser. By this point, the reader should have acquired a sufficient understanding of laser operation to go on to a study of the properties of the output beam (coherence, monochromaticity, directionality, brightness, noise). These properties are considered in Chapter 7. Finally, Chapter 8 has its basis in the fact that, before being put to use, a laser beam is generally transformed in some way. This includes (1) spatial transformation of the beam due to its propagation in

free-space or through a lens system; (2) amplitude transformation due to propagation through an amplifier; (3) wavelength transformation, or frequency conversion, due to a number of nonlinear phenomena (second harmonic generation, parametric generation).

PROBLEMS

1.1. That part of the e.m. spectrum that is of interest in the laser field starts from the submillimeter wave region and goes down in wavelength to the x-ray region. This covers the following regions in succession: (1) far infrared; (2) near infrared; (3) visible; (4) ultraviolet (uv); (5) vacuum ultraviolet (vuv); (6) soft x-ray; (7) x-ray. From standard textbooks find the wavelength intervals of the above regions. Memorize or record these intervals since they are frequently used in this book.

1.2. As a particular case of Problem 1.1, memorize or record the wavelengths corresponding to blue, green, and red light.

1.3. If levels 1 and 2 of Fig. 1.1 are separated by an energy $E_2 - E_1$ such that the corresponding transition frequency falls in the middle of the visible range, calculate the ratio of the populations of the two levels in thermal equilibrium at room temperature.

1.4. When in thermal equilibrium (at $T = 300$ K), the ratio of the level populations N_2/N_1 for some particular pair of levels is given by $1/e$. Calculate the frequency ν for this transition. In what region of the e.m. spectrum does this frequency fall?

1.5. A laser cavity consists of two mirrors with reflectivity $R_2 = 1$ and $R_1 = 0.5$. The length of the active material is $l = 7.5$ cm and the transition cross section is $\sigma = 3.5 \times 10^{-19}$ cm^2. Calculate the threshold inversion.

1.6. The beam from a ruby laser ($\lambda = 0.694\ \mu$m) is sent to the moon after passing through a telescope of 1 m diameter. Calculate the beam diameter D on the moon assuming that the beam has perfect spatial coherence (the distance between earth and moon is approximately 384,000 km).

7

Properties of Laser Beams

7.1. INTRODUCTION

In Chapter 1 it was stated that the most characteristic properties of laser beams are (1) monochromaticity, (2) coherence (spatial and temporal), (3) directionality, (4) brightness. The material presented in earlier chapters allows us to now examine these properties in more detail and compare them with the properties of conventional light sources (thermal sources).

7.2. MONOCHROMATICITY

If the laser is oscillating on a single mode and if the output is constant in time, the theoretical limit of monochromaticity arises from zero-point fluctuations and is given by (5.66). This limit, however, gives a very low value for the oscillating bandwidth $\Delta\nu_{osc}$ (a value of $\Delta\nu_{osc}/\nu_{osc} \simeq 10^{-15}$ was calculated in Section 5.3.7 for a laser power of 1 mW) which has never been reached in practice. Vibrations and thermal expansion of the cavity indeed limit $\Delta\nu_{osc}$ to much higher values. If a sufficiently massive structure made of material with a low expansion coefficient (e.g., Invar) is used to support the laser cavity, $\Delta\nu_{osc}$ can be reduced to a value in the range of 1-10 kHz. For a low-pressure gas laser (e.g., He-Ne) locked to the center of the absorption line of an appropriate gas, one can obtain[1] $\Delta\nu_{osc} = 50$-500 Hz (i.e., $\Delta\nu_{osc}/\nu = 10^{-12}$-$10^{-13}$). In pulsed operation the minimum linewidth is obviously limited by the inverse of the pulse duration τ_p. For example, for a single-mode giant pulse laser, assuming $\tau_p \simeq 10$ ns one has $\Delta\nu_{osc} \simeq 100$ MHz.

In the case of a laser oscillating on many modes, the monochromaticity is obviously related to the number of oscillating modes. For a solid-state laser (ruby, neodymium, semiconductor), where it is usually difficult to obtain single-mode oscillation (because of the large linewidth $\Delta\nu_0$) the oscillation

bandwidth is often of the order of gigahertz. Of course, one does not always want to have a very narrow oscillating bandwidth. We recall, for example, that in order to get very short pulses of light (mode locking), it is desirable to obtain oscillation over as wide a bandwidth as possible.

7.3. COMPLEX REPRESENTATION OF POLYCHROMATIC FIELDS

Before going any further with our discussion of the properties of laser beams it is appropriate to introduce a very useful complex representation for polychromatic fields (due to Gabor[2]). For the sake of simplicity we will consider a linearly polarized e.m. wave. This can then be specified by a single real scalar quantity $V^{(r)}(\mathbf{r}, t)$ (e.g., $|E|$ or $|H|$ or the modulus of the vector potential $|A|$). This quantity, which is a function of position $\mathbf{r}$ and time t, can be expressed as a Fourier integral, i.e.,

$$V^{(r)}(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} V(\mathbf{r}, \omega) \exp(-i\omega t)\, d\omega \tag{7.1}$$

Equation (7.1) has the well-known inverse relationship

$$V(\mathbf{r}, \omega) = \int_{-\infty}^{+\infty} V^{(r)}(\mathbf{r}, t) \exp(i\omega t)\, dt \tag{7.2}$$

Since $V^{(r)}$ is real, we see from (7.2) that

$$V(\mathbf{r}, -\omega) = V^*(\mathbf{r}, \omega) \tag{7.2a}$$

Hence, the negative frequency spectrum does not add any further information about the field to that already contained in the spectrum of positive frequencies. So, instead of $V^{(r)}$, we can consider the complex quantity $V(\mathbf{r}, t)$ defined by

$$V(\mathbf{r}, t) = \frac{1}{2\pi} \int_{0}^{\infty} V(\mathbf{r}, \omega) \exp(-i\omega t)\, d\omega \tag{7.3}$$

$V(\mathbf{r}, t)$ is called the complex analytic signal associated with $V^{(r)}$. Obviously there is a unique relation between the two functions. In fact, given V, we find from (7.1), (7.2a), and (7.3) that

$$V^{(r)} = 2\,\mathrm{Re}(V) \tag{7.4}$$

Conversely, it is easy to see that, if $V^{(r)}$ is given, then V is uniquely determined. In fact, given $V^{(r)}$, then from (7.2) we obtain $V(\mathbf{r}, \omega)$. With the help of (7.3), we then get $V(\mathbf{r}, t)$.

The analytic signal V proves much more convenient than the real signal as a way of representing the e.m. field. If, for example, the real signal is monochromatic, we can write $V^{(r)} = E \cos \omega t$. Hence, from (7.2) and (7.3) we have $V = E \exp(-i\omega t)/2$. In this case the analytic signal representation is just the well-known exponential representation for sinusoidal functions, whose advantages are well known. For all cases of interest to us, the spectrum of the analytic signal has an appreciable value only in an internal $\Delta\omega$ that is very small compared to the mean frequency $\langle\omega\rangle$ of the spectrum (quasimonochromatic wave). In this case we can write

$$V(t) = E(t) \exp\{i[\phi(t) - \langle\omega\rangle t]\} \tag{7.5}$$

where $E(t)$ and $\phi(t)$ are both slowly varying, i.e.,

$$\left[\left|\frac{dE}{E\,dt}\right|, \left|\frac{d\phi}{dt}\right|\right] \ll \langle\omega\rangle \tag{7.6}$$

For a quasimonochromatic wave, one can readily express other relevant quantities as functions of the analytic signal. In particular, one can define the intensity $I(\mathbf{r}, t)$ of the beam by the relation

$$I(\mathbf{r}, t) = V(\mathbf{r}, t) V^*(\mathbf{r}, t) \tag{7.7}$$

In fact it can be readily shown that $I(\mathbf{r}, t)$ is equal to the mean value of $[V^{(r)2}/2]$ averaged over a few optical cycles.

7.4. STATISTICAL PROPERTIES OF LASER LIGHT AND THERMAL LIGHT

Before embarking on a discussion of the coherence properties of a light beam, it is worth briefly comparing the statistical properties of laser light with those of a conventional light source.

Let us consider the case of a cw laser oscillating on a single transverse and longitudinal mode. As already pointed out in Section 5.3.1, the output intensity of this laser is determined, for a given pump rate, by the condition that the downward transitions due to stimulated emission must exactly balance the upward transitions due to pumping. It was also pointed out that the output intensity is only influenced to a very small degree by amplitude fluctuations arising from spontaneous emission. The oscillation bandwidth of a single-mode laser can thus be considered to arise predominantly from phase fluctuations,

$\phi(t)$, rather than amplitude fluctuations of the laser field. These fluctuations are due either to the phase fluctuations arising from zero-point fluctuations or, more commonly, from cavity length variations induced by thermal changes or acoustic vibrations. This means that, if we write the analytic signal $V(t)$ at a given point in space as

$$V(t) = E(t)\exp\{i[\phi(t) - \omega t]\} \tag{7.8}$$

the fractional amplitude fluctuations of $E(t)$, $|dE/E\,dt|$, will be much smaller than the phase changes $|\dot{\phi}|$. We can now make use of a very useful three-dimensional representation in which the probability of measuring a given value of V is plotted versus the real and imaginary parts, $E^{(r)}$, and, $E^{(i)}$, respectively, of the phasor $\tilde{E}(t) = E(t)\exp[i\phi(t)]$. Since amplitude fluctuations are very small, the representation will appear as shown in Fig. 7.1a. Note that the quantity $p(E)$ in the figure has the meaning that $p(E)\,dE^{(r)}dE^{(i)}$ gives the elemental probability that the measurement yields a value for $E^{(r)}$ between $E^{(r)}$ and $E^{(r)} + dE^{(r)}$ and a value for $E^{(i)}$ between $E^{(i)}$ and $E^{(i)} + dE^{(i)}$. Alternatively we can say that $p(E)(E\,dE\,d\phi)$ is the probability that the

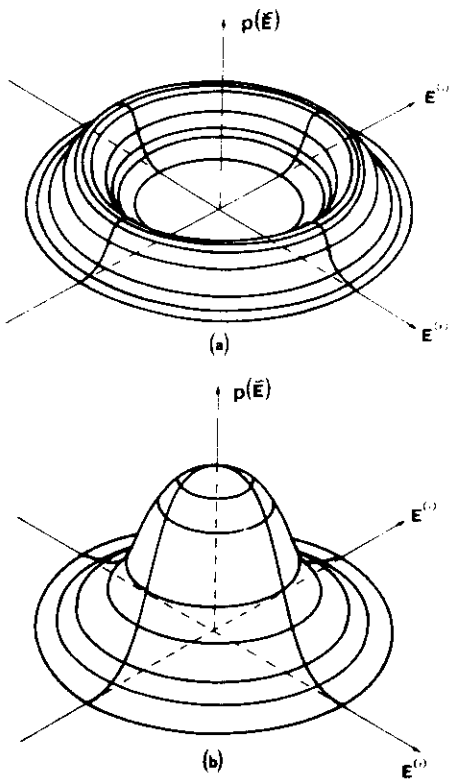

FIG. 7.1. Plot of the probability distribution, $P(\tilde{E})$, of the signal $\tilde{E}$ of a light beam versus the real and imaginary parts, $E^{(r)}$, and, $E^{(i)}$, respectively of the signal. (a) The case of a coherent signal, such as that of a single-mode laser. (b) The case of thermal light, such as that emitted by a conventional light source.

measurement gives a value for E between E and $E + dE$ and a value of ϕ between ϕ and $\phi + d\phi$. Note that the amplitude fluctuations of $|\tilde{E}| = E(t)$ have been greatly exaggerated in the figure, and in fact, for a laser just above threshold, we can write the probability distribution $p(E)$ as

$$p(E) \propto \delta(E - E_0) \tag{7.9}$$

where δ is the Dirac δ function and E_0, according to (7.7) and (7.8), is related to the beam intensity I by the equation $E_0^2 = I$. Thus, the point that represents $\tilde{E}(t)$ in the phasor plane will travel in time essentially along the circumference of radius $|\tilde{E}| = E_0$. Because of the statistical nature of the phase fluctuations, this movement will take the form of a random walk whose angular speed, in terms of the phase angle $\phi(t)$, determines the laser bandwidth.

