QUANTUM THEORY OF HETERODYNE DETECTION†

M. C. TEICH

Department of Electrical Engineering, Columbia University, New York, New York, U.S.A.

ABSTRACT

A quantum theory of photomixing is presented which applies to nonstationary as well as stationary fields of an arbitrary statistical nature. The primary distinction from the classical theory is that double-frequency and sum-frequency components do not appear in the heterodyne signal, to good approximation, when $h\nu \gg kT$. The simple conditions for optimum photomixing with a sinusoidal heterodyne signal are first-order spatial coherence of the total incident field, and first-order temporal coherence and stationarity of the constituent beams. In the limit $h\nu \gg kT$, the heteodyne process may be interpreted as the annihilation of a single nonmonochromatic photon. The theory is valid for arbitrarily small photon numbers.

1. INTRODUCTION

Heterodyne or coherent detection has been used in many regions of the electromagnetic spectrum.[1,2] We present here a quantum theory of coherent detection which differs from both the classical and the semiclassical treatments which have appeared in the literature. The theory is valid in the region $h\nu \gg kT$, i.e. the infrared and optical. It is shown to reduce to the classical result in the limit of low radiation frequencies ($h\nu \ll kT$) and, for a certain class of fields, to the semiclassical result for high radiation frequencies ($h\nu \gg kT$). The main deviation from the classical theory is that double-frequency and sum-frequency components of the heterodyne signal do not appear, to good approximation, when $h\nu \gg kT$. The theory is valid for fields of an arbitrary statistical nature, and for arbitrarily low photon numbers.

2. THEORY

A generalized schematic for an optical or infrared heterodyne receiver is given in Fig. 1. Two plane parallel electromagnetic waves of angular frequencies ω_1 and ω_2 impinge normally on an ideal quantum-mechanical photodetector in its ground state. It is assumed that the photon energy $h\nu$ is much greater than the thermal excitation energy of the detector, kT. Glauber[3,4] has shown that the average count-rate for such an absorption detector (to good approximation),

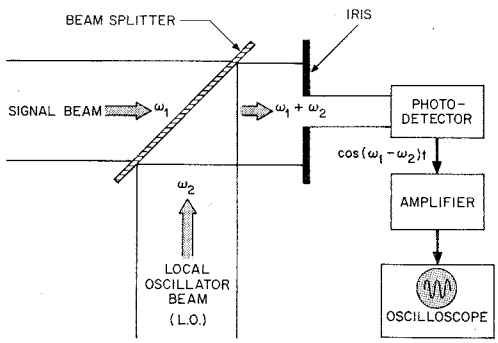

Fig. 1. The generalized infrared or optical heterodyne receiver.

at the space-time point $\chi = \mathbf{r}, t$, may be expressed as the first-order correlation function $G^{(1)}_{\mu\nu}(\chi, \chi)$, where

$$G^{(1)}_{\mu\nu}(\chi, \chi) = \mathrm{tr}\{\rho E_\mu^-(\chi) E_\nu^+(\chi)\}. \qquad (1)$$

Here, ρ is the density operator for the field,[4,5] and E^- and E^+ are the negative- and positive-frequency portions of the electric field operator E respectively. The subscripts μ, ν label Cartesian components, and the symbol tr stands for the trace. For simplicity, only projections of the field along a single (possibly complex) unit vector are considered at this point, so that the correlation function above may be written as a scalar quantity rather than as a tensor.

Coherent detection experiments are frequently performed using a given beam and a

†This work was supported by the National Science Foundation.

time-delayed form of the same beam[6] (so-called homodyne detection), so that it is more convenient to discuss time correlations of the field relative to the radiation source rather than to the detector.[7] That is, the output of a detector illuminated by a single beam is proportional to $G^{(1)}(\chi',\chi')$, where $\chi' = \mathbf{r}, t'$. When illuminated by a phase-retarded form of the same beam, the output of the detector *at time* t' may be written as $G^{(1)}(\chi'',\chi'')$ where $\chi'' = \mathbf{r}, t''$ and $t'' > t'$. Thus, phase retardation is equivalent to time-displacement at the detector. The detection of the total incident field (consisting of two component beams) at the time t is therefore expressed in terms of two individual time parameters t_1 and t_2.

