UNIVERSITY OF LONDON

Imperial College of Science and Technology

Department of Physics

Optics Section

SOME APPLICATIONS OF HOLOGRAPHIC INTERFEROMETRY AND SPECKLE CORRELATION TECHNIQUES TO THE STUDY OF PLANT GROWTH AND PHYSIOLOGY

John David Briers

Thesis submitted for the degree of Doctor of Philosophy
of the University of London

October 1975

ABSTRACT

Part 1 of this thesis describes an investigation of the possible use of time-varying speckle patterns as a measurement tool in biology. The statistical properties of laser speckle patterns are discussed, and the extension of these results to fluctuating speckle patterns is presented. The close relationship between the analysis of speckle fluctuations and the techniques of intensity fluctuation spectroscopy is pointed out, and a brief survey of the latter field is given. Some experiments on the speckle fluctuations produced when botanical specimens are illuminated with laser light are described, and the wavelength dependence of these fluctuations is emphasised. A model is proposed to explain the effects observed, and suggestions are made for further work in this field.

Part 2 deals with the application of holographic interferometry to the measurement of plant growth. A brief historical review of holographic interferometry is followed by a description of experiments carried out on various plants. The difficulties of using living organisms as holographic subjects are discussed, and in particular the speckle fluctuations which led to the work described in Part 1 of this thesis. These fluctuations impose severe restrictions on the usefulness of holographic interferometry as a biological tool. The more promising areas of application are identified, and proposals are made for future work.

A perennial problem in the field of holographic interferometry is that of the quantitative interpretation of the fringe patterns observed. Part 3 of this thesis is an attempt to remove some of the confusion that has arisen on this subject, and starts with a critical review of the literature. The problem of fringe localisation is discussed in some detail, and an attempt is made to reconcile the various conflicting views. The many

different interpretation techniques are rationalised into four main classes, and their fundamental equivalence is stressed. Finally, recommendations are made regarding the choice of technique to employ in different types of application.

"....and most things twinkled after that...."

- the Mad Hatter, in 'Alice's Adventures in Wonderland', Lewis Carroll (1865)

CONTENTS

NOTATION ADOPTED IN THIS THESIS

Roman upper case

- A complex amplitude
- C autocovariance
- D distance from object to far field
- D_T diffusion constant
- E electric field ($E = E^+ + E^-$)
- G autocorrelation function
- I intensity
- J Bessel function
- L distance from hologram to image
- N (1) number of data points
  (2) number of fringes
  (3) number of scatterers
- P (1) probability density function
  (2) object point
- Q object point
- R distance from object (or image) to fringe plane
- S intensity distribution of light leaving scattering surface
- T (1) exposure time
  (2) time resolution
- V arbitrary real variable
- W Wiener spectrum

Symbols

- $\underline{\ }$ vector
- $|\ |$ modulus
- $\langle\ \rangle$ ensemble average
- $\langle\ \rangle_t$ time average
- $\langle\ \rangle_s$ spatial average
- $\hat{\ }$ best estimate of

Roman lower case

- a arbitrary constant
- c (1) normalised autocovariance
  (2) velocity of light
- d displacement of object point
- f arbitrary function
- g normalised autocorrelation function
- h fringe spacing
- i $\sqrt{-1}$
- k (1) clipping level
  (2) sensitivity factor (for holo-diagram)
- m fringe number, or order
- n refractive index
- q $= \frac{4\pi}{\lambda}\sin\frac{\alpha}{2}$
- r intensity ratio, background/speckle
- s (1) scattering area radius
  (2) fringe radius
- t time variable
- u spatial frequency in x directio
- v (1) " " " y "
  (2) velocity
- v_c half-width velocity
- x coordinate direction in either far field or hologram plane
- y coordinate direction in either far field or hologram plane
- z coordinate direction along line of sight

Superscripts

- $*$ complex conjugate
- (1) first order, pertaining to field
- (2) second order, pertaining to intensity

Greek upper case

Γ half-width of Wiener spectrum (Lorentzian distribution)

Δ (1) small change in a variable
(2) change in optical path

Π product of terms

Σ sum of terms

Greek lower case

α scattering angle

β phase of single component of scattered light

γ angle between displacement vector and some other direction, e.g. line of sight

δ (1) small increment in variable
(2) delta function

ε angle between displacement vector and x-z (horizontal) plane

ζ $= q.\tau$

η coordinate in object (scattering surface) plane

θ (1) angular coordinate
(2) angle between illumination direction and line of sight

θ_1 angle of incidence of illumination

θ_2 angle between line of sight and normal to object surface

λ wavelength

μ semi-angle subtended at the image by the two measurement points on the hologram (HF technique)

ν frequency of light

ξ coordinate in object (scattering surface) plane

ρ intensity ratio, background/total

σ standard deviation of intensity

σ^2 variance of intensity

τ time lag

τ_c time constant, or decorrelation time

Greek lower case (ctd.)

ϕ resultant phase of scattered light

χ angular fringe spacing

χ_m angular subtense of fring

ψ angle of rotation

ω angular frequency

$\Delta\omega$ Doppler shift

Subscripts

A complex amplitude

c (1) clipped function
(2) τ_c = time constant
v_c = half-width velocity

D background beam (Dainty)

dc double-clipped function

i imaginary component

M moving scatterers

m m'th fringe

N speckle intensity (Dainty)

n typical member of set

o stationary scatterers

r real component

s spatial

sc single-clipped function

t temporal

x component in x-direction

y " " y- "

z " " z- "

PREAMBLE

0.1 - Background of the Project

The project described in this thesis arose from a suggestion made by Dr M C Probine (Director of the Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Lower Hutt, New Zealand), that holographic interferometry might be used to monitor the growth rate of plants. Some initial work was carried out in New Zealand in 1972-3, and the results were fairly encouraging. It was decided to continue this work at Imperial College, and the original aim of the project was simply the investigation of holographic interferometry as a viable tool in the measurement of plant growth.

Soon after the commencement of the project in June 1973, Professor Welford suggested that an attempt should be made to unravel the complications of the interpretation of holographic interferograms, and that this analysis should form part of the thesis. Much confusion had arisen in the literature on the subject, and it was thought that a critical review of the field and a rationalisation of interpretation techniques would form a useful contribution to the art of holography. At this stage, then, it was envisaged that the thesis would consist of two separate but linked parts - an account of experimental work on the measurement of plant growth by means of holographic interferometry, and a rationalisation of techniques for interpreting the fringe patterns obtained in holographic interferograms.

Experiments carried out during late 1973 and early 1974 revealed problems in using living plants as holographic subjects. When fringes were obtained they could be interpreted by means of standard techniques, and gave results which were consistent with the long-term growth rates measured by traditional methods.

However, it proved difficult to achieve any degree of repeatability and consistency in the obtaining of fringe patterns, and even of acceptable reconstructions of the holograms. Further experiments eliminated lack of stability as a source of the trouble, and it became apparent that the problem arose from the nature of the specimens themselves. At about this time, in March 1974, Dr Nils Abramson, of the Royal Institute of Technology, Stockholm, Sweden, paid a visit to the Imperial College Optics Group, and showed some interest in the work we were doing. During the visit, he casually asked if we had noticed that "when an apple is illuminated with laser light, the speckles move!" This chance remark immediately suggested a possible cause of the lack of success with holograms of plants, and experiments soon confirmed that there was some correlation between speckle "movement" and the ability of a particular plant to produce good fringes, or even good holograms. Thus the "moving speckle" can be used to determine in advance whether a particular botanical specimen is suitable for holographic interferometry, and is a limiting factor in the usefulness of the technique.

However, attention was now turned towards the speckle movement itself, and it was soon established that the "movement" was really a fluctuation in the intensity of each individual speckle. The fact that different specimens, and even different parts of the same specimen, exhibited different degrees of fluctuation, suggested that the phenomenon might be able to tell us something about what was happening inside the plant, and an investigation of the effect was initiated. This paved the way for a third part to the thesis, dealing with the statistics and possible applications of the speckle fluctuations observed when biological specimens are illuminated with laser light.

