In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

7/25/68

THE EFFECTS OF PARTIALLY COHERENT RADIATION ON THE PERFORMANCE OF AN OPTICAL HETERODYNE RECEIVER

A THESIS

Presented to

The Faculty of the Graduate Division

by

Charles Franklin Morris, Jr.

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in the School of Electrical Engineering

Georgia Institute of Technology

January, 1970

THE EFFECTS OF PARTIALLY COHERENT RADIATION ON THE PERFORMANCE OF AN OPTICAL HETERODYNE RECEIVER

Approved:

Chairman

Date approved by Chairman: January 30 1970

ACKNOWLEDGMENTS

Dr. William B. Jones, Jr. was my thesis advisor for the initial stages of my Doctoral research and provided much helpful guidance. Many worthwhile discussions with Bill Mayo, at the time a Ph.D. student and now graduated, contributed a great deal to the formulation of the problem and the statistical approach presented in Appendix II. When Dr. Jones left Georgia Tech, Dr. Aubrey M. Bush became my advisor. He provided many helpful suggestions and much guidance during the intermediate and final stages of the research. Dr. F. Kenneth Hurd also provided much assistance during this period and he with Dr. Bush helped me in getting all the equipment necessary for the experimental work. Dr. Albert P. Sheppard and Dr. Ray H. Pettit served as members of my reading committee and provided many worthwhile suggestions for the experimental work.

My first three years of graduate school were financed by a NASA fellowship; the last two by a Teaching and Research Assistantship provided by Dr. B. J. Dasher, then Director of the School of Electrical Engineering and continued by Dr. D. T. Paris, the present Director. The experimental facilities were provided by the School of Electrical Engineering. The computer facilities were provided by the Rich Electronic Computer Center. The staff of the computer center, and in particular Mr. S. P. Lenoir, Mr. Jerry Segers, and Mr. Leon Kinard were

always very considerate in assisting me with the problems that arose in this area.

My wife, Elizabeth, spent many long hours typing rough drafts; Betty Sims typed the final version of the thesis.

My wife, Elizabeth, and daughter, Martha, shared with me the life of a graduate student.

To these people, I express my deepest appreciation and most heartfelt thanks.

TABLE OF CONTENTS

APPENDIX Page

LIST OF TABLES

LIST OF ILLUSTRATIONS

Figure Page

Figure Page

SUMMARY

Heterodyne detection of optical signals offers, potentially, several advantages over other methods of detection, including inherent gain and high discrimination against noise. In many cases, to realize these advantages, the optical heterodyning must be done efficiently. This places some rather stringent requirements on the alignment of the optical system and the coherence of the signal and local oscillator fields.

This research is a study of the effects of partial coherence of the signal and local oscillator fields on the performance of an optical heterodyne receiver; the heterodyne receiver efficiency and signal-to-noise ratio are taken as performance indicators. An analytical approach is taken where the heterodyne receiver efficiency and signal-to-noise ratio are defined mathematically for a general heterodyne system. Each is, then, related to the signal and local oscillator field coherence in a single expression. These expressions are applied to three specific heterodyne systems: (1) a communications link through a turbulent atmosphere, (2) a laser heterodyne rough surface velocity measuring system, and (3) a laser heterodyne fluid flow velocity measuring system. In each system, the signal field coherence function is derived based on a statistical description of the system, and the heterodyne efficiency and signal-to-noise ratio are calculated and studied as functions of the receiver variables such as receiver aperture size and shape, local

oscillator field intensity distribution, and local oscillator coherence.

From this study, the conclusions made are that for the heterodyne systems studied, efficient operation and good signal-to-noise ratios cannot be obtained simultaneously; the most important parameters are the size of the local oscillator intensity distribution and the size of the receiver aperture. The shape of the local oscillator intensity distribution and the shape of the receiver aperture have little effect on system performance, nor do slight reductions in local oscillator coherence.

The performance of the surface velocity and fluid flow velocity measuring systems was studied experimentally. The experimentally measured efficiency and signal-to-noise ratio were compared to that predicted theoretically and found to be in close agreement.

CHAPTER I

INTRODUCTION

Heterodyne detection has been used in many communication and measurement systems for some time and its principles and advantages have long been understood. Prior to the availability of a coherent light source, the use of heterodyning was confined to RF and microwave frequencies. With the development of the laser, a coherent source of light, an almost immediate extension of heterodyne principles to optical frequencies was attempted by many in the hope of realizing the advantages heterodyning has in certain applications. It was soon apparent, however, that heterodyning at optical frequencies presented many unexpected and previously unencountered problems. These problems lay primarily in the stringent requirements made on the alignment of the optical system and on the incoming signal field coherence. In some systems, these problems were severe enough to make heterodyning impractical.

These problems originate from the fact that heterodyning at optical frequencies is basically a distributed phenomenon; an incoming signal and a locally produced signal (called a local oscillator) exist as fields that are superimposed over a photoelectric detector surface. The heterodyne signal in the photodetector output current arises from the mutual interference of the signal and local oscillator fields on the photosurface. The interference pattern formed will fluctuate in intensity at a rate equal to the frequency difference in the two fields.

The output current is proportional to the intensity and will likewise fluctuate at the difference frequency.

The total photodetector current is the sum of contributions from each elemental area of the photosurface. Therefore, for a maximum heterodyne signal, these currents and hence the intensity fluctuations at each point on the photosurface must all be exactly in phase. This requires the two interfering fields to have the same relative phase relationship over the photosurface; that is, the wavefronts of both fields must be coincident. This, in turn, requires two things:

1. The signal and local oscillator fields must have the same direction of propagation.

2. The signal and local oscillator fields must be individually spatially coherent over the photosurface.*

The first condition is met by proper alignment of the two fields. The second is met by the use of spatially coherent signal and local oscillator fields. Due to the smallness of an optical wavelength, both conditions are frequently very difficult if not impossible to achieve.

The alignment problems have been extensively studied and by now are very well understood. The coherence problems have received only minor attention, however, and none using the terminology of coherence theory. There are many systems where the signal field coherence is a limiting factor in the performance of the system: in an earthbound optical communications link or radar system where the signal field

---

*Except in a special case where both beams are distorted in the same way, i.e. individually incoherent but coherent with each other.

coherence is degraded by propagating through a turbulent atmosphere; in a measurement system employing light scattering from a random medium as a signal. Because heterodyning can provide improved signal-to-noise ratios in the communication uses and improved sensitivity and selectivity in certain measurement systems, it is important to know the effects of signal field coherence on the heterodyne system operation.

This research is a study of the effects of partial coherence of the signal and local oscillator fields on the performance of an optical heterodyne detector. The heterodyne system efficiency and the output signal-to-noise ratio serve as performance indicators. An analytical approach is taken where the heterodyne efficiency and signal-to-noise ratio are each defined for a general system and related to the signal and local oscillator field coherence in single expressions. These expressions are applied to three specific systems in which coherence is a problem: (1) an optical communications link through a turbulent atmosphere, (2) a laser heterodyne fluid flow measuring system, and (3) a laser heterodyne rough surface measuring system. In each system, the signal field coherence function is derived based on a statistical description of the system and the heterodyne efficiency and signal-to-noise ratio are calculated and studied as functions of receiver variables such as receiver aperture size and shape, local oscillator intensity function, etc.

Some of these results are studied experimentally in the fluid flow and rough surface measuring systems. The principle objective is to verify the applicability of the analytical work.

The organization is as follows: The necessary background material on coherence, interference, and heterodyne detection theory is presented in Chapter II. The derivation of the heterodyne efficiency and signal-to-noise ratio expressions in terms of the signal and local oscillator coherence functions is made in Chapter III. The three systems mentioned above are studied as examples in Chapter IV, the experimental work is presented in Chapter V, and general conclusions as well as some suggestions for future work are presented in Chapter VI.

CHAPTER II

BACKGROUND

Introduction

The interrelation of the coherence of radiation, its interference, and heterodyne detection form the basis of this study. In this chapter, the central concepts of the coherence and interference of radiation and optical heterodyne detection are presented and discussed. The theory of coherence is first discussed, then the theory of interference and its relation to coherence theory. Finally heterodyne detection is discussed in terms of interference and coherence theory. Only the theory necessary to the understanding of this problem and its development is presented. For a more detailed study, the reader is referred to the literature cited.

The Coherence of Optical Fields

The concept of a monochromatic wave is very useful in electromagnetic field theory and optics. A necessary starting point in many problems is the examination of the monochromatic case. For such fields, the amplitude and phase of the wave are deterministic functions of space and time.

Every physical optical field, however, has some fluctuations associated with it. These fluctuations may be small and slow so that the field is essentially constant in amplitude and phase over several

optical periods. In this case, the field looks like a monochromatic field except for small fluctuations about an average amplitude and phase. For this reason, this type of field is called quasi-monochromatic. On the other hand, these fluctuations may be large and rapid. In this case, no average amplitude or phase is discernible. This type of field is called polychromatic.

Optical coherence theory is concerned, in general, with a statistical description of these fields. Most of the work to date has been on second order effects, that is, on the correlation between these fluctuations in the field at two different points. These second order effects are related to the interference phenomenon and, therefore, are of central importance in this work.

Consider a real classical wave function $\bar{V}(\bar{R},t)$ characterizing the field at a point $\bar{R}$ in space at time t. $\bar{V}$ can describe, for example the electric field, the magnetic field, or the vector potential of the field. $\bar{V}(\bar{R},t)$ will be a fluctuating function of time due to random effect in the source, along its path of propagation, or both. It can, therefore, be considered as a member of the ensemble of all possible realizations of the field. Now, as is usually done in electromagnetic field theory, we can associate a complex field $\bar{\mathcal{V}}(\bar{R},t)$ with this field for which $\bar{V} = \mathrm{Re}\{\bar{\mathcal{V}}\}$.* Generally for quasi-monochromatic fields

$$\bar{V}(\bar{R},t) = \bar{A}(\bar{R},t)\cos(\bar{\omega}t + \phi(\bar{R},t))$$

---

*In most approaches, the complex analytic signal is used here for which $\bar{V} = 2\,\mathrm{Re}\{\bar{\mathcal{V}}\}$ but this simpler notation is used here.

and

$$\bar{V}(\bar{R},t) = \bar{A}(\bar{R},t)e^{j\bar{\omega}t},$$

where

$$\bar{A}(\bar{R},t) = \bar{A}(\bar{R},t)e^{j\phi(\bar{R},t)}$$

is the complex envelope of $\bar{V}$ and $\bar{\omega}$ is the average of center radian frequency of the wave. The field, thus, looks like a narrow band random process in $\bar{R}$ and t. No such simple characterization exists for polychromatic fields in general but they may still be represented by the more general form $\bar{V}(\bar{R},t)$.