For the heterodyne experiment, we may write the total electric field operator as a superposition of the operators for the constituent waves.[4] The positive-frequency component of the field present at the photodetector, $E^+(\mathbf{r},t)$, may therefore be written

$$E^+(\mathbf{r},t) = \lambda_1 E^+(\mathbf{r},t_1) + \lambda_2 E^+(\mathbf{r},t_2). \tag{2}$$

The complex coefficients λ_1 and λ_2 contain the relative strengths of the two waves, and are taken to be independent of the properties of the field. The count-rate R may then be expressed as

$$\begin{aligned} R &= \mathrm{tr}\{\rho E^-(\mathbf{r},t)E^+(\mathbf{r},t)\} \\ &= \mathrm{tr}\{\rho[\lambda_1^* E^-(\chi_1) + \lambda_2^* E^-(\chi_2)] \\ &\quad \times [\lambda_1 E^+(\chi_1) + \lambda_2 E^+(\chi_2)]\}, \end{aligned} \tag{3}$$

with $\chi_i = \mathbf{r}, t_i$. Again, the quantity t_i is taken relative to the radiation source, and $\mathbf{r}$ represents a point on the detector surface.

We now assume that the *total* incident radiation field would display maximum-contrast fringes[8] in a hypothetical spatial interference experiment performed with the radiation at any two points on the detector surface, at an arbitrary time t. The first-order correlation function $G^{(1)}(r't, r''t)$, contained in the interference term, may then be written as

$$\begin{aligned} G^{(1)}(r't, r''t) = \mathrm{tr}\{\rho[&\lambda_1^* E^-(r't) + \lambda_2^* E^-(r't)] \\ &\times [\lambda_1 E^+(r''t) + \lambda_2 E^+(r''t)]\}. \end{aligned} \tag{4}$$

Thus, at the detection time t, this function may be written as

$$\begin{aligned} G^{(1)}(r't, r''t) = \mathrm{tr}\{\rho[&|\lambda_1|^2 E^-(r't)E^+(r''t) \\ &+ |\lambda_2|^2 E^-(r't)E^+(r''t) \\ &+ 2\,\mathrm{Re}\,\{\lambda_1^*\lambda_2\}E^-(r't)E^+(r''t)]\}. \end{aligned} \tag{5}$$

But, for first-order coherence *at all times* t, as described above, the correlation function factors into the scalar field product $\epsilon^*(r't)\epsilon(r''t)$ so that in this case,

$$\begin{aligned} G^{(1)}(r't, r''t) = &[\lambda_1^*\epsilon^*(r't) + \lambda_2^*\epsilon^*(r't)] \\ &\times [\lambda_1\epsilon(r''t) + \lambda_2\epsilon(r''t)]. \end{aligned} \tag{6}$$

Therefore, from (5) and (6), we obtain

$$\begin{aligned} &\mathrm{tr}\{\rho[|\lambda_1|^2 E^-(r't)E^+(r''t) + |\lambda_2|^2 E^-(r't)E^+(r''t) \\ &\quad + 2\mathrm{Re}\{\lambda_1^*\lambda_2\}E^-(r't)E^+(r''t)]\} \\ &= |\lambda_1|^2\epsilon^*(r't)\epsilon(r''t) + |\lambda_2|^2\epsilon^*(r't)\epsilon(r''t) \\ &\quad + 2\mathrm{Re}\{\lambda_1^*\lambda_2\}\epsilon^*(r't)\epsilon(r''t). \end{aligned} \tag{7}$$

Since this equation holds for abritrary λ_1 and λ_2, it is seen that

$$\mathrm{tr}\{\rho|\lambda_1|^2 E^-(r't)E^+(r''t)\} = \lambda_1^*\epsilon^*(r't)\lambda_1\epsilon(r''t) \tag{8}$$

and

$$\mathrm{tr}\{\rho\lambda_1^*\lambda_2 E^-(r't)E^+(r''t)\} = \lambda_1^*\epsilon^*(r't_1)\lambda_2\epsilon(r''t_2). \tag{9}$$

Equation (8) shows that first-order coherence for the total incident radiation field implies first-order coherence for the individual beams as well. Equation (9) indicates that for such a first-order coherent field, maximum fringe visibility would, in addition, be obtained in an interference experiment similar to that described above, but using the constituent waves instead of the total fields.