The light from a conventional lamp, on the other hand, can be considered as arising from the superposition of uncorrelated light emitted, by spontaneous emission, from the atoms of the material. Note that, since this emission occurs under essentially thermal equilibrium conditions, this light is also called thermal light. In this case, since the number of these uncorrelated emitters is very large, it follows, from the central limit theorem of statistics, that the amplitude distribution of both the real and imaginary parts of $\tilde{E}$ must follow a Gaussian law. We can thus write $p(E) \propto \exp[-E^2/C]$, where C is a constant, which can readily be shown to be equal to the average beam intensity $\langle I \rangle$. According to the definition of I given by (7.7), we can in fact see that $\langle I \rangle = \iint E^2 p(E)(E\,dE\,d\phi)/\iint p(E)(E\,dE\,d\phi) = \int E^2 p(E)\,dE^2/\int p(E)\,dE^2 = C$. Thus $p(E)$ can be written as

$$p(E) \propto \exp(-E^2/\langle I \rangle) \tag{7.10}$$

This function is plotted in Fig. 7.1b versus the real and imaginary parts of the field $\tilde{E}(t)$. Note that the average value of both $E^{(r)}$ and $E^{(i)}$ is now zero while the average value of E^2 is just the beam intensity. In the $E^{(r)}$, $E^{(i)}$ plane, the movement of the point that represents $\tilde{E}(r)$ can be viewed as a random walk around the origin. The speed of this movement in terms of both amplitude and phase changes ($dE/E\,dt$, and $d\phi/dt$, respectively) determines the bandwidth of the thermal light source.

A comparison of Figs. 7.1a and 7.1b, makes the profound differences between laser light and thermal light clearly apparent.

7.5. FIRST-ORDER COHERENCE[3]

In Chapter 1 the concept of coherence of an e.m. wave was introduced in an intuitive fashion, with two types of coherence being distinguished: (1) spatial and (2) temporal coherence. In this section we give more detailed

discussion of these two concepts. In fact, as will be seen better by the end of this chapter, it turns out that spatial and temporal coherence describe the coherence properties of an e.m. wave only to first order.

7.5.1. Degree of Spatial and Temporal Coherence

In order to describe the coherence properties of a light source, we can introduce a whole class of correlation functions for the analytic signal. For the moment, however, we will limit ourselves to looking at the first-order functions.

Suppose that a measurement is performed of the analytic signal at some point $\mathbf{r}_1$ in a time interval between 0 and T. We can then obtain the product $V(\mathbf{r}_1, t_1)V^*(\mathbf{r}_1, t_2)$ where t_1 and t_2 are given time instants within the time interval 0–T. If the measurement is now repeated a large number of times, we can calculate the average of the above product over all the measurements. This average is called the ensemble average and written as

$$\Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, t_1, t_2) = \langle V(\mathbf{r}_1, t_1)V^*(\mathbf{r}_1, t_2)\rangle \tag{7.11}$$

For the remainder of this section as well as in the next two sections, we will consider the case of a stationary beam,† as would for instance apply to a single-mode cw laser or to a cw laser oscillating on many modes that are not locked in phase or to a cw thermal light source. In these cases, by definition, the ensemble average will only depend upon the time difference $\tau = t_1 - t_2$ and not upon the particular times t_1 and t_2. We can then write

$$\Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, t_1, t_2) = \Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, \tau) = \langle V(\mathbf{r}_1, t + \tau)V^*(\mathbf{r}_1, t)\rangle \tag{7.12}$$

where we have set $t = t_2$ and where $\Gamma^{(1)}$ only depends upon τ. If now the analytic signal, besides being stationary is also ergodic (a condition that also usually applies to the cases considered above), then, by definition, the ensemble average is the same as the time average. We can then write

$$\Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, \tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T V(\mathbf{r}_1, t + \tau)V^*(\mathbf{r}_1, t)\, dt \tag{7.13}$$

Note that, the definition of $\Gamma^{(1)}$ in terms of a time average is perhaps easier to deal with than that based on ensemble averages. However, the definition of $\Gamma^{(1)}$ in terms of an ensemble average is more general and, in the form given

† A process is said to be stationary when the ensemble average of any variable that describes it (e.g., the analytic signal or the beam intensity, in our case), is independent of time.

by (7.11), can be applied also to nonstationary beams, as we shall see in Section 7.5.4.

Having defined the first-order correlation function $\Gamma^{(1)}$ at a given point $\mathbf{r}_1$, we can define a normalized function $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, \tau)$ as follows:

$$\gamma^{(1)} = \frac{\langle V(\mathbf{r}_1, t+\tau) V^*(\mathbf{r}_1, t)\rangle}{\langle V(\mathbf{r}_1, t) V^*(\mathbf{r}_1, t)\rangle^{1/2} \langle V(\mathbf{r}_1, t+\tau) V^*(\mathbf{r}_1, t+\tau)\rangle^{1/2}} \tag{7.14}$$

Note that, for a stationary beam, the two ensemble averages in the denominator of (7.14) are equal to each other and, according to (7.7), are both equal to the average beam intensity $\langle I(\mathbf{r}_1, t)\rangle$. The function $\gamma^{(1)}$ as defined by (7.14) is called the *complex degree of temporal coherence* while its modulus $|\gamma^{(1)}|$ is called the *degree of temporal coherence.* Indeed $\gamma^{(1)}$ gives the degree of correlation between the analytic signals at the same point $\mathbf{r}_1$ at two instants separated by a time τ. The function $\gamma^{(1)}$ has the following main properties: (1) $\gamma^{(1)} = 1$ for $\tau = 0$, as evident from (7.14). (2) $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, -\tau) = \gamma^{(1)*}(\mathbf{r}_1, \mathbf{r}_1, \tau)$ as can readily be seen from (7.14) with the help of (7.5). (3) $|\gamma \quad (\mathbf{r}_1, \mathbf{r}_1, \tau)| \leq 1$, which follows from applying the Schwarz inequality to (7.14).

We can now say that a beam has perfect temporal coherence when $|\gamma^{(1)}| = 1$ for any τ. For a cw beam this essentially implies that both amplitude and phase fluctuations of the beam are zero so that the signal reduces to that of a sinusoidal wave, i.e., $V = E(\mathbf{r}_1)\exp(-i\omega t)$. Indeed, the substitution of this expression into (7.14) shows that $|\gamma^{(1)}| = 1$ in this case. The opposite case of complete absence of temporal coherence occurs when $\langle V(\mathbf{r}_1, t+\tau),\ V(\mathbf{r}_1, t)\rangle$ and hence $\gamma^{(1)}$ vanishes for $\tau > 0$. That would be the case of a thermal light source of very large bandwidth (e.g., a blackbody source; see Fig. 2.3). In more realistic situations $|\gamma^{(1)}|$ is generally expected to be a decreasing function of τ as indicated in Fig. 7.2 [note that, according to the property stated in (2) above, $|\gamma^{(1)}|$ is a symmetric function of τ]. We can therefore define a characteristic time τ_{co} (called the coherence time) as, for instance, the time for which $|\gamma^{(1)}| = 1/2$. For a perfectly coherent wave, we obviously have $\tau_{co} = \infty$, while for a completely incoherent wave we have $\tau_{co} = 0$. Note that we can also define a temporal coherence length L_c as $L_c = c\tau_{co}$.

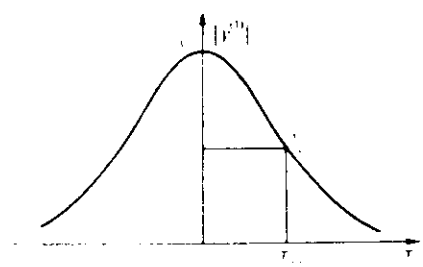

FIG. 7.2. Example of possible behavior of the degree of temporal coherence $|\gamma^{(1)}(\tau)|$. The coherence time can be defined as the half-width of the curve at half-height.

In a similar way, we can define a first-order correlation function between two points $\mathbf{r}_1$ and $\mathbf{r}_2$ at the same instant as

$$\Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, 0) = \langle V(\mathbf{r}_1, t) V^*(\mathbf{r}_2, t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T V(\mathbf{r}_1, t) V^*(\mathbf{r}_2, t)\, dt \quad (7.15)$$

We can also define the corresponding normalized function $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, 0)$ as

$$\gamma^{(1)} = \frac{\langle V(\mathbf{r}_1, t) V^*(\mathbf{r}_2, t) \rangle}{\langle V(\mathbf{r}_1, t) V^*(\mathbf{r}_1, t) \rangle^{1/2} \langle V(\mathbf{r}_2, t) V^*(\mathbf{r}_2, t) \rangle^{1/2}} \quad (7.16)$$

The quantity $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, 0)$ is called the *complex degree of spatial coherence* and its modulus the *degree of spatial coherence.* Indeed $\gamma^{(1)}$ in this case gives the degree of correlation between the analytic signals at the two space points $\mathbf{r}_1$ and $\mathbf{r}_2$ at the same instant. Note that from the Schwarz inequality we again find that $|\gamma^{(1)}| \leq 1$. A wave will be said to have a perfect spatial coherence if $|\gamma^{(1)}| = 1$ for any two points $\mathbf{r}_1$ and $\mathbf{r}_2$ (provided that they lie on the same wave front or on wave fronts whose separation is much smaller than the coherence length L_c). Often, however, one has a situation of partial spatial coherence. This means that, for a fixed value of $\mathbf{r}_1$, $|\gamma^{(1)}|$, as a function of $\mathbf{r}_2$, decreases from the value 1 (which occurs for $\mathbf{r}_2 = \mathbf{r}_1$) to zero as $|\mathbf{r}_2 - \mathbf{r}_1|$ increases. Thus $|\gamma^{(1)}|$ will be greater than some prescribed value (e.g., 1/2) over a certain characteristic area on the wave front around the point P_1, described by the vector $\mathbf{r}_1$. This will be called the coherence area of the wave at point P_1 of the wave front.