Next, for simplicity, consider a linearly polarized field so that it may be represented as

$$\bar{V}(\bar{R},t) = \bar{a}_u V(\bar{R},t),$$

where $\bar{a}_u$ is a unit vector in the direction of $\bar{V}$ and $V(\bar{R},t)$ is a scalar quantity. The correlation* in the fluctuations of the field at two space-time points $\bar{R}_1,t_1$ and $\bar{R}_2,t_2$ then is

$$\Gamma_e(\bar{R}_1,\bar{R}_2,t_1,t_2) = E[V(\bar{R}_1,t_1)V^*(\bar{R}_2,t_2)], \qquad (2\text{-}1)$$

---

*The definition varies and depends on what literature source is used.

where E[ ] denotes the expectation or ensemble average over all possible realizations of the field. This function is called the mutual coherence function.[1-3] For the case of quasi-monochromatic fields, the mutual coherence function can be written more simply as the correlation of the complex envelopes

$$\Gamma_e(\bar{R}_1,\bar{R}_2,t_1,t_2) = E[A(\bar{R}_1,t_1)A^*(\bar{R}_2,t_2)].$$

One is often concerned with stationary fields in which case, the ensemble average is independent of time origin. For such fields

$$\Gamma_e(\bar{R}_1,\bar{R}_2,t_1,t_2) = E[V(\bar{R}_1,t_1)V^*(\bar{R}_2,t_1-\tau)] = \Gamma_e(\bar{R}_1,\bar{R}_2,\tau),$$

where $\tau = t_1 - t_2$.

A normalized coherence function can be defined as

$$\gamma_e(\bar{R}_1,\bar{R}_2,\tau) = \frac{\Gamma_e(\bar{R}_1,\bar{R}_2,\tau)}{[\Gamma_e(\bar{R}_1,\bar{R}_1,0)\Gamma_e(\bar{R}_2,\bar{R}_2,0)]^{\frac{1}{2}}}, \qquad (2\text{-}2)$$

for which $0 \le |\gamma_e| \le 1$. This function is called the complex degree of coherence or sometimes just the degree of coherence and provides a measure of the correlation in the field at the two space-time points. Values of γ_e near one indicate good correlation in the field at the two points, values near zero indicate poor correlation. For monochromatic fields, the correlation will be perfect for all points and $|\gamma_e| = 1$ since the amplitude and phase are completely deterministic.

A monochromatic field is, therefore, completely coherent.[3] For most practical fields, however, there will be a region of space-time points $\bar{R}_2,t_2$ about the point $\bar{R}_1,t_1$ for which $|\gamma|$ will be significantly greater than zero. The size of this region depends on the characteristics of the fluctuations and the random effects producing them. But within this region, the fluctuations are correlated to some degree while outside the region, the fluctuations are in general uncorrelated.

For each member of the ensemble of wave functions, a time correlation can be defined as

$$\Gamma_t(\bar{R}_1,\bar{R}_2,\tau) = \langle V(\bar{R}_1,t)V^*(\bar{R}_2,t-\tau)\rangle, \tag{2-3}$$

where the brackets $\langle\ \rangle$ indicate an infinite time average defined as

$$\langle f(t)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)dt.$$

In normalized form, the time correlation is

$$\gamma_t(\bar{R}_1,\bar{R}_2,\tau) = \frac{\Gamma_t(\bar{R}_1,\bar{R}_2,\tau)}{[\Gamma_t(\bar{R}_1,\bar{R}_1,0)\Gamma_t(\bar{R}_2,\bar{R}_2,0)]^{\frac{1}{2}}},$$

where again $0\le|\gamma_t|\le 1$.

They are referred to as the time mutual coherence function and degree of coherence, respectively. If the fields are ergodic then

$$\Gamma_e(\bar{R}_1,\bar{R}_2,\tau) = \Gamma_t(\bar{R}_1,\bar{R}_2,\tau)$$

and

$$\gamma_e(\bar{R}_1,\bar{R}_2,\tau) = \gamma_t(\bar{R}_1,\bar{R}_2,\tau).$$

This is useful in physical situations where it may be easier to measure one average, but more convenient to work analytically with the other.

The results for a linearly polarized wave are easily generalized to non-linearly polarized fields.[2,3] The mutual coherence function and degree of coherence can be defined to represent the correlation between various Cartesian components of the vector fields. For the purposes of this work, however, the scalar definition is suitable since linearly polarized fields will be used for the most part. Also, the electric field will be used throughout rather than the H field or vector potential so the coherence functions will be defined in terms of the E field. That is

$$\Gamma_e(\bar{R}_1,\bar{R}_2,\tau) = E[\bar{E}(\bar{R}_1,t)\bar{E}^*(\bar{R}_2,t-\tau)],$$

etc., where $\bar{E}$ is the complex representation of the real field E.

The Interference of Two Waves

The concept of second order coherence has long been associated with the interference phenomenon, for it is from interference that coherence can best be understood and physically described. The correlation in two fields can best be exhibited in an interference experiment. To see this, let us examine the interference between two waves;

first, consider the ideal case of monochromatic waves to establish some basic results, then, the more practical case of quasi-monochromatic waves.

Interference of Monochromatic Waves

Consider the superposition of two linearly polarized monochromatic waves whose electric fields at some observation point $\bar{R}$ are

$$E_1(\bar{R},t) = \bar{A}_1(\bar{R})\cos(\omega_1 t - \bar{k}_1(\bar{R})\cdot\bar{R})$$

and

$$E_2(\bar{R},t) = \bar{A}_2(\bar{R})\cos(\omega_2 t - k_2(\bar{R})\cdot\bar{R}), \qquad (2\text{-}4)$$

where $\bar{k}_1$ and $\bar{k}_2$ are the wave vectors of magnitudes ω_1/c and ω_2/c, respectively, and in the direction of the propagation of $\bar{E}_1$ and $\bar{E}_2$, respectively. The associated magnetic fields are

$$\bar{H}_1(\bar{R},t) = \frac{\bar{k}_1 \times \bar{E}_1(\bar{R},t)}{\omega_1 \mu}$$

and

$$\bar{H}_2(\bar{R},t) = \frac{\bar{k}_2 \times \bar{E}_2(\bar{R},t)}{\omega_2 \mu} . \qquad (2\text{-}5)$$

The total fields resulting from the superposition of these fields are

$$\bar{E}(\bar{R},t) = \bar{E}_1(\bar{R},t) + \bar{E}_2(\bar{R},t) \tag{2-6}$$

and

$$\bar{H}(\bar{R},t) = \bar{H}_1(\bar{R},t) + \bar{H}_2(\bar{R},t). \tag{2-7}$$

The intensity at the observation point $\bar{R}$ is defined in terms of the average power flow per unit time to be[1]

$$I(\bar{R},t) = \langle \bar{S}(\bar{R},t)_T \rangle \cdot \bar{n}. \tag{2-8}$$

$\bar{S}(\bar{R},t)$ is the Poynting vector at $\bar{R}$ and $\bar{n}$ is a unit vector normal to the observation surface. The brackets $\langle\ \rangle_T$ represent an average over a resolving time T of the observer which is long compared to a period, $2\pi/\omega_1$ or $2\pi/\omega_2$. The exact nature of this average depends on the observer but will, in general, be approximately of the form

$$\langle f(t) \rangle_T \cong \frac{1}{T} \int_t^{t+T} f(t')dt'.$$

The vectors $\bar{k}_1$, $\bar{k}_2$, and $\bar{n}$ are illustrated in Figure 1.

Substituting equations 2-6 and 2-7 into equation 2-8 results in

$$I(\bar{R},t) = \langle (\bar{E}_1+\bar{E}_2) \times (\bar{H}_1+\bar{H}_2) \rangle_T \cdot \bar{n} \tag{2-9}$$

$$= \langle \bar{E}_1 \times \bar{H}_1 + \bar{E}_1 \times \bar{H}_2 + \bar{E}_2 \times \bar{H}_1 + \bar{E}_2 \times \bar{H}_2 \rangle_T \cdot \bar{n}.$$

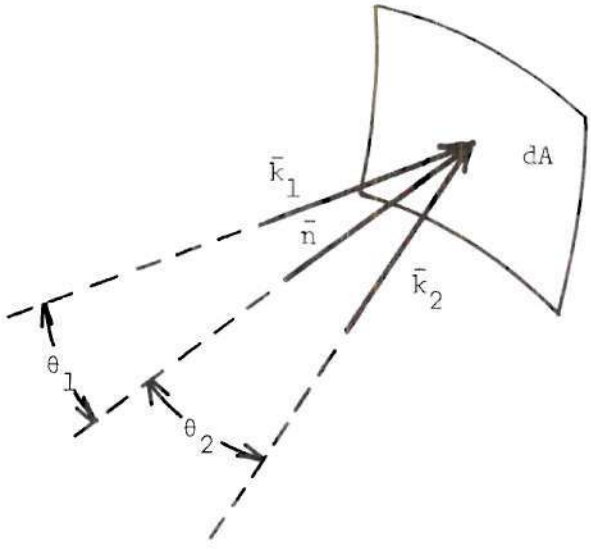

Figure 1. An Illustration of the Vectors $\bar{k}_1$, $\bar{k}_2$, and $\bar{n}$

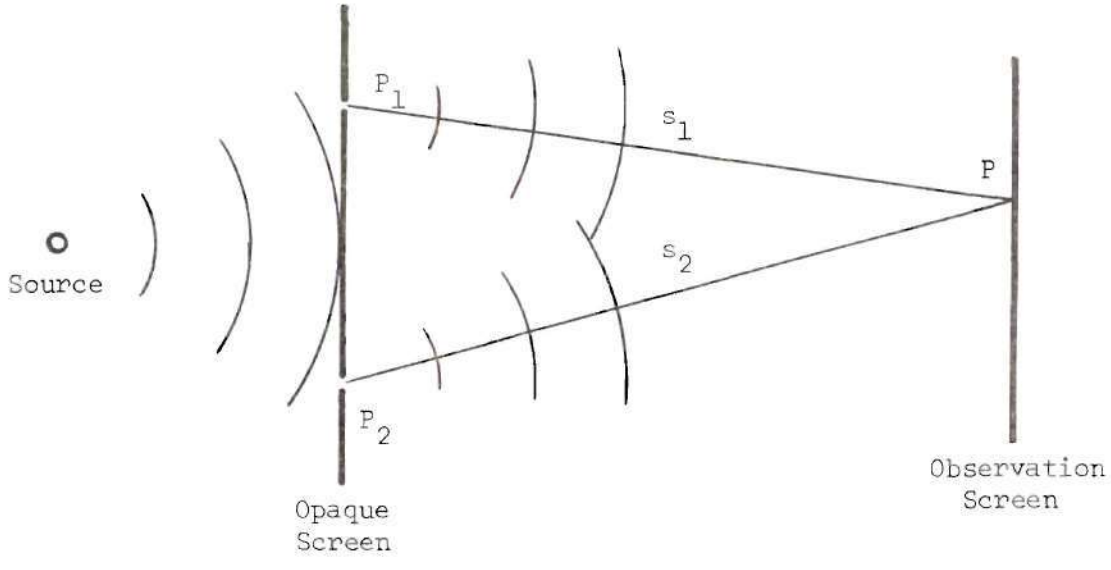

Figure 2. Young's Interference Experiment

Using equations 2-5 for the $\bar{H}$ fields along with the vector identity,[4]

$$\bar{A} \times (\bar{B}\times\bar{C}) = (\bar{A}\cdot\bar{C})\bar{B} - (\bar{A}\cdot\bar{B})\bar{C},$$

we can simplify equation 2-9 to

$$I(\bar{R},t) = \langle \bar{E}_1\cdot\bar{E}_1 + \bar{E}_1\cdot\bar{E}_2 \rangle_T \frac{\bar{k}_1\cdot\bar{n}}{\omega_1\mu} + \left\langle \frac{(\bar{k}_2\cdot\bar{E}_1)(\bar{E}_2\cdot\bar{n})}{\omega_2\mu} \right\rangle_T$$

$$+ \left\langle \frac{(\bar{k}_1\cdot\bar{E}_2)(\bar{E}_1\cdot\bar{n})}{\omega_1\mu} \right\rangle_T + \langle \bar{E}_2\cdot\bar{E}_1 + \bar{E}_2\cdot\bar{E}_2 \rangle_T \frac{\bar{k}_2\cdot\bar{n}}{\omega_2\mu} . \qquad (2\text{-}10)$$