Using this assumption, then, and the correlation function identity $[G^{(1)}(\chi_1,\chi_2)]^* = G^{(1)}(\chi_2,\chi_1)$, the count-rate becomes

$$\begin{aligned} R = &|\lambda_1|^2 G^{(1)}(\chi_1,\chi_1) + |\lambda_2|^2 G^{(1)}(\chi_2,\chi_2) \\ &+ 2\mathrm{Re}\{\lambda_1^*\epsilon^*(\chi_1)\lambda_2\epsilon(\chi_2)\}. \end{aligned} \tag{10}$$

The first two terms on the right represent the intensities which would be contributed by each beam independently of the other. The last term represents the interference for constituent fields which are first-order coherent.

We now direct our attention to component beams which are stationary. This condition, coupled with individual first-order temporal coherence for these fields, implies monochromaticity of the individual beams. The function $\epsilon^*(\chi_i)$ for a well-collimated, fully-polarized beam of frequency ω_i may then be expressed as[4]

$$\epsilon^*(\chi_i) = [G^{(1)}(\chi_i, \chi_i)]^{1/2} e^{j(\omega_i t_i - \mathbf{k}_i \cdot \mathbf{r})} \tag{11}$$

where the space-time point χ_i has its usual definition. Inserting this into Eq. (10), a small amount of algebra leads to a count-rate expressed as

$$\begin{aligned} R = {} & |\lambda_1|^2 G^{(1)}(\chi_1, \chi_1) + |\lambda_2|^2 G^{(1)}(\chi_2, \chi_2) \\ & + 2[G^{(1)}(\chi_1, \chi_1) G^{(1)}(\chi_2, \chi_2)]^{1/2} \\ & \times \mathrm{Re}\{\lambda_1^* \lambda_2 e^{j(\omega_1 - \omega_2)t_1} e^{j\omega_2 \tau}\}, \end{aligned} \tag{12}$$

where the quantity $\omega_2 \tau = \omega_2(t_1 - t_2)$ may be thought of as a phase difference between the beams. The spatially dependent exponential portions of ϵ and ϵ^* have been supressed in writing Eq. (12) because the first-order coherence requirement for the total field, imposed across the photodetector surface, insures parallel constituent beams, and we have assumed normal incidence.

Since we do not have advance information about the phase of a particular beam in any experiment, however, in using this theory we should properly choose states which are averaged over phase. Although the interference term in Eq. (12) above will vanish through the ensemble average in this case, the interference would be present in any individual experiment. We assume that we can select an ensemble by considering only experiments with the same phase difference. This permissible procedure is entirely analogous to that used for spatial interference.[4] For convenience, we shall choose the phase difference $\omega_2 \tau$ in such a manner as to precisely cancel the phase factors arising from λ_1^* and λ_2.

The count rate for a restricted ensemble such as that discussed above, and for a field posessing first-order coherence with stationary and temporally coherent constituent beams, may therefore finally be written as

$$\begin{aligned} R = {} & |\lambda_1|^2 G^{(1)}(\chi_1, \chi_1) + |\lambda_2|^2 G^{(1)}(\chi_2, \chi_2) \\ & + 2[|\lambda_1|^2 G^{(1)}(\chi_1, \chi_1) |\lambda_2|^2 G^{(1)}(\chi_2, \chi_2)]^{1/2} \\ & \times \cos(\omega_1 - \omega_2)t. \end{aligned} \tag{13}$$

The phase difference has been conveniently chosen as described above, and the quantity t_1 has been written as t in the interference term. We note that $G^{(1)}(\chi_1, \chi_1)$ and $G^{(1)}(\chi_2, \chi_2)$ are count rates which are constant in time and do not possess any fluctuating components. In terms of the classical intensities I_1 and I_2 for the individual beams, this is equivalent to

$$R = I_1 + I_2 + 2\sqrt{(I_1 I_2)} \cos(\omega_1 - \omega_2)t. \tag{14}$$

This expression differs from the usual classical result[6] in that it does not contain sum- and double-frequency components of ω_1 and ω_2. No contradiction with experiment has been noted by using the classical theory in the optical and infrared, however.[1,6] This is because although these additional terms appear in the classical theory, they are generally disregarded because of the "inability of the detector to follow such rapid fluctuations." However, it is clear from the quantum analysis that these rapidly varying terms never appear for the usual absorption detector when $h\nu \gg kT$, and therefore would not be observed even with detectors of arbitrarily small resolving time.

