taking account of the finite size of the atom. Instead of having atomic matrix elements simply occurring as constant factors in the transition probabilities we would have to integrate products of the atomic wave functions and the correlation functions. The transition probabilities, in other words, would be integrals which involve the correlation functions for finite spatial as well as temporal intervals. Fortunately these unilluminating complications are not too necessary quantitatively at optical and lower frequencies.

Lecture V THE n-ATOM PHOTON DETECTOR

The photon counter we have thus far discussed has as its sensitive element only a single atom. Since that is hardly a very realistic picture of an actual detector, we must generalize our arguments to deal with detectors containing arbitrarily many atoms which may undergo photoabsorption processes. We shall carry out this generalization in two stages. In the present lecture we consider detectors which consist of a relatively modest number of atoms and show how these can be used to investigate the higher order correlation properties of the fields. We shall postpone until the last lecture a full discussion of the statistical properties of actual photon counting experiments, since it will be useful to discuss the coherence properties of fields first.

The one-atom detector, as we have seen, furnishes us with measurements of the first-order correlation function of the field, $G^{(1)}$. There exist, however, more general correlation properties of fields; some of these are related, for example, to experiments in which we measure coincidences of photon absorption processes taking place at different points in space and time. Such an experiment has been performed for example by Hanbury Brown and Twiss, and we shall discuss it in some detail in the later lectures.

Let us suppose that n similar atoms are placed at different positions $\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_n$ in the field. These atoms, we assume, form the sensitive element of a species of compound detector. A shutter in front of all of the atoms will be opened during the time interval from t_0 to t. We ask for the probability that each of the atoms has absorbed a photon from the field during that time interval. Though this problem is still rather artificial in nature, its solution will be an essential part of the general discussion of photon counting we shall undertake later.

The process in question involves the absorption of n photons, and therefore, to calculate its probability, we must, strictly speaking, apply n-th order perturbation theory. Needless to say, a number of simplifications are available to us in doing this.

In order to solve the Schrödinger equation in the interaction representation

$$i\hbar \frac{\partial}{\partial t} |t> = \mathscr{H}_I(t) |t> , \tag{5.1}$$

we have already introduced the unitary time development operator $U(t, t_0)$ which transforms the states according to the scheme

$$|t> = U(t, t_0) |t_0>.$$

A formal solution for $U(t, t_0)$ may be written in the form

$$U(t, t_0) = \left\{ e^{-\frac{i}{\hbar}\int_{t_0}^{t} \mathscr{H}_I(t')dt'} \right\}_+ \tag{5.2a}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{-i}{\hbar}\right)^n \int_{t_0}^{t} \cdots \int_{t_0}^{t} \{\mathscr{H}_I(t_1) \cdots \mathscr{H}_I(t_n)\}_+ \prod_{p=1}^{n} dt_p , \tag{5.2b}$$

where the bracket symbol $\{\ \}_+$ stands for a time ordering operation to be carried out on all the operators inside the bracket. It requires that the products of operators be rearranged so that their time arguments increase from right to left. The representations (5.2a and b) for the solution are perhaps most easily derived by writing the Schrödinger equation (5.1) as an integral equation and solving the integral equation by means of a power series.

The interaction Hamiltonian $\mathscr{H}_I(t)$ for the n atoms interacting with the field is given by

$$\mathscr{H}_I(t) = \sum_{j=1}^{n} \mathscr{H}_{I,j}(t) , \tag{5.3}$$

where $\mathscr{H}_{I,j}(t)$ represents the coupling of the j-th atom to the field. The individual coupling terms take the form

$$\mathscr{H}_{I,j}(t) = -e \sum_{\gamma} q_{j\gamma}(t) \cdot \mathbf{E}(\mathbf{r}_j, t) , \tag{5.4}$$

which we have already discussed. We shall assume, for simplicity, that the atoms are dynamically independent of one another, i.e., that their zeroth order Hamiltonians are separable and commute.

The n-fold absorption process is described, to lowest order, by the n-th order term $U^{(n)}(t, t_0)$ of $U(t, t_0)$, i.e., the n-th order term of the series in Eq. (5.2b). By inserting the Hamiltonian given by Eq. (5.3) into Eq. (5.2b), we obtain for $U^{(n)}(t, t_0)$ an expression containing n^n terms, which represent all of the ways in which n atoms can participate in an n-th order process. Many of these terms, however, have nothing to do with the process we are considering, since we require each atom to participate by absorbing a photon once and only once. Terms involving repetitions of the Hamiltonian for a given atom describe processes other than those we are interested in. The only terms which do contribute are those in which each of the $\mathscr{H}_{I,j}$ appears only once. There are n! such terms, and all of them contribute equally since the bracket $\{\ \}_+$ is a symmetric function of the operators it contains. Therefore, the part of $U^{(n)}(t, t_0)$ we must consider reduces to

$$\left(\frac{-i}{\hbar}\right)^n \int_{t_0}^{t} \cdots \int_{t_0}^{t} \{\mathscr{H}_{I,1}(t_1)\mathscr{H}_{I,2}(t_2) \cdots \mathscr{H}_{I,n}(t_n)\}_+ \prod_{p=1}^{n} dt_p . \tag{5.5}$$

Since none of the n atoms can emit a photon (each of them is in the ground state initially), only the positive frequency part of the electric field operator in each $\mathscr{H}_{I,j}$ will contribute to the transition amplitude. When the electric field operator in Eq. (5.4) is replaced by $\mathbf{E}^{(+)}(\mathbf{r}_j, t)$ we shall write the resulting interaction Hamiltonian as $\mathscr{H}_{I,j}^{(+)}$. The operators $\mathscr{H}_{I,j}^{(+)}$ commute with each other since the atoms are dynamically independent and the fields $\mathbf{E}^{(+)}(\mathbf{r}_j, t)$ commute. We can therefore drop the ordering bracket $\{\ \}_+$ in the expression (5.5), and write the desired part of $U^{(n)}(t, t_0)$ as an n-fold product of single integrals

$$\left(\frac{-i}{\hbar}\right)^n \prod_{j=1}^{n} \int_{t_0}^{t} \mathscr{H}_{I,j}^{(+)}(t')\, dt' . \tag{5.6}$$

The result is a simple one. The operator which induces the transitions which interest us is simply a product of the operators which induce the individual absorption processes. This does not mean, however, that the matrix of the transition operator factorizes.

In evaluating the matrix element of the operator (5.6) between two states of the entire system we must note that the individual atoms which are all in the same ground state initially may make transitions to final states a_j which are different for different atoms. If we indicate these initial and final states for the atoms with

$|\{g\}>$ and $|\{a_j\}>$, and use the symbols $|i,\{g\}>$ and $|f,\{a_j\}>$ for the initial and final states of the entire system, then the matrix element of (5.6) or of $U^{(n)}(t,t_0)$ takes the form

$$<f,\{a_j\}|U^{(n)}(t,t_0)|i,\{g\}> =$$

$$(\frac{ie}{\hbar})^n \int_{t_0}^{t}\cdots\int_{t_0}^{t} e^{i(\omega_{f1}+\omega_{f2}+\cdots+\omega_{fn})t} <f|E^{(+)}(r_n t_n)\cdots E^{(+)}(r_1 t_1)|i> \prod_{j=1}^{n} M_{a_j g}\prod_{p=1}^{n} dt_p , \quad (5.7)$$

where we have introduced notation for the atomic matrix elements and frequencies analogous to that of the preceding lecture, and have eliminated tensor indices by assuming the field to have a unique polarization as in Eq. (4.21).

We must next carry out upon the amplitude (5.7) the now familiar procedures of squaring, summing over final states of the field and averaging over initial states of the field. The expression we derive in that way is a transition probability for each of the atoms to reach a specified final state $|a_j>$. Since each of these final states is in general part of a continuum we must sum the probability we have derived over all the relevant final atomic states. We shall again assume that our counting device does not record all of these final states with equal likelihood, but is characterized by a certain probability $R(a_j)$ that any particular photoabsorption process is recorded. For simplicity we shall take this recording probability to be the same function for each of the n atoms of the detector. We may then carry out the final state summations for the atoms by introducing the same sensitivity functions we discussed in Eqs. (4.10) and (4.12) of the preceding lecture. When these simple sums and averages are all carried out we find for the n-fold counting probability

$$p^{(n)}(t) = \int_{t_0}^{t}\cdots\int_{t_0}^{t}\prod_{j=1}^{n} S(t'_j - t''_j)\, G^{(n)}(r_1 t'_1 \cdots r_n t'_n, r_n t''_n \cdots r_1 t''_1)\times \prod_{j=1}^{n} dt'_j\, dt''_j . \quad (5.8)$$

In this expression $G^{(n)}$ is the n-th order correlation function for the field defined by

$$G^{(n)}(x_1 \cdots x_{2n}) = \mathrm{Tr}\{\rho E^{(-)}(x_1)\cdots E^{(-)}(x_n) E^{(+)}(x_{n+1})\cdots E^{(+)}(x_{2n})\}$$

with $x_j = \{r_j, t\}$.

For broad band detectors eq. (5.8) reduces to the simpler form

$$p^{(n)}(t) = s^n \int_{t_0}^{t}\cdots\int_{t_0}^{t} G^{(n)}(r_1 t'_1 \cdots r_n t'_n, r_n t'_n \cdots r_1 t'_1)\prod_{j=1}^{n} dt'_j . \quad (5.9)$$

An ideal n-atom counter thus measures a time integral of the n-th order correlation function.