In general, $\bar{k}_1$, $\bar{k}_2$, $\bar{E}_1$, and $\bar{E}_2$ are all functions of $\bar{R}$. For small θ_1 and θ_2 (see Figure 1), however, the quantities

$$k_1\cdot\bar{E}_2 \cong 0, \qquad \bar{E}_2\cdot\bar{n} \cong 0,$$

$$\bar{k}_2\cdot\bar{E}_1 \cong 0, \qquad \bar{E}_1\cdot\bar{n} \cong 0,$$

$$\bar{k}_1\cdot\bar{n} \cong \frac{\omega_1}{c}, \qquad \bar{k}_2\cdot\bar{n} \cong \frac{\omega_2}{c} . \qquad (2\text{-}11)$$

Using equations 2-11 in equation 2-10, we can simplify

$$I(\bar{R},t) \cong \frac{1}{c} \langle \bar{E}_1\cdot\bar{E}_1 + 2\bar{E}_1\cdot\bar{E}_2 + \bar{E}_2\cdot\bar{E}_2 \rangle_T$$

$$= \sqrt{\epsilon/\mu}\, \langle \bar{E}(\bar{R},t)\cdot\bar{E}(\bar{R},t) \rangle_T . \qquad (2\text{-}12)$$

Now, let us look for the moment at the use of a complex representation for $\bar{E}$. That is,

$$\bar{E}_1(\bar{R},t) = \tfrac{1}{2}[\bar{E}_1(\bar{R},t) + \bar{E}_1^*(\bar{R},t)]$$

or

$$\bar{E}_1(\bar{R},t) = \mathrm{Re}\{\bar{E}_1(\bar{R},t)\},$$

where

$$\bar{E}_1(\bar{R},t) = \bar{A}_1(\bar{R})e^{j\omega_1 t}$$

and

$$\bar{A}_1(\bar{R}) = \bar{A}_1(\bar{R})e^{j\bar{k}_1(\bar{R})\cdot\bar{R}},$$

and similarly for $\bar{E}_2$. Using these in equation 2-12, we get

$$I(\bar{R},t) = \sqrt{\epsilon/\mu}\left\langle\left\{\frac{\bar{E}_1+\bar{E}_1^*+\bar{E}_2+\bar{E}_2^*}{2}\right\}\cdot\left\{\frac{\bar{E}_1+\bar{E}_1^*+\bar{E}_2+\bar{E}_2^*}{2}\right\}\right\rangle_T$$

$$= \sqrt{\epsilon/\mu}\left\langle\frac{\bar{E}_1\cdot\bar{E}_1^*}{2} + \frac{\bar{E}_2\cdot\bar{E}_2^*}{2} + \frac{\bar{E}_1\cdot\bar{E}_2^*}{2} + \frac{\bar{E}_1^*\cdot\bar{E}_2}{2}\right.$$

$$+ \frac{\bar{E}_1\cdot\bar{E}_1}{4} + \frac{\bar{E}_1^*\cdot\bar{E}_1^*}{4} + \frac{\bar{E}_2\cdot\bar{E}_2}{4} + \frac{\bar{E}_2^*\cdot\bar{E}_2^*}{4}$$

$$\left.+ \frac{\bar{E}_1\cdot\bar{E}_2}{2} + \frac{\bar{E}_1^*\cdot\bar{E}_2^*}{2}\right\rangle_T.$$

Examination of the first four terms shows that they are of zero, zero, $\omega_1 - \omega_2$, and $-(\omega_1 - \omega_2)$ frequency variation, respectively, while the last six are of $2\omega_1, -2\omega_1, 2\omega_2, -2\omega_2, \omega_1+\omega_2$, and $-(\omega_1 + \omega_2)$ frequency variation. The averaging process will eliminate the last six terms but not affect the first four if $2\pi/|\omega_1-\omega_2| >> T$. This equation can be rewritten as

$$I(\bar{R},t) = \tfrac{1}{2}\sqrt{\epsilon/\mu}\ (\bar{E}_1+\bar{E}_2)\cdot(\bar{E}_1^*+\bar{E}_2^*) = \tfrac{1}{2}\sqrt{\epsilon/\mu}\ \bar{E}(\bar{R},t)\cdot\bar{E}^*(\bar{R},t). \qquad (2\text{-}13)$$

The use of the complex representation for the fields, therefore, eliminates the need to retain the time average if

$$\frac{2\pi}{|\omega_1-\omega_2|} >> T >> \frac{2\pi}{\omega_1} \text{ or } \frac{2\pi}{\omega_2}.$$

The results of equations 2-12 and 2-13 are approximately the same.

Substituting the complex forms of equation 2-4 into equation 2-13, we get

$$I(\bar{R},t) = \tfrac{1}{2}\sqrt{\epsilon/\mu}\ [\bar{E}_1\cdot\bar{E}_1^* + \bar{E}_2\cdot\bar{E}_2^* + 2\mathrm{Re}\{\bar{E}_1\cdot\bar{E}_2^*\}]$$

$$= I_1(\bar{R}) + I_2(\bar{R}) + \sqrt{\epsilon/\mu}\ \mathrm{Re}\{\bar{E}_1(\bar{R},t)\cdot\bar{E}_2^*(\bar{R},t)\}, \qquad (2\text{-}14)$$

where I_1 and I_2 are the intensity at R due to $\bar{E}_1$ and $\bar{E}_2$ alone, respectively. Looking just at the last term in this equation, we can simplify it to

$$\sqrt{\epsilon/\mu}\ \mathrm{Re}\{\bar{E}_1\cdot\bar{E}_2^*\} = \sqrt{\epsilon/\mu}\ \mathrm{Re}\{\bar{A}_1\cdot\bar{A}_2 e^{j[(\omega_1-\omega_2)t+(\bar{k}_2-\bar{k}_1)\cdot\bar{R}]}\}$$

$$= \sqrt{\epsilon/\mu}\ \bar{A}_1\cdot\bar{A}_2\cos[(\omega_1-\omega_2)t + (\bar{k}_2-\bar{k}_1)\cdot\bar{R}].$$

This is a time varying term of radian frequency $\omega_1-\omega_2$. The amplitude depends on $\bar{A}_1\cdot\bar{A}_2$. If $\bar{E}_1$ and $\bar{E}_2$ are mutually orthogonal, then $\bar{A}_1\cdot\bar{A}_2 = 0$. On the other hand, if $\bar{E}_1$ and $\bar{E}_2$ are both polarized in the same direction, then $\bar{A}_1\cdot\bar{A}_2 = A_1A_2$. For this case, we have

$$\sqrt{\epsilon/\mu}\ A_1A_2 = \sqrt{\sqrt{\epsilon/\mu}\ (\bar{E}_1\cdot\bar{E}_1^*)\ \sqrt{\epsilon/\mu}\ (\bar{E}_2\cdot\bar{E}_2^*)} = 2\sqrt{I_1(\bar{R})I_2(\bar{R})}.$$

Two monochromatic waves linearly polarized in the same direction, will therefore interact or interfere to produce a time varying intensity of

$$I(\bar{R},t) = I_1(\bar{R}) + I_2(\bar{R}) + 2\sqrt{I_1(\bar{R})I_2(\bar{R})}\cos[(\omega_1-\omega_2) + (\bar{k}_2-\bar{k}_1)\cdot\bar{R}]$$

at a frequency of $\omega_1-\omega_2$ about an average of I_1+I_2. Since the amplitude and phase of this term are functions of $\bar{R}$ only, the complex amplitude

$$A_1(\bar{R})A_2^*(\bar{R})$$

is a stationary function.

When the interfering waves are of the same frequency ($\omega_1=\omega_2$), the intensity expression becomes

$$I(\bar{R}) = I_1(\bar{R}) + I_2(\bar{R}) + 2\sqrt{I_1(\bar{R})I_2(\bar{R})}\cos[(\bar{k}_2-\bar{k}_1)\cdot\bar{R}] \qquad (2\text{-}15)$$

and is no longer time variable. The intensity is solely a function of $\bar{R}$ and can be any value between

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}$$

and

$$I = I_1 + I_2 - 2\sqrt{I_1 I_2}.$$

The result will be a standing interference pattern whose bright and dark fringe intensities are given by these equations.

These results are directly applicable to Young's interference experiment.[1-3] In this experiment, the radiation from a monochromatic source is incident on an opaque screen with two pinholes (see Figure 2). The interaction of the radiation passing through these pinholes is observed on a screen located some distance from the pinhole.

For a linearly polarized source, the fields from pinholes P_1 and P_2 will be polarized in the same direction. Thus, only the magnitudes are significant and the fields may be considered as scalar quantities. The complex fields at points P_1 and P_2 in Figure 2 can be represented by

$$E(\bar{R}_1,t) \quad \text{and} \quad E(\bar{R}_2,t),$$

respectively. At P on the observation screen, the resulting field is given by

$$\bar{E}(\bar{R},t) = K_1\bar{E}(\bar{R}_1,t-t_1) + K_2\bar{E}(\bar{R}_2,t-t_2),$$

where K_1 and K_2 are complex constants which depend on the pinhole size and the geometry. The times $t_1 = s_1/c$ and $t_2 = s_2/c$ are the propagation times for the radiation in going from P_1 to P and P_2 to P, respectively.