In the low frequency limit, where $h\nu \ll kT$, the quantum theory for the heterodyne detector reduces to the *classical* result, since in this region photons are emitted as readily as they are absorbed. The two processes, which are described by E^- and E^+, respectively, therefore occur with equal magnitude and we must sum the effects of both. The real field strength E is obtained as a consequence, so that the classical result for the heterodyne detector, including sum- and double-frequency components, is recovered. The *semiclassical* theory,[9] which makes use of the analytic signal, gives the correct form for the heterodyne signal only when

the fields considered are stationary, possess a positive-definite weight function in the P-representation, and are in the realm $h\nu \gg kT$.

In the usual classical or semiclassical formulation of optical or infrared heterodyne detection, a separate requirement for optimum photomixing is that the wave normals of the constituent beams be aligned to within an angle λ/a (λ is the radiation wavelength and a is the detector aperture).[1,9] It is interesting that this requirement need not be stated separately in the quantum theory; rather, it may be seen to be a simple consequence of the general condition of first-order coherence of the total incident radiation field over the detector aperture.

To consider the case where the incident radiation field is not fully coherent (to first-order), or the constituent beams are non-stationary, we write Eq. (10) in its more general form

$$R = |\lambda_1|^2 G^{(1)}(\chi_1, \chi_1) + |\lambda_2|^2 G^{(1)}(\chi_2, \chi_2) + 2|\lambda_1|\,|\lambda_2|\,|G^{(1)}(\chi_1, \chi_2)| \times \cos\{\phi(\chi_1, \chi_2) + \theta\}, \quad (15)$$

where $\phi(\chi_1, \chi_2)$ is a phase function derived from $G^{(1)}(\chi_1, \chi_2)$. The phase angle θ depends on the geometry of the experiment.

Using the correlation-function equality[3] for radiation with maximum visibility fringes

$$|G^{(1)}(\chi_1, \chi_2)| = [G^{(1)}(\chi_1, \chi_1) G^{(1)}(\chi_2, \chi_2)]^{1/2}, \quad (16)$$

we obtain another expression for R,

$$R = |\lambda_1|^2 G^{(1)}(\chi_1, \chi_1) + |\lambda_2|^2 G^{(1)}(\chi_2, \chi_2) + 2[|\lambda_1|^2 G^{(1)}(\chi_1, \chi_1)\,|\lambda_2|^2 G^{(1)}(\chi_2, \chi_2)]^{1/2} \times \cos\{\phi(t_1, t_2)\}, \quad (17)$$

which is equivalent to Eq. (10) except for the (unimportant) suppression of θ. This result is valid for a general first-order coherent field (nonstationary as well as stationary) with arbitrary statistical properties (since only first-order correlation functions appear).

It is observed from Eq. (17) that for non-stationary beams with a first-order coherent field, the intereference term exists but is *not* sinusoidal. The result in Eq. (15) is valid even when there is not maximum fringe contrast (first order coherence). In that case, however, the equality in Eq. (16) no longer holds, and must be replaced by the inequality[3]

$$|G^{(1)}(\chi_1, \chi_2)| < [G^{(1)}(\chi_1, \chi_1) G^{(1)}(\chi_2, \chi_2)]^{1/2}. \quad (18)$$

From this, it is evident that the photomixing term will be reduced below its maximum value when there is a departure from precise first-order coherence of the incident radiation. Therefore in the quantum theory, optimum photomixing with a sinusoidal beat signal arises from the simple conditions of first-order coherence of the total incident field, and monochromatiticy of the constituent fields. The count-rate is seen to be independent of the statistical properties of the radiation in that only first-order correlation functions enter the equations. If, however, knowledge about information other than count rates is desired, e.g. the frequency spectrum or the statistical properties of the beat signal, then knowledge about correlation functions higher than first order would be necessary.[10]

It is, perhaps, worthy of mention that a consideration of the tensor properties of the correlation functions[3] shows that orthogonally polarized beams normally incident on a photodetector will not give rise to a photomixing signal. This has been experimentally confirmed some time ago by Javan, Ballik and Bond.[11]

3. CONCLUSION

If the radiation incident on the detector possesses precise first-order coherence, two interesting consequences follow. The first relates to constraints on the correlation functions,[3] and has provided us with the magnitude of the heterodyne signal. The second concerns the density operator for the radiation field,[4] and allows a physical interpretation for the beating process.