We have thus far considered the n atoms which undergo photoabsorption to be part of a single detector. But a detector constructed in this way is not very different, really, from a set of n detectors of the one-atom variety we discussed in the last lecture. If we regard the n atoms as the sensitive elements of a set of n independent detectors, then the n-fold photoabsorption process we have been discussing furnishes the basis of a primitive technique for n-fold coincidence counting of photons.

The technique may be refined a little if we imagine that there is a separate shutter in front of each one-atom detector. Then we may assume that all the shutters open at the same time t_0 but that the time at which each of them is closed may be varied arbitrarily. Let us suppose that the time at which the j-th shutter is closed is t_j. Then the j-th atom only sees the field from time t_0 to t_j. The effect of closing the shutter may be simulated by assuming that the atom is decoupled from

the field at time t_j. For this purpose we may introduce the step function

$$\theta(t) = \begin{cases} 0 \text{ for } t<0 \\ 1 \text{ for } t>0 \end{cases} \quad (5.10)$$

and write an effective interaction Hamiltonian (i.e., one which takes account of the closing of all the shutters) as

$$\mathcal{H}_I(t) = \sum_{j=1}^{n}\theta(t_j - t)\,\mathcal{H}_{I,j}(t) . \quad (5.11)$$

The calculation of the probability that a photoabsorption takes place in each detector is essentially the same with the effective Hamiltonian (5.11) as the calculation we have described earlier. The only real difference, besides the one of interpretation, is that the answer for the total detection probability is now an n-fold time integral in which the upper limits of integration are the times t_j. For the broad band case the answer is, for example

$$p^{(n)}(t_1\cdots t_n) = s^n \int_{t_0}^{t_1} dt'_1 \cdots \int_{t_0}^{t_n} dt'_n\; G^{(n)}(r_1 t'_1 \cdots r_n t'_n, r_n t'_n \cdots r_1 t'_1) . \quad (5.12)$$

The times $t_1 \cdots t_n$ may be varied independently. An n-fold delayed coincidence rate, i.e., a counting rate per (unit time)^n, may therefore be defined as

$$w^{(n)}(t_1\cdots t_n) = \frac{\partial^n}{\partial t_1 \cdots \partial t_n}\, p^{(n)}(t_1\cdots t_n) = s^n G^{(n)}(r_1 t_1 \cdots r_n t_n, r_n t_n \cdots r_1 t_1) . \quad (5.13)$$

This result verifies the statement we made earlier that coincidence experiments performed with ideal detectors furnish measurements of the higher order correlation functions.

It may be worth emphasizing that the kinds of measurement processes we have been describing differ both in method and in spirit from those that are customarily discussed in the formal quantum mechanical theory of measurement. The formal theory of measurement has been useful in establishing the physical interpretation of quantum mechanical expressions. But because there are few areas in which exact statements meeting the required assumptions of the theory can be made, the applications of the formal theory have been quite restricted to date.

The kinds of field measurements we have discussed are, by contrast, explicitly approximate in character. We have only calculated the transition probabilities to the lowest order in which the transitions occur. While this approximation would not be too difficult to remedy for individual atomic transitions, the higher order effects in multi-atom detectors would be found to have quite a complicated mathematical structure. It is implicit in the approximation we have used that the electromagnetic influences (as well as other influences) of one atom on another are ignored. That can be seen, for example, from the fact that the $E^{(+)}$ operators which occur in the correlation function $G^{(n)}$ all commute. The transition rate (5.13), for example, does not depend on the ordering of the times $t_1 \cdots t_n$ even though the points $r_j t_j$ may have time-like relationships to one another and electromagnetic disturbances can indeed pass from one point to another.

While the atoms may influence one another electromagnetically in ways not described by our lowest-order results, those influences are typically extremely small and are sometimes of a kind that can be eliminated experimentally. To take a specific example, let us suppose, that instead of a simple photoabsorption process in atom 1, we have a type of Raman effect which produces another photon as well as a photoelectron (Fig. 6). The emmited photon may then be absorbed by atom 2, producing a second photoelectron. Not only does this type of process have an

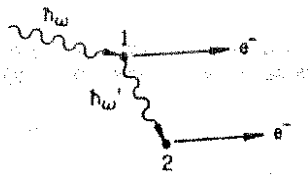

Figure 6

extremely small cross section, but it may be eliminated entirely by choosing detector atoms with ionization potentials greater than $(1/2)\hbar\omega$.

We have mentioned the electromagnetic influences of the atoms upon one another just to underscore the fact that we have not been describing an exact theory of measurement. It may none the less be an extremely useful and accurate theory.

Lecture VI PROPERTIES OF THE CORRELATION FUNCTIONS

The n-th order correlation function was defined as the expectation value

$$G^{(n)}(x_1 \cdots x_{2n}) = \mathrm{Tr}\{\rho E^{(-)}(x_1) \cdots E^{(-)}(x_n) E^{(+)}(x_{n+1}) \cdots E^{(+)}(x_{2n})\} . \tag{6.1}$$

The averaging process we carry out to evaluate this expression is the quantum mechanical analogue of the classical procedure introduced in the first lecture. There we spoke of averages over a set of random Fourier coefficients. The resemblance between the two approaches is not yet a very persuasive one, but it will become more so as we proceed.

As a first property of the correlation functions we note that when we have an upper bound on the number of photons present in the field then the functions $G^{(n)}$ vanish identically for all orders higher than a fixed order M. To state the property more explicitly, if $|n>$ is an n-quantum state and the density operator is written in the form

$$\rho = \sum_{m,n} c_{nm} |n><m| , \tag{6.2}$$

then if we have $c_{nm} = 0$ whenever $n > M$ or $m > M$, it follows from the nature of the annihilation operators $E^{(+)}$, that

$$E^{(+)}(x_1) \cdots E^{(+)}(x_p)\rho = 0 \qquad \text{for } p > M . \tag{6.3}$$

Furthermore, the conjugate relation

$$\rho E^{(-)}(x_1) \cdots E^{(-)}(x_p) = 0 \tag{6.4}$$

also holds for $p > M$. Thus it follows that

$$G^{(p)} \equiv 0 \qquad \text{for } p > M. \tag{6.5}$$

This property of the correlation functions must be regarded as a rather strange one when viewed from the standpoint of classical theory. There the correlation functions are essentially sums of moments of the probability distribution for the Fourier coefficients, and it would be quite difficult to imagine a case for which the moments higher than a certain order vanish identically. We have, in fact, constructed states which have no classical analogue by imposing an upper bound on the number of photons present. However, that should not be surprising since in the limit $\hbar \to 0$ these are states whose total energy goes to zero.

A further property of the correlation functions can be derived from the general statement

$$\mathrm{Tr}(A^\dagger) = (\mathrm{Tr}\, A)^* , \tag{6.6}$$

which holds for all linear operators A. Applying this identity to the correlation function (6.1), we find

$$\begin{aligned}[G^{(n)}(x_1 \cdots x_{2n})]^* &= \mathrm{Tr}\{E^{(-)}(x_{2n}) \cdots E^{(-)}(x_{n+1}) E^{(+)}(x_n) \cdots E^{(+)}(x_1)\rho^\dagger\} \\ &= \mathrm{Tr}\{\rho E^{(-)}(x_{2n}) \cdots E^{(-)}(x_{n+1}) E^{(+)}(x_n) \cdots E^{(+)}(x_1)\} = G^{(n)}(x_{2n} \cdots x_1) .\end{aligned} \tag{6.7}$$

Here we have made use of the Hermitian character of ρ and of the invariance of the trace of a product of operators under a cyclic permutation.

As a consequence of the commutation properties of the $E^{(+)}$ and $E^{(-)}$ we can freely permute the arguments $(x_1 \cdots x_n)$ and $(x_{n+1} \cdots x_{2n})$ without altering the value of $G^{(n)}(x_1 \cdots x_n, x_{n+1} \cdots x_{2n})$. We cannot, however, interchange any of the first n arguments with any of the remaining n, unless suitable terms are added, since the corresponding operators do not commute.

A number of interesting inequalities can be derived from the general statement

$$\mathrm{Tr}\{\rho A^\dagger A\} \geq 0. \tag{6.8}$$

This relation, which follows from the positive definite character of the operator in the brackets, holds for any linear operator A. To prove the inequality we note that ρ is Hermitian and therefore can be diagonalized. Thus, in some representation it has the form

$$<k|\rho|m> = \delta_{km} p_k . \tag{6.9}$$

It follows immediately from the definition of the density operator that

$$p_k = <k|\rho|k> = \{<k|i><i|k>\}_{av.} = \{|<i|k>|^2\}_{av.} \geq 0. \tag{6.10}$$

(Furthermore, since $\mathrm{Tr}\,\rho = \sum_k p_k = 1$, not all the p_k vanish.) Now a simple application of the completeness relation gives

$$\begin{aligned}\mathrm{Tr}\{\rho A^\dagger A\} &= \sum_k p_k <k|A^\dagger A|k> \\ &= \sum_k p_k \sum_m <k|A^\dagger|m><m|A|k> = \sum_k p_k \sum_m |<m|A|k>|^2 \geq 0.\end{aligned} \tag{6.11}$$

Of course this value for the trace is independent of the particular representation used. Hence the proof of the inequality (6.8) is completed.