The intensity at P is given by

$$I(\bar{R}) = \tfrac{1}{2}\sqrt{\epsilon/\mu}\ \bar{E}(\bar{R},t)\bar{E}^*(\bar{R},t)$$

$$= \tfrac{1}{2}\sqrt{\epsilon/\mu}\ [K_1K_1^*\bar{E}(\bar{R}_1,t-t_1)\bar{E}^*(\bar{R}_1,t-t_1) + K_2K_2^*\bar{E}(\bar{R}_2,t-t_2)\bar{E}^*(\bar{R}_2,t-t_2)$$

$$+ 2\mathrm{Re}\{K_1K_2^*\bar{E}(\bar{R}_1,t-t_1)\bar{E}^*(\bar{R}_2,t-t_2)\}]$$

$$= I_1(\bar{R}) + I_2(\bar{R}) + \sqrt{\epsilon/\mu}\ \mathrm{Re}\{K_1K_2^*\bar{E}(\bar{R}_1,t-t_1)\bar{E}^*(\bar{R}_2,t-t_2)\}. \qquad (2\text{-}16)$$

Now, for the geometry in Figure 2, $K_1 = K_2 = K$ and

$$\bar{E}(\bar{R}_1,t) = \bar{E}(\bar{R}_2,t) = Ae^{j\omega t},$$

so that

$$\bar{E}(\bar{R}_1,t-t_1) = KAe^{j[\omega t-ks_1]}$$

and

$$E(\bar{R}_2, t-t_2) = KAe^{j[\omega t - ks_2]},$$

where $k = \omega/c$. Putting these into equation 2-16, we get

$$I(\bar{R}) = I_1(\bar{R}) + I_2(\bar{R}) + Re\{\sqrt{\epsilon/\mu}\ |K|^2 A^2 e^{jk(s_2-s_1)}\}$$

$$= I_1(\bar{R}) + I_2(\bar{R}) + 2\sqrt{I_1(\bar{R})I_2(\bar{R})}\cos k(s_2-s_1). \qquad (2\text{-}17)$$

Equations 2-17 and 2-15 correspond exactly with

$$(\bar{k}_2-\bar{k}_1)\cdot\bar{R} = k(s_2-s_1).$$

We can see that depending on s_2-s_1, we will get bright or dark interference fringes.

A quantity related to the fringe contrast can be defined. The fringe visibility, first introduced by Michelson, is defined as

$$\nu = \frac{I_{max} - I_{min}}{I_{max} + I_{min}},$$

where I_{max} and I_{min} are the intensities of adjacent bright and dark fringes, in the vicinity of point P. For the interference pattern given by equation 2-17, ν becomes

$$v = \frac{2\sqrt{I_1(\bar{R})I_2(\bar{R})}}{I_1(\bar{R}) + I_2(\bar{R})} . \qquad (2\text{-}18)$$

As we shall later see, this is the maximum visibility we can expect.

The Interference of Non-monochromatic Radiation

As we have seen, two beams of monochromatic radiation will always interfere as long as they are not mutually orthogonally polarized. When more than one frequency is present in the radiation, the conditions for interference are more severe. These conditions are determined primarily by the spectral width of the radiation, and the correlation between the fluctuations in amplitude and phase of the interfering waves.

The position of the bright and dark fringes in an interference experiment are frequency dependent as can be seen from equation 2-17. Therefore, when the radiation is composed of many frequencies, the bright fringes produced by one frequency component may coincide with the dark fringes produced by another. The visibility of the fringes may, therefore, be seriously reduced.

Consider the superposition of two linearly polarized quasi-monochromatic waves whose complex representations are

$$\bar{E}_1(\bar{R},t) = \bar{A}_1(\bar{R},t)e^{j\bar{\omega}_1 t}$$

and

$$\bar{E}_2(\bar{R},t) = \bar{A}_2(\bar{R},t)e^{j\bar{\omega}_2 t},$$

where $\bar{A}_1$ and $\bar{A}_2$ are the complex envelopes and $\bar{\omega}_1$ and $\bar{\omega}_2$ are the radian center frequencies of $\bar{E}_1$ and $\bar{E}_2$, respectively. The total field is

$$\bar{E}(\bar{R},t) = \bar{E}_1(\bar{R},t) + \bar{E}_2(\bar{R},t).$$

For quasi-monochromatic fields the magnetic field is approximately orthogonally polarized to the electric field so equation 2-13 can be used to compute the intensity if the fields are nearly normal to the observer. It is given by

$$I(\bar{R},t) = \tfrac{1}{2}\sqrt{\epsilon/\mu}\ (\bar{E}_1+\bar{E}_2)\cdot(\bar{E}_1^*+\bar{E}_2^*)$$

$$= \tfrac{1}{2}\sqrt{\epsilon/\mu}\ [\bar{E}_1\cdot\bar{E}_1^* + \bar{E}_2\cdot\bar{E}_2^* + 2\mathrm{Re}\{\bar{E}_1\cdot\bar{E}_2^*\}$$

$$= I_1(\bar{R},t) + I_2(\bar{R},t) + \sqrt{\epsilon/\mu}\ \mathrm{Re}\{\bar{E}_1(\bar{R},t)\cdot\bar{E}_2^*(\bar{R},t)\}, \qquad (2\text{-}19)$$

where I_1 and I_2 are the intensities at point $\bar{R}$ due to $\bar{E}_1$ and $\bar{E}_2$, individually. Looking just at the last term, we can simplify

$$\sqrt{\epsilon/\mu}\ \mathrm{Re}\{\bar{E}_1\cdot\bar{E}_2^*\} = \sqrt{\epsilon/\mu}\ \mathrm{Re}\{e^{j(\bar{\omega}_1-\bar{\omega}_2)t}\bar{A}_1(\bar{R},t)\cdot\bar{A}_2^*(\bar{R},t)\}$$

$$= \sqrt{\epsilon/\mu}\ \bar{A}_1(\bar{R},t)\cdot\bar{A}_2(\bar{R},t)\cos[(\bar{\omega}_1-\bar{\omega}_2)t + (\bar{k}_2(\bar{R},t)-\bar{k}_1(\bar{R},t))\cdot\bar{R}].$$

When $\bar{E}_1$ and $\bar{E}_2$ are polarized in the same direction, this is maximized to

$$\sqrt{\epsilon/\mu}\, A_1(\bar{R},t)A_2(\bar{R},t)\cos[(\bar{\omega}_1-\bar{\omega}_2)t + (\bar{k}_2(\bar{R},t)-\bar{k}_1(\bar{R},t))\cdot\bar{R}]$$

$$= 2\sqrt{I_1(\bar{R},t)I_2(\bar{R},t)}\, \cos[(\bar{\omega}_1-\bar{\omega}_2)t + (\bar{k}_2(\bar{R},t) - \bar{k}_1(\bar{R},t))\cdot\bar{R}].$$

This is a time varying term of radian frequency $\bar{\omega}_1-\bar{\omega}_2$ whose amplitude and phase fluctuates with time.

Unlike the monochromatic case, the complex magnitude of this oscillating term is time varying. It is given by

$$A_1(\bar{R},t)A_2^*(\bar{R},t).$$

In general A_1 and A_2 will both contain some fluctuations in value about an average magnitude and phase. If E_1 and E_2 are generated separately, these fluctuations will not be correlated.

In order for there to be an average stationary pattern in $A_1A_2^*$ over some region, as there was in the monochromatic case, its fluctuations at two points $\bar{R}_1$ and $\bar{R}_2$ within this region must be correlated. This means the fluctuations in A_1 at $\bar{R}_1$ and $\bar{R}_2$ must be correlated to some extent and likewise for A_2 at $\bar{R}_1$ and $\bar{R}_2$. In other words, the two fields must be nearly perfectly coherent within this region. If the fields were not coherent, the intensity fluctuations from point-to-point would not be correlated.

Next let us consider Young's interference experiment using a non-monochromatic source. In this case, the intensity at the observation point P is given by

$$I(\bar{R},t) = I_1(\bar{R},t) + I_2(\bar{R},t) + \sqrt{\epsilon/\mu}\ \mathrm{Re}\{K_1 K_2^{*} E(R_1,t-t_1)E^{*}(\bar{R}_2,t-t_2)\},$$

where

$$\begin{aligned} E(\bar{R}_1,t-t_1) &= A(\bar{R}_1,t-t_1)e^{j\bar{\omega}t} \\ &= A(\bar{R}_1,t-t_1)e^{j[\bar{\omega}t-\bar{k}(\bar{R}_1,t)\cdot(\bar{R}-\bar{R}_1)]} \\ &= A(\bar{R}_1,t-t_1)e^{j[\bar{\omega}t-ks_1-\alpha(\bar{R}_1,t)]} \end{aligned}$$

and

$$\begin{aligned} E(\bar{R}_2,t-t_2) &= A(\bar{R}_2,t-t_2)e^{j\bar{\omega}t} \\ &= A(\bar{R}_2,t-t_2)e^{j[\bar{\omega}t-\bar{k}(\bar{R}_2,t)\cdot(\bar{R}-\bar{R}_2)]} \\ &= A(\bar{R}_2,t-t_2)e^{j[\bar{\omega}t-ks_2-\alpha(\bar{R}_2,t)]}, \end{aligned}$$

where

$$\alpha(\bar{R}_i,t) = \bar{k}(\bar{R}_i,t)\cdot(\bar{R}-\bar{R}_i) - ks_i \qquad \text{for } i=1,2,$$

and k is an average wave vector. The expected value of the intensity is given by

$$E[I(\bar{R},t)] = E[E_1(\bar{R},t)] + E[I_2(\bar{R},t)]$$

$$+ \sqrt{\epsilon/\mu}\ \mathrm{Re}\{K_1 K_2^{*} E[E(\bar{R}_1,t-t_1)E^{*}(\bar{R}_2,t-t_2)]\},$$

which from equation 2-1 is

$$E[I(\bar{R},t)] = E[I_1(\bar{R},t)] + E[I_2(\bar{R},t)]$$

$$+ \sqrt{\epsilon/\mu}\ \mathrm{Re}\{K_1 K_2^{*} \Gamma_e(\bar{R}_1,\bar{R}_2,t-t_1,t-t_2)\}.$$

For the case of a stationary field, this is

$$E[I(\bar{R})] = E[I_1(\bar{R})] + E[I_2(\bar{R})]$$

$$+ \sqrt{\epsilon/\mu}\ \ \mathrm{Re}\{K_1 K_2^{*} \Gamma_e(\bar{R}_1,\bar{R}_2,\tau)\},$$

which using equation 2-2 is

$$E(I(\bar{R})] = E[I_1(\bar{R})] + E[I_2(\bar{R})]$$

$$+ \sqrt{\epsilon/\mu}\ \mathrm{Re}\{K_1 K_2^{*} \sqrt{\Gamma_e(\bar{R}_1,\bar{R}_1,0)\Gamma_e(\bar{R}_2,\bar{R}_2,0)}\ \gamma_e(\bar{R}_1,\bar{R}_2,\tau)\}.$$

Now for the geometry in Figure 2, $K_1 = K_2$ and

$$K_1 K_2^* \sqrt{\Gamma_e(\bar{R}_1,\bar{R}_1,0)\Gamma_e(\bar{R}_2,\bar{R}_2,0)} = \sqrt{|K_1|^2\Gamma_e(\bar{R}_1,\bar{R}_1,0)|K_2|^2\Gamma_e(\bar{R}_2,\bar{R}_2,0)}$$

$$= 2\sqrt{E[I_1(\bar{R})]E[I_2(\bar{R})]} \qquad (2\text{-}20)$$

$$E[I(\bar{R})] = E[I_1(\bar{R})] + E[I_2(\bar{R})] + 2\sqrt{E[I_1(\bar{R})]E[I_2(\bar{R})]}\ \mathrm{Re}\{\gamma_e(\bar{R}_1,\bar{R}_2,\tau)\}. \qquad (2\text{-}21)$$