Titulaer and Glauber[12] have generalized the definition of a mode to include nonmonochromatic solutions to the wave equation, and have thereby derived a density operator for the most general type of field possessing first-order coherence. This operator may be obtained by replacing the creation operator $a_k\dagger$ in the single-mode density operator by a more general creation operator $b\dagger$. This latter quantity creates a photon in a particular superposition of modes

which may be considered as specifying a particular type of photon wave packet. Therefore, a field which has first-order coherence may be regarded as consisting of photons of only a single (in general nonmonochromatic) variety. It has also been shown that if a field possesses such a density operator, it is first-order coherent. Since any field expressible in Glauber's P-representation may be separated into a coherent and an incoherent portion, it may be seen that for this type of field only the coherent portion will contain photons of the variety that give rise to a heterodyne signal.

The heterodyne detection process may then be considered as the annihilation of a single photon of this variety. Thus even in the presence of a single one of these photons, a heterodyne signal may still be observed. Dirac's[13] well-known comment, "... Each photon interferes only with itself. Interference between two different photons never occurs," therefore applies to the heterodyne experiment, as it applies in the case of spatial interference experiments.[14] This is not surprising since we are considering a type of interference experiment which is a one-quantum process. For multiple-photon processes, such as two-quantum photodetection[7,15] or the Hanbury Brown–Twiss effect, however, this is not necessarily true.[4]

The uncertainty principle also shows that it is not useful to consider the photons of the constituent beams separately. In fact, in a heterodyne experiment, we are unable to determine from which beam a photon is absorbed in a given time interval. Consider a description in which there are two alternate ways in which the system can evolve from its initial state to the final state: by absorption of a photon from beam 1 or by absorption of a photon from beam 2. In order to ascertain which beam gave rise to the ejection of a particular photoelectron, its energy would have to be measured to within a value ΔE given by $\Delta E < \hbar|\omega_1 - \omega_2|$. From the uncertainty principle,

$$\Delta E \Delta\tau \geqslant \hbar, \tag{19}$$

the time $\Delta\tau$ required for such a measurement would be

$$\Delta\tau \geqslant \frac{\hbar}{\Delta E} = |\omega_1 - \omega_2|^{-1}. \tag{20}$$

The required measurement time is greater than the period of the beat frequency and such a measurement would therefore wash out the time interference. Thus, one cannot ascribe a detected photon to one or the other of the constituent beams. An analogous argument has been applied by Pfleegor and Mandel[14] to independent-beam spatial interference at the single-photon level.

REFERENCES

1. TEICH, M. C., *Proc. IEEE* **56**, 37 (1968).
2. TEICH, M. C., In *Semiconductors and Semimetals* (Edited by R. K. WILLARDSON and A. C. BEER), Vol. 5, *Infrared Detectors*, Academic Press, New York (1970), p. 361. *Appl. Phys. Letters* **14**, 201 (1969).
3. GLAUBER, R. J., *Phys. Rev.* **130**, 2529 (1963).
4. GLAUBER, R. J., In *Quantum Optics and Electronics* (Edited by C. DEWITT, A. BLANDIN, and C. COHEN-TANNOUDJI), p. 65. Gordon and Breach Science Publishers, New York (1965).
5. GLAUBER, R. J., *Phys. Rev.* **131**, 2766 (1963).
6. JACOBS, S. and RABINOWITZ, P., In *Quantum Electronics III* (Edited by P. GRIVET and N. BLOEMBERGEN), p. 481. Columbia University Press, New York (1964).
7. TEICH, M. C. and WOLGA, G. J., *Phys. Rev. Letters* **16**, 625 (1966).
8. TITULAER, U. M. and GLAUBER, R. J., *Phys. Rev.* **140**, 3676 (1965).
9. MANDEL, L. and WOLF, E., *Rev. Mod. Phys.* **37**, 231 (1965).
10. TEICH, M. C., *Proc. IEEE* **57**, 786 (1969).
11. JAVAN, A., BALLIK, E. A. and BOND, W. L., *J. Opt. Soc. Am.* **52**, 96 (1962).
12. TITULAER, U. M. and GLAUBER, R. J., *Phys. Rev.* **145**, 1041 (1966).
13. DIRAC, P. A. M., *Quantum Mechanics*, 4th Ed., Chap. I, p. 9. Oxford University Press, London (1958).
14. PFLEEGOR, R. L. and MANDEL, L., *J. Opt. Soc. Am.* **58**, 946 (1968); *Phys. Rev.* **159**, 1084 (1967); *Phys. Letters* **24A**, 766 (1967).
15. TEICH, M. C. and WOLGA, G. J., *Phys. Rev.* **171**, 809 (1968).