0.2 - Summary of the Thesis

The thesis is divided into three separate but linked parts, each dealing with one of the three main lines of work outlined above. For reasons of continuity, the fluctuating speckle effect is considered first. This is followed by an account of the application of holographic interferometry to the measurement of plant growth, and finally by the investigation into the interpretation of holographic interferograms.

Part 1 starts with a chapter on the mathematics of laser speckle patterns. Well-established results for the statistics of static speckle patterns are quoted, formal proofs being given in an Appendix to this thesis. The statistics of fluctuating speckle patterns are treated in more detail, and the case of speckle patterns produced by a mixture of moving and stationary scatterers is given particular prominence. Chapter 1.2 covers the theory and practice of intensity fluctuation spectroscopy, a technique which is very useful in the analysis of speckle fluctuations. The mathematical relationships used in the application of the technique are quoted, the reader being referred to the literature for formal proofs. The development of autocorrelators for analysing fluctuating intensity fields is reviewed, and specific applications of such instruments are mentioned. The third chapter describes experiments carried out on the speckle fluctuations from botanical specimens illuminated with laser light. The wavelength dependence of these fluctuations is stressed, and a model is proposed to explain the observed effects. Some quantitative analyses of the fluctuations are presented, and suggestions are made for further work.

The first chapter of Part 2 presents a brief and qualitative review of holography and holographic interferometry. This is followed by the main chapter on the measurement of plant growth by means of holographic interferometry. The problems encountered in obtaining fringes with double-exposure holograms of growing plants, and even good reconstructions from single-exposure holograms, are discussed. Particular attention is paid to the problems caused by the fluctuating speckle effect, which led to the work described in Part 1 of this thesis, and it is pointed out how this effect can be used to predict the suitability of a particular specimen for holographic measurements. The results of some successful experiments are presented, and it is shown that growth rates calculated from double-exposure holographic interferograms with a time lapse of a few seconds between the exposures agree quite well with the long-term growth rates measured over a period of several hours by traditional techniques. Finally, we review the advantages and the limitations of holographic interferometry as a means of measuring plant growth, and suggest situations in which the technique might be of practical use. Some particular botanical investigations are proposed, as is the extension of the technique by the incorporation of some degree of magnification into the optical system to produce a low-powered holographic interference microscope.

The third and final part of the thesis opens with a historical review of holographic interferometry. The various interpretation techniques are classified in Chapter 3.2 into four main approaches, and the basic theory of each method is given. The fundamental equivalence of the different approaches is stressed. One of the main sources of confusion and conflict in the literature is the concept of fringe localisation, and this is the subject of Chapter 3.3. Part 3 closes with some recommendations concerning the choice of interpretation scheme for particular types of application.

PART 1 - TIME-VARYING SPECKLE AS A MEASUREMENT TOOL IN BIOLOGY

CHAPTER 1.1

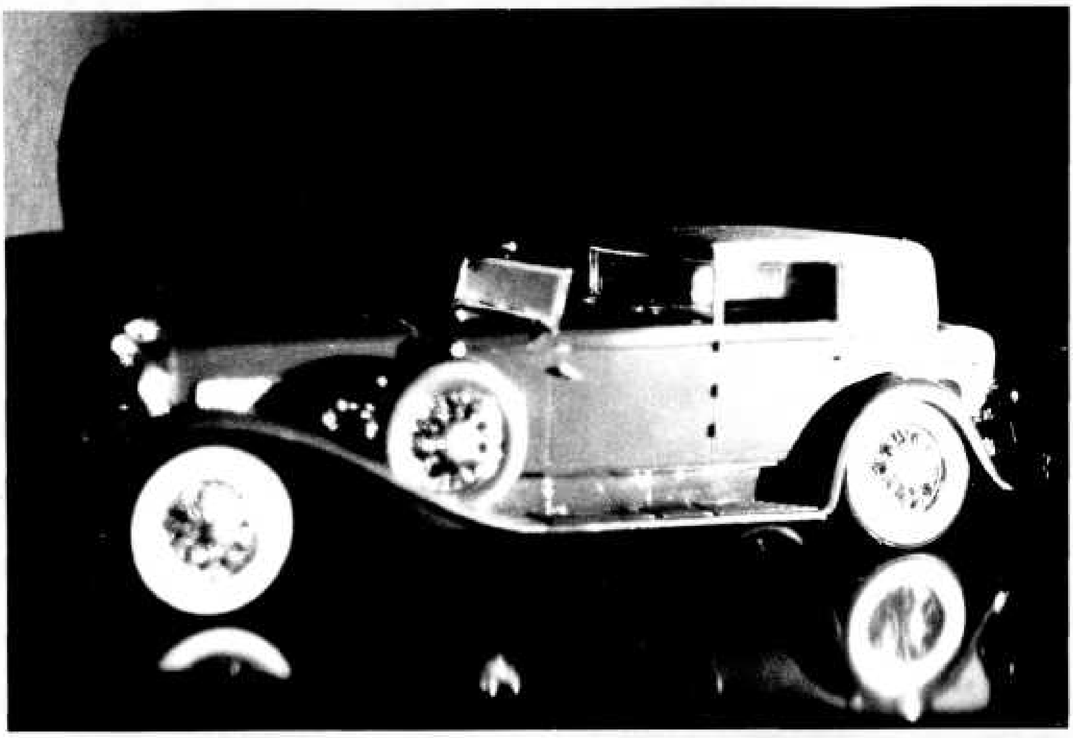
THE MATHEMATICS OF LASER SPECKLE PATTERNS

1.1.1 - A Qualitative Account of Laser Speckle Phenomena

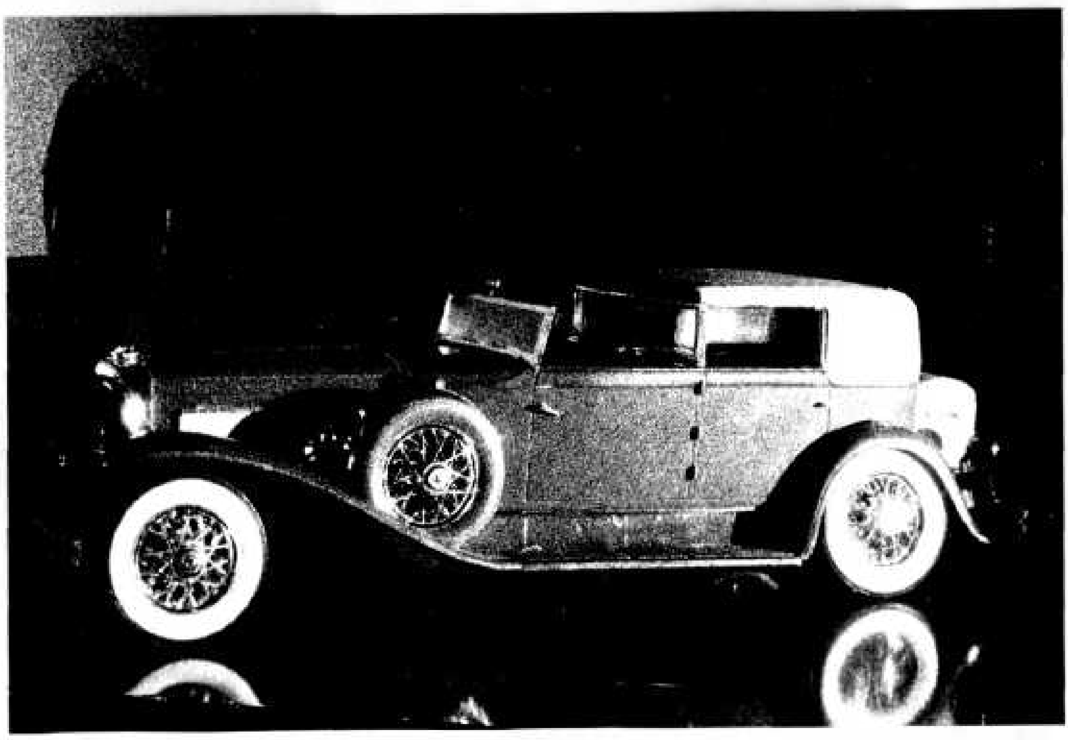
When a diffusely reflecting object is illuminated with laser light, a high-contrast "speckle pattern" is produced. This phenomenon, illustrated in Figure 1.1 , was first reported in the literature by Rigden and Gordon (1962). At first, the effect was regarded purely as a nuisance, since it severely affected resolution when laser light was used, and much effort has been directed towards reducing speckle in images formed in laser light (McKechnie 1975).