Using the results of equation 2-20, we can rewrite equation 2-21 as

$$E[I(\bar{R})] = E[I_1(\bar{R})] + E[I_2(\bar{R})]$$

$$+ 2\sqrt{E[I_1(\bar{R})]E[I_2(\bar{R})]}\ |\gamma_e(\bar{R}_1,\bar{R}_2,\tau)|\cos[k(s_2-s_1)-\alpha'(\bar{R}_1,\bar{R}_2,\tau)],$$

where

$$\alpha'(\bar{R}_1,\bar{R}_2,\tau) = \alpha(\bar{R}_2,t-t_2) - \alpha(\bar{R}_1,t-t_1).$$

The maximum and minimum intensities around $\bar{R}$ are

$$I_{max} = E[I_1(\bar{R})] + E[I_2(\bar{R})] + 2\sqrt{E[I_1(\bar{R})]E[I_2(\bar{R})]}\ |\gamma_e(\bar{R}_1,\bar{R}_2,\tau)|$$

and

$$I_{min} = E[I_1(\bar{R})] + E[I_2(\bar{R})] - 2\sqrt{E[I_1(\bar{R})]E[I_2(\bar{R})]}\ |\gamma_e(\bar{R}_1,\bar{R}_2,\tau)|, \qquad (2\text{-}22)$$

so the fringe visibility is

$$v = \frac{2\sqrt{E[I_1(\bar{R})]E[I_2(\bar{R})]}}{E[I_1(\bar{R})] + E[I_2(\bar{R})]} \left|\gamma_e(\bar{R}_1,\bar{R}_2,\tau)\right|.$$

This corresponds to equation 2-18 for the monochromatic case but with the added factor of $\left|\gamma_e(\bar{R}_1,\bar{R}_2,\tau)\right|$. If the field is non-monochromatic, the field variations at $\bar{R}_1$ and $\bar{R}_2$ will not be perfectly correlated thus causing a reduction in fringe contrast or visibility from the monochromatic case. If the field variations are not correlated, then no average interference pattern will result. Young's interference experiment can therefore be used to measure the coherence of a particular field.

Principles of Optical Heterodyning

In this section, a discussion of the theory of optical heterodyning is presented along with the development of the equations describing its operation. In addition, heterodyning will be compared to ordinary optical photodetection and the relative merits of each are discussed.

A basic heterodyne system shown in block diagram in Figure 3 consists of a signal field on which a local oscillator field is superimposed usually by a beam splitter, both being made incident on a photoelectric detector surface. Under appropriate conditions discussed later, the fields interact with the photosurface to produce a signal of intermediate frequency in the photodetector output. This intermediate

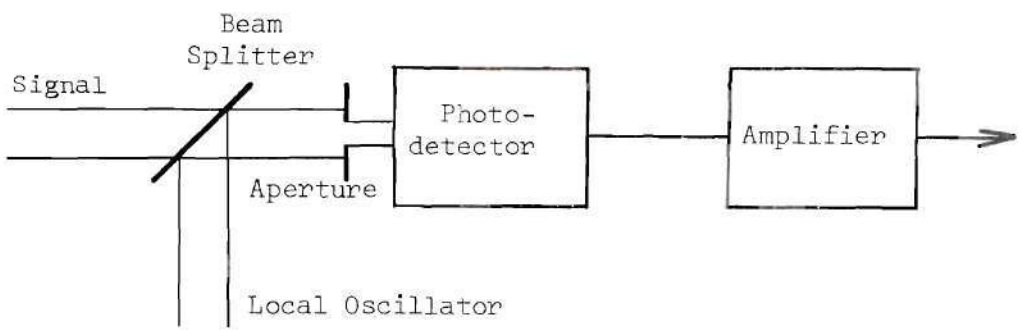

Figure 3. A Basic Optical Heterodyne System

intermediate frequency is equal to the difference in frequency between the local oscillator and signal fields.

There are several different optical configurations used for optical heterodyning. Three of the most common ones are shown in Figure 4. The first employs just a beam splitter and aperture before photodetection, the second, a lens to focus the radiation after superposition but before photodetection, and the third, focussing before superposition and photodetection. The last two have the advantage of being easier to align optically, but all three have been shown to be equivalent.[5,6] For analysis purposes, however, the configuration of Figure 4a is the simplest and will be used here

The incident electric field $\bar{E}$ at a point $\bar{R}$ on the photosurface is given by

$$\bar{E}(\bar{R},t) = \bar{E}_L(\bar{R},t) + \bar{E}_s(\bar{R},t),$$

where $\bar{E}_L(\bar{R},t)$ and $\bar{E}_s(\bar{R},t)$ are the local oscillator field and signal field at $\bar{R}$, respectively. If $\bar{E}_L$ and $\bar{E}_s$ are quasi-monochromatic and near normal to the photosurface, the intensity at $\bar{R}$ as defined by equation 2-13 is

$$I(\bar{R},t) = \tfrac{1}{2}\sqrt{\epsilon/\mu}\ \bar{E}_L(\bar{R},t) \cdot \bar{E}_L^*(\bar{R},t) + \tfrac{1}{2}\sqrt{\epsilon/\mu}\ \bar{E}_s(\bar{R},t) \cdot \bar{E}_s^*(\bar{R},t)$$

$$+ \sqrt{\epsilon/\mu}\ \mathrm{Re}\{\bar{E}_L(\bar{R},t) \cdot \bar{E}_s^*(\bar{R},t)\}$$

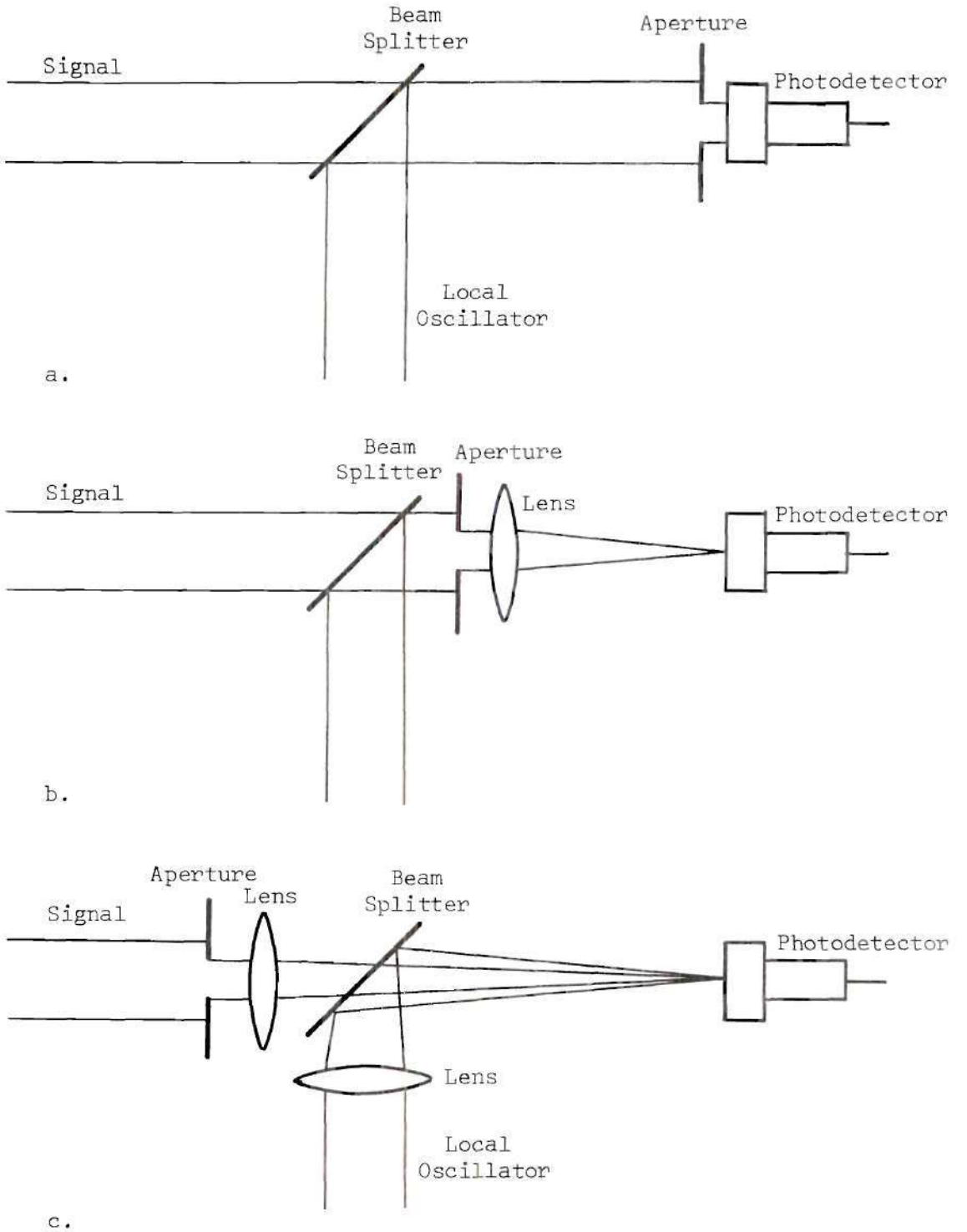

Figure 4. Three Optical Configurations for Optical Heterodyning

$$= I_L(\bar{R},t) + I_s(\bar{R},t) + \sqrt{\epsilon/\mu}\ \mathrm{Re}\{\bar{E}_L(\bar{R},t) \cdot \bar{E}_s(\bar{R},t)\}.$$

For quasi-monochromatic fields

$$\bar{E}_L(\bar{R},t) = \bar{A}_L(\bar{R},t)e^{j\bar{\omega}_L t}$$

and

$$\bar{E}_s(\bar{R},t) = \bar{A}_s(\bar{R},t)e^{j\bar{\omega}_s t},$$

where A_L and A_s, $\bar{\omega}_L$ and $\bar{\omega}_s$ are the complex envelopes and radian center frequencies of the local oscillator and signal fields. Putting equation 24 into 23 gives us

$$I(\bar{R},t) = I_L(\bar{R},t) + I_s(\bar{R},t) + \sqrt{\epsilon/\mu}\ \mathrm{Re}\{e^{j\omega_D t}\bar{A}_L(\bar{R},t) \cdot \bar{A}_s^*(\bar{R},t)\}, \quad (2\text{-}23)$$

where $\bar{\omega}_D = \bar{\omega}_L - \bar{\omega}_s$. This corresponds to equation 2-19 derived in the last section describing the interference between $\bar{E}_L$ and $\bar{E}_s$. The first two terms, as we have seen, are the constant components of the total intensity due to the local oscillator and signal fields individually, while the last is the fluctuating component varying at frequency $\bar{\omega}_D = \bar{\omega}_L - \bar{\omega}_s$, the difference frequency due to interference.