The concepts of spatial and temporal coherence can be combined by means of the so-called mutual coherence function, defined as

$$\Gamma^{(1)}(r_1, r_2, \tau) = \langle V(r_1, t + \tau) V^*(r_2, t) \rangle \quad (7.16a)$$

which can also be normalized as follows:

$$\gamma^{(1)}(r_1, r_2, \tau) = \frac{\langle V(\mathbf{r}_1, t + \tau) V^*(\mathbf{r}_2, t) \rangle}{\langle V(\mathbf{r}_1, t) V^*(\mathbf{r}_1, t) \rangle^{1/2} \langle V(\mathbf{r}_2, t) V^*(\mathbf{r}_2, t) \rangle^{1/2}} \quad (7.17)$$

This function, called the complex degree of coherence, provides a measure of the coherence between two different points of the wave at different instants of time. For a quasimonochromatic wave, it follows from (7.5) and (7.14) that we can write

$$\gamma^{(1)}(\tau) = |\gamma^{(1)}| \exp\{i[\phi(\tau) - \langle \omega \rangle \tau]\} \quad (7.18)$$

where $|\gamma^{(1)}|$ and $\phi(\tau)$ are both slowly varying functions of τ, i.e.,

$$\left[\frac{d|\gamma^{(1)}|}{|\gamma^{(1)}|\, d\tau}, \left| \frac{d\phi}{d\tau} \right| \right] \ll \langle \omega \rangle \quad (7.19)$$

7.5.2. Measurement of Spatial and Temporal Coherence

One very simple way of measuring the degree of spatial coherence between two points P_1 and P_2 on the wave front of a light wave is by using Young's interferometer (Fig. 7.3). This simply consists of a screen 1, in which two small holes have been made at positions P_1 and P_2, and a screen 2 on which an interference pattern is produced by the light diffracted from the two holes. More precisely, the interference at point P and time t will arise from the superposition of the waves emitted from points P_1 and P_2 at times $[t - (L_1/c)]$ and $[t - (L_2/c)]$, respectively. One will therefore see interference fringes on screen 2 around point P that are more distinct the better the correlation between the two analytic signals of the light wave, $V[\mathbf{r}_1, t - (L_1/c)]$ and $V[\mathbf{r}_2, t - (L_2/c)]$, where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the coordinates of points P_1 and P_2. Note that the integration time T appearing in the correlation function [see (7.13)] is now equal to the time taken for the measurement of the fringes (e.g., the exposure time of a photographic plate). If now the point P on the screen is chosen so that $L_1 = L_2$, the visibility of the fringes around P will give a measure of the degree of spatial coherence between points P_1 and P_2. To be more precise we define the visibility $V_{(P)}$ of the fringes at point P as

$$V_P = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{7.20}$$

where $I_{\max}$ and $I_{\min}$ are, respectively, the maximum intensity of a bright fringe and the minimum intensity of a dark fringe in the region of P. If the two holes 1 and 2 produce the same illumination at point P and if the wave has perfect spatial coherence, then $I_{\min} = 0$ and therefore $V_P = 1$. For the case in which the signals at the two points P_1 and P_2 are completely uncorrelated (i.e., incoherent), the fringes disappear (i.e., $I_{\max} = I_{\min}$) and therefore $V_{(P)} = 0$. From what we have said in the previous section, it seems clear that V_P must be related to the modulus of the function $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, 0)$. More generally, for any point P on the screen, we expect V_P to be related to the modulus of the function $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, \tau)$, where $\tau = (L_2 - L_1)/c$. At the end of this section we will indeed show that, if the two holes produce the same illumination at point

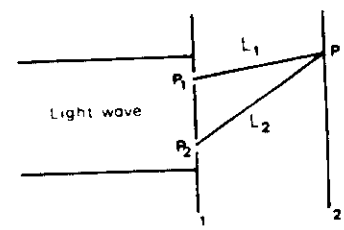

FIG. 7.3. The use of Young's interferometer for the measurement of the degree of spatial coherence of an e.m. wave between points P_1 and P_2.

P, we have

$$V_P(\tau) = |\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, \tau)| \tag{7.21}$$

Thus, by measuring the fringe visibility V_P at a point P such that $L_1 = L_2$, the degree of spatial coherence between points P_1 and P_2 is obtained.

The Michelson interferometer (Fig. 7.4) provides a very simple method of measuring the temporal coherence. Let P be the point where the temporal coherence of the wave is to be measured. A combination of a small hole placed at P and a lens with its focus at P transforms the incident wave into a plane wave (see Fig. 7.9 also). This wave then falls on a partially reflecting mirror S_1 (reflectivity $R = 50\%$) which splits it into two waves A and B. These waves are reflected back by mirrors S_2 and S_3 ($R = 1$) and recombine to form the wave C. Since the waves A and B interfere, the illumination in the direction of C will be either light or dark according to whether $2(L_3 - L_2)$ is an even or odd number of half-wavelengths. Obviously this interference will only be observed as long as the difference $L_3 - L_2$ does not become so large that the two beams A and B are uncorrelated in phase. Thus, for a partially coherent wave, the intensity I_c of beam C as a function of $2(L_3 - L_2)$ will behave as shown in Fig. 7.4b. We can again define a fringe visibility $V_P(\tau)$ for some given value of the difference $L_3 - L_2$ between the lengths of the interferometer arms, i.e., for a given value of the delay $\tau = 2(L_3 - L_2)/c$ between the two reflected waves, with $V_P(\tau)$ as in (7.20), and the quantities I_{max} and I_{min} as shown in Fig. 7.4b. Just as in the case of Young's interferometer, it can now

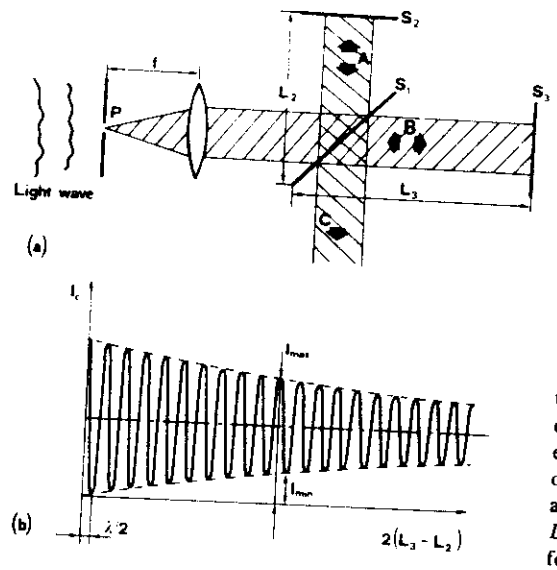

FIG. 7.4. (a) Michelson interferometer for the measurement of the degree of temporal coherence of an e.m. wave at point P; (b) behavior of the light output in the direction C as a function of the difference $L_3 - L_2$ between the lengths of the interferometer arms.

be shown that

$$|\gamma^{(1)}(\mathbf{r}, \mathbf{r}, \tau)| = V_P(\tau) \tag{7.22}$$

where $\mathbf{r}$ is the coordinate of point P. Therefore the measurement of fringe visibility in this case gives a value for the degree of temporal coherence of the wave at point P. Once $|\gamma^{(1)}|$ is known, the value of the coherence time τ_{co} can be found and hence the coherence length $L_c = c_0\tau_{co}$. Using the definition τ_{co} adopted in Fig. 7.2, we then see that L_c is equal to twice the difference $L_3 - L_2$ between interferometer arms at which the visibility falls to $V_P = 1/2$.

We conclude this section with a proof of (7.21). This also serves as an exercise in the use of analytic signals. A similar sort of argument can be used to prove (7.22). Let us call $V(t')$ the analytic signal at point P of Fig. 7.3 at time t'. Since it is due to the superposition of signals coming from each of the two holes of Fig. 7.3, it can be written as

$$V = K_1 V(\mathbf{r}_1, t' - t_1) + K_2 V(\mathbf{r}_2, t' - t_2) \tag{7.23}$$

where $t_1 = L_1/c$, $t_2 = L_2/c$. The factors K_1 and K_2 are inversely proportional to L_1 and L_2 and also depend on the hole dimensions and the angle between the incident wave and the wave diffracted from P_1 and P_2. Since the diffracted secondary wavelets are always a quarter of a period out of phase with the incident wave[5] [see also the discussion of "Huyghens wavelets" appearing in Section 4.4.2] it follows that

$$K_1 = |K_1| \exp(-i\pi/2) \tag{7.24a}$$

$$K_2 = |K_2| \exp(-i\pi/2) \tag{7.24b}$$

If we now define $t = t' - t_2$ and $\tau = t_2 - t_1$, equation (7.23) can be written as

$$V = K_1 V(\mathbf{r}_1, t + \tau) + K_2 V(\mathbf{r}_2, t) \tag{7.25}$$

The intensity at the point P therefore has the value

$$I = VV^* = I_1(t + \tau) + I_2(t) + 2\,\mathrm{Re}[K_1 K_2^* V(\mathbf{r}_1, t + \tau) V^*(\mathbf{r}_2, t)] \tag{7.26}$$

where I_1 and I_2 are the intensities at point P due to the emission from point P_1 alone and point P_2 alone, respectively, and are given by

$$I_1 = |K_1|^2 |V(\mathbf{r}_1, t + \tau)|^2 = |K_1|^2 I(\mathbf{r}_1, t + \tau) \tag{7.27a}$$

$$I_2 = |K_2|^2 |V(\mathbf{r}_2, t)|^2 = |K_2|^2 I(\mathbf{r}_2, t) \tag{7.27b}$$

where $I(\mathbf{r}_1, t + \tau)$ and $I(\mathbf{r}_2, t)$ are the intensities at points P_1 and P_2. Taking the time average of both sides of (7.26) and using equation (7.16a), we find

$$\langle I \rangle = \langle I_1 \rangle + \langle I_2 \rangle + 2|K_1||K_2|\,\mathrm{Re}[\Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, \tau)] \tag{7.28}$$

where equations (7.24) have also been used. Equation (7.28) can be expressed in terms of $\gamma^{(1)}$ by noting that from (7.17) we have

$$\Gamma^{(1)} = \gamma^{(1)}[\langle I(\mathbf{r}_1, t+\tau)\rangle\langle I(\mathbf{r}_2, t)\rangle]^{1/2} \tag{7.29}$$

Substituting (7.29) in (7.28) and using (7.27) we get

$$\begin{aligned}\langle I\rangle &= \langle I_1\rangle + \langle I_2\rangle + 2(\langle I_1\rangle\langle I_2\rangle)^{1/2}\,\mathrm{Re}[\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, \tau)] \\ &= \langle I_1\rangle + \langle I_2\rangle + 2(\langle I_1\rangle\langle I_2\rangle)^{1/2}|\gamma^{(1)}|\cos[\phi(\tau) - \langle\omega\rangle\tau]\end{aligned} \tag{7.30}$$

where we have used (7.18). Now, since both $|\gamma^{(1)}|$ and $\phi(\tau)$ are slowly varying, it follows that the variation of intensity $\langle I\rangle$ as P is changed is due to the rapid variation of the cosine term with its argument $\langle\omega\rangle\tau$. So, in the region of P, we have

$$I_{\max} = \langle I_1\rangle + \langle I_2\rangle + 2(\langle I_1\rangle\langle I_2\rangle)^{1/2}|\gamma^{(1)}| \tag{7.31a}$$

$$I_{\min} = \langle I_1\rangle + \langle I_2\rangle - 2(\langle I_1\rangle\langle I_2\rangle)^{1/2}|\gamma^{(1)}| \tag{7.31b}$$

and therefore, from equation (7.20)

$$V_P = \frac{2(\langle I_1\rangle\langle I_2\rangle)^{1/2}}{\langle I_1\rangle + \langle I_2\rangle}\cdot|\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2, \tau)| \tag{7.32}$$

For the case where $\langle I_1\rangle = \langle I_2\rangle$ equation (7.32) reduces to (7.21).