A number of results may be derived from the general inequality (6.8) by means of various substitutions. For example the choice $A = E^{(+)}(x)$ gives at once

$$G^{(1)}(x,x) \geq 0. \tag{6.12}$$

Similarly the substitution $A = E^{(+)}(x_1) \cdots E^{(+)}(x_n)$ give us

$$G^{(n)}(x_1 \cdots x_n, x_n \cdots x_1) \geq 0. \tag{6.13}$$

These two relations are also evident from the physical meaning of the "diagonal" forms of the $G^{(n)}$. The forms are interpretable as photon intensities and coincidence rates respectively, and are thus intrinsically positive.

These results and all of our later ones can be generalized immediately to deal with vector fields $E_\mu^{(+)}(x)$ rather than the scalar field $E^{(+)}(x)$. We need only associate a vector index μ_j with each coordinate x_j. We can thus consider x_j as a shorthand for the set of variables $\{r_j, t_j, \mu_j\}$ instead of simply $\{r_j, t_j\}$.

Another possible choice for the operator A is

$$A = \sum_{j=1}^{n} \lambda_j E^{(+)}(x_j) \quad , \tag{6.14}$$

where the λ_j are a set of arbitrary complex numbers. For this case (6.8) gives us

$$\sum_{ij} \lambda_i^* \lambda_j G^{(1)}(x_i, x_j) \geq 0. \tag{6.15}$$

Thus the set of correlation functions $G^{(1)}(x_i, x_j)$ forms a matrix of coefficients for a positive definite quadratic form. Such a matrix has, of course, a positive determinant,

$$\det [G^{(1)}(x_i, x_j)] \geq 0. \tag{6.16}$$

For n = 1 this is simply the relation (6.2). For n = 2 we find

$$G^{(1)}(x_1, x_1)\, G^{(1)}(x_2, x_2) \geq |G^{(1)}(x_1 x_2)|^2, \tag{6.17}$$

which is a simple generalization of the Schwarz inequality.

By proceeding along the same line we can derive an infinite sequence of inequalities. We shall confine ourselves however, to mentioning the quadratic ones for the higher order correlation functions. If we write

$$A = \lambda_1 E^{(+)}(x_1) \cdots E^{(+)}(x_n) + \lambda_2 E^{(+)}(x_{n+1}) \cdots E^{(+)}(x_{2n}), \tag{6.18}$$

then the positive-definiteness of the related quadratic form requires that we have

$$G^{(n)}(x_1 \cdots x_n, x_n \cdots x_1)\, G^{(n)}(x_{n+1} \cdots x_{2n}, x_{2n} \cdots x_{n+1}) \geq |G^{(n)}(x_1 \cdots x_n, x_{n+1} \cdots x_{2n})|^2. \tag{6.19}$$

SPACE AND TIME DEPENDENCE OF THE CORRELATION FUNCTIONS

We note that the operators $E^{(+)}(r,t)$ occurring in the correlation functions, obey the Maxwell equations and furthermore satisfy whatever boundary conditions we ordinarily require of the electric field vector (e.g., periodic boundary conditions or the conditions for conducting walls). As a result the functions $G^{(n)}(x_1 \cdots x_{2n})$ obey 2n wave equations and 2n sets of boundary conditions, one for each of the space-time variables.

Let us now consider the structure of the functions $G^{(n)}$ in stationary fields. The best way to define stationarity in quantum mechanics is to require that the density operator commute with the Hamiltonian. This criterion is equivalent to the statement that ρ is independent of time in the Schrodinger picture. (In the Heisenberg picture, however, the density operator for isolated systems is always time-independent.) If we use this definition and the familiar interpretation of the Hamiltonian as an infinitesimal time-displacement operator we may write

$$\begin{aligned} \mathrm{Tr}\{\rho E^{(-)}(x_1) \cdots E^{(+)}(x_{2n})\} &= \mathrm{Tr}\left\{e^{\frac{i\mathcal{H}\tau}{\hbar}} \rho E^{(-)}(x_1) \cdots E^{(+)}(x_{2n})\, e^{\frac{-i\mathcal{H}\tau}{\hbar}}\right\} \\ &= \mathrm{Tr}\left\{e^{\frac{i\mathcal{H}\tau}{\hbar}} \rho e^{\frac{-i\mathcal{H}\tau}{\hbar}} e^{\frac{i\mathcal{H}\tau}{\hbar}} E^{(-)}(x_1)\, e^{\frac{-i\mathcal{H}\tau}{\hbar}} \cdots e^{\frac{i\mathcal{H}\tau}{\hbar}} E^{(+)}(x_{2n})\, e^{\frac{-i\mathcal{H}\tau}{\hbar}}\right\} \\ &= \mathrm{Tr}\{\rho E^{(-)}(r_1, t_1 + \tau) \cdots E^{(+)}(r_{2n}, t_{2n} + \tau)\}, \end{aligned}$$

where τ is an arbitrary time parameter. We have thus shown that for stationary fields the correlation functions obey the identity

$$G^{(n)}(r_1 t_1 \cdots r_{2n} t_{2n}) = G^{(n)}(r_1 t_1 + \tau, \cdots r_{2n} t_{2n} + \tau), \tag{6.20}$$

i.e. they are not changed by a common time displacement of all the arguments. As a result, the $G^{(n)}$ may be thought of as depending only on (2n-1) time differences. The same sort of argument can also be constructed for dealing with spatial displacements. When the density operator commutes with the components of the momentum of the field, the correlation functions are invariant under displacement of the spatial coordinates in the corresponding directions.

One further mathematical property of the correlation functions is a consequence of the way in which the functions are constructed from the positive and negative frequency parts of the fields. The function $G^{(n)}(t_1 \cdots t_n, t_{n+1} \cdots t_{2n})$ has a time dependence which, according to our convention, contains only positive frequencies for the variables $t_{n+1} \cdots t_{2n}$ and only negative frequencies for $t_1 \cdots t_n$. Thus, for example, if we ignore the spatial dependences we may write

$$G^{(1)}(t, t') = \sum_{kk'} c_{kk'}\, e^{i\omega_k t}\, e^{-i\omega_{k'} t'} \tag{6.21}$$

with ω_k and $\omega_{k'} > 0$.

Now if we consider $G^{(1)}(t, t')$ as a function of two complex time variables, t and t', it is clearly an analytic function of t' in the half plane $\mathrm{Im}\, t' \leq 0$, and an analytic function of t in the half-plane $\mathrm{Im}\, t \geq 0$.

We can therefore use the Cauchy theorem of complex function theory to construct identities such as

$$G^{(1)}(t, t') = \frac{1}{2\pi i} \int_C \frac{G^{(1)}(t, t'')}{t'' - t'}\, dt'' \quad , \tag{6.22}$$

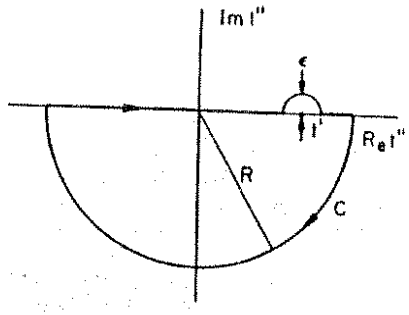

Figure 7.

where C is the contour in the complex t''-plane which is shown in Fig. 7.

Now from the boundedness of the coefficients $c_{kk'}$ in Eq. (6.21) we may see that the semi-circular part of the contour in the lower half plane gives no contribution in the limit as the radius R goes to infinity. Furthermore we note that the contribution of the infinitesimal semi-circular contour in the upper half-plane is just $-\pi i$ times the residue at the pole. In this way we find

$$G^{(1)}(t,t') = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{G^{(1)}(t,t'')}{t''-t'} dt'' \quad , \tag{6.23}$$

where the integration is performed along the real axis and the symbol P denotes the Cauchy principal value. When we take the real and imaginary parts of Eq. (6.23), we obtain the pair of relations

$$\mathrm{Im}\, G^{(1)}(t,t') = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\mathrm{Re}\, G^{(1)}(t,t'')}{t''-t'} dt'' \tag{6.24}$$

$$\mathrm{Re}\, G^{(1)}(t,t') = -\frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\mathrm{Im}\, G^{(1)}(t,t'')}{t''-t'} dt'' . \tag{6.25}$$

These relations enable us in principle to calculate the imaginary part of the correlation functions once we know the real part and vice versa.

Hilbert transform relationships of this type have received a considerable amount of attention in physics and electrical engineering in connection with the requirement that linearly responding systems behave causally. The relations such as (6.24) and (6.25) which are obeyed by the correlation functions, however, have nothing to do with causality. They are simply consequences of the way in which the functions have been defined.

Lecture VII DIFFRACTION AND INTERFERENCE

From a mathematical standpoint, the quantum mechanical treatment of diffraction problems need not differ too greatly from the classical treatment. The field operators are required in general to obey the same linear differential equations and boundary conditions as the classical fields. The problem of constructing such operators may be reduced to the problem of finding a suitable set of mode functions in which to expand them (i.e., a set of mode functions which satisfies the wave equation together with suitable boundary conditions on any surfaces present). To find these modes we naturally resort to the familiar methods of the classical theory of boundary value problems. The solution for the mode functions is not a quantum dynamical problem at all. On the other hand, the fact that it is a well-explored "classical" problem does not mean, as we all know, that it is necessarily a simple one.