However, human behaviour - or scientist behaviour - being what it is, it was not long before people started to study speckle for its own sake, and practical applications of the phenomenon began to be reported in the literature. The two main approaches have been aimed at measuring surface roughness by the effect that this has on the statistics of the speckle pattern (Crane 1970, Fujii and Asakura 1974), and at measuring displacements and deformations of a surface by means of a technique known as speckle interferometry (Archbold et al. 1970, Burch 1971). At the same time, the statistical properties of laser speckle patterns have been investigated, starting with the pioneering work of Goodman (1963).

Even an uninitiated person, seeing a speckle pattern for the first time, realises that there is something odd about it. He gets the impression that he cannot focus his eyes properly on the surface exhibiting the speckle pattern, and he notices that the speckles move as he moves his head. If he were to try a few simple experiments he would discover even stranger effects. For



(a)



(b)

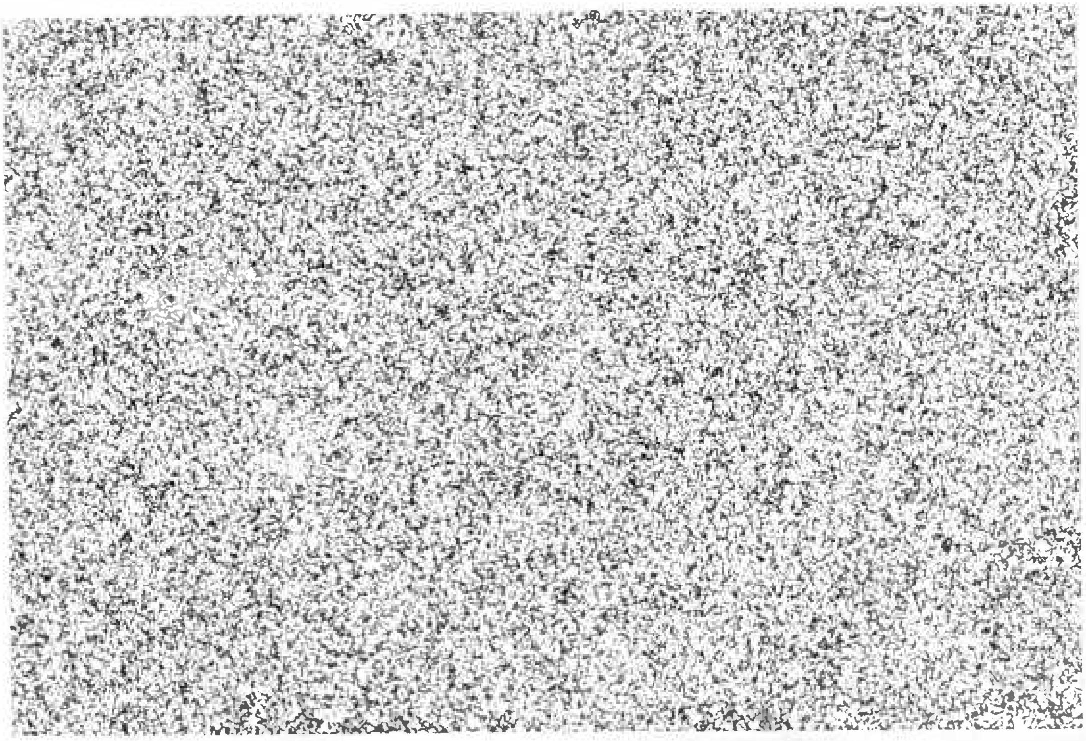
Figure 1.1 - Speckle produced on an image by laser illumination, (b) showing the effect of reducing the aperture of the viewing system.

example, if he were to observe the speckle pattern through a pinhole held in front of his eye, he would find that the speckles appeared larger (see Figure 1.1). If he were to look at the light scattered by the object and falling on a screen, he would find that this, too, produced a speckle pattern, and that in this case the speckle size depended on the area of the object surface illuminated - the smaller this area, the larger the speckles (see Figure 1.2). This latter type of speckle pattern will be referred to in this thesis as "far-field speckle", while the speckle observed on an illuminated surface will be called "image speckle". (In Gabor's terminology (Gabor 1970) they are referred to as "objective" and "subjective" speckle).

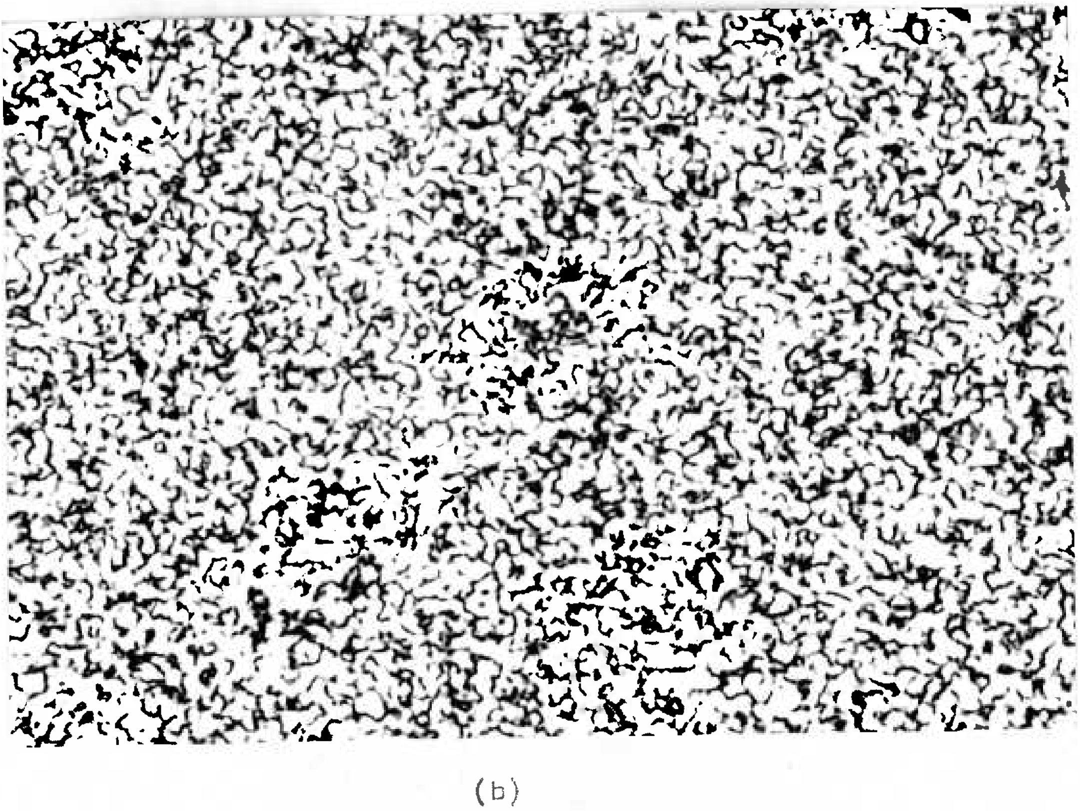
1.1.2 - A Qualitative Explanation of the Speckle Effect

The fact that speckle patterns only came into prominence with the invention of the laser suggests that the cause of the phenomenon is the high degree of coherence of the laser. Further investigation shows that this is, indeed, the case, although it is, in fact, possible to observe speckle patterns with other, less coherent, sources such as mercury lamps, and even, in certain circumstances, with sunlight.