7.5.3. *Relation Between Temporal Coherence and Monochromaticity*

From the previous paragraphs it is clear that, for a stationary beam, the concept of temporal coherence is intimately connected with the monochromaticity. For example, the more monochromatic the wave is, the greater its temporal coherence. So it is clear that the coherence time must be inversely proportional to the oscillation bandwidth. In this section we wish to discuss this relationship in more depth.

We start by noting that the spectrum of an e.m. wave as measured by a spectrograph is proportional to the power spectrum $W(\mathbf{r}, \omega)$ of the signal $V(\mathbf{r}, t)$. Since the power spectrum W is equal to the Fourier transform of the autocorrelation function $\Gamma^{(1)}$, either one of these quantities can be obtained once the other is known. To give a precise expression for the relation between τ_{co} and $\Delta\nu_{osc}$ we need to redefine these two quantities in an appropriate way. So we will define τ_{co} as the mean square width of the function $|\Gamma^{(1)}(\tau)|^2$, i.e., such that $(\tau_{co})^2 = \int_{-\infty}^{+\infty}(\tau - \langle\tau\rangle)^2|\Gamma(\tau)|^2\,d\tau / \int_{-\infty}^{+\infty}|\Gamma(\tau)|^2\,d\tau$. As a short-hand notation for the above expression, we will write

$$(\tau_{co})^2 = \langle(\tau - \langle\tau\rangle)^2\rangle \tag{7.33}$$

where the mean value $\langle\tau\rangle$ is defined by $\langle\tau\rangle = \int \tau|\Gamma(\tau)|^2\,d\tau / \int |\Gamma(\tau)|^2\,d\tau$. Since $|\Gamma(-\tau)| = |\Gamma(\tau)|$, we see from this definition that $\langle\tau\rangle = 0$ and (7.33) reduces to

$$(\tau_{co})^2 = \langle\tau^2\rangle \tag{7.34}$$

The coherence time defined in this way is conceptually simpler (although sometimes more lengthy to calculate) than that defined earlier (i.e., the half-width at half-height of the curve $|\Gamma(\tau)|$, see Fig. 7.2). If the curve in Fig. 7.2 were oscillatory, τ_{co}, as we first defined it, could not be uniquely determined. Similarly we define the oscillation bandwidth $\Delta\nu_{osc}$ as the mean square width of $W^2(\nu)$, i.e., such that

$$(\Delta\nu_{osc})^2 = \langle(\nu - \langle\nu\rangle)^2\rangle \tag{7.35}$$

where $\langle\nu\rangle$, the mean frequency of the spectrum, is given by $\langle\nu\rangle = \int \nu W^2\,d\nu / \int W^2\,d\nu$. Now, since W and Γ are related by a Fourier transform, it can be shown that $\Delta\nu_{osc}$ and τ_{co}, as we have just defined them, satisfy the condition

$$\tau_{co}\Delta\nu_{osc} \geq 1/4\pi \tag{7.36}$$

The relation (7.36) is closely analogous to the Heisenberg uncertainty relation and can be proved using the same procedure as used in the derivation of the uncertainty relation.[6] The equality sign in (7.36) applies when $|\Gamma^{(1)}(\tau)|$ [and hence $W(\omega)$] are Gaussian functions. This case is obviously the analogue of the minimum uncertainty wave packet.[6]

7.5.4. *Nonstationary Beams*†

We will now briefly consider the case of a nonstationary beam. In this case, by definition, the function $\Gamma^{(1)}$ in (7.11) depends on both t_1 and t_2 and not only on their difference $\tau = t_1 - t_2$. Examples would include an amplitude-modulated laser, an amplitude-modulated thermal light source, a Q-switched or a mode-locked laser. For a nonstationary beam the correlation function can be obtained as the ensemble average of many measurements of the analytic signal in a time interval between 0 and T, where the origin of the time interval is synchronized to the driving signal (e.g., synchronized to the amplitude modulator for a mode-locked laser or the Pockels cell driver for a Q-switched laser). The degree of temporal coherence at a given point $\mathbf{r}$ can then be defined as

$$\gamma^{(1)}(t_1, t_2) = \frac{\langle V(t_1)V^*(t_2)\rangle}{\langle V(t_1)V^*(t_1)\rangle^{1/2}\langle V(t_2)V^*(t_2)\rangle^{1/2}} \tag{7.37}$$

† The author wishes to acknowledge some enlightening discussions on this topic with Professor V. Degiorgio.

where t_1 and t_2 are two given times in the interval 0-T and where all signals are measured at point $\mathbf{r}$. We can now say that the beam has a perfect temporal coherence if $|\gamma^{(1)}(t_1, t_2)| = 1$ for all times t_1 and t_2. According to this definition we can see that a nonstationary beam *without amplitude and phase fluctuations* has a *perfect temporal coherence*.† In the absence of fluctuations, in fact, the products $V(t_1)V^*(t_2)$, $V(t_1)V^*(t_1)$, and $V(t_2)V^*(t_2)$ appearing in (7.37) remain the same for all measurements of the ensemble. These products are thus equal to the corresponding ensemble averages and $\gamma^{(1)}(t_1, t_2)$ reduces to

$$\gamma^{(1)}(t_1, t_2) = \frac{V(t_1)V^*(t_2)}{|V(t_1)||V(t_2)|} \tag{7.38}$$

From (7.38) we then immediately see that $|\gamma^{(1)}| = 1$. According to this definition of temporal coherence, the coherence time of a nonstationary beam, e.g., of a mode-locked laser, is infinite if the beam does not fluctuate in amplitude or phase. This shows that the coherence time of a nonstationary beam is not related to the inverse of the oscillating bandwidth. In a practical situation, however, if we correlate, e.g., one pulse of a mode-locked train with some other pulse of the train, i.e., if we choose $t_2 - t_1$ to be larger than the pulse repetition time, some lack of correlation will be found due to fluctuations. This means that $|\gamma^{(1)}|$ will decrease at $t_2 - t_1$ increases beyond the pulse repetition time. Thus the coherence time is expected to be finite although not related to the inverse of the oscillating bandwidth but to the *inverse of the fluctuation bandwidth.*

7.5.5. *Spatial and Temporal Coherence of Single-Mode and Multimode Lasers*

Consider first a cw laser oscillating on a single transverse and longitudinal mode. Just above threshold, the amplitude fluctuations may be neglected and the analytic signals of the wave, at the two points $\mathbf{r}_1$ and $\mathbf{r}_2$, can be written as

$$V(\mathbf{r}_1, t) = E(\mathbf{r}_1)\exp\{i[\phi(t) - \omega t]\} \tag{7.39a}$$

$$V(\mathbf{r}_2, t) = E(\mathbf{r}_2)\exp\{i[\phi(t) - \omega t]\} \tag{7.39b}$$

where $E(\mathbf{r})$ is the mode amplitude and ω is the angular frequency at band center. The substitution of (7.39) into (7.16) then gives $\gamma^{(1)} = E(\mathbf{r}_1)E^*(\mathbf{r}_2)/|E(\mathbf{r}_1)||E(\mathbf{r}_2)|$, which shows that $|\gamma^{(1)}| = 1$. Thus, a laser oscillating on a single mode has perfect spatial coherence. Its temporal coherence, on the other hand, is established by the oscillating bandwidth $\Delta\nu_{\text{osc}}$. As an

† This is indeed the concept that a radio-engineer has in mind where he thinks about a coherent signal in the radiofrequency field.

example, a frequency stabilized laser of good frequency stability in the visible range may have $\Delta\nu_{osc} \cong 1$ kHz and hence the coherence time will be $\tau_{co} \cong 1/\Delta\nu_{osc} \cong 1$ ms. Note that, in this case, the coherence length is very long ($L_c = c\tau_{co} \cong 300$ km!).

Let us now consider a laser oscillating on a single transverse mode and on many longitudinal modes. The analytic signals at two points $\mathbf{r}_1$ and $\mathbf{r}_2$ belonging to the same wave front can generally be written, in terms of the fields of the cavity modes, as

$$V(\mathbf{r}_1, t) = \sum_k a_k U_k(\mathbf{r}_1) \exp\{i[\phi_k(t) - \omega_k t]\} \tag{7.40a}$$

$$V(\mathbf{r}_2, t) = \sum_k a_k U_k(\mathbf{r}_2) \exp\{i[\phi_k(t) - \omega_k t]\} \tag{7.40b}$$

where a_k are constant factors, U_k, ϕ_k, and ω_k are, respectively, the amplitude, phase, and frequency of the kth mode and where the sum is taken over all oscillating modes. Note now that, since the transverse field configuration is the same for all modes (e.g., that of a TEM_{00} mode) the amplitude U_k is independent of mode index k. We can then write

$$U_k(\mathbf{r}_1) = U(\mathbf{r}_1) \tag{7.41a}$$

$$U_k(\mathbf{r}_2) = U(\mathbf{r}_2) \tag{7.41b}$$

Substitution of (7.41) into (7.40) then leads to the following result

$$V(\mathbf{r}_2, t) = [U(\mathbf{r}_2)/U(\mathbf{r}_1)]V(\mathbf{r}_1, t) \tag{7.42}$$

This means that, whatever time variation $V(\mathbf{r}_1, t)$ is observed in $\mathbf{r}_1$, the same time variation will be observed at $\mathbf{r}_2$ except for a proportionality constant. Substitution of (7.42) into (7.16) then readily gives $|\gamma^{(1)}| = 1$. Thus the beam still has perfect spatial coherence. The temporal coherence, if the phases of all modes are random, is again equal to the inverse of the oscillating bandwidth. If no frequency-selecting elements are used in the cavity, the oscillating bandwidth may now be comparable to the laser bandwidth and hence the coherence time may be much shorter than in the example considered previously [nanoseconds to picoseconds]. With mode-locking, however, the temporal coherence may become very long: thus a mode-locked laser can in principle have perfect spatial and temporal coherence.