Let us return, for example, to the discussion of Young's experiment, illustrated in Fig. 2. When we said that the field at points on the screen Σ_2 is simply a linear combination of the fields at the two pinholes P_1 and P_2, evaluated at appropriate times, we were not solving the diffraction problem exactly, but making a number of physical approximations. One approximation, for example, was an implicit neglect of the fact that transmission of light through the pinholes has a slightly dispersive character. (This effect can be quite small if the bandwidth of the incident radiation is not too broad.) Approximations such as these are essentially classical in character. They are present simply because we have not taken the trouble to solve the classical diffraction problem more precisely.

With this understanding we can now discuss Young's experiment in fully quantum mechanical terms. The positive frequency part of the field $E^{(+)}(r,t)$ when evaluated on the screen Σ_2 will be given, just as in the classical theory, by a linear combination of the fields $E^{(+)}$ evaluated at the pinholes and having the form of Eq. (2.1). The only difference is that the fields $E^{(+)}$ are now operators. If we assume that the two pinholes are not only quite tiny compared with their separation but equal in size then we shall have $\lambda_1 = \lambda_2$ in Eq. (2.1) and we may let the constant λ stand for both coefficients. Now if our observations of the interference pattern on the screen Σ_2 are made with an ideal photon detector, the counting rate of the detector will be proportional to $G^{(1)}(r\,t, r\,t)$. In other words, the intensity observed will be proportional to

$$I = \mathrm{Tr}\{\rho E^{(-)}(rt)\, E^{(+)}(rt)\} = \mathrm{Tr}\{\rho|\lambda|^2 [E^{(-)}(x_1) + E^{(-)}(x_2)][E^{(+)}(x_1) + E^{(+)}(x_2)]\} \quad , \tag{7.1}$$

where we have again let x_j stand for the point (r_j, t_j). This intensity may be expressed in terms of first order correlation functions by expanding the product in Eq. (7.1). We then find

$$I = |\lambda|^2 \{G^{(1)}(x_1,x_1) + G^{(1)}(x_2,x_2) + 2\,\mathrm{Re}\, G^{(1)}(x_1,x_2)\} \quad . \tag{7.2}$$

The first two terms on the right side of this equation are the intensities which would be contributed by either pinhole in the absence of the other. These are, according to the assumptions we have made, rather slowly varying functions of x_1 and x_2. The third term on the right side of Eq. (7.2) is the interference term, as we have already noted in the classical discussion. The correlation function for $x_1 \neq x_2$ in general takes on complex values. If we write it as

$$G^{(1)}(x_1x_2) = |G^{(1)}(x_1x_2)|\, e^{i\varphi(x_1,x_2)} \quad ,$$

then the intensity becomes

$$I = |\lambda|^2 \{G^{(1)}(x_1x_1) + G^{(1)}(x_2x_2) + 2|G^{(1)}(x_1x_2)|\cos\varphi(x_1x_2)\} \quad , \tag{7.3}$$

and we see in the oscillation of the cosine term the origin of the familiar interference fringes.

SOME GENERAL REMARKS ON INTERFERENCE

The discussion we have given of Young's experiment is so closely related to the usual classical analysis that it may not be too clear in what way the interference phenomenon is a quantum mechanical one. A few general remarks about the quantum mechanical interpretation of interferences may therefore be in order. Interference phenomena characteristically occur in quantum mechanics whenever the probability amplitude for reaching a given final state from a given initial one is the sum of two or more partial amplitudes which have well defined phase relations. The individual partial amplitudes are usually contributed by alternative ways in which the system can evolve from its initial state to the final one.

The Young experiment furnishes a simple illustration of these generalities. We may consider as the initial state of the system one in which a wave packet representing a single incident photon lies to the left of the first screen σ (Fig. 2.) which has the single pinhole. We assume that initially all atoms of our photodetector are in the ground state. The final state of the system will be taken to be one in which the photon has been absorbed and one of the atoms of the counter has been correspondingly excited. The amplitude for reaching this final state is the sum of

two amplitudes, each associated with the passage of the photon through one of the two pinholes in the screen Σ_1.

It is interesting to note that the existence of the interference effect is linked quite essentially with our inability to tell which of the possible paths the photon actually takes. Neils Bohr has shown, in a famous argument, that any attempt to determine which of the two paths the photon has followed will wipe out the interference fringes. One way of making such an attempt, for example, is by trying to measure the recoil of the screen Σ_1 when it deflects the photon. The photon may transfer either of two different recoil momenta to the screen (if it excites the counter). However, if we are to make sufficiently accurate measurements of the momentum of the screen we must be prepared to accept an uncertainty in its position which will mean that no fringes appear when the experiment is performed repeatedly.

This lesson is one which can be generalized to apply to all of the quantum mechanical situations we have described earlier. The different paths by which a system may evolve will contribute amplitudes with well-defined phase relations only as long as we have no way of telling which path the system takes. When we make observations to determine the path we characteristically alter the system by making the phases of the partial amplitudes random relative to one another, i.e., we wipe out any interference of the amplitudes on the average.

The alternative paths we have been speaking of are evolutionary paths or histories. For single particle systems such histories may often be identified with spatial trajectories, but for systems with many particles or variable numbers of particles the concept is a much more general one. It is important to emphasize that the quantities which interfere in quantum mechanics are amplitudes associated with particular histories, since the terminology which has been used has often invited confusion on this score.

An example of a statement which is often quoted and easily misinterpreted is made by Dirac in the first chapter of his classic text, The Principles of Quantum Mechanics (Oxford, Clarendon Press, 3rd edition, 1947, p. 9.) There Dirac points out that the interference of the two component beams of the Michelson Interferometer cannot be interpreted as taking place because the photons of one beam sometimes annihilate photons from the other and sometimes combine to produce four photons. "This would contradict the conservation of energy. The new theory, which connects the wave functions with probabilities for one photon, gets over the difficulty by making each photon go partly into each of the two components. Each photon then interferes only with itself. Interference between two different photons never occurs." These remarks were only intended to refer to an experimental situation generically similar to that of Young's experiment, one in which the interference pattern is revealed by detecting single photons. To attempt to apply Dirac's remarks as a general doctrine for dealing with other types of interference experiments may lead to contradictions, as we shall presently see.

FIRST-ORDER COHERENCE

The word "coherence" is used not only in optics, but in a variety of quantum mechanical and communication theoretical contexts as well. We shall not attempt to construct an encyclopedia of these usages here. We shall try instead to give the term a precise meaning when applied to electromagnetic fields. The meaning we shall adopt is in fact one which links several of these conventional usages together.

The familiar concept of optical coherence is associated with the possibility of producing interference fringes when two fields are superposed. Let us return to the expression (7.3) for the intensity observed in Young's experiment. It is clear that no fringes will be observed if the correlation function $G^{(1)}(x_1, x_2)$ vanishes, and we may describe that condition by saying that the fields at x_1 and x_2 are incoherent.

It is only natural, on the other hand, to associate the highest degree of coherence with a field which exhibits the strongest possible interference fringes. Now, in the last lecture, we have derived a general inequality, (Eq. 6.17), which states

$$|G^{(1)}(x_1x_2)| \leq \{G^{(1)}(x_1x_1)\, G^{(1)}(x_2x_2)\}^{\frac{1}{2}} .$$

When we keep the intensities $G^{(1)}(x_1x_1)$ and $G^{(1)}(x_2x_2)$ fixed, the strongest contrast of the fringe intensities which is possible corresponds to using the equality sign in this relation. Thus we have established the necessary condition for coherence

$$|G^{(1)}(x_1, x_2)| = \{G^{(1)}(x_1,x_1)\, G^{(1)}(x_2, x_2)\}^{\frac{1}{2}} . \tag{7.4}$$

If we introduce the normalized correlation function

$$g^{(1)}(x_1,x_2) = \frac{G^{(1)}(x_1, x_2)}{\{G^{(1)}(x_1, x_1)\, G^{(1)}(x_2, x_2)\}^{\frac{1}{2}}} , \tag{7.5}$$

the condition (7.4) becomes

$$|g^{(1)}(x_1, x_2)| = 1 \tag{7.6}$$

or, in other words,

$$g^{(1)}(x_1, x_2) = e^{i\varphi(x_1, x_2)} .$$

Substitution in (7.3) now gives for the intensity in Young's experiment

$$\begin{aligned} |\lambda|^{-2} I &= G^{(1)}(x_1, x_1) + G^{(1)}(x_2, x_2) + 2\{G^{(1)}(x_1, x_1)\, G^{(1)}(x_2, x_2)\}^{\frac{1}{2}} \cos\varphi(x_1, x_2) \\ &= \left|\{G^{(1)}(x_1, x_1)\}^{\frac{1}{2}} e^{i\varphi(x_1, x_2)} + \{G^{(1)}(x_2, x_2)\}^{\frac{1}{2}}\right|^2 . \end{aligned} \tag{7.7}$$

This intensity varies between the limits

$$I_{min} = (\{G^{(1)}(x_1, x_1)\}^{\frac{1}{2}} - \{G^{(1)}(x_2, x_2)\}^{\frac{1}{2}})^2 \tag{7.8}$$

and

$$I_{max} = (\{G^{(1)}(x_1, x_1)\}^{\frac{1}{2}} + \{G^{(1)}(x_2, x_2)\}^{\frac{1}{2}})^2 . \tag{7.9}$$

The parameter which is usually called the visibility of the fringes is given by

$$v = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{2\{G^{(1)}(x_1, x_1)\, G^{(1)}(x_2, x_2)\}^{\frac{1}{2}}}{G^{(1)}(x_1, x_1) + G^{(1)}(x_2, x_2)} \tag{7.10}$$

If the fields incident on the two pinholes have equal intensity, i.e., if $G^{(1)}(x_1, x_1) = G^{(1)}(x_2, x_2)$, then the intensity varies between zero and $4G^{(1)}(x_1, x_1)$ and the visibility is $v = 1$.