The photodetector, although basically a quantum device providing a current output proportional to the incident photon flux, can be considered from a classical viewpoint. The output current has been shown

to be proportional to the average intensity of the fields incident on its photosensitive surface.[7] That is

$$i(t) = \frac{qe}{hf} \int_S I(\bar{R},t)dA, \qquad (2\text{-}24)$$

where $I(\bar{R},t) = \sqrt{\epsilon/\mu}\ \langle \bar{E}(\bar{R},t)\cdot\bar{E}(\bar{R},t)\rangle_T$ is the intensity at $\bar{R}$ on the photosurface S, and T is the response time of the photodetector (usually less than 10^{-8} sec.). For analysis purposes, S is taken to be that portion of the photosurface that is exposed to incident radiation as limited by the receiving aperture. The constant q is the quantum efficiency of the photosurface (electrons/incident proton), e is the magnitude of the electronic charge, h is Planck's constant and f is the frequency of the field. The average over T eliminates any optical frequency variations in i(t). Now we have

$$i(t) = \frac{qe}{hf}\sqrt{\epsilon/\mu} \int_S \langle \bar{E}(\bar{R},t)\cdot\bar{E}(\bar{R},t)\rangle_T dA$$

$$\simeq \frac{qe}{2hf}\sqrt{\epsilon/\mu} \int_S \bar{E}(\bar{R},t)\cdot\bar{E}^*(\bar{R},t)dA.$$

Putting equation 2-23 into 2-24, we get

$$i(t) = \frac{qe}{2hf}\sqrt{\epsilon/\mu} \int_S \bar{E}_L(\bar{R},t)\cdot\bar{E}_L^*(\bar{R},t)dA$$

$$+ \frac{qe}{2hf}\sqrt{\epsilon/\mu} \int_S \bar{E}_s(\bar{R},t)\cdot\bar{E}_s^*(\bar{R},t)dA$$

$$+ \frac{qe}{hf} \sqrt{\epsilon/\mu}\ \mathrm{Re}\{e^{j\bar{\omega}_D t} \int_S \bar{A}_L(\bar{R},t)\cdot\bar{A}_S^*(\bar{R},t)dA\}$$

$$= i_{DC_L} + i_{DC_S} + i_s(t). \qquad (2\text{-}25)$$

The first two terms are DC or low frequency components of the photocurrent due to the field intensities of the signal and local oscillator individually. The last term is the heterodyne signal current of the difference frequency $\bar{\omega}_D$. Note that any amplitude or phase modulation in the signal field will result in proportional changes in $\bar{A}_L \cdot \bar{A}_S^*$ and hence in i_s. The heterodyne detection system is, therefore best suited for processing of signals containing amplitude, phase or frequency modulation. The amplitude, phase, or frequency modulation of the signal is preserved in the conversion of the signal down to the lower difference frequency.

The integral in the last term of equation 2-25 gives the complex amplitude of the heterodyne signal. The integrand gives the complex amplitude of fluctuating intensity of the interfering waves. We see that for interference to take place, the fields must overlap ($\bar{E}_L \neq 0$ and $\bar{E}_S \neq 0$ at all $\bar{R}$ over S) and not be mutually orthogonally polarized. For maximum interference effect, the signal and local oscillator beams must be coaxial[8] and if they are linearly polarized, they should be polarized in the same direction.

For a maximum value of the integral and hence a maximum heterodyne signal, the integrand must, in addition to satisfying the above

requirements, be of constant phase for all $\bar{R}$ on S. This requires a constant phase difference between $A_L(\bar{R},t)$ and $A_s(\bar{R},t)$ for all $\bar{R}$ on S, a condition satisfied only when the fluctuations of $A_L A_s^*$ are correlated over S and when the wavefronts of the signal and local oscillator fields have identical shape and are perfectly aligned. This requires the fields to be nearly coherent over S and to have the same direction of propagation.

This last requirement is equivalent to that of having the intensity fluctuations produced by the interfering wave to all be in phase. If the wavefronts were distorted or tilted with respect to each other, the resulting intensity fluctuations at different points on S will not be in phase and in some cases not be correlated and, hence, may cancel, a condition that clearly will not yield a maximum heterodyne signal.

In summary, therefore, there are four requirements for efficient heterodyning. The signal and local oscillator fields must

(1) be identically polarized,

(2) overlap on the photosurface,

(3) have the same direction of propagation sometimes referred to as angular alignment, and

(4) be nearly perfectly coherent, i.e. approximately monochromatic.

These four requirements are not easily satisfied and are sometimes so difficult as to make heterodyne detection of doubtful advantage.

As a standard for comparing all heterodyne systems, let us consider the simple ideal case of uniform plane, linearly polarized,

monochromatic signal and local oscillator fields superimposed and normally incident on the photosurface so that all four requirements are ideally satisfied. The incident field is given by

$$E(\bar{R},t) = A_L e^{j\omega_L t} + A_s e^{j\omega_s t},$$

where A_L and A_s are the amplitudes of the local oscillator and signal carrier fields and are real (taking the photosurface as a phase reference). The resulting intensity on S is

$$I(\bar{R},t) = I_L + I_s + \sqrt{\epsilon/\mu}\, A_L A_s \cos\omega_D t,$$

where $\omega_D = \omega_L - \omega_s$ and $I_L = A_L^2/2$ and $I_s = A_s^2/2$. The intensity fluctuations at all points on S are in phase. The photocurrent is

$$i(t) = \frac{qe}{hf} I_L A + \frac{qe}{hf} I_s A + \frac{qe}{hf}\sqrt{\epsilon/\mu}\, A_L A_s A \cos\omega_D t = i_{DC_L} + i_{DC_s} + i_s(t),$$

where A is the photosurface area. The carrier signal power delivered to a unit resistance after amplification is

$$S_h = \langle G^2 i_s^2(t)\rangle = \frac{q^2 e^2 \epsilon}{2h^2 f^2 \mu} G^2 A_L^2 A_s^2 A^2 = 2G^2 \frac{q^2 e^2}{h^2 f^2} I_L I_s A^2, \qquad (2\text{-}26)$$

where G is the gain of the following amplifier. The RMS shot noise current is given after amplification by[14]

$$i_{n_{rms}} = \langle G^2 i_n^2(t) \rangle^{\frac{1}{2}} = G\sqrt{2eB(i_{DC_L} + i_{DC_s})} ,$$

where B is the equivalent rectangular bandwidth of the amplifier. The shot noise power is therefore given by

$$N_h = 2G^2 eB(i_{DC_L} + i_{DC_s}) = \frac{2qe^2BG^2}{hf}(I_L + I_s)A. \qquad (2\text{-}27)$$

The carrier signal-to-shot noise ratio (SNR), then, is

$$SNR_h = \frac{S_h}{N_h} = \frac{qe}{hfB}\frac{I_L I_s A}{I_L + I_s} .$$

As the local oscillator intensity is increased

$$N_h \rightarrow \frac{2qe^2BG^2}{hf} I_L A$$

and

$$SNR_h \rightarrow \frac{qe}{hfB} I_s A. \qquad (2\text{-}28)$$

This results in a signal-to-noise ratio dependent on the signal power only, not on local oscillator power. I_L can therefore be increased to the point that the shot noise produced by it completely predominates over any other source of noise generated in the detection process. The signal-to-noise ratio is unaffected because the signal power increases at the same rate with I_L as the noise.

To examine the performance of a heterodyne system in comparison to other detection systems, let us consider a direct detection such as the one shown in Figure 5. In this system, the signal field $\bar{E}_s(\bar{R},t)$ is incident on a photodetector. The photocurrent is given by

$$i_s(t) = \frac{qe}{2hf}\sqrt{\epsilon/\mu}\int_S \bar{E}_s(\bar{R},t)\cdot\bar{E}_s^*(\bar{R},t)dA.$$

Any intensity modulation of the signal field will, therefore, result in proportional changes in i_s. The direct detection system is, thus, best suited for detection of intensity modulated signals. Since there is no local oscillator, there are no alignment requirements in a direct detection system.

Examine the ideal case of a uniform plane, linearly polarized signal carrier normally incident on the photosurface. The incident field is given by

$$E_s(\bar{R},t) = A_s e^{j\omega_s t}$$

where A_s is amplitude of the signal field and is real. The signal output is

$$i_s = \frac{qe}{2hf}\sqrt{\epsilon/\mu}\, A_s^2 A = \frac{qe}{hf} I_s A .$$

The carrier signal power, thus, is given by

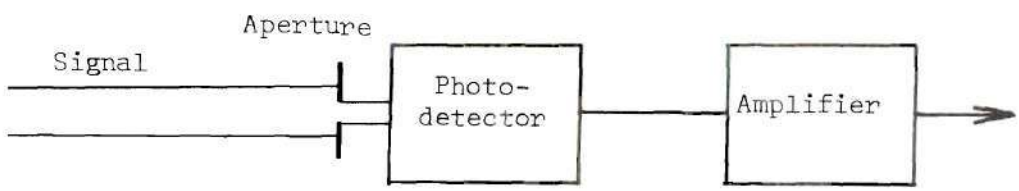

Figure 5. A Basic Direct Detection System

$$S_d = \langle G^2 i_s^2(t) \rangle = \frac{q^2 e^2 G^2}{h^2 f^2} I_s^2 A^2. \tag{2-29}$$

The signal shot noise power is given by

$$N_d = 2G^2 eBi_{DC} = \frac{2qe^2 BG^2}{hf} I_s A, \tag{2-30}$$

where again B is the bandwidth of the photodetection process or the signal amplifier that follows. The ideal signal-to-noise ratio is

$$SNR_d = \frac{S_d}{N_d} = \frac{qe}{2hfB} I_s A. \tag{2-31}$$

Comparing this with equation 2-28 we note a twofold improvement in signal-to-noise ratio for the heterodyne case, an advantage for heterodyning but not a distinct one. In comparing equation 2-29 to 2-26 we have

$$\frac{S_h}{S_d} = 2 \frac{I_L}{I_s},$$

a conversion gain in the heterodyne process not exhibited by direct detection that increases with I_L. The shot noise also increases with I_L, however, since from equations 2-27 and 2-30, we have

$$\frac{N_h}{N_d} = \frac{I_L + I_s}{I_s} \rightarrow \frac{I_L}{I_s} \quad \text{for} \quad I_L >> I_s,$$

but only at half the rate of the signal power, hence the twofold improvement in signal-to-noise ratio.

The real advantage of heterodyning is more obvious, however, when one considers the effects of other sources of noise. In practice, the noise for the direct detection case is usually much greater than the signal shot noise due to noise created by background light, photodetector dark current, and most important, thermal effects. The signal-to-noise ratio, in this case, will be much less than that given in equation 2-31. Such reductions even for ideal signals are typical in many direct detection systems and can limit the usefulness for certain applications.

In a heterodyne system, I_L can be made large enough for the local oscillator shot noise to predominate over these other noise sources. Because of the conversion gain in signal that accompanies a large I_L, the signal-to-noise ratio will still be given by equation 2-28 and, therefore, will not be affected by these other noise sources.

Heterodyne Efficiency

Heterodyne detection, as we have just seen, is generally more advantageous than direct detection provided it is efficient, that is, when its performance is comparable to that theoretically possible under ideal conditions. This is possible only with the exact fulfillment of the four requirements. Should any or all of these requirements not be exactly fulfilled, a reduction in efficiency and a subsequent reduction in signal-to-noise ratio from the ideal case will result. This reduction is customarily measured by a parameter η_h usually called the heterodyne efficiency. It is usually given by

$$\eta_h = \frac{\langle i_s^2(t)\rangle_{actual}}{\langle i_s^2(t)\rangle_{ideal}}, \qquad (2\text{-}32)$$

that is, the ratio of the signal power obtainable under actual operating conditions to the signal power obtainable under ideal conditions. Values near one indicate high efficiency heterodyning (near ideal operation); values near zero indicate low efficiency heterodyning.