The last case we shall consider is that of a laser oscillating on many transverse modes. It can be shown (see Problem 7.16) that such a laser has only partial spatial coherence. This result comes about because the modes differ in both their transverse profile and their resonance frequency.

7.6. DIRECTIONALITY

The directional properties of a laser beam are strictly related to its spatial coherence. We will therefore discuss first the case of an e.m. wave with perfect spatial coherence and then the case of partial spatial coherence.

7.6.1. Beams with Perfect Spatial Coherence

Let us first consider a wave with perfect spatial coherence consisting of a plane wave beam of circular cross section having constant intensity over the cross section (Fig. 7.5a). As a result of diffraction, this beam has an intrinsic divergence θ_d. This can be understood with the help of Fig. 7.5a, which shows the wave front $A'B'$ obtained from AB by applying the Fresnel–Huygens principle. It can be shown that the divergence θ_d is given by

$$\theta_d = 1.22 \frac{\lambda}{D} \tag{7.43}$$

where D is the beam diameter. To see what is meant by this divergence, let us see what happens when the beam we are considering is focused by a lens (Fig. 7.5b). Since, as we have seen, the beam has some spread, it can be shown that the beam can be decomposed into a set of plane waves propagating in slightly different directions. One of these, inclined at an angle θ, is indicated by dotted lines in the figure. Now, as we can see, this wave will be focused to the point P in the focal plane and (for small θ) at a distance

$$r = f\theta \tag{7.44}$$

from the beam axis. Hence, a knowledge of the intensity distribution $I(r)$ in the focal plane gives the angular distribution of the original beam. Now, it is

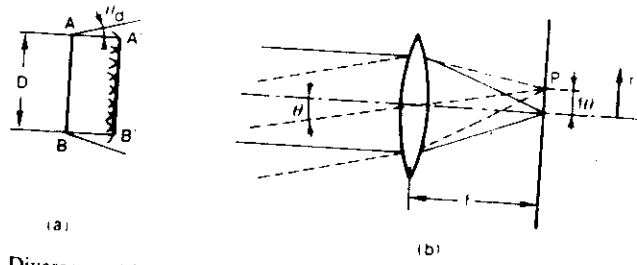

FIG. 7.5. (a) Divergence (due to diffraction) of a plane e.m. wave of constant intensity across its circular cross section; (b) method of measuring the divergence of the plane wave of (a).

known from diffraction theory[7] that the function $I(r)$ is given by the Airy formula

$$I = \left[\frac{2J_1(krD/2f)}{krD/2f}\right]^2 I_0 \tag{7.45}$$

where $k = 2\pi/\lambda$, J_1 is the first-order Bessel function, and I_0 (the intensity at the center of the focal spot) has the value

$$I_0 = P_i\left(\frac{\pi D^2}{4\lambda^2 f^2}\right), \tag{7.46}$$

where P_i is the power of the beam incident on the lens.

The behavior of the intensity I is shown in Fig. 7.6 as a function of

$$x = \frac{krD}{2f} \tag{7.47}$$

Consequently, the diffraction pattern formed at the focal plane of the lens consists of a circular central zone (the Airy disk) surrounded by a series of rings of rapidly decreasing intensity. Now the divergence θ_d of the original beam is conventionally defined to correspond to the radius of the first minimum shown in Fig. 7.6. So, from Fig. 7.6, with the help of (7.47) and (7.44) we obtain (7.43). It can be seen then that the expression (7.43) for θ_d has a certain arbitrariness.

As a second example of the propagation of a spatially coherent beam, we consider the case of a Gaussian beam (TEM_{00}) such as can be obtained from a stable laser cavity consisting of two spherical mirrors. If we let w_0 be the spot size at the beam waist, the spot size w and the radius of curvature R

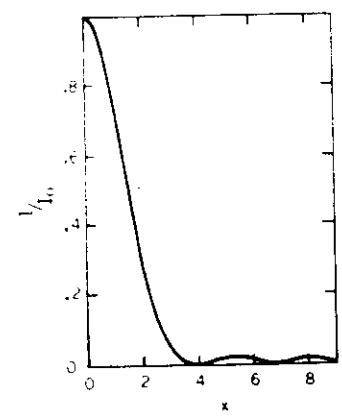

FIG. 7.6. Distribution of light intensity in the focal plane of Fig. 7.5b as a function of radial distance r (normalized, i.e., $x = krD/2f$).

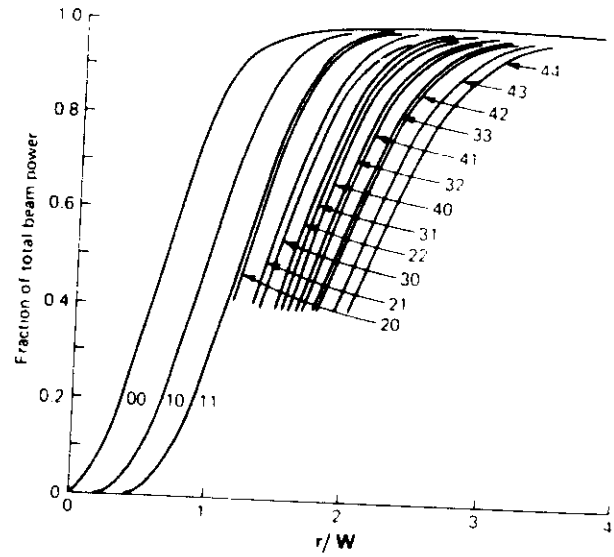

FIG. 7.7. Fraction of total power of a given $TEM_{l,m}$ mode that falls within a circular cross section of radius r. In the figure, w is the spot size of the TEM_{00} mode and the number on each curve gives the mode indexes l, m.

of the equiphase surface at a distance z from the waist can be obtained from (4.105) and (4.106), respectively. To calculate the divergence behavior of this beam, we consider both (4.105) and (4.106) at a large distance from the waist (i.e., for $\lambda z/\pi w_0^2 \gg 1$). We see that, at large distances, $w = \lambda z/\pi w_0$ and $R = z$, and since both w and R increase linearly with distance, the wave can be considered to be a spherical wave having its origin at the waist. Its divergence can then be obtained as

$$\theta_d = \frac{w}{z} = \frac{\lambda}{\pi w_0} \tag{7.48}$$

We can now compare (7.48) and (7.43). If, for the purpose of comparison, we put $D = 2w_0$, we see that, for the same diameter, a Gaussian beam has a divergence about half that of a plane beam.

We next consider the case of a higher-order, $TEM_{l,m}$ Gaussian mode. To calculate its divergence we need to define an effective spot size $w_{l,m}$ of this mode. This can be done with the help of Fig. 7.7, which shows the calculated fraction of the total power for each transverse mode that falls within a circular aperture of radius r. The radius r is normalized to w, the spot size of the TEM_{00} mode in the plane of the aperture. We can now define the effective spot size, $w_{l,m}$, as the radius which contains, e.g., 90% of the beam power. This spot size can then be written as

$$w_{l,m} = C_{l,m} w \tag{7.49}$$

where $C_{l,m}$ is a numerical factor, always larger than 1, which depends on the mode indexes l and m and whose value can readily be obtained from Fig. 7.7. Note that, according to this definition, one has $C_{0,0} \simeq 1.16$ and the effective spot size of a TEM_{00} mode is about $1.16w$. Note also that the effective spot size increases with increasing mode indexes l and m. We now define the beam divergence as

$$\theta_{l,m} = \lim_{z \to \infty} \frac{w_{l,m}}{z} = C_{l,m} \lim_{z \to \infty} \frac{w}{z} \tag{7.50}$$

where (7.49) has been used. Since at large distances from the waist one has $w \cong \lambda z / \pi w_0$, from (7.50) we get

$$\theta_{l,m} = C_{l,m} \frac{\lambda}{\pi w_0} \tag{7.51}$$

which shows that the divergence of a higher-order Gaussian mode is always larger than that of the corresponding TEM_{00} mode. Note that, according to the definition adopted for the effective spot size, the divergence of the TEM_{00} mode is a factor about 1.16 larger than that given by (7.48). Note also that, if we define an effective spot size for the TEM_{00} mode, $w_{0,0}$, in the waist plane as $w_{0,0} = 1.16 w_0$, then the beam divergence of this mode could be written as $\theta_{0,0} = 1.16\lambda / \pi w_0 = (1.16)^2 \lambda / \pi w_{0,0}$.

As a conclusion to this section we can say that the divergence θ_d of a spatially coherent beam can always be written as

$$\theta_d = \beta \lambda / D \tag{7.52}$$

where D is a suitably defined beam diameter and β is a numerical factor of the order of unity whose exact value depends on the field amplitude distribution as well as on the way in which both θ_d and D are defined. Such a beam is commonly referred to as being *diffraction-limited.*

7.6.2. *Beams with Partial Spatial Coherence*

For an e.m. wave with partial spatial coherence the divergence is greater than for a spatially coherent wave having the same intensity distribution. This can, for example, be understood from Fig. 7.5a: if the wave is not spatially coherent, the secondary wavelets emitted over the cross section AB would no longer be in phase and the wave front produced by diffraction would have a larger divergence than that given by equation (7.43). A rigorous treatment of this problem (i.e., the propagation of partially coherent waves) is beyond the scope of this book, and the reader is therefore referred to more specialized texts.[8] We will limit ourselves to considering first a particularly simple case of a beam of diameter D (Fig. 7.8a), which is made up of many smaller beams

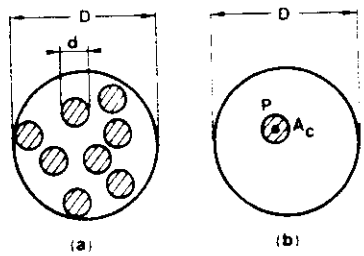

FIG. 7.8. Examples to illustrate the different divergence properties of coherent and partially coherent waves: (a) beam of diameter D made of the superposition of several smaller and coherent beams of diameter d; (b) beam of diameter D and coherence area A_c at point P.