The condition (7.4) is only a condition on the fields at two space-time points x_1 and x_2. When it is satisfied we might speak of the fields at those two points as being coherent with one another. That would correspond to the usage adopted by Born and Wolf in their discussion of classical fields on the basis of time-averaged correlation functions.

In quantum mechanics one characteristically thinks of the entire field as a dynamical system. It will be rather more convenient, therefore, for many analytical and statistical purposes to think of coherence as an idealized property of whole fields. That property can be described in terms of the condition (7.4), but an equivalent and mathematically more useful description can be given in terms

of the requirement that the first order correlation function factorize. Let us suppose that the correlation function $G^{(1)}(x_1, x_2)$ separates into a product of two functions $A(x_1)$ and $B(x_2)$. Then from

$$G^{(1)}(x_1, x_2) = A(x_1)\, B(x_2) \tag{7.11}$$

we conclude via the symmetry relation, Eq. (6.7), that the functions A and B obey the identity

$$A(x_2)\, B(x_1) = A^*(x_1)\, B^*(x_2)$$

or

$$\frac{A(x_2)}{B^*(x_2)} = \frac{A^*(x_1)}{B(x_1)} \tag{7.12}$$

Since in the latter relation a function of x_1 is equated to one of x_2 both functions must be constant. Furthermore the constant, let us call it μ, must be real as we can see by equating x_1 and x_2. We thus have

$$A(x) = \mu\, B^*(x), \tag{7.13}$$

and from the fact that $G^{(1)}(x, x)$ is positive it becomes evident that μ is positive. Hence, if we define the function

$$\mathcal{E}(x) = \sqrt{\mu}\; B(x), \tag{7.14}$$

we see that the first order correlation function falls into the form

$$G^{(1)}(x_1, x_2) = \mathcal{E}^*(x_1)\, \mathcal{E}(x_2). \tag{7.15}$$

This explicit construction of the factorized form of the correlation function shows that, when factorization does take place, the function $\mathcal{E}(x)$ is almost uniquely determined. The only ambiguity which remains is that of a constant multiplicative phase factor.

We shall find it most convenient to use the factorization property (7.15) as our definition of optical coherence or first-order coherence of the field. It is immediately evident that this condition implies the conditions (7.4) and (7.6) on the absolute values of the correlation functions. In fact, it is also true that the latter conditions, if they hold at all points in the field, imply in turn the factorization condition (7.15). We shall demonstrate that shortly and thereby show that the two ways of discussing coherence are equivalent. But first let us discuss some examples of coherent fields.

The most elementary example of a field for which $G^{(1)}$ factorizes is any classical field for which the Fourier coefficients C_k are precisely determined, i.e., any field for which the probability distribution $P(\{C_k\})$ reduces to a product of delta-functions. In that case the function $\mathcal{E}(x)$ is simply the classical field $E^{(+)}(x)$ itself. We perceive here a first hint of the close association which exists between coherence and noiselessness, an association which we shall presently explore further. The absence of randomness or noise in the specification of the Fourier coefficients of a field has long been the criterion used by communication engineers for speaking of a "coherent" signal.

To see another illustration of coherence let us note that one of the possible ways of performing Young's experiment, though perhaps not the most practical one, is to begin with a single photon wave-packet incident upon the first pinhole. Then if we repeat the experiment many times, duplicating the wave packet precisely in each repetition, we should expect to see the familar interference fringes in the statistical distribution of photons received on the final screen. That pure states for single photons are always capable of giving rise to fringes, in this statistical sense, may be seen by examining the first-order correlation function. Let us suppose that the field is in some pure single-photon state which we denote by

$|1\ \text{phot.}\rangle$. Then the density operator for the field is

$$\rho = |1\ \text{phot.}\rangle\langle 1\ \text{phot.}| \tag{7.16}$$

and the first order correlation function reduces to

$$G^{(1)}(x_1, x_2) = \langle 1\ \text{phot.}|E^{(-)}(x_1)\, E^{(+)}(x_2)|1\ \text{phot.}\rangle \tag{7.17}$$

Now since $E^{(+)}$ is a photon annihilation operator, the state $E^{(+)}(x_2)\,|1\ \text{phot.}\rangle$ can only be a multiple of the vacuum state which we denote as $|0\rangle$. It is therefore possible to insert the projection operator upon the vacuum state, $|0\rangle\langle 0|$, between the $E^{(-)}$ and $E^{(+)}$ operators in Eq. (7.17) without altering the value of the correlation function. When we do that we find

$$G^{(1)}(x_1, x_2) = \langle 1\ \text{phot.}|E^{(-)}(x_1)|0\rangle\langle 0|E^{(+)}(x_2)|1\ \text{phot.}\rangle, \tag{7.18}$$

which is exactly the factorized form required by Eq. (7.15). Hence any pure state in which the field is occupied by a single photon possesses first order coherence. (In this way the optical definition of coherence makes contact with some of the ways in which the term is used quantum mechanically in connection with pure states.)

We have, of course, only proved that a pure one photon state is coherent. If, for example, we repeat our hypothetical one-photon interference experiment without duplicating the same wave packet each time, i.e., if we consider a mixture of pure states, then we can not expect in general to observe intensity fringes of maximum contrast. Certain particular mixtures of one photon states may, however, preserve the factorization property (7.15) of the correlation function and thereby preserve the coherence property. Hence we must not think of pure states as the only ones which bring about coherence.

To give an example, let us suppose that only one mode of the field is excited, say the k - th. Then, since the other modes all remain in their ground states, it is easily seen that we may ignore them altogether in calculating the correlation function. Now if the density operator for the k - th mode assumes the general form

$$\rho = \sum_{n,m} c_{n,m}\, |n\rangle\langle m|, \tag{7.19}$$

where $|n\rangle$ is the n-th quantum state for the mode, we may write the first-order correlation function as

$$G^{(1)}(r_1 t_1, r_2 t_2) = \frac{1}{2}\hbar\omega_k \sum_{n,m} c_{nm} \langle m|a_k^\dagger a_k|n\rangle\, u_k^*(r_1) u_k(r_2)\, e^{i\omega_k(t_1 - t_2)}$$

$$= C^2\; u_k^*(r_1) e^{i\omega_k t_1}\; u_k(r_2) e^{-i\omega_k t_2}, \tag{7.20}$$

where in the first of these expressions we have anticipated some of the notation of Eq. (8.21) and in the second we have used the definition

$$C^2 = \frac{1}{2}\hbar\omega_k \sum_n n\, c_{nn}. \tag{7.21}$$

It is clear from the possibility of writing

$$\mathcal{E}(r, t) = C\, u_k(r) e^{-i\omega_k t} \tag{7.22}$$

that the correlation function (7.20) falls into the factorized from (7.15). Hence the excitation of a single mode, whether it is in a pure state or an arbitrary mixture, leads to fields with first-order coherence.

Although we have been able to give some simple examples of fields which possess first order coherence, it is worth pointing out that the factorization condition (7.15) is quite a restrictive one. It is, for example, not satisfied by pure

states of the field in general as one may easily verify by calculating the correlation function for a state in which two or more photons are present and occupy different modes. Initial states such as these may lead to fringes in Young's experiment but the fringes will not, as a rule, satisfy the condition of maximum contrast. While the coherence condition is a restrictive one, we shall show presently that there exists a much broader class of states which satisfy it than those we have considered thus far.

Let us note particularly that no statement has been made requiring that coherent fields be monochromatic. The fields which satisfy the factorization condition (7.15), or for which interference fringes of maximum (instantaneous) contrast occur, can have arbitrary time dependences. The functions $\mathcal{E}(r, t)$ which determine the correlation functions of these fields may consequently have arbitrary Fourier spectra. What seems perhaps curious about these statements is that the experimental effort to produce nearly coherent beams of light has chiefly been a struggle to produce highly monochromatic ones. The reason for this connection has been that all of the effort has involved the use of stationary light sources. Such sources lead to fields for which the first order correlation function depends only on the difference of two times,

$$G^{(1)}(t_1, t_2) = G^{(1)}(t_1 - t_2) \quad . \tag{7.23}$$

If such fields are to be coherent the correlation function must factorize to the form

$$G^{(1)}(t_1 - t_2) = \mathcal{E}^*(t_1)\,\mathcal{E}(t_2), \tag{7.24}$$

but this is a functional equation which has only exponential solutions. Since the dependence of $G^{(1)}$ on the variable t_2, can only contain positive frequencies we must have $\mathcal{E}(t) \sim e^{-i\omega t}$ for some $\omega > 0$. In other words, a coherent field which is stationary can only be monochromatic.

After giving so precise a definition to first order coherence we must add that it is a rather idealized condition, as is nearly any condition one places upon quantum mechanical states. We must not expect correlation functions for actual fields to obey the factorization condition (7.15) over unlimited ranges of the variables x_1 and x_2. In practice we define coherence lengths and times to describe the ranges of the spatial and temporal variables over which the factorization holds to a good approximation.