Speckle can be regarded as an interference pattern produced by coherent light reflected from different parts of the illuminated surface. If the surface is rough compared with the wavelength of the light, rays reflected from different parts of the surface within a resolution cell (the size of which is determined by the limiting aperture of the observing system) will traverse different optical pathlengths in reaching the eye of the observer (or the observing screen in the case of far-field speckle), and the resultant intensity at a given point on the object (or on the observing screen) will be determined by the coherent addition of the complex amplitudes associated with each of these rays. If the resultant amplitude is zero, a dark "speckle" will be seen at the point, while if all the rays arrive at the



(a)



(b)

Figure 1.2 - Speckle produced in the far field of an object illuminated with laser light, (b) showing the effect of reducing the area illuminated.

point in phase an intensity maximum will be observed.

In the case of far-field speckle (Figure 1.3), the whole of the illuminated area of the object contributes to each speckle observed on the screen. The intensity at any point on the screen is given by the squared modulus of the resultant amplitude obtained by the coherent addition of complex amplitudes of rays arriving at that point from all parts of the object. It is intuitively reasonable to deduce that the larger the surface contributing to the pattern, the more rapid will be the variation in resultant amplitude across the far-field speckle pattern, and hence the smaller will be each individual speckle. In fact, it can be shown that the average speckle size is equal to that of the Airy disc produced in the far field by diffraction from an aperture the size of the illuminated area.

In image speckle (Figure 1.4), the size of a speckle is also determined by an aperture, but in this case it is the limiting aperture of the observing system which is important. When an object is viewed with the unaided eye, the limiting aperture is the pupil of the eye itself. The speckle size in the case of image speckle is equal to the resolution limit of the optical system used, and hence once again the speckle size increases as the aperture of the system is reduced. In this case, of course, the light rays contributing to the intensity of a particular speckle can only come from that same resolution cell; this is an important difference from far-field speckle, in which light from the whole illuminated surface contributes to every speckle. However, it can be shown that the same basic mathematical theory governs the properties of both types of speckle, providing a large number of scattering centres contribute to each speckle. In the case of far-field speckle, this condition is met simply by arranging for the illuminated area to be very large compared with the roughness of the surface.

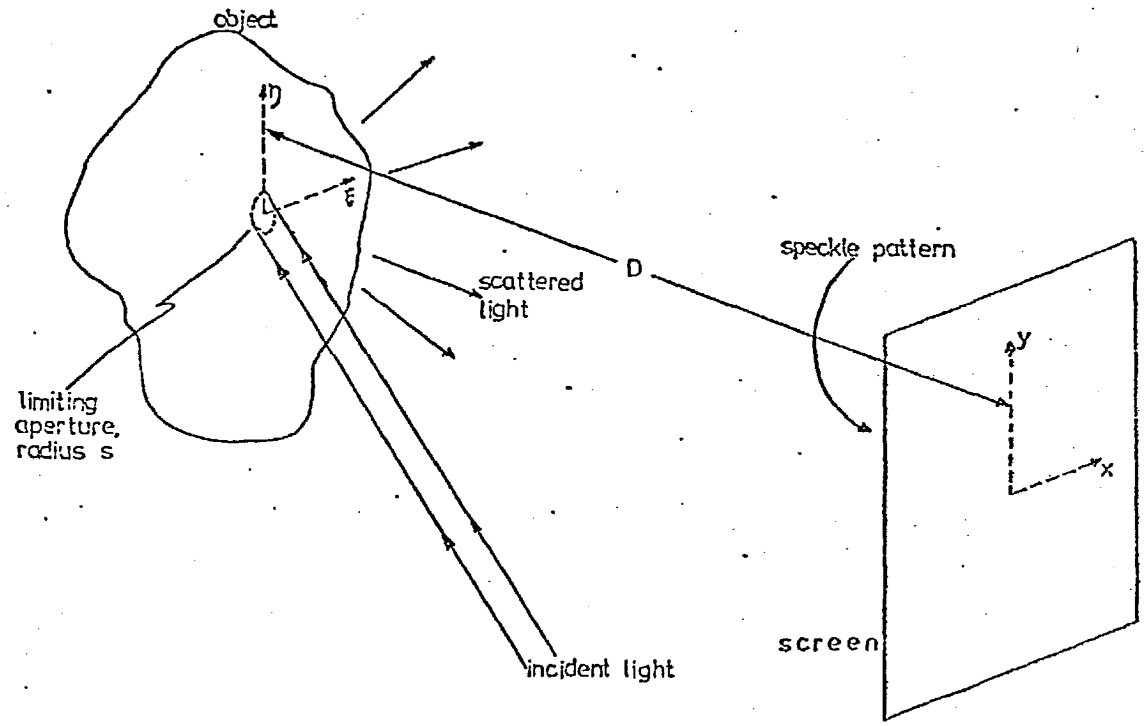


Figure 1.3 - Experimental arrangement for observing far-field speckle patterns.

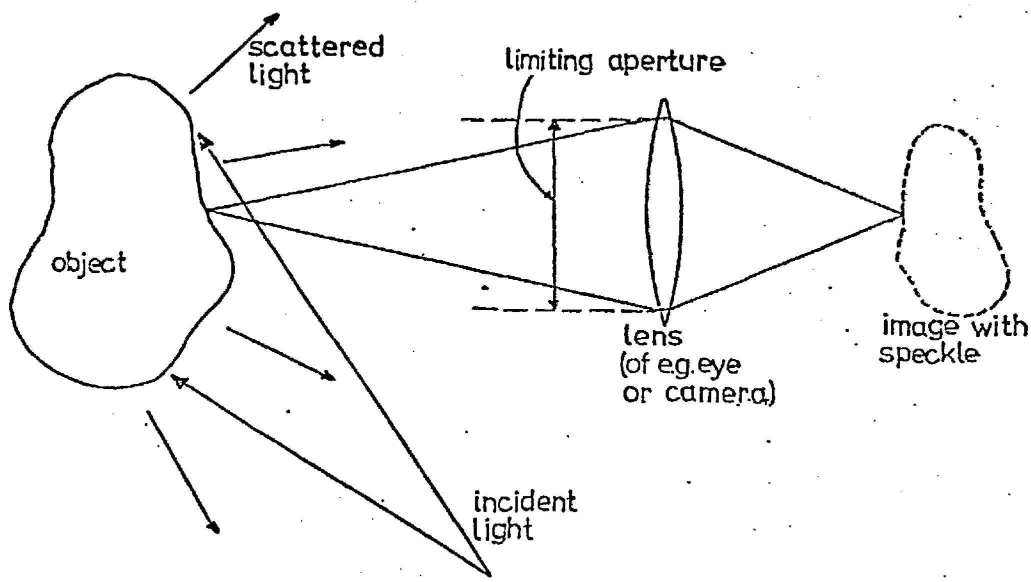


Figure 1.4 - Experimental arrangement for observing image speckle.

1.1.3 - First and Second Order Statistics of Laser Speckle Patterns

In this section, the mathematical properties of speckle patterns will be stated without proof; formal derivations of most of the formulae are given in the Appendix.

We consider specifically the case of far-field speckle, though the same formulation applies equally to image speckle. We shall assume that the surface producing the speckle pattern is rough compared with the wavelength of the light, and that both the coherence length of the light and the size of the scattering area are much larger than the path differences introduced by the roughness of the surface.