(shaded in the figure) of diameter d. We will assume that each of these smaller beams is diffraction limited (i.e., spatially coherent). Now, if the various beams are mutually uncorrelated, the divergence of the beam as a whole will be equal to $\theta_d = \beta\lambda/d$. If, on the other hand, the various beams were correlated, the divergence would be $\theta_d = \beta\lambda/D$. This last case is actually equivalent to a number of antennas (the small beams) all emitting in phase with each other. After this simple case we can consider the general case in which the partially coherent beam has a given intensity distribution over its diameter D and a given coherence area A_c at each point P (Fig. 7.8b). By analogy with the previous case we can readily understand that, in this case, $\theta_d = \beta\lambda/[A_c]^{1/2}$, where the numerical factor β is of the order of unity, whose value depends both on the particular intensity distribution and on the way in which A_c is defined. The concept of directionality is thus intimately related to that of spatial coherence.

After these general remarks about a beam with partial spatial coherence we can go on to consider the particularly important case of a laser oscillating on many transverse modes. Thus we consider a stable laser cavity in which the transverse dimension $2a$ of the active laser material is appreciably larger than the spot size of the TEM_{00} mode within the material. A relevant example would be that of a cw or pulsed solid-state laser and we will therefore refer to the case shown in Fig. 5.14. The considerations that follow, are, however, of general application to any multimode laser using a stable cavity. For simplicity we shall assume that the spot size w in the material is approximately equal to the spot size w_0 at the beam waist. Since a is appreciably larger than w_0, it is expected that many transverse modes will be excited, so filling the available cross section of the laser material. The highest-order mode that is expected to be excited is limited to a size that is not significantly truncated by the aperture of the material. The transverse indexes of this mode can then be estimated from Fig. 7.7 once the maximum allowed loss for a mode to oscillate is known. Assuming, for example, that this loss is 10%, then 90% of the power of this highest-order mode must pass through the laser aperture. In this case, the effective spot size $w_{l,m}$ as defined in the previous section must be equal to the radius a of the material, i.e., $w_{l,m} = a$. With the help of (7.49) we then get

$$a = C_{l,m}w = C_{l,m}w_0 \tag{7.53}$$

For given values of a and w_0, (7.53) allows one to calculate $C_{l,m}$, which can then be used in (7.51) to obtain the divergence of the mode. Since this mode has the highest divergence, we can get a rough estimate of the overall divergence of the beam θ_d, by assuming it to be equal to the divergence of this mode $\theta_{l,m}$. From (7.51) and (7.53) we then get

$$\theta_d \cong \frac{a\lambda}{\pi w_0^2} \tag{7.54}$$

Equation (7.54) is useful in a number of ways. If w_0 is in fact known, (7.54) can be used to estimate the expected divergence of the multimode laser. If w_0 is not known and if a measurement of θ_d has been performed, (7.54) can be used in reverse to get an estimate of w_0. Note that, according to the expression (7.54), the beam divergence of a multimode laser is expected to decrease with decreasing values of the cavity aperture, a, and with increasing values of the TEM_{00} mode spot size, w_0.

Given an incoherent lamp S, one can obtain a spatially coherent wave, i.e., drastically decrease its divergence by using the arrangement of Fig. 7.9. The light from S is imaged on a pinhole of diameter d situated in the focal plane of the lens L'. The light passing through the pinhole will fill a large cone of angles (solid lines in Fig. 7.8) corresponding to the focusing cone of the lens L. The beam produced by diffraction from the pinhole, however, has a much smaller divergence given by $\theta = 1.22\lambda/d$ and will thus occupy the hatched area of Fig. 7.9. If now the aperture D of the collecting lens L' satisfies the condition $D = 2\theta f = 2.44\lambda f/d$, where f is the focal length of the lens, this lens would collect just the light resulting from diffraction from the hole, thus producing a coherent output beam. In fact, however, this argument is somewhat oversimplified as it uses equation (7.43), which is only valid when the hole is illuminated by light that is already coherent. To deal with this problem in a more rigorous fashion requires a treatment of the propagation of partially coherent e.m. waves.[8] Let us suppose for simplicity (and also because this is frequently the case in practice) that the wave arriving at the hole has no spatial coherence. Then, for this case, it follows from the well-known Van Cittert–Zernike theorem[8] that, if the exit beam from the lens L' of Fig. 7.8 is to have some particular value of spatial coherence, the diameter D of the lens must have the value $D = \beta\lambda f/d$, where β is a numerical factor, which depends on the degree of coherence we stipulate. If, for instance, we require that the

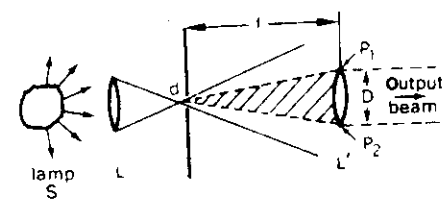

FIG. 7.9. Method for obtaining a coherent output beam from an incoherent lamp.

degree of spatial coherence between two extreme points P_1 and P_2 on the lens circumference has the value $|\gamma(P_1, P_2, 0)| = 0.88$, the corresponding value of β turns out to be $\beta = 0.32$. This gives

$$D = \frac{0.32\lambda f}{d} \tag{7.55}$$

which is of the same form as that established by the earlier simplified argument but with a different (and in fact significantly smaller) numerical factor.

7.7. LASER SPECKLE[9,10]

Following what has been said in the previous sections about first-order coherence, we now mention a very striking phenomenon characteristic of laser light, known as laser speckle. Laser speckle is apparent when one observes laser light scattered from a wall or transparent diffuser. The scattered light is seen to consist of a random collection of alternately bright and dark spots (or speckles) (Fig. 7.10a). Despite the randomness, one can distinguish an average speckle (or grain) size. This phenomenon was quickly recognized by early workers in the field as being due to constructive and destructive interference of radiation coming from the small scattering centers on the surface of the wall or of the transparent diffuser. Since the phenomenon depends on there being a high degree of first-order coherence, it is an inherent feature of laser light.

The physical origin of the observed granularity can be readily understood both for free-space propagation (Fig. 7.10b) and for an imaging system (Fig. 7.10c), when it is realized that the surfaces of most materials are extremely rough on the scale of an optical wavelength. For free-space propagation, the resulting optical wave at any moderately distant point from the scattering surface consists of many coherent components or wavelets, each arising from a different microscopic element of the surface. Referring to Fig. 7.10b, we note that the distances traveled by these various wavelets may differ by many wavelengths. Interference of the phase-shifted but coherent wavelets results in the granular intensity (or speckle pattern, as it is called). When the optical arrangement is that of an imaging system (Fig. 7.10c) an explanation of the observed pattern must take account of diffraction as well as interference. Even for a perfectly corrected imaging system, the intensity at a given image point can result from the coherent addition of contributions from many independent parts of the surface. It is only necessary that the point-spread function of the imaging system be broad in comparison to the microscopic surface variations to ensure that many phase-shifted coherent contributions add at each image point.

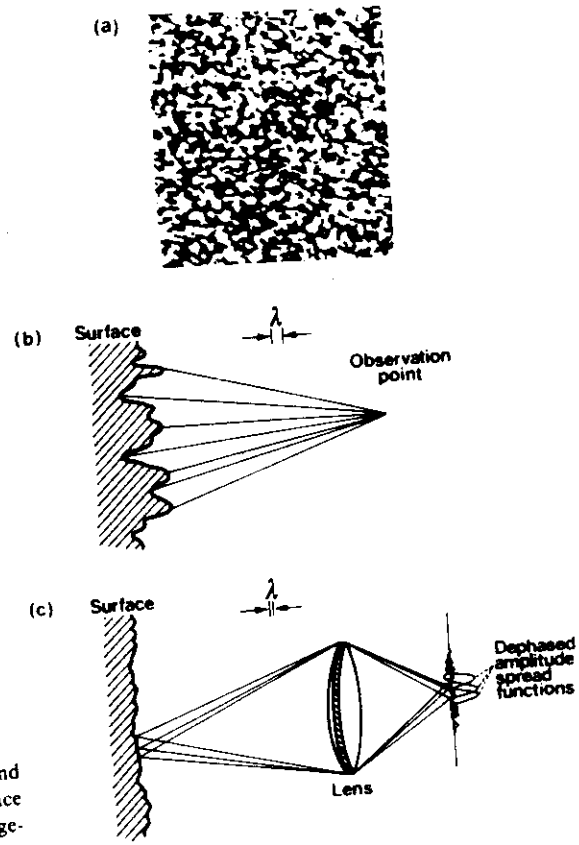

FIG. 7.10. (a) Speckle pattern and its physical origin (b) for free-space propagation, and (c) for an image-forming system.

We can readily obtain an order-of-magnitude estimate for the grain size d_g (i.e., the average size of the spots in the speckle pattern) for the two cases just considered. In the first case (Fig. 7.11a) the scattered light is recorded on photographic film at a distance L from the diffuser with no lens between film and diffuser. Suppose that, at some point P in the recording plane, there is a bright speckle. This means that the light diffracted by all points of the diffuser will interfere at point P in a predominantly constructive way so as to give an overall peak for the field amplitude. In a heuristic way we can then say that the diffractive contributions at point P from the waves scattered from points P_1, P_1', P_1'', etc. add (on the average) in phase with those from points P_2, P_2', P_2'' etc. We now ask how far the point P must be moved along the x axis in the recording plane in order to destroy this constructive interference. This will happen when the contributions of, e.g., the diffracted waves from points P_1 and P_2 at the new point P' interfere destructively rather than constructively. In this case we will show in fact that the contributions from points P_1', P_2' will also interfere destructively, as do those from points P_1'' and P_2'' etc., and

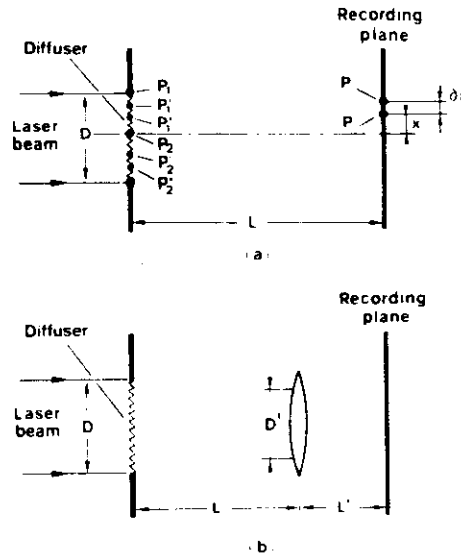

FIG. 7.11. Grain-size calculation (a) for free-space propagation and (b) for an image-forming system.