FRINGE CONTRAST AND FACTORIZATION

In the foregoing section we have defined coherence, mainly for reasons of mathematical convenience, in terms of a factorization property of the correlation function. That factorization property, we then showed, implies the condition (7.4) on the absolute value of the correlation function, i.e., the condition that the fringes show maximum contrast. Now it is possible to show that the latter condition, provided it holds for all space-time points, also implies the factorization property. The proof we present is taken from a forthcoming paper by U. Titulaer and the author.

When the relation

$$|G^{(1)}(x_1, x_2)| = G^{(1)}(x_1, x_1)\, G^{(1)}(x_2, x_2) \tag{7.25}$$

holds it places severe constraints upon the density operator for the field. These constraints may be found by first noting that Eq. (7.25) implies the existence of operators A such that

$$\mathrm{Tr}(\rho\, A^\dagger A) = 0 \quad . \tag{7.26}$$

To exhibit such operators A we choose an arbitrary space-time point x_0 at which the intensity of the field is non-vanishing, $G^{(1)}(x_0, x_0) \neq 0$, and write

$$A = E^{(+)}(x) - \frac{G^{(1)}(x_0, x)}{G^{(1)}(x_0, x_0)}\, E^{(+)}(x_0) \quad . \tag{7.27}$$

It then follows that

$$\mathrm{Tr}(\rho\, A^\dagger A) = G^{(1)}(x\, x_0) - \frac{|G^{(1)}(x_0 x)|^2}{G^{(1)}(x_0\, x_0)} = 0 \tag{7.28}$$

for all points x. Now the density operator ρ can be written as an average of products of the state vectors of the system having the form

$$\rho = \sum_i p_i\, |i><i| \quad , \tag{7.29}$$

where the probabilities p_i are all positive. The vanishing of the trace given by Eq. (7.26) means that

$$\sum_i p_i <i|A^\dagger A|i> = 0 \quad . \tag{7.30}$$

Since all the terms entering the sum are intrinsically positive, we may conclude that

$$<i|\, A^\dagger A|i> = 0 \tag{7.31}$$

for all states $|i>$ for which $p_i \neq 0$. But this relation implies in turn that these states $|i>$ are eigenstates of A with eigenvalue zero

$$A|i> = 0 \quad . \tag{7.32}$$

What we have shown is that the vanishing of the trace (7.26) implies the pair of operator relations

$$A\rho = \rho A^\dagger = 0 \quad . \tag{7.33}$$

Since these relations hold when the operator A takes on the value given by Eq. (7.27), the density operator must obey the pair of identities

$$E^{(+)}(x)\, \rho = \frac{G^{(1)}(x_0, x)}{G^{(1)}(x_0, x_0)}\, E^{(+)}(x_0)\, \rho \tag{7.34}$$

$$\rho\, E^{(-)}(x) = \frac{G^{(1)}(x, x_0)}{G^{(1)}(x_0, x_0)}\, \rho\, E^{(-)}(x_0) \quad . \tag{7.35}$$

These identities may now be used to shift the arguments of correlation functions to a common reference point x_0. If we let $x = x_2$ in the first of these identities and $x = x_1$ in the second of them we may then use them to construct the relation

$$\mathrm{Tr}\{\rho\, E^{(-)}(x_1)\, E^{(+)}(x_2)\} = \frac{G^{(1)}(x_1, x_0)}{G^{(1)}(x_0, x_0)}\, \mathrm{Tr}\{\rho\, E^{(-)}(x_0)\, E^{(+)}(x_0)\} \frac{G^{(1)}(x_0, x_2)}{G^{(1)}(x_0, x_0)},$$

which can also be written as the functional identity

$$G^{(1)}(x_1, x_2) = \frac{G^{(1)}(x_1, x_0)\, G^{(1)}(x_0, x_2)}{G^{(1)}(x_0, x_0)} \quad .$$

Now we have only to define the function $\mathcal{E}(x)$ as

$$\mathcal{E}(x) = \frac{G^{(1)}(x_0, x)}{\{G^{(1)}(x_0, x_0)\}^{\frac{1}{2}}} \tag{7.36}$$

in order to see that the first order correlation function takes on the factorized form

$$G^{(1)}(x_1, x_2) = \mathcal{E}^*(x_1)\,\mathcal{E}(x_2) \quad . \tag{7.37}$$

There is no need to repeat this demonstration in order to deal with the tensor structure of the correlation functions for fields which are not fully polarized. All we need to do is to consider each coordinate x as specifying a tensor index as well as a position and time.

Lecture VIII INTERPRETATION OF INTENSITY INTERFEROMETER EXPERIMENTS

In the preceding lecture we have discussed Young's experiment at some length as an example typical of the interference experiments which are based upon the measurement of a first order correlation function. While all of the older interference experiments share this charact r, we have discussed in the second lecture some more recent experiments which are of a fundamentally different type. These are the intensity interferometry experiments of Hanbury Brown, and Twiss which measure, in effect, the second order correlation function of the incident field.

We have given a simple classical discussion of the way in which the correlation fringes appear in the intensity interferometer when the field is produced by a pair of sources with small angular separation. It is interesting, therefore, to investigate the quantum mechanical origin of these same fringes. If we remember that the intensity interferometer functions by first detecting the incident fields in each of two receivers, we see immediately that pairs of photons must be involved in the interference effect, i.e., nothing is recorded at all unless different photons are incident on each of the two detectors at more or less the same time. It is at precisely this point that one is confronted by a serious dilemma if he attaches too great a generality to Dirac's statement that "interference between two different photons never occurs."

The general discussion of interference which we gave in the last lecture should make it clear that no such dilemma need exist. The things which should be regarded as interfering are not, strictly speaking, the photons, but alternative "histories" of the system as a whole. Let us imagine that the initial state of the system is one in which two (generally overlapping) single-photon wave packets are present in the field and the atoms of the two detectors (represented by photon counters) are in the ground state. We may take the final state of the system to be one in which both photons have been absorbed and one atom in each of the counters is correspondingly excited. If we label the photons 1 and 2, and the two counters a and b, we see that there are two alternative ways in which the final state may be reached. Either photon 1 is absorbed by counter a and 2 by b, or 1 is absorbed by b and 2 by a.

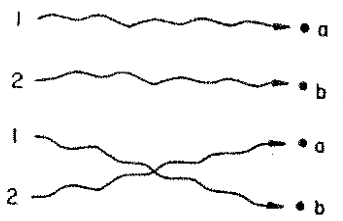

Figure 8

If the packets had altogether different average propagation vectors these alternative histories would be distinguishable by means of careful measurements

made in the counters. But the circumstances in which the fringes are observable are precisely those in which the packets have nearly the same average propagation vectors (e.g., packets with the same frequencies, small angular separation of the sources). In other words the fringes appear once again just when the alternative histories of the system become indistinguishable. Since the amplitudes for the two histories interfere, it becomes meaningless to ask which counter absorbed which photon.

HIGHER ORDER COHERENCE AND PHOTON COINCIDENCES

We recall from our classical discussions of the second lecture that the intensity interferometer measures the second order correlation function of the incident field. Radiation fields generated by natural sources tend to have a chaotic quality which allows us to construct these correlation functions from a knowledge of the first order functions. However, no such constructions are available in general for dealing with radiation from man-made sources such as the laser or radio transmitters. The fields, generated by these sources can have much higher regularity than is ever possible for natural sources. It will be useful, therefore, to sharpen the concept of coherence by defining higher order analogues of optical coherence.

We begin once more by stating conditions on the absolute values of the correlation functions. For full coherence we shall require that the normalized form of the n-th order correlation function,

$$g^{(n)}(x_1 \cdots x_{2n}) = \frac{G^{(n)}(x_1 \cdots x_{2n})}{\prod_{j=1}^{n} \{G^{(1)}(x_j, x_j)\}^{\frac{1}{2}}} \quad , \tag{8.1}$$

have modulus unity for all n and all combinations of arguments x. If the functions have unit modulus only for $n \leq M$ we shall speak of M-th order coherence.

The concept of M-th order coherence has a simple interpretation in terms of n-fold (delayed) coincidence experiments. We know that $G^{(n)}(x_1 \cdots x_n, x_n \cdots x_1)$ is an average coincidence rate for n ideal photo-detectors registering at the points $x_1 \cdots x_n$. Since this value of the function is real and positive the condition that $g^{(n)}$ have unit modulus for $n \leq M$ implies that

$$g^{(n)}(x_1 \cdots x_n, x_n \cdots x_1) = 1$$

for $n \leq M$. Hence for fields with M-th order coherence, it is clear from the definition of $g^{(n)}$ that we have

$$G^{(n)}(x_1 \cdots x_n, x_n \cdots x_1) = \prod_{j=1}^{n} G^{(1)}(x_j\, x_j) \tag{8.2}$$

for $n \leq M$.

Expressed in experimental terms, this means that the n-fold coincidence rate is just the product of the counting rates which would be measured by each counter individually in the absence of the others. Thus there is no tendency toward statistical correlation of the photon counts. In a field with coherence of order $M \geq n$ the n photon counters register in a statistically independent way.

Several investigations of light beams using coincidence counting of photons or equivalent experimental procedures have in fact been carried out during the last few years. The first of these to detect a tendency toward statistical correlation of the arrival times of photons was performed (in addition to the other experiments we have mentioned) by Hanbury Brown, and Twiss. [1] In the experiment light from a source S (Fig. 9) passes through a pinhole P and then reaches a half-silvered mirror m, which splits it into two beams. Detectors D_1 and D_2 are placed symmetrically with respect to the mirror. Their photocurrents are multiplied together by the correlator C whose average output is the quantity measured. We may

consider the half-silvered mirror m as a device, which permits us, in effect, to place two different photodetectors at essentially the same position in the beam.