We shall adopt the following notation (see also Figure 1.3):

x, y	are coordinate directions in the far-field speckle pattern;
$\delta x, \delta y$	are small increments in x and y;
ξ, η	are coordinate directions in the plane of the scattering surface;
u, v	are spatial frequencies in the x and y directions in the speckle pattern;
D	is the distance from the scattering surface to the observing screen;
s	is the radius of the scattering area, when considered circular;
I	is the intensity at a point in the far field;
$\langle I \rangle$	is the ensemble average of intensity: in practice, this can be taken to be the mean intensity in the speckle pattern;
σ	is the standard deviation of intensity in the speckle pattern;
σ^2	is the variance of the intensity;
$P(I)$	is the probability density function of intensity in the speckle pattern: this is the probability that the intensity at a point in the far field will have the value I;

$G^{(2)}(\delta x,\delta y)$ is the autocorrelation function of intensity, defined by the equation:

$$G^{(2)}(\delta x,\delta y) = \langle I(x,y)I(x+\delta x,y+\delta y)\rangle ;$$

$g^{(2)}(\delta x,\delta y)$ is the normalised autocorrelation function:

$$g^{(2)}(\delta x,\delta y) = \frac{G^{(2)}(\delta x,\delta y)}{\langle I(x,y)\rangle^2} ;$$

$C^{(2)}(\delta x,\delta y)$ is the autocovariance of intensity, defined by the equation:

$$C^{(2)}(\delta x,\delta y) = \langle (I(x,y)-\langle I\rangle)(I(x+\delta x,y+\delta y)-\langle I\rangle)\rangle ;$$

$c^{(2)}(\delta x,\delta y)$ is the normalised autocovariance:

$$c^{(2)}(\delta x,\delta y) = \frac{C^{(2)}(\delta x,\delta y)}{\sigma^2} ;$$

$S(\xi,\eta)$ is the intensity distribution of light leaving the scattering surface;

J_1 is a first-order Bessel function;

$W(u,v)$ is the Wiener spectrum (power spectrum) of spatial frequencies in the speckle pattern;

λ is the wavelength of the light used.

We also make the following assumptions:

(i) there is a very large number of scattering centres contributing to the speckle pattern;

(ii) they are distributed randomly in the scattering area;

(iii) the resulting speckle pattern is hence random, and is stationary (its properties have no dependence on x and y, but are identical at all points in the speckle pattern).

(a) First order statistics

The first order statistics describe the properties of a speckle pattern point by point: they are not concerned with the relationship between intensities at different points in the speckle pattern.

The probability density function of intensity is found to follow a negative exponential law:

$$P(I) = \frac{1}{\langle I\rangle}\exp(-I/\langle I\rangle) \quad \ldots\ldots\ldots\ldots \quad (1.1)$$

This leads to two apparently paradoxical results:

(i) the most probable intensity at any point is zero;

(ii) it is possible to have an infinite intensity at some points.

The first of these properties is, in fact, true, and has been confirmed experimentally (Condie 1966, Dainty 1970, McKechnie 1974). The second cannot, of course, be true, and the intensity at any point in the speckle pattern is limited absolutely by the total amount of energy scattered by the surface, and by the fact that this energy must be spread over at least one speckle, whose minimum size is limited by diffraction theory. However, it has been shown by Parry (1975) that the intensity at a point in the far-field speckle pattern can exceed the mean intensity by a factor of several thousand. (The situation is a little more complicated in the case of image speckle, and is still under consideration).

The other important first order property of the speckle pattern is its variance, which describes the spread of values of intensity about the mean, and hence is a measure of the contrast of the speckle pattern. The variance of an ideal speckle pattern is, in fact, equal to the square of the mean intensity:

$$\sigma^2 = \langle I\rangle^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1.2)$$

The simplicity of this relationship leads to a very useful parameter for describing the contrast of speckle patterns, namely the ratio $\sigma/\langle I\rangle$. For an ideal speckle pattern, formed in perfectly coherent light, this ratio is equal to unity, but for a so-called reduced speckle pattern (McKechnie 1974) its value will be something less than unity.

(b) Second order statistics

The second order statistics of a speckle pattern describe how rapidly the intensity varies from point to point in the pattern. They thus give an indication of the size of the speckles, and the distribution of speckle size in the pattern.

We consider first the autocorrelation function of intensity. There is some confusion in the literature over the definition of this function. Workers in the fields of photon counting and intensity fluctuation spectroscopy (see, for example, Jakeman 1973) usually define the autocorrelation function as follows:

$$G^{(2)}(\delta x,\delta y) = \langle I(x,y)I(x+\delta x,y+\delta y)\rangle \qquad 1.3$$

(We are using the index (2), following the practice of workers in photon counting and intensity fluctuation spectroscopy, to indicate that this is the second order autocorrelation, i.e. the autocorrelation of intensity; the first order autocorrelation, $G^{(1)}(\delta x,\delta y)$, is the autocorrelation of the field, i.e. of the amplitude).

On the other hand, workers in speckle statistics tend to regard the autocovariance of the intensity as a more useful function, and to define the autocorrelation function as the normalised autocovariance:

$$c^{(2)}(\delta x,\delta y) = \frac{\langle (I(x,y)-\langle I\rangle)(I(x+\delta x,y+\delta y)-\langle I\rangle)\rangle}{\sigma^2} \qquad 1.4$$

In this thesis we shall follow the first of these alternatives, and we define the following functions:

(i) Autocorrelation function

$$G^{(2)}(\delta x,\delta y) = \langle I(x,y)I(x+\delta x,y+\delta y)\rangle \qquad 1.5$$

(ii) Normalised autocorrelation function

$$g^{(2)}(\delta x,\delta y) = \frac{G^{(2)}(\delta x,\delta y)}{\langle I\rangle^2} \qquad 1.6$$

$$\left(= \frac{G^{(2)}(\delta x,\delta y)}{\sigma^2} \quad \text{from Equation 1.2}\right)$$

(iii) Autocovariance

$$C^{(2)}(\delta x,\delta y) = \langle (I(x,y) - \langle I\rangle)(I(x+\delta x,y+\delta y) - \langle I\rangle)\rangle \qquad 1.7$$

It is easily shown that, for a stationary process, in which $\langle I(x,y)\rangle = \langle I(x+\delta x,y+\delta y)\rangle$, this is equivalent to:

$$C^{(2)}(\delta x,\delta y) = \langle I(x,y)I(x+\delta x,y+\delta y)\rangle - \langle I\rangle^2 \qquad 1.8$$

$$= G^{(2)}(\delta x,\delta y) - \langle I\rangle^2 \qquad 1.9$$

(iv) Normalised autocovariance

$$c^{(2)}(\delta x,\delta y) = \frac{C^{(2)}(\delta x,\delta y)}{\sigma^2} \qquad 1.10$$

In all of the above expressions the angular brackets, $\langle\ \rangle$, again indicate an ensemble average, which for the purposes of the present discussion can be taken to be the spatial average over the whole speckle pattern.

The limiting properties of the above functions are compared in Table 1.1.

TABLE 1.1

Limiting Properties of Autocorrelation Functions and Autocovariances of Speckle Patterns

function	zero lag ($\delta x = \delta y = 0$)	infinite lag ($\delta x, \delta y \rightarrow \infty$)
$G^{(2)}(\delta x,\delta y)$	$\langle I^2 \rangle$	$\langle I \rangle^2$
$g^{(2)}(\delta x,\delta y)$	2	1
$C^{(2)}(\delta x,\delta y)$	$\langle I^2 \rangle - \langle I \rangle^2 = \sigma^2$	0
$c^{(2)}(\delta x,\delta y)$	1	0

The advantage of using the autocovariance rather than the autocorrelation function (as defined in this thesis) can now be seen to lie in the fact that its value at large lags goes to zero. Normalising the function merely facilitates the comparison of different autocorrelation curves.