the overall light intensity will have a minimum value. Taking, for example, points P_1 and P_2, we require that the change δx in the x coordinate of point P is such that the corresponding change $\delta(P_2P - P_1P)$ in the path difference $P_2P - P_1P$ be equal to $\lambda/2$. Since $P_2P = (x^2 + L^2)^{1/2}$ and $P_1P = \{[(D/2) - x]^2 + L^2\}^{1/2}$, we find (for $D \ll L$) that $\delta(P_2P - P_1P) \cong (D/2L)\delta x$. If we require $\delta(P_2P - P_1P) = \lambda/2$, we get

$$\delta x = \lambda L/D \tag{7.56}$$

By following a similar calculation it can readily be shown that the same result is obtained by considering points P_1' and P_2' (or points P_1'' and P_2'' etc.) rather than points P_1 and P_2, and all the corresponding contributions will now (on average) add destructively rather than constructively. Thus, we can write for the grain size, d_g, the following approximate expression:

$$d_g = 2\delta x = 2\lambda L/D \tag{7.57}$$

It should be noted that a similar argument can be used to calculate the beam diameter of the spot in the focal plane of a lens. Consider in fact the case where the diffuser in Fig. 7.11a is replaced by a lens of focal length $f = L$. An intensity maximum will then exist at $x = 0$ (i.e., at the center of the recording plane) since, as a result of the spherical wave front produced by the lens, the contributions from points P_1, P_1', P_1'', etc. all add in phase with those from points P_2, P_2', P_2'', etc. The size of the spot in the focal plane is again given approximately by (7.57), i.e., for the case considered, $d_g = 2\lambda f/D$. This result

should be compared with the exact value $d = 2.44\lambda f/D$ as obtained from Fig. 7.6. We can therefore now understand the following general property of a diffracted wave: whenever the whole aperture of diameter D contributes coherently to the diffraction to one or more spots in a plane located at a distance L, the minimum spot size in this recording plane is always approximately given by $2\lambda L/D$.* Note that, in the case of a diffuser, this coherent contribution from the whole aperture D occurs provided that (1) the size d_s of the individual scatters is much smaller than the aperture D; (2) there is an appreciable overlap, at the recording plane, between wavelets diffracted from various scattering centers. This implies that the dimension of each of these wavelets at the recording plane ($\sim \lambda L/d_s$) is larger than their mean separation ($\sim D$). The length L must therefore be such that $L > d_s D/\lambda$. For $d_s = 10\ \mu$m and $\lambda = 0.5\ \mu$m, for instance, $L > 20D$.

The second case we will consider is that of scattered light recorded on a photographic film after it passes through a lens which images the diffuser on the film. An aperture of diameter D' is placed in front of the lens (Fig. 7.11b). If the length L is again such that $L > d_s D/\lambda$, the grain size d_g on the lens will be given by (7.57). As in the previous case, we will assume that (1) this grain size d_g is much smaller than the aperture D'; (2) there is an appreciable overlap, at the recording plane, of wavelets diffracting from these various grains. This implies that the dimension of each of these wavelets at the recording plane ($\lambda L'/d_g$) is larger than their mean separation (D'). By use of (7.57) this is seen to imply $D' < D(L'/L)$. Under the above two assumptions, the grain size d'_g at the recording plane is given by

$$d'_g = 2\lambda L'/D' \tag{7.58}$$

Now it is the whole beam of aperture D' that acts coherently in its contribution of diffracted light to each individual spot. Note that the arrangement of Fig. 7.11b also describes the case where one looks directly at a diffusing surface. In this case the lens and the recording plane correspond to the lens of the eye and the retina. Accordingly, d'_g given by (7.58) is the grain size on the retina. Note that the apparent grain size on the diffuser d_{ag} is then given by $d_{ag} = d'_g(L/L') = 2\lambda L/D'$. This increases with increasing L, i.e., with increasing distance between the observer and the diffuser. It also decreases with increased aperture of the iris (i.e., when the eye is dark-adapted). Both these predictions are indeed confirmed by experimental observations.

Speckle noise often constitutes an undesirable feature of coherent light. The spatial resolution of an object illuminated by laser light is in fact often limited by speckle noise. Speckle noise is also apparent in the reconstructed image of a hologram, again limiting the spatial resolution of this image. Some

* Since, for $L \gg D$, the field distribution in the recording plane is the spatial Fourier transform of that in the input plane,[11] this property emerges as a general property of the Fourier transform.

techniques have therefore been developed to reduce speckle in coherently illuminated objects.[10] Speckle noise is not always a nuisance, however. In fact techniques have been developed that exploit the speckle behavior (speckle interferometry) to show up, in a rather simple way, the deformation of large objects due, for instance, to stresses or vibrations.[10]

7.8. BRIGHTNESS

The brightness B at a given point of a light source for a given direction of emission has already been defined in Chapter 1 [see Fig. 1.7 and equation (1.13)]. It is important to note that the most significant parameter of a laser beam (and in general of any light source) is neither power nor intensity, but brightness. In fact let us compare, for example, two lasers 1 and 2 having the same diameter and output power, one having a beam divergence θ_1, the other θ_2, where $\theta_2 > \theta_1$. From what was said in connection with Fig. 7.5b it can be seen that the first of these beams produces the higher intensity at the focus of a lens. Since the solid angle of emission Ω is proportional to the square of the divergence, the first beam is brighter than the second. It follows therefore that the intensity that can be produced at the focus of a lens is proportional to the beam brightness. Since, in most applications, one is interested in the beam intensity that can be produced by focusing with a lens, it follows that brightness is the significant quantity. This is further demonstrated by the fact that although the intensity of a beam can be increased, its brightness cannot. The simple arrangement of confocal lenses shown in Fig. 7.12 can be used to decrease the beam diameter if $f_2 < f_1$. The intensity of the exit beam is therefore greater than that of the entrance beam. However, the divergence of the exit beam ($\sim\lambda/D_2$) is also greater than that ($\sim\lambda/D_1$) of the entrance beam, and so one can see that the brightness remains invariant. This property, seen here for a particular case, is of general validity (for incoherent sources also): Given some light source and an optical imaging system, the image cannot be brighter than the original source[12] (this is true provided the source and image are surrounded by media of the same refractive index).

The brightness of lasers is several orders of magnitude greater than that of the most powerful incoherent sources. This is due to the extreme directionality of a laser beam. Let us compare, for example, a He-Ne laser oscillating

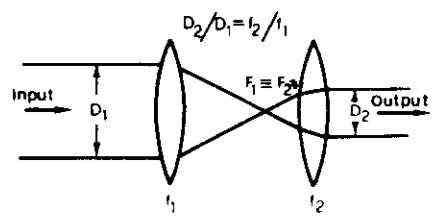

FIG. 7.12. Method for increasing the intensity of a plane wave.

on a single mode at a wavelength $\lambda \cong 0.63\ \mu$m with a modest output power of 1 mW with what is probably the brightest conventional source. This would be a high-pressure mercury vapor lamp (PEK Labs type 107/109), with an output power of ~100 W and a brightness B of ~95 W/cm^2 sr in its most intense green line ($\lambda = 546$ nm, $\Delta\lambda = 10$ nm). To obtain a diffraction-limited beam we can use the arrangement of Fig. 7.9. The solid angle of the light emitted by the pinhole and collected by the lens L' is $\Omega = \pi D^2/4f^2$ while the emitting area is $A = \pi d^2/4$. Since the brightness of the lamp image on the pinhole cannot be greater than that of the lamp, the output beam power is, at most,

$$P = B\Omega A \cong (\lambda/4)^2 B \cong 1.7 \times 10^{-8}\ \text{W} \tag{7.59}$$

where we have used (7.55). The output power turns out to be about five orders of magnitude less than that of the He-Ne laser. We also note from (7.59) that the diffraction-limited power obtainable from a lamp depends only on its brightness. This further illustrates the importance of the concept of brightness.

7.9. COMPARISON BETWEEN LASER LIGHT AND THERMAL LIGHT

Using the system of Fig. 7.9 we have arranged it so that the two beams (from the laser and from the mercury lamp) have the same degree of spatial coherence. To obtain the same degree of temporal coherence it is necessary to insert a filter in the arrangement of Fig. 7.9 so as to pass a very narrow band, i.e., equal to the oscillation bandwidth $\Delta\nu_{osc}$ of the He-Ne laser. Assuming $\Delta\nu_{osc} \cong 1$ kHz, since the linewidth of the mercury lamp under consideration is $\Delta\nu = 10^{13}$ Hz, it follows that this second operation further reduces the output power by ten orders of magnitude ($P \cong 10^{-18}$ W). We recall that the lamp power that we started with was 100 W! This also shows how much more difficult it is to produce interference phenomena (which require sources with good coherence) starting with incoherent sources.

This output beam from the mercury lamp now has the same spatial and temporal coherence characteristics as a He-Ne laser. It is therefore natural to ask whether this light now has exactly the same coherence characteristics as a laser beam. The answer, however, is negative. Despite having paid such a heavy penalty in terms of output power, the laser light is still more coherent than the "filtered" thermal light of this source. The difference essentially arises from the statistically different properties of the two light sources as discussed in Section 7.4. In that section we have indeed shown that the fluctuations of a cw laser beam essentially consist of a random walk of its phase around the whole 2π angle (Fig. 7.1a) while the fluctuations of a thermal source consist of a random movement around the origin by the point representing $\tilde{E}(t)$ in

the $E^{(1)} - E^{(1)}$, plane. If the two beams are now made to have the same temporal coherence, the speed of movement of this representative point is the same for the two cases of Figs. 7.1a and 7.1b. If the two beams are then made to have the same degree of spatial coherence, this speed of movement will then be the same, for the two beams, at any point of the wave front. Suppose, furthermore, that the intensity of the two beams be made the same, as can be done in principle either by attenuating the laser beam by a linear attenuator or amplifying the thermal beam by a linear amplifier (by a factor 10^{15} in the example of the previous section!). This would simply mean that the quantity E_0 appearing in (7.9) is such that E_0^2 is equal to the average intensity $\langle I \rangle$ of the thermal light source. Despite this, the statistical properties of the laser light and of the thermal light source remain different since a linear attenuation or amplification of a beam does not alter its statistical properties.