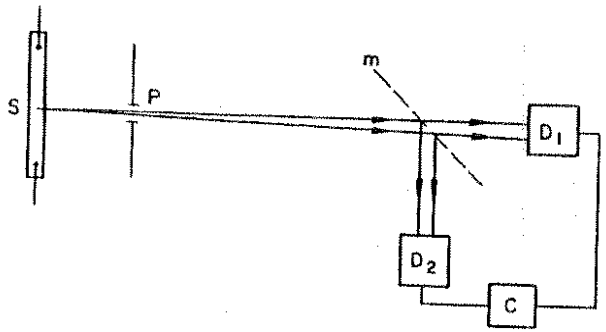

Figure 9

Shortly after the original experiment had been performed another version of it with a slightly more direct interpretation was performed by Rebka and Pound.[2] In the latter experiment D_1 and D_2 are counters of individual photons, and C is a device for registering delayed coincidences. The experiment measures the average coincidence rate as a function of delay time while the counters remain fixed in their symmetrical positions relative to the mirror. Now, even if the photon beams incident on the two counters were statistically independent of one another, there would be a certain background counting rate of accidental coincidences. This rate would, however, be independent of any time delay. Thus any observed dependence of the coincidence rate on the time delay indicates a lack of statistical independence.

The result of the experiments is indicated in Fig. 10. If the responses of the counters were statistically independent the coincidence rate would be independent of time delay. The observation of a small "bump" in the experimental curve

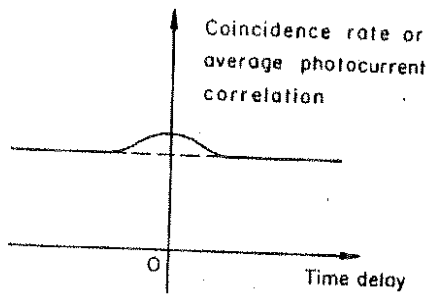

Figure 10

indicates that the photons have a distinct tendency to arrive in pairs. Although the effect was at first difficult to observe it is, as we shall show, not necessarily a small one at all. The small magnitude of the observed "bump" and its particular shape in these experiments were determined almost entirely by the relatively slow response times of the counters.

Let us note that, if the counters are placed symmetrically with respect to the mirror, the fields which are incident upon them are essentially identical, apart from a constant multiplicative factor. It follows then that if r_1 and r_2 are mirror-image points in the two detectors we have

$$|g^{(1)}(r_1 t,\ r_2 t)| = 1 \tag{8.3}$$

i.e., the fields which fall on the two detectors have essentially perfect first order coherence. The observation of a positive correlation in the coincidence rate demonstrates, on the other hand, that the fields are not coherent in the second order sense. We shall show presently that this result is a characteristic one for all experiments performed with natural light sources. These have a random character which destroys second order coherence.

FURTHER DISCUSSION OF HIGHER ORDER COHERENCE

Let us return now to the definition of higher order coherence. We have, by analogy with first order coherence, defined M-th order coherence in terms of the succession of conditions

$$|G^{(n)}(x_1 \cdots x_{2n})|^2 = \prod_{j=1}^{2n} G^{(1)}(x_j,\ x_j) \tag{8.4}$$

on the absolute values of the correlation functions for $n \leq M$. Just as in the first order case we found it convenient to express the coherence condition in an alternative way, as a factorization property of the correlation function, we shall find it even more convenient here to do much the same thing. We shall therefore state as an alternative definition the requirement that there exist a single complex function $\mathcal{E}(x)$ such that

$$G^{(1)}(x_1 \cdots x_{2n}) = \prod_{j=1}^{n} \mathcal{E}^*(x_j) \prod_{j=n+1}^{2n} \mathcal{E}(x_j) \tag{8.5}$$

for all $n \leq M$. If this factorization holds for all n we shall speak of full coherence.

If we note that the definition (8.5) contains the statement

$$G^{(1)}(x,\ x) = |\mathcal{E}(x)|^2, \tag{8.6}$$

then we see immediately that it requires that the correlation functions obey the absolute value conditions (8.4).

It is possible, on the other hand, to show that the absolute value conditions also imply the factorization properties. To do that we note that M-th order coherence always requires first order coherence. We may therefore make use of the identities which were shown in the last lecture to be consequences of first order coherence. In particular, since the operators $E^{(-)}(x_j)$ for $j = 1, \cdots n$ all commute with one another, as do the operators $E^{(+)}(x_j)$ for $j = n+1, \cdots 2n$, we can use each of the two identities (7.34) and (7.35) n times in order to shift all of the arguments of the n-th order correlation function to a particular reference point x_0. More specifically, we write

$$\mathrm{Tr}\{\rho\, E^{(-)}(x_1) \cdots E^{(-)}(x_n)\, E^{(+)}(x_{n+1}) \cdots E^{(+)}(x_{2n})\}$$

$$= \prod_{j=1}^{n} \frac{G^{(1)}(x_j, x_0)}{G^{(1)}(x_0, x_0)}\, \mathrm{Tr}\{\rho\, E^{(-)}(x_0) \cdots E^{(-)}(x_0)\, E^{(+)}(x_0)\, E^{(+)}(x_0)\}$$

$$\prod_{j=n+1}^{2n} \frac{G^{(1)}(x_0, x_j)}{G^{(1)}(x_0, x_0)},$$

which is the identity

$$G^{(n)}(x_1 \cdots x_{2n}) = \frac{G^{(n)}(x_0 \cdots x_0)}{\{G^{(1)}(x_0, x_0)\}^n} \cdot \frac{\prod_{j=1}^{n} G^{(1)}(x_j, x_0) \prod_{j=n+1}^{2n} G^{(1)}(x_0, x_j)}{\{G^{(1)}(x_0, x_0)\}^n}$$

If we introduce the function $\mathcal{E}(x)$ which is defined by Eq. (7.37), and make use of the normalized form of the correlation function, we may write the latter identity in the form

$$G^{(n)}(x_1 \cdots x_{2n}) = g^{(n)}(x_0 \cdots x_0) \prod_{j=1}^{n} \mathcal{E}^*(x_j) \prod_{j=n+1}^{2n} \mathcal{E}(x_j). \qquad (8.7)$$

Now, as we have shown earlier, the functions $\mathcal{E}(x)$ can only depend on the choice of the arbitrary reference point x_0 through a constant phase factor. Since that phase factor cancels out of the product which occurs in Eq. (8.7), it follows that for fields with first order coherence the functions $g^{(n)}(x_0 \cdots x_0)$ are independent of x_0. In other words, the condition of first order coherence alone is sufficient to bring all of the higher order correlation functions into a factorized form, although not exactly the form, in general, which is required for higher order coherence. The difference is that Eq. (8.7) contains the constant factors $g^{(n)}(x_0 \cdots x_0)$ which should be unity if higher order coherence is to hold. Now the higher order coherence conditions (8.4) do require these coefficients to have unit absolute value for $n \leq M$. Then, since the $g^{(n)}(x_0 \cdots x_0)$ must be real and positive, they must be equal to one.

Hence the conditions (8.4) do indeed imply the factorization condition (8.5).

TREATMENT OF ARBITRARY POLARIZATIONS

From a mathematical standpoint, very little need be added to our earlier discussions in order to treat fields with arbitrary polarization properties rather than the fully polarized fields we have been discussing. All we need do, as we have already noted, in order to deal with the general tensor character of the correlation functions, is to think of every coordinate in the formulae we have derived as specifying a tensor index as well as a position and time.

Thus the relations (6.7) for $n = 1$ and (6.17), for example, may be generalized to read

$$\{G^{(1)}_{\mu\nu}(x_1, x_2)\}^* = G^{(1)}_{\nu\mu}(x_2, x_1) \qquad (8.8)$$

and

$$|G^{(1)}_{\mu\nu}(x_1, x_2)|^2 \leq G^{(1)}_{\mu\mu}(x_1, x_1)\, G^{(1)}_{\nu\nu}(x_2, x_2). \qquad (8.9)$$

It may be worth noting that all information about the state of polarization of the field is contained in the correlation tensor $G^{(1)}_{\mu\nu}(x, x)$. Let us denote this tensor by $\mathcal{G}_{\mu\nu}$. We see immediately that $\mathcal{G}_{\mu\nu}$ is a Hermitian matrix, $\mathcal{G}_{\mu\nu}{}^* = \mathcal{G}_{\nu\mu}$. If we substitute $A = \sum_{\mu=1}^{3} \lambda_\nu E^{(+)}_\nu(x)$ in the general inequality $\mathrm{Tr}\{\rho A^\dagger A\} \geq 0$ we find

$$\sum_{\mu,\nu=1}^{3} \lambda_\mu{}^* \lambda_\nu \mathcal{G}_{\mu\nu} \geq 0. \qquad (8.10)$$

Thus $\mathcal{G}_{\mu\nu}$ is also positive definite. Because of its Hermitian character $\mathcal{G}_{\mu\nu}$ can be diagonalized, that is to say there exist three real and positive eigenvalues λ_p and three (generally complex) eigenvectors $e^{(p)}$, such that

$$\mathcal{G} \cdot e^{(p)*} = \lambda_p e^{(p)*} \quad ; \quad e^{(p)} \cdot \mathcal{G} = \lambda_p e^{(p)}. \qquad (8.11)$$

Note that both the λ_p and the $e^{(p)}$ depend in general on the space-time point x, that occurs in the definition of $\mathcal{G}$.