The autocovariance of intensity in the far field can be expressed in terms of the Fourier transform of the intensity distribution of the light scattered by the surface:

$$C^{(2)}(\delta x,\delta y) = \left| \iint_{-\infty}^{\infty} S(\xi,\eta) \exp\left(\frac{2\pi i}{\lambda D}(\xi.\delta x + \eta.\delta y)\right) d\xi\, d\eta \right|^2 \qquad 1.11$$

Now let us consider the scattering area to be circular in shape, with radius s, and uniformly illuminated. In this case, $S(\xi,\eta) \rightarrow \langle I \rangle$, and the autocovariance is given in terms of a first order Bessel function:

$$C^{(2)}(\delta x,\delta y) = \langle I\rangle^2 \left[\frac{2J_1\left(\frac{2\pi s}{\lambda D}\sqrt{(\delta x)^2 + (\delta y)^2}\right)}{\frac{2\pi s}{\lambda D}\sqrt{(\delta x)^2 + (\delta y)^2}}\right]^2 \qquad 1.12$$

Finally, let us consider the Wiener spectrum, which is analogous to the power spectrum of electronics theory. This has the same form as the optical transfer function (OTF), and in fact the Wiener spectrum of the speckle pattern produced by a uniformly illuminated, circular scattering area takes exactly the same form as the OTF of an aberration-free lens (Goldfischer 1965). The function is related to the autocorrelation function by a Fourier transform relationship:

$$W(u,v) = \iint G^{(2)}(\delta x,\delta y)\ \exp(-2\pi i(\delta x.u + \delta y.v))\ d(\delta x)d(\delta y) \qquad 1.13$$

It should be noted that if the autocovariance, $C^{(2)}(\delta x,\delta y)$, is used in this expression in place of the autocorrelation function, $G^{(2)}(\delta x,\delta y)$, the effect is merely to subtract a d.c. term ($\langle I\rangle^2$) from the autocorrelation function, and hence to remove the delta function from the origin in the Wiener spectrum.

(c) Effect of a finite aperture

It should be emphasised that the above account of speckle statistics is purely theoretical, and is based on the concept of intensity measurements being taken at a "point" in the speckle pattern. When practical measurements of speckle patterns are carried out, the intensity measurements must be made through finite apertures, and are thus integrated over an area, albeit a very small area, of space. This has the effect of smoothing out the speckle pattern and modifying the above statistics. This effect has been investigated by several authors, and appropriate corrections have been derived (Goodman 1965, Condie 1966, Dainty 1971, Barakat 1973, McKechnie 1974). Since our main concern in this thesis is with relative rather than absolute measurements on speckle patterns, we shall not concern ourselves too much with the effect of finite apertures, but merely note that the statistics will, in practice, be modified.

1.1.4 - The Statistics of Speckle Patterns added to Uniform Background Fields

(a) Coherent background field

The statistics of the light field resulting from the coherent addition of a speckle pattern and a uniform background have been investigated by Dainty (1972). Using a result previously obtained by Goodman (1967), Dainty gave the following expression for the standard deviation, σ, of the resultant pattern:

$$\frac{\sigma}{\langle I \rangle} = \frac{(2r+1)^{\frac{1}{2}}}{r+1} \qquad 1.14$$

where $\langle I \rangle$ = average total intensity of the combined beams,

r = ratio of the intensity of the background beam to the mean intensity of the speckle pattern,

$= I_D / \langle I_N \rangle$ in Dainty's notation.

Squaring Equation 1.14 gives the following expression for the variance of the pattern resulting from the coherent addition of a speckle pattern and a uniform background:

$$\frac{\sigma^2}{\langle I \rangle^2} = \frac{2r+1}{(r+1)^2} \qquad 1.15$$

We have proposed a simpler and more convenient form of this expression (Briers 1975e), which has also been used recently by Pedersen (1974). This is obtained merely by using the ratio of the intensity of the background beam to the average total intensity of the combined beams as the parameter, instead of the ratio used by Dainty. Using Dainty's notation, we define

$$\rho = I_D / \langle I \rangle = I_D / (I_D + \langle I_N \rangle) \qquad 1.16$$

and we use this parameter instead of $r = I_D / \langle I_N \rangle$ as used by Dainty.

Making this change results in the following expression for the variance:

$$\frac{\sigma^2}{\langle I \rangle^2} = 1 - \rho^2 \qquad 1.17$$

This should be compared with the rather more complex Equation 1.15. Equation 1.17 has the additional advantage that it is more conveniently inverted to give ρ (and hence the contribution of the background beam to the pattern) in terms of the variance:

$$\rho = \left(1 - \frac{\sigma^2}{\langle I \rangle^2}\right)^{\frac{1}{2}} \qquad 1.18$$

(b) Incoherent background field

The case of a uniform background intensity being added incoherently to a speckle pattern can best be regarded from a purely physical viewpoint. It is intuitively obvious that the effect is merely to add a constant intensity I_D to the whole area of the speckle pattern, and hence that the variance σ^2 will remain unchanged while the mean intensity of the combined pattern will be increased by an amount I_D. Since for the original speckle pattern alone $\sigma^2 = \langle I_N \rangle^2$, it follows that the variance for the combined field is given by:

$$\frac{\sigma^2}{\langle I \rangle^2} = \frac{\sigma^2}{(\langle I_N \rangle + I_D)^2} = \frac{\langle I_N \rangle^2}{(\langle I_N \rangle + I_D)^2} \qquad 1.19$$

Substituting $\rho = I_D / (\langle I_N \rangle + I_D)$ as before, this becomes:

$$\frac{\sigma^2}{\langle I \rangle^2} = (1 - \rho)^2 \qquad 1.20$$

This equation can also be derived analytically from the expression given by Burch (1970) for the probability distribution of intensity in the combined field. In Dainty's notation this expression takes the form:

$$P(I) = \frac{1}{\langle I_N \rangle} \exp\left[- \left(\frac{I - I_D}{\langle I_N \rangle} \right) \right] \qquad 1.21$$

Equation 1.20 should be compared with the coherent case given by Equation 1.17:

$$\frac{\sigma^2}{\langle I \rangle^2} = 1 - \rho^2$$

Equations 1.17 and 1.20 are plotted for comparison in Figure 1.5.

1.1.5 - The Spatial Statistics of Time-Averaged Fluctuating Speckle Patterns

So far we have been concerned only with time-independent speckle patterns produced by stationary diffusers. The question of moving diffusers has been considered by several authors (e.g. Arsenault and Lowenthal 1970, Lowenthal et al. 1970) in connection with speckle reduction techniques. We shall now consider the case, not of the scattering surface moving as a whole, but of the individual scattering centres moving independently of each other. In this case, the instantaneous far-field speckle pattern will be a normal speckle pattern with negative exponential distribution of intensities and $\sigma / \langle I \rangle = 1$ (see Section 1.1.3). However, if all the scatterers are moving at random relative to each other, this