7.10. HIGHER-ORDER COHERENCE*[4]

A complementary way of describing the difference between laser light and thermal light is to introduce some suitably defined higher-order coherence functions for the corresponding fields. Indeed, the coherence properties of a wave were defined in Section 7.5 in terms of the correlation function $\Gamma^{(1)}$. Since this function involves a product between the signals at just two different times or two different locations, the function was called a first-order correlation function. Correspondingly, the degree of coherence defined by these functions describes the statistical properties of the wave only to first order. In fact, to provide a complete characterization of the field, one can introduce a whole class of higher-order correlation functions. For the sake of brevity we will use the symbol $x_i = (\mathbf{r}_i, t_i)$ for the space and time coordinates of a point. We can then define the nth-order correlation function as

$$\Gamma^{(n)}(x_1, x_2, \ldots, x_{2n}) = \langle V(x_1) \cdots V(x_n) V^*(x_{n+1}) \cdots V^*(x_{2n}) \rangle \tag{7.60}$$

which involves the product of $2n$ terms, these being the function V evaluated at $2n$ space-time points $x_1, x_2, \ldots, x_{2n}$. The corresponding normalized quantity is then given by

$$\gamma^{(n)}(x_1, x_2, \ldots, x_{2n}) = \frac{\langle V(x_1) \cdots V(x_n) V^*(x_{n+1}) \cdots V^*(x_{2n}) \rangle}{\prod_{r=1}^{2n} \langle V(x_r) V^*(x_r) \rangle^{1/2}} \tag{7.61}$$

where $\prod$ is the symbol for product. Obviously these expressions reduce to (7.16a) and (7.17) for the case $n = 1$. Note that, in the experiment considered

* The author wishes to acknowledge some enlightening discussions on this topic with Professor V. Degiorgio.

in the previous section, the He–Ne laser beam and the beam from the mercury lamp were made to have the same degree of spatial and temporal coherence, i.e., to have the same first-order correlation function $\Gamma^{(1)}$. Since, however, the statistical properties of the two signals are completely different, we can now expect that the higher-order correlation functions $\Gamma^{(n)}$ will be different for the two cases and can therefore be used to distinguish between coherent and incoherent waves. First we need to define, in terms of higher-order correlation functions, what we mean by a completely coherent beam. We begin by noting that, if a wave is perfectly coherent to first order (i.e., if $|\gamma^{(1)}(x_1, x_2)| = 1$), then

$$\Gamma^{(1)}(x_1, x_2) = V(x_1)V^*(x_2) \tag{7.62}$$

i.e., $\Gamma^{(1)}$ can be separated into a product of the analytic signals at x_1 and x_2. Indeed, if field fluctuations are completely absent, the ensemble averages of, e.g., (7.11) or (7.16a) will be simply the product of the corresponding signals. By analogy we can define a perfectly coherent e.m. wave as one for which $\Gamma^{(n)}$ factorizes for all n. This means that

$$\Gamma^{(n)}(x_1, x_2, \ldots, x_{2n}) = \prod_{r=1}^{n} V(x_r) \prod_{k=n+1}^{2n} V^*(x_k) \tag{7.63}$$

Indeed, when field fluctuations are completely absent, the ensemble average of (7.60) will again be simply the product of the analytic signals. In this case we find from (7.61) that

$$|\gamma^{(n)}(x_1, x_2, \ldots, x_{2n})| = 1 \tag{7.64}$$

for all orders n. For the particular case where $x_1 = x_2 \cdots = x_{2n} = x$, we find from (7.63) that

$$\Gamma^{(n)}(x, x, \ldots, x) = |V(x)|^{2n} = I^n(x) = [\Gamma^{(1)}(x, x)]^n \tag{7.65}$$

since, in this case, $I(x) = |V(x)|^2 = \Gamma^{(1)}(x, x)$.

The signal of a cw laser oscillating in a single mode can, to a good approximation, be considered to have only phase fluctuations. For a frequency stabilized laser the rate change of this phase is rather slow, however. For example, in the case of a laser with a bandwidth of $\Delta\nu_{\text{osc}} \cong 1$ kHz the phase change will occur in about $\tau_c = 1/\Delta\nu_{\text{osc}} = 1$ ms [so that $|d\phi/\phi\, dt| \approx \Delta\nu_{\text{osc}}$]. This means that, for time intervals much smaller than τ_c or, separations between the equiphase surfaces of the $2n$ space-time points much smaller than $c\tau_c = 300$ km, phase fluctuations can also be neglected. In this case the beam does not show any fluctuation and it can thus be considered to be coherent to all orders. Note that, according to the discussion in Section 7.5.4, the field of a nonstationary laser beam (e.g., a mode-locked or single mode Q-switched

laser) can also, in principle, be made coherent to all orders if fluctuations are again eliminated. In both cases, therefore, for $x_1 = x_2 = \cdots x$, equation (7.65) applies.

A thermal light source, on the other hand, has completely different statistical properties and the higher-order correlation functions can be shown to be different from those of a coherent light source. Let us, in particular, consider again the case $x_1 = x_2 = \cdots x_{2n} = x$. Then $\Gamma^{(n)}(x, x, \ldots, x)$ can be calculated from the following expression

$$\Gamma^{(n)} = \int E^{2n} p(E)\, dE \Big/ \int p(E)\, dE \tag{7.66}$$

where $E = E(x)$ is the field amplitude at coordinate x [see (7.8)] and $p(E)$ is the probability density introduced in Section 7.4. If now the expression (7.10) for $p(E)$ is used in (7.66), it can be shown that one obtains

$$\Gamma^{(n)}(x, x, \ldots, x) = n!\langle I\rangle^n = n![\Gamma^{(1)}(x, x)]^n \tag{7.67}$$

since, in this case, $\langle I\rangle = \langle V(x), V(x)\rangle = \Gamma(x, x)$. A comparison of (7.67) with (7.65) then shows that, for the same value of $\Gamma^{(1)}(x, x)$, i.e., for the same value of the (average) intensity, the nth order correlation function of a thermal light source is $n!$ larger than that of a coherent source. Substituting (7.67) into (7.61) then gives

$$\gamma^{(n)} = n! \tag{7.68}$$

Comparison of (7.68) with (7.64) shows that a thermal source can satisfy the coherence condition only for $n = 1$, i.e., only to first order. It follows that one can, at most, arrange that a thermal light source has a perfect (first-order) spatial and temporal coherence, as indeed described in the previous section.

PROBLEMS

7.1. Show that, for a quasimonochromatic e.m. wave, the relationship between the intensity $I(\mathbf{r}, t)$ as defined by (7.7) and $V^{(r)}$ is given by $2I = \langle V^{(r)2}\rangle$, where the average is taken over a few optical cycles [hint: use (7.5)].

7.2. Calculate $\Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, \tau)$ for a sinusoidal wave.

7.3. Calculate $\Gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, \tau)$ for a sinusoidal wave undergoing phase jumps as in Fig. 2.5 with a probability p_τ as in (2.52). Plot the corresponding $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_1, \tau)$ versus τ and compare this curve with that of Fig. 7.2.

7.4. Prove equation (7.18).

7.5. Prove equation (7.22).

7.6. For a Michelson interferometer, find the analytical relation between I_t and $2(L_2 - L_3)$ for the e.m. wave of Problem 7.3. Calculate the corresponding fringe visibility $V_P(\tau)$.

7.7. A laser operating at $\lambda = 10.6\,\mu$m produces an output having a Gaussian line shape with a bandwidth of 10 kHz [$\Delta\nu_{osc}$ is defined according to (7.35)]. With reference to Fig. 7.4b, calculate both the distance ΔL between two successive maxima of the intensity curve and the coherence length L_c.

7.8. A plane e.m. wave of circular cross section, uniform intensity, and perfect spatial coherence is focused by a lens. What is the increase in intensity at the focus compared to that of the incident wave?

7.9. A Nd:YAG laser beam with a diameter of $D \simeq 6$ mm and approximately a constant intensity distribution over its cross section has a divergence $\theta_d \cong 3$ mrad. Show that the laser is not diffraction-limited and estimate the spot size w_0 of the cavity TEM_{00} mode.

7.10. How much would you need to reduce the aperture of the active rod in the previous problem to reduce the corresponding divergence by a factor of 2?

7.11. How would you measure the beam divergence of the laser in Problem 7.9?

7.12. Suppose the beam of Problem 7.9 passes through an attenuator whose (power) transmission T varies with radial distance r according to the law $T = \exp[-(r/w_1)^2]$ with $w_1 = 0.5$ mm. Thus the beam, after the attenuator, has a Gaussian intensity profile. Does this mean that the beam is now a Gaussian beam of (intensity) spot size w_1?

7.13. The laser beam of Problem 7.9 is passed through a telescope as in Fig. 7.12. Calculate the diameter of the pinhole that must be inserted at the common focus $F_1 = F_2$ in order to produce a diffraction-limited output beam. Note that, since the beam already has a fairly good spatial coherence, one should use the equation appropriate to a coherent beam, rather than that for an incoherent beam [i.e., (7.55)].

7.14. Show that (7.62) holds for a perfect sinusoidal wave.

7.15. Show that (7.63) holds for a perfect sinusoidal wave.

7.16. Consider a laser beam oscillating in l transverse modes. Write down the corresponding analytic signals at the two points $\mathbf{r}_1$ and $\mathbf{r}_2$ as in (7.40). Assuming that the mode frequencies are all different, show that

$$\langle V(\mathbf{r}_1, t), V(\mathbf{r}_2, t)\rangle = \sum_1^l a_j a_j^* U_j(\mathbf{r}) U_j^*(\mathbf{r})$$

where the sum extends over the l oscillating modes. Then show that

$$\gamma^{(1)} = \frac{\sum_1^l a_j a_j^* U_j(\mathbf{r}_1) U_j^* \mathbf{r}(\mathbf{r}_2)}{[\sum_1^l |a_j|^2 |U_j(\mathbf{r})|^2]^{1/2} [\sum_1^l |a_k|^2 |U_k(\mathbf{r})|^2]^{1/2}}$$

If we now define two l-dimensional vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ with components $[a_1U_1(\mathbf{r}_1), \ldots, a_lU_l(\mathbf{r}_1)]$ and $[a_1U_1(\mathbf{r}_2), \ldots, a_lU_l(\mathbf{r}_2)]$ respectively, show that $\gamma^{(1)}$ can be written as $\gamma^{(1)} = \mathbf{R}_1 \cdot \mathbf{R}_2 / R_1 \cdot R_2$, where R_1 and R_2 are the magnitudes of the vectors $\mathbf{R}_1$ and $\mathbf{R}_2$. Show that according to this last expression, since $\mathbf{R}_2 \neq \mathbf{R}_1$, one always has $|\gamma^{(1)}| < 1$, i.e., the beam has only partial spatial coherence.

REFERENCES

1. V. P. Chebotaiev, Super-High Resolution Spectroscopy, in *Laser Handbook*, ed. by M. Bass and M. L. Stitch (North-Holland, Amsterdam, 1985), Vol. 5, pp. 289–404.
2. D. Gabor, *J. Inst. Elec. Eng.* **93**, 429 (1946).
3. M. Born and E. Wolf, *Principles of Optics*, 6th edn. (Pergamon Press, Oxford, 1980), pp. 491–544.
4. R. J. Glauber, Optical Coherence and Photon Statistics, in *Quantum Optics and Electronics*, ed. by C. De Witt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York, 1965), pp. 71, 94–98, 103, 151–155.
5. Reference 3, pp. 370–375.
6. W. H. Louisell, *Radiation and Noise in Quantum Electronics* (McGraw-Hill, New York, 1964), pp. 47–53.
7. Reference 3, pp. 395–398.
8. Reference 3, pp. 508–518.
9. *Laser Speckle and Related Phenomena*, ed. by J. C. Dainty (Springer-Verlag, Berlin, 1975).
10. M. Françon, *Laser Speckle and Applications in Optics* (Academic, New York, 1979).
11. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968), Chap. 5.
12. Reference 3, pp. 189–190.