The $e^{(p)}$ are either found to be mutually orthogonal if the λ's have no degeneracy,

or they can be chosen orthogonal if the λ's are degenerate. Hence we may assume

$$e^{(p)} \cdot e^{(q)*} = \delta_{pq}. \qquad (8.12)$$

Since the tensor product

$$e^{(p)} \cdot \mathcal{G} \cdot e^{(q)*} = \lambda_p \delta_{pq} \qquad (8.13)$$

expresses the correlation of the field components in the directions of $e^{(p)}$ and $e^{(q)}$ there are three "directions" (i.e., complex directions) in which the field components are mutually uncorrelated. Any field may thus be regarded as a superposition of three orthogonally polarized fields whose amplitudes are (instantaneously) uncorrelated.

The eigenvalues $\lambda^{(p)}$ are the intensities corresponding to the three polarizations. The total intensity is given by

$$\mathrm{Tr}\, \mathcal{G} = \sum_p \lambda_p. \qquad (8.14)$$

A set of normalized intensities can be defined as

$$I_p = \frac{\lambda_p}{\sum_{j=1}^{3} \lambda_j} \qquad (p = 1, 2, 3).$$

These numbers can be interpreted as specifying the degree of polarization of the field. In an isotropic radiation field we must have $I_p = 1/3$, $(p = 1, 2, 3)$. If the field is stationary i.e., $[\rho, H] = 0$ then $\mathcal{G}$ is time independent and the λ_p and I_p and $e^{(p)}$ become fixed at any spatial position $\mathbf{r}$.

If we are considering a beam with a single direction of propagation $\hat{k}$, then clearly $\hat{k} \cdot \mathcal{G} = \mathcal{G} \cdot \hat{k} = 0$ (since light is a transverse wave). Hence $\hat{k}$ is an eigenvector of $\mathcal{G}$ corresponding to the eigenvalue $\lambda = 0$. Then there are two remaining eigenvalues λ_p, $p = 1, 2$. The net polarization of the beam is usually defined as $|I_1 - I_2| = |\lambda_1 - \lambda_2|/(\lambda_1 + \lambda_2)$. The two polarizations $e^{(p)}$ for $p = 1, 2$ clearly lie in the plane perpendicular to $\hat{k}$.

The higher order correlation tensors are defined by

$$G^{(n)}_{\mu_1 \cdots \mu_{2n}}(x_1 \cdots x_{2n}) = \mathrm{Tr}\{\rho E_{\mu_1}{}^{(-)}(x_1) \cdots E^{(-)}_{\mu_n}(x_n) E^{(+)}_{\mu_{n+1}}(x_{n+1}) \cdots E^{(+)}_{\mu_{2n}}(x_{2n})\}. \qquad (8.15)$$

The coherence condition, Eq. (8.5), may evidently be restated for fields of arbitrary polarization by requiring that there exist a vector function $\mathcal{E}_\mu(x)$ such that

$$G_{\mu_1 \cdots \mu_{2n}}(x_1 \cdots x_{2n}) = \prod_{j=1}^{n} \mathcal{E}^*_{\mu_j}(x_j) \prod_{j=n+1}^{2n} \mathcal{E}_{\mu_j}(x_j) \qquad (8.16)$$

for $n \leq M$.

As a last remark on polarizations we note that first order coherence implies full polarization of the field, i.e., if we have

$$G^{(1)}_{\mu\nu}(xx) = \mathcal{G}_{\mu\nu} = \mathcal{E}_\mu{}^*(x)\, \mathcal{E}_\nu(x), \qquad (8.17)$$

then clearly the vector $\mathcal{E}_\mu(x)$ itself is an eigenvector. The corresponding intensity is $\sum_{\mu=1}^{3} |\mathcal{E}_\mu(x)|^2$, which is the full intensity of the field present.

COHERENT STATES OF THE FIELD - INTRODUCTION

Let us try to construct states in which the fields have full coherence, that is to say, states in which all the correlation functions $G^{(n)}$ factorize according to Eqs. (8.5) or (8.16). If there existed simultaneous eigenstates of the operators $E^{(+)}$ and $E^{(-)}$, such eigenstates would clearly bring about the desired factorization. However, since $E^{(+)}$ and $E^{(-)}$ do not commute (and have a commutator which is a c-number) it is clear that no such eigenstates exist. We may reduce our demand to a more plausible level by noting that in the correlation functions the field operators always occur in normal order. Therefore, it is sufficient to secure coherence if the state of the field is simply an eigenstate of $E^{(+)}$ in the restricted sense

$$E^{(+)}_{\mu}(x)|> \;=\; \mathcal{E}_{\mu}(x)|> \quad . \tag{8.18}$$

This is true because the adjoint relation is

$$<|E^{(-)}_{\mu}(x) \;=\; \mathcal{E}_{\mu}^{*}(x)<| \quad , \tag{8.19}$$

and together the two relations lead to the desired factorization of the correlation functions.

Since the operator $E^{(+)}$ is neither Hermitian nor normal (i.e., it does not commute with its Hermitian adjoint), there is no a priori reason why eigenstates of this form should exist. Indeed it is easily shown that the similar relation

$$<|E^{(+)}(x) \;=\; \mathcal{E}(x)<| \tag{8.20}$$

can have no normalizable solution at all. The simplest way to show that Eq. (8.18) has solutions is to construct them.

If any solution of Eq. (8.18) is to exist then it is clear that the function $\mathcal{E}_{\mu}(x)$ must satisfy the same wave equation and boundary conditions as the operator $E^{(+)}_{\mu}(x)$. The latter has the Fourier expansion

$$\begin{aligned} E^{(+)}(r,t) &= -\frac{1}{c}\frac{\partial A^{(+)}}{\partial t} \\ &= i\sum_{k}\left\{\frac{\hbar\omega_k}{2}\right\}^{1/2} a_k u_k(r) e^{-i\omega_k t} \end{aligned} \quad . \tag{8.21}$$

Here the time independent operators a_k are described completely by means of their commutation relations

$$\begin{aligned} [a_k, a_{k'}] \;&=\; [a_k^{\dagger}, a_{k'}^{\dagger}] \;=\; 0 \\ [a_k, a_{k'}^{\dagger}] \;&=\; \delta_{kk'} \quad . \end{aligned} \tag{8.22}$$

For $\mathcal{E}(r,t)$ we must have a corresponding expansion

$$\mathcal{E}(r,t) \;=\; i\sum_{k}\left\{\frac{\hbar\omega_k}{2}\right\}^{1/2} \alpha_k u_k(r) e^{-i\omega_k t} \quad , \tag{8.23}$$

where the coefficients α_k are a set of numbers which can take on arbitrary complex values.

Now if we substitute the expansions (8.21) and (8.23) in the equation which determines the eigenstates, we see that the coefficients of each mode function must separately be equal. Hence the eigenstate must satisfy the conditions

$$a_k|> \;=\; \alpha_k\,|> \tag{8.24}$$

for all modes k.

The coefficients α_k correspond in a simple way to the classical Fourier coefficients C_k which we introduced in the first lecture. More specifically if we compare Equations (1.8) and (8.23) we see that the correspondence is

$$C_k \;=\; i\left\{\frac{\hbar\omega_k}{2}\right\}^{1/2} \alpha_k \quad . \tag{8.25}$$

This relation shows that to describe classical fields we shall have to deal with parameters α_k of large modulus, i.e., if we let $\hbar \to 0$ then α_k increases as $\hbar^{-1/2}$.

To construct the desired eigenstate we can begin with the construction of a state $|\alpha_k>_k$ for the single mode k, such that

$$a_k|\alpha_k>_k \;=\; \alpha_k|\alpha_k>_k \quad . \tag{8.26}$$

The state for the entire system is then given by the direct product

$$|> \;=\; \prod_k |\alpha_k>_k \quad . \tag{8.27}$$

We shall call these states the coherent states. From the fact that they remain the same, up to a numerical factor, when we apply an annihilation operator a_k, it follows immediately that they cannot be eigenstates of the photon number operator.

The sense in which states of the type (8.27) are coherent includes, of course, optical coherence (they secure factorization of the first order coherence function). But it also includes a sense used in communication theory which we have mentioned earlier. There a coherent signal is a pure signal, one that has no noise. A classical signal of this type is ideally one with a precisely defined set of Fourier coefficients C_k. But this is exactly the kind of field we are talking about in the more general quantum mechanical context. Our precise specification of the Fourier coefficients α_k means, as we shall see, that we are as close as possible to having no noise in the signal. It can not mean, however, that there is no noise at all. Unpredictably fluctuating fields are present even in the vacuum. Our detectors detect individual photons, and photons tend to arrive randomly. Even when we specify the field as accurately as we can, we can only make predictions about the response of our counter in statistical terms; there will be some inevitable noise, and the coherent states of the field only tend to reduce that noise to a minimum.

REFERENCES

(1) R. Hanbury Brown and R.Q. Twiss Nature 177, 27 (1956)
Proc. Roy. Soc. (London) A242 300 (1957)
A243 291 (1957)

(2) G.A. Rebka and R.V. Pound Nature 180, 1035 (1957)