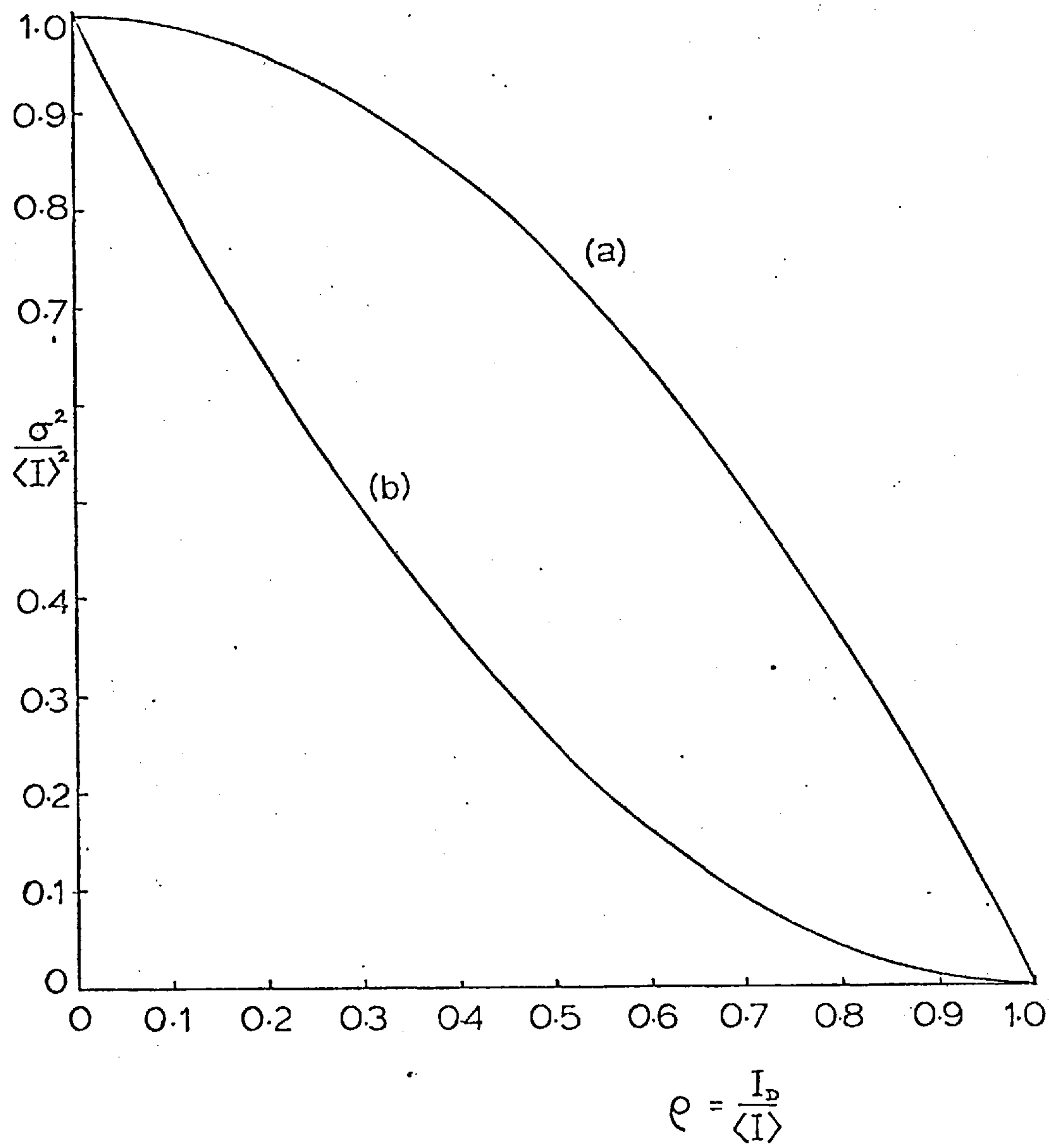


Figure 1.5 - Variation of $\sigma^2/\langle I\rangle^2$ with ρ for addition of speckle pattern and uniform background: (a) coherent addition; (b) incoherent addition.

speckle pattern will be constantly changing, and will decorrelate after a time long enough for the majority of the scatterers to have moved a distance greater than the wavelength of the light used. If the integration time of the detector is long compared with this decorrelation time, the speckle will be completely smoothed out and no intensity variations will be apparent over the far field. This, in fact, is the explanation of the observation that no speckle pattern is produced when milk (or any similar suspension) is illuminated with laser light. The Brownian motion of the scattering centres causes rapid fluctuations of the instantaneous speckle pattern, and the decorrelation time is extremely short - much shorter than the integration time of the eye.

If the speckle pattern degenerates into a uniform intensity in this way, the ratio $\sigma/\langle I\rangle$ has, of course, become zero, and it is interesting to ask how the ratio changes from unity to zero as the integration time of the detector is increased from zero. In fact it can be shown (Goodman 1965) that the spatial variance in the time-averaged speckle pattern is given by the average value of the temporal autocovariance of intensity (averaged over the integration time), i.e.:

$$\sigma^2(T) = \frac{1}{T}\int_0^T C^{(2)}(\tau)\,d\tau \qquad 1.22$$

This relationship can be used as a possible method, albeit a clumsy one, for measuring the temporal autocovariance of the speckle pattern. However, this function will be discussed in more detail in Chapter 1.2, when more appropriate methods of its measurement will be introduced.

Of more interest is the case of speckle patterns produced by mixtures of moving and stationary scatterers. In this case, the speckle pattern will not be smoothed out completely, even after long integration times, but the contrast, specified by

$\sigma/\langle I\rangle$, will tend to a minimum value as the integration time is increased. We shall now consider this situation in some detail (see also Briers 1975e).

Consider a scattering medium consisting of a mixture of stationary and moving scatterers, and illuminated with laser light. Assume that the motion of the moving scatterers is random. The resulting far-field speckle pattern produced by the light scattered by the medium can be considered as consisting of two independent components, one from the stationary scatterers and one from the moving scatterers. At any instant in time these two speckle patterns will add coherently to produce a third speckle pattern with normal speckle statistics (negative exponential distribution of intensities, $\sigma = \langle I\rangle$, etc.). Let the mean intensities of the individual speckle patterns from the stationary and moving scatterers be $\langle I_o\rangle$ and $\langle I_M\rangle$ respectively, and let their complex amplitudes at a given time and at a given point in the far field be A_o and A_M respectively. Since the two beams combine coherently, we must add complex amplitudes rather than intensities in order to determine the resultant intensity. Thus, at the given point in the far field, the instantaneous intensity is given by:

$$I(x,y,t) = \left|A_o + A_M\right|^2$$

where x,y designates a point in the far field and t designates a particular time.

The time-averaged intensity at the given point, such as would be obtained by exposing a photographic plate to the far-field speckle pattern for a time which is long compared with the decorrelation time of the pattern, is given by:

$$\langle I(x,y)\rangle_t = \left\langle \left|A_o + A_M\right|^2 \right\rangle_t \qquad 1.23$$

where $\langle\,\rangle_t$ denotes time-average.

Since A_o and A_M are complex quantities, Equation 1.23 can be written:

$$\langle I(x,y)\rangle_t = \langle (A_o + A_M)(A_o^* + A_M^*)\rangle_t$$

where * denotes the complex conjugate.

Expanding this expression, and noting that $A_o A_o^* = I_o$, $A_M A_M^* = I_M$, and $\langle A_M^*\rangle_t = \langle A_M\rangle_t = 0$, we obtain:

$$\langle I(x,y)\rangle_t = I_o(x,y) + \langle I_M(x,y)\rangle_t \qquad 1.24$$

Since the addition of intensities, rather than amplitudes, is involved, it follows that the time-averaged speckle pattern is produced by the <u>incoherent</u> addition of the speckle patterns from the stationary and the moving scatterers. Further, since $I_M(x,y)$ is varying randomly over all permissible values (providing the exposure or integration time of the detector is long enough), it is apparent that:

$$\langle I_M(x,y)\rangle_t = \langle I_M(t)\rangle_s \qquad (\langle\ \rangle_s = \text{spatial average})$$

$$= \langle I_M\rangle \qquad (\text{the ensemble average})$$

$$= \text{constant at all points in the far field.}$$

Hence the resultant time-averaged speckle pattern is equivalent to the incoherent addition of the speckle pattern due to the stationary scatterers (mean intensity $\langle I_o(x,y)\rangle_s$) and a uniform background intensity $\langle I_M\rangle$. (This result is also used in speckle interferometry (Archbold et al. 1970)).

From Equation 1.20, therefore, the variance of the time-averaged speckle pattern is given by:

$$\frac{\sigma^2}{\langle I \rangle^2} = (1 - \rho)^2$$

Thus the ratio, ρ, of the mean intensity of the light from the moving scatterers to the total intensity of the scattered light is given in terms of the contrast of the time-averaged resultant speckle pattern by the following expression:

$$\rho = 1 - \frac{\sigma}{\langle I \rangle} \qquad 1.25$$

If it can further be assumed that the moving and stationary scatterers are identical in scattering properties, and that all the light scattered into the far field is scattered by these particles, then ρ also represents the ratio of the <u>number</u> of moving scatterers to the total number of scatterers involved. Even if these last two assumptions are not valid, comparative measurements of ρ, made on the same or similar systems, will give an indication of how the relative number of moving scatterers varies from system to system, or from time to time in the same system. As explained above, the temporal autocovariance of intensity can also be determined, in principle, from measurements of the spatial variance of the time-averaged speckle patterns averaged over varying integration times. In the case of a mixture of moving and stationary scatterers, therefore, it is possible to obtain information about both the number of moving scatterers (from the variance of the time-averaged speckle pattern with a long integration time) and the velocity distribution of these scatterers (from the autocovariance, obtained via the variances of patterns with various integration times from zero upwards) - both from the spatial statistics of time-averaged speckle patterns. We shall see later, however, that at least one of these properties - the velocity distribution - is much more easily obtained from the temporal statistics of the fluctuating speckle pattern.