The obtainable SNR ratio is given in terms of the ideal SNR by

$$SNR_{actual} = \eta_h SNR_{ideal}.$$

Therefore, for realization of near ideal signal-to-noise ratios, the heterodyne detector must operate efficiently.

The effects of polarization, overlap, and propagation direction misalignments on heterodyne operation have been studied extensively.[8,12,13] The general conclusions are that these alignment requirements are very strict and sometimes cannot be met. The effects of the reductions in signal or local oscillator field coherence, however, have received little attention. This is the performance limiting factor in many systems.

CHAPTER III

THE DERIVATION OF THE HETERODYNE EFFICIENCY AND SNR EQUATIONS

The equations describing ideal operation of a heterodyne detection system are given in the preceding chapter. In this chapter, the equations for the efficiency and signal-to-noise ratio are derived for a non-ideal system where the amplitude and phase variations of the signal and local oscillator fields are not perfectly correlated. This will be shown to be mathematically related to the degree of coherence of each field.

Consider a heterodyne system where the signal and local oscillator fields are polarized linearly in the same direction. The signal current is given from equation 2-25 by

$$i_s(t) = \frac{qe}{hf}\sqrt{\epsilon/\mu}\ \mathrm{Re}\{\int_S E_L(\bar{R},t)E_s^*(\bar{R},t)dA\},$$

where

$$E_L(\bar{R},t) = A_L(\bar{R},t)e^{j\omega_L t}$$

and

$$E_s(\bar{R},t) = A_s(\bar{R},t)e^{j\omega_s t}, \qquad (3\text{-}1)$$

so that the average power delivered by the photodetector to a unit resistance is given by

$$\langle i_s^2(t)\rangle = \left\langle \left[\frac{qe}{hf} \sqrt{\epsilon/\mu}\, \mathrm{Re}\{ \int_S E_L(\bar{R},t) E_s^*(\bar{R},t) dA \} \right]^2 \right\rangle .$$

The squared integral can be written as an iterated integral;

$$\begin{aligned} i_s^2(t) &= \left\langle \frac{q^2 e^2 \epsilon}{h^2 f^2 \mu} \mathrm{Re}\{ \int_{S_1} E_L(\bar{R}_1,t) E_s^*(\bar{R}_1,t) dA_1 \} \right. \\ &\quad \left. \cdot\, \mathrm{Re}\{ \int_{S_2} E_L(\bar{R}_2,t) E_s^*(\bar{R}_2,t) dA_2 \} \right\rangle \\ &= \left\langle \frac{q^2 e^2 \epsilon}{h^2 f^2 \mu} \int_{S_1} \mathrm{Re}\{ E_L(\bar{R}_1,t) E_s^*(\bar{R}_1,t) \} dA_1 \right. \\ &\quad \left. \cdot \int_{S_2} \mathrm{Re}\{ E_L(\bar{R}_2,t) E_s^*(\bar{R}_2,t) \} dA_2 \right\rangle \\ &= \left\langle \frac{q^2 e^2 \epsilon}{h^2 f^2 \mu} \int_{S_1} \int_{S_2} \mathrm{Re}\{ E_L(\bar{R}_1,t) E_s^*(\bar{R}_1,t) \} \right. \\ &\quad \left. \cdot\, \mathrm{Re}\{ E_L(\bar{R}_2,t) E_s^*(\bar{R}_2,t) \} dA_1 dA_2 \right\rangle, \end{aligned}$$

where $S_1 = S_2 = S$.

The time average is given by the integral

$$\langle i_s^2(t)\rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} i_s^2(t) dt.$$

Only the integrand is a function of time so the order of integration may be interchanged to give

$$i_s^2(t) = \frac{q^2 e^2 \epsilon}{h^2 f^2 \mu} \int_{S_1} \int_{S_2} \langle \mathrm{Re}\{E_L(\bar{R}_1,t)E_s^*(\bar{R}_1,t)\}$$

$$\cdot \mathrm{Re}\{E_L(\bar{R}_2,t)E_s^*(\bar{R}_2,t)\}\rangle dA_1 dA_2.$$

Let us consider just the integrand for the moment.

$$\langle \mathrm{Re}\{E_L(\bar{R}_1,t)E_s^*(\bar{R}_1,t)\}\mathrm{Re}\{E_L(\bar{R}_2,t)E_s^*(\bar{R}_2,t)\}\rangle$$

$$= \langle \tfrac{1}{2}(E_L(\bar{R}_1,t)E_s^*(\bar{R}_1,t) + E_L^*(\bar{R}_1,t)E_s(\bar{R}_1,t))$$

$$\cdot \tfrac{1}{2}(E_L(\bar{R}_2,t)E_s^*(R_2,t) + E_L^*(\bar{R}_2,t)E_s(\bar{R}_2,t))\rangle$$

$$= \tfrac{1}{4}[\langle E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)\rangle$$

$$+ \langle E_L^*(\bar{R}_1,t)E_L(\bar{R}_2,t)E_s(\bar{R}_1,t)E_s^*(\bar{R}_2,t)\rangle$$

$$+ \langle E_L(\bar{R}_1,t)E_L(\bar{R}_2,t)E_s^*(\bar{R}_1,t)E_s^*(\bar{R}_2,t)\rangle$$

$$+ \langle E_L^*(\bar{R}_1,t)E_L^*(\bar{R}_2,t)E_s(\bar{R}_1,t)E_s(\bar{R}_2,t)\rangle].$$

By substituting equations 3-1 into this, we see that before the time average is performed, the first two terms are of low frequency variation while the last two are of $2\bar{\omega}_D$ variation. The time average, therefore, eliminates the last two terms and

$$\langle \mathrm{Re}\{E_L(\bar{R}_1,t)E_s^*(\bar{R}_1,t)\}\mathrm{Re}\{E_L(\bar{R}_2,t)E_s(\bar{R}_2,t)\}\rangle$$

$$= \tfrac{1}{4}[\langle E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)\rangle$$

$$+ \langle E_L^*(\bar{R}_1,t)E_L(\bar{R}_2,t)E_s(\bar{R}_1,t)E_s^*(\bar{R}_2,t)\rangle]$$

$$= \tfrac{1}{2}\,\mathrm{Re}\{\langle E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)\rangle\}.$$

In many practical systems, the field processes can be assumed to be ergodic.[2] Making this assumption here, we have

$$\tfrac{1}{2}\,\mathrm{Re}\{\langle E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)\rangle\}$$

$$= \tfrac{1}{2}\,\mathrm{Re}\{E[E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)]\}.$$

In addition, if the signal and local oscillator fields are statistically independent, as they would be if generated separately, this equation becomes

$$\tfrac{1}{2}\,\mathrm{Re}\{E[E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)]E[E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)]\}$$

$$= \tfrac{1}{2}\,\mathrm{Re}\{\langle E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)\rangle\langle E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)\rangle\}.$$

Comparing the time average in this last term to those of equation 2-3, we see that

$$\langle E_L(\bar{R}_1,t)E_L^*(\bar{R}_2,t)\rangle = \Gamma_{L_t}(\bar{R}_1,\bar{R}_2,0) = \Gamma_{L_e}(\bar{R}_1,\bar{R}_2,0)$$

$$= \Gamma_L(\bar{R}_1,\bar{R}_2)$$

and

$$\langle E_s^*(\bar{R}_1,t)E_s(\bar{R}_2,t)\rangle = \Gamma_{s_t}^*(\bar{R}_1,\bar{R}_2,0) = \Gamma_{s_e}^*(\bar{R}_1,\bar{R}_2,0)$$

$$= \Gamma_s^*(\bar{R}_1,\bar{R}_2),$$

the mutual coherence functions for the local oscillator and signal field. The t and e subscripts are dropped since the averages are equivalent under the ergodic assumptions and the τ variable is dropped since it is always zero. The expression for the average signal power, therefore, is

$$\langle i_s^2(t)\rangle = \frac{q^2e^2\epsilon}{2h^2f^2\mu}\int_{S_1}\int_{S_2} \mathrm{Re}\{\Gamma_L(\bar{R}_1,\bar{R}_2)\Gamma_s^*(\bar{R}_1,\bar{R}_2)\}dA_1dA_2.$$

The mutual coherence functions, in terms of the degree of coherence functions are given from equations 2-2 and 2-13 by

$$\Gamma_L(\bar{R}_1,\bar{R}_2) = [\Gamma_L(\bar{R}_1,\bar{R}_1)\Gamma_L(\bar{R}_2,\bar{R}_2)]^{\frac{1}{2}}\gamma_L(\bar{R}_1,\bar{R}_2)$$

$$= 2\sqrt{\mu/\epsilon}\,[I_L(\bar{R}_1)I_L(\bar{R}_2)]^{\frac{1}{2}}\gamma_L(\bar{R}_1,\bar{R}_2)$$

and

$$\Gamma_s(\bar{R}_1,\bar{R}_2) = [\Gamma_s(\bar{R}_1,\bar{R}_1)\Gamma_s(\bar{R}_2,\bar{R}_2)]^{\frac{1}{2}}\gamma_s(\bar{R}_1,\bar{R}_2)$$

$$= 2\sqrt{\mu/\epsilon}\,[I_s(\bar{R}_1)I_s(\bar{R}_2)]^{\frac{1}{2}}\gamma_s(\bar{R}_1,\bar{R}_2),$$

so

$$\langle i_s(t)\rangle = \frac{2q^2e^2}{h^2f^2}\int_{S_1}\int_{S_2}[I_L(\bar{R}_1)I_L(\bar{R}_2)I_s(\bar{R}_1)I_s(\bar{R}_2)]^{\frac{1}{2}}$$

$$\cdot\,\mathrm{Re}\{\gamma_L(\bar{R}_1,\bar{R}_2)\gamma_s^*(\bar{R}_1,\bar{R}_2)\}dA_1dA_2. \qquad (3\text{-}2)$$

The signal power is thus related to the respective degree of coherence of the signal and local oscillator fields through the above integral.

In the ideal case, the signal and local oscillator must be perfectly coherent. In addition, if the signal and local oscillator are perfectly aligned and normal to the photosurface.

$$\gamma_L(\bar{R}_1,\bar{R}_2) = \gamma_s(\bar{R}_1,\bar{R}_2) = 1 \quad \text{for } \bar{R}_1,\bar{R}_2 \text{ on S.}$$

This is because E_L and E_s are perfectly monochromatic and $\bar{R}_1$ and $\bar{R}_2$ lies along an equiphase front. The signal power for this case is

$$\langle i_s(t)\rangle_{\text{ideal}} = \frac{2q^2e^2}{h^2f^2}\int_{S_1}\int_{S_2}[I_L(\bar{R}_1)I_L(\bar{R}_2)I_s(\bar{R}_1)I_s(\bar{R}_2)]^{\frac{1}{2}}dA_1dA_2.$$

The heterodyne efficiency, therefore, is given from equation 2-32 by

$$\eta_h = \frac{\langle i_s^2(t)\rangle_{actual}}{\langle i_s^2(t)\rangle_{ideal}}$$

$$= \frac{\int_{S_1}\int_{S_2} [I_L(\bar{R}_1)I_L(\bar{R}_2)I_s(\bar{R}_1)I_s(\bar{R}_2)]^{\frac{1}{2}} \mathrm{Re}\{\gamma_L(\bar{R}_1,\bar{R}_2)\gamma_s^*(\bar{R}_1,\bar{R}_2)\}dA_1dA_2}{\int_{S_1}\int_{S_2} [I_L(\bar{R}_1)I_L(\bar{R}_2)I_s(\bar{R}_1)I_s(\bar{R}_2)]^{\frac{1}{2}} dA_1dA_2} . \quad (3\text{-}3)$$

The numerator and denominator are identical except for the coherence functions whose maximum values are unity. The heterodyne efficiency for the perfect coherence case is, as might be expected, equal to one. Since

$$[I_L(\bar{R}_1)I_L(\bar{R}_2)I_s(\bar{R}_1)I_s(\bar{R}_2)]^{\frac{1}{2}}$$

is always positive and $0 \le |\gamma_L| \le 1$ and $0 \le |\gamma_s| \le 1$, reduction in either the signal field coherence or the local oscillator coherence result in a decrease in efficiency.

The SNR for this case assuming local oscillator shot noise limited operation is given by

$$SNR = \frac{\langle i_s^2(t)\rangle}{\langle i_n^2(t)\rangle} = \frac{qe}{hfB}$$

$$\cdot \frac{\int_{S_1}\int_{S_2} [I_L(\bar{R}_1)I_L(\bar{R}_2)I_s(\bar{R}_1)I_s(\bar{R}_2)]^{\frac{1}{2}} \mathrm{Re}\{\gamma_L(\bar{R}_1,\bar{R}_2)\gamma_s^*(\bar{R}_1,\bar{R}_2)\}dA_1dA_2}{\int_S I_L(\bar{R})dA} .$$

(3-4)

Both the heterodyne efficiency and signal-to-noise ratio of a heterodyne system can, thus, be determined as functions of the signal and local oscillator field coherence functions, γ_L and γ_S, at the photosurface, the intensity distribution, I_L and I_S, of the local oscillator and signal field, and the photosurface size and shape S. We see that with everything else fixed, a reduction in the coherence of either the signal or local oscillator will cause a reduction in η_h and SNR.

In summary, the efficiency and SNR of any heterodyne detection system are given by equation 3-3 and 3-4. The only conditions placed on the system are:

1. The system is ergodic in the mean and correlation.

2. The local oscillator and signal are statistically independent.

These are not severe restrictions and apply to many practical and useful heterodyne systems. Although the effects of imperfect overlap and angular alignment are not included in this study, their effects can be studied from these equations. These equations can, therefore, be very useful in a thorough analysis of most heterodyne systems regardless of their structure.

CHAPTER IV

THE ANALYTICAL STUDY

Thus far, only the most general type of heterodyne system has been considered. In this chapter, the general equations that have been developed are applied to three specific systems to evaluate their performance as heterodyne systems. The systems are (1) a communications link through a turbulent atmosphere, (2) a laser Doppler surface motion measuring system, and (3) a laser Doppler fluid flow measuring system. The basic approach, for each system, is to (1) calculate the signal field coherence function based on a statistical description of the system, and (2) calculate the heterodyne efficiency and signal-to-noise ratio from equation 3-3 and 3-4 for different receiving apertures, different local oscillator intensity functions, and different degrees of local oscillator coherence. The main objective is to determine conditions at the receiver for efficient operation and maximum SNR. The effects of changing signal parameters are not considered here.

An Optical Communications Link

The development of the laser has established a great new interest in coherent light as a carrier for information in communications systems. Not only does this provide wider bandwidth and greater information capacity in a highly colimated beam, but also the possibility of the use of coherent detection techniques and the resulting improved

noise immunity. Ironically, for earthbound communication links, the coherence of such a signal can be seriously reduced by random atmospheric effects, limiting the usefulness of this type of optical communication system.

The random atmospheric effects are primarily of the form of random variations in the index of refraction from point-to-point caused by atmospheric turbulence. These variations produce, in turn, random fluctuations in amplitude and phase in an optical signal propagating through the atmosphere. This is illustrated in Figure 6. The subsequent reduction in signal coherence can severely reduce the effectiveness of the optical heterodyne receiver as a coherent detector.

The performance of an optical heterodyne communications system as it is affected by atmospheric turbulence has been investigated by many authors in different ways. Gardner studied power loss in heterodyne signal due to turbulence compared to no turbulence for various size circular receiver apertures.[6] Fried studied the signal-to-noise ratio as a function of the turbulence intensity and receiver aperture diameter.[5] Rosner investigates the IF signal power for various receiver aperture diameters.[15] In this thesis, the heterodyne system will be re-examined in terms of heterodyne efficiency and signal-to-noise ratio as defined in equations 3-3 and 3-4.

Derivation of the Coherence Function

In deriving the degree of coherence function for the signal field, the second order statistics of the amplitude and phase fluctuations must be known. These depend on the statistics of the refractive

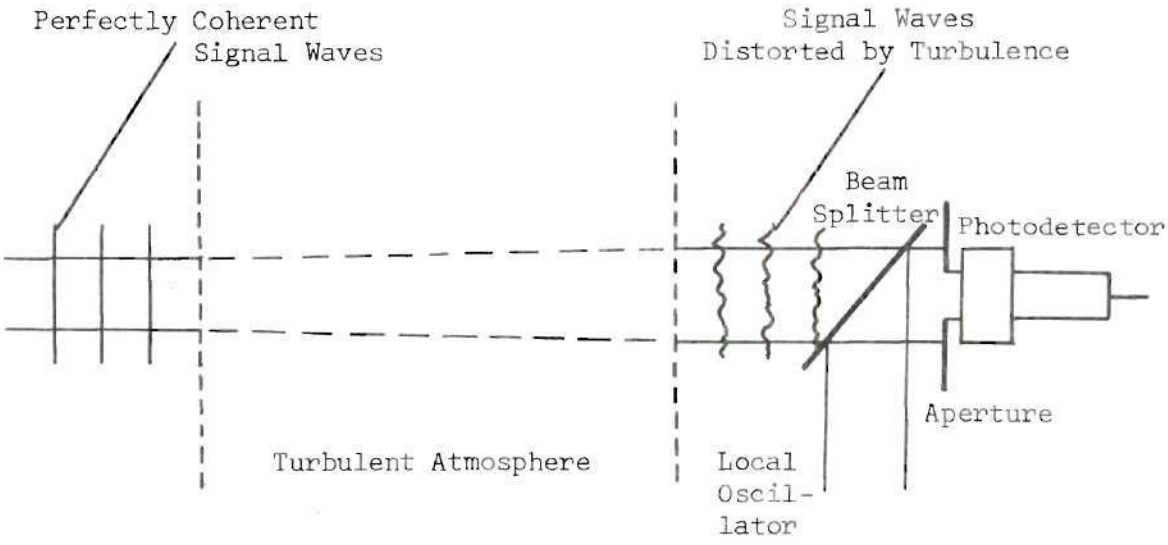

Figure 6. Optical Heterodyne Communications Link Through a Turbulent Atmosphere

index fluctuations. These statistics have been the subject of investigations by many authors, most notably, Kolmogoroff[16] and Tatarski.[17] They have found that the most convenient way to describe these fluctuations is with a structure function given by

$$D_n(|\bar{R}_2-\bar{R}_1|) = E[(\Delta n(\bar{R}_2)-\Delta n(\bar{R}_1))^2],$$

where Δn is the variation in index of refraction from the average, i.e., $\Delta n = n - E[n]$. The structure function in contrast to the correlation function, is stationary over a certain period of time (usually an hour or two) and largely unaffected by changes in the variance of Δn from point-to-point. Kolmogoroff has shown that the structure function for separations $|\bar{R}_2-\bar{R}_1|$ up to a few meters is given by

$$D_n(|\bar{R}_2-\bar{R}_1|) = C_n^2|\bar{R}_2-\bar{R}_1|^{2/3},$$

where C_n^2 is known as the structure constant and depends primarily on the variance of the refractive index fluctuation. It is usually determined experimentally.[5]

The corresponding variations in amplitude and phase of an optical wave are also described by a structure function. This is given by Tartarski[17] as

$$\begin{aligned} D(|\bar{R}_2-\bar{R}_1|) &= D_1(|\bar{R}_2-\bar{R}_1|) + D_\phi(|\bar{R}_2-\bar{R}_1|) \\ &= \mathcal{D}|\bar{R}_2-\bar{R}_1|^{5/3}, \end{aligned} \tag{4-1}$$

where $\mathcal{D}$ is a constant determined by the path of propagation, turbulence along the path, and wavelength of the optical wave. It is given for paths just above ground level by

$$\mathcal{D} = 2.91\left(\frac{2\pi}{\lambda}\right)^2 LC_n^2,$$

where λ is the wavelength of the light, L is the communication path length and C_n^2 is the structure constant. In a typical example of $\lambda = 0.63$ μm (He-Ne laser wavelength), L = 1 km, and $C_n^2 = 10^{-15}$ (mid-day turbulence), $\mathcal{D} = 291$. D_1 and D_ϕ are the log-amplitude and phase structure functions and are given by

$$D_1(|\bar{R}_2-\bar{R}_1|) = E[(\ln(A(\bar{R}_2)/E[A^2(\bar{R}_2)]^{\frac{1}{2}} - \ln(A(\bar{R}_1)/E[A^2(\bar{R}_1)]^{\frac{1}{2}})^2]$$

$$= E[(l(\bar{R}_2) - l(\bar{R}_1))^2]$$

and

$$D_\phi(|\bar{R}_2-\bar{R}_1|) = E[(\phi(\bar{R}_2)-\phi(\bar{R}_1))^2],$$

where

$$l(\bar{R}) = \ln(A(\bar{R})/E[A^2(\bar{R})]^{\frac{1}{2}}).$$

These functions are also stationary for a period of about an hour or two, homogeneous, and isotropic. In addition, the log-amplitude and phase variations are Gaussian in nature.[5,17] With these results,

the ensemble degree of coherence function can be derived in terms of this structure function.

Let us assume that the structure function is ergodic. This assumption is not always justifiable because the constant V in equation M--1 varies with the intensity of the turbulence which, in turn, varies with time of day. Over a short time, however, the structure function as stated above, is stationary, and one might expect to see all possible configurations of refractive index variation along the propagation path. Under these conditions, the ergodic assumption is reasonable,

The degree of coherence of a linearly polarized signal **carrier** field is given from equation 2-2 by

$$V^{s}i'V \qquad \frac{E[E_g(R_{19}t)E\ (R_2,t)]}{[\pounds[E_s(_iL,t)E_s(R_{19}t)]E[E_s(R_0,t)E_s''(R_0,t)]]'^5}$$

where

$$E_s(R,t) = A_s(R,t)e^{jw_S t},$$

and

$$A_s(R_n,t) = A_s(R_5t)e^{\wedge\{R,t\}}$$

in this case. This may be rewritten, dropping the t, as