1.1.6 - The Temporal Statistics of Fluctuating Speckle Patterns

We now consider the temporal statistics of a randomly fluctuating speckle pattern. Again we shall consider the fluctuations as being due to the fact that the scattering medium contains a mixture of stationary and moving scatterers. When we considered the spatial statistics of the time-averaged speckle pattern due to this arrangement, we found that the result of the averaging was the same as adding an incoherent background intensity, equal to the mean intensity of the speckle pattern due to the moving scatterers, to the stationary speckle pattern produced by the stationary scatterers. The fact that such an arrangement would involve incoherent addition might have been foreseen, since the time-averaging process could be expected to destroy the effective coherence of the light. In the case of temporal statistics, however, we are concerned with the intensity fluctuations at one point in the speckle pattern as a function of time, and hence all combinations of light beams are to be performed instantaneously in time. This lack of time-averaging should lead us to anticipate that the addition in this case should be coherent, and indeed this does turn out to be the case.

Consider one point in the far-field speckle pattern produced by a mixture of stationary and moving scatterers. At any instant, the complex amplitude at the point will be given by the coherent addition of a fixed complex amplitude, A_o, due to the stationary speckle pattern produced by the stationary scatterers, and a randomly varying complex amplitude, A_M, due to the fluctuating speckle pattern produced by the moving scatterers.

Thus the resultant intensity at the point at any instant of time is given by:

$$I = \left|A_o + A_M\right|^2 \qquad 1.26$$

$$= (A_o + A_M)(A_o^* + A_M^*)$$

i.e.:

$$I = I_o + I_M + A_oA_M^* + A_MA_o^* \qquad 1.27$$

and, as before, the time-averaged intensity at the point is given by:

$$\langle I\rangle_t = I_o + \langle I_M\rangle_t \qquad 1.28$$

However, we are interested this time in the temporal variance of the intensity, rather than the spatial variance of the time-averaged speckle pattern. This can be obtained by remembering that the variance, from its definition, can be written as follows:

$$\sigma_t^2 = \langle I^2\rangle_t - \langle I\rangle_t^2 \qquad 1.29$$

From Equation 1.28 we have:

$$\langle I\rangle_t^2 = I_o^2 + 2I_o\langle I_M\rangle_t + \langle I_M\rangle_t^2 \qquad 1.30$$

Also, from Equation 1.27 we have the instantaneous intensity:

$$I = I_o + I_M + A_oA_M^* + A_MA_o^*$$

Squaring this and taking the time average gives:

$$\langle I^2\rangle_t = I_o^2 + \langle I_M^2\rangle_t + 2I_o\langle I_M\rangle_t + 2I_o\langle I_M\rangle_t$$

(all other terms go to zero on time-averaging)

i.e.: $$\langle I^2\rangle_t = I_o^2 + \langle I_M^2\rangle_t + 4I_o\langle I_M\rangle_t \qquad 1.31$$

Further, if the fluctuating part of the speckle pattern is produced by a large number of scatterers moving randomly with a normal distribution of velocities, then the temporal statistics of the fluctuations must be identical to the spatial statistics of the instantaneous speckle pattern, and in particular the variance will be equal to the square of the mean:

$$\sigma_{M_t}^2 = \langle I_M \rangle_t^2 \qquad 1.32$$

But by definition we have:

$$\sigma_{M_t}^2 = \langle I_M^2 \rangle_t - \langle I_M \rangle_t^2$$

Hence, $$\langle I_M^2 \rangle_t = 2\langle I_M \rangle_t^2 \qquad 1.33$$

Substituting this in Equation 1.31 gives:

$$\langle I^2 \rangle_t = I_o^2 + 2\langle I_M \rangle_t^2 + 4I_o\langle I_M \rangle_t \qquad 1.34$$

From Equations 1.29, 1.30 and 1.34 we now have:

$$\sigma_t^2 = 2I_o\langle I_M \rangle_t + \langle I_M \rangle_t^2 \qquad 1.35$$

(This expression is the temporal analogue of an equation derived by Goodman (1967) for the spatial variance of a speckle pattern added coherently to a uniform background intensity).

The presence of the cross-term in Equation 1.35 is due to the coherent nature of the combination of the two speckle patterns, and tells us that the temporal variance of the fluctuating speckle pattern depends on the value of I_o, and hence depends on the point in the speckle pattern at which the fluctuating intensity is monitored. Hence if any meaning is to be attached to the temporal variance, it must first be spatially averaged over the speckle pattern.

Let us now consider the effect of such a spatial averaging on Equation 1.35:

$$\sigma_t^2 = 2I_o\langle I_M\rangle_t + \langle I_M\rangle_t^2$$

Taking the spatial average of this gives:

$$\langle\sigma_t^2\rangle_s = 2\langle I_o\rangle_s\langle I_M\rangle_s + \langle I_M\rangle_s^2 \qquad 1.36$$

(due to ergodicity, we have: $\langle I_M\rangle_s = \langle I_M\rangle_t = \langle I_M\rangle$, the ensemble average).

Hence, $$\langle\sigma_t^2\rangle_s = (\langle I_o\rangle_s + \langle I_M\rangle_s)^2 - \langle I_o\rangle_s^2 \qquad 1.37$$

Since I_o and I_M are independent, this can be written:

$$\langle\sigma_t^2\rangle_s = \langle I_o + I_M\rangle_s^2 - \langle I_o\rangle_s^2 \qquad 1.38$$

and hence,

$$\frac{\langle\sigma_t^2\rangle_s}{\langle I_o + I_M\rangle_s^2} = 1 - \frac{\langle I_o\rangle_s^2}{\langle I_o + I_M\rangle_s^2}$$

$$= 1 - (1-\rho)^2$$

i.e. $$\frac{\langle\sigma_t^2\rangle_s}{\langle I_o + I_M\rangle_s^2} = \rho(2-\rho) \qquad 1.39$$

Hence ρ, the ratio of the intensity of the light scattered from the moving scatterers to the total scattered intensity, and hence (if all the scatterers are identical) the proportion of scatterers which is moving, can also be obtained from the temporal variance of the intensity fluctuations in the far-field speckle pattern, providing the variance is averaged over the speckle pattern.

In order to obtain information about the velocity distribution of the moving scatterers, it is necessary to go to the second order statistics of the temporal fluctuations, and to measure the temporal autocorrelation function (or the autocovariance) of the intensity at a point in the far-field speckle pattern. Since the second order statistics of a stationary process do not vary from point to point, spatial averaging of this function over the speckle pattern is not necessary, and any one point in the far field can be used. This brings us into the realm of intensity fluctuation spectroscopy, which is the province of the next chapter.

CHAPTER 1.2

INTENSITY FLUCTUATION SPECTROSCOPY

1.2.1 - The Background to Intensity Fluctuation Spectroscopy

Intensity fluctuation spectroscopy, or light beating spectroscopy as it is called by some authors, came into prominence with the advent of the laser. Prior to that, classical spectroscopy, using gratings or Fabry-Perot etalons, had proved adequate for use with conventional sources. The latter had such large bandwidths, due usually to Doppler broadening, that there was little point in extending the resolving power of the technique beyond the 10^{-6} easily attainable by traditional methods. The arrival of the laser, however, with its very narrow spectral lines, soon placed classical spectroscopy and interferometry well out of court. The reason for this can be seen from the following argument (Pike 1973).

In classical spectroscopy the optical field is split by the spectrometer into its spectral components, the Wiener spectrum being related to the optical field by the Fourier transform equation of the Wiener-Khintchine theorem:

$$W(\omega) = \int_{-\infty}^{\infty} \langle E^+(\tau)E^-(0)\rangle e^{-i\omega t} d\tau \qquad 1.40$$

where ω = angular frequency

$E(t) = E^+(t) + E^-(t)$

= instantaneous value of the field at time t. This is related to the more familiar quantity, the intensity I, by the equation

$$I(t) = E^+(t)E^-(t) \qquad 1.41$$

(Glauber 1963)

Thus the Fourier transform pair are the Wiener spectrum, $W(\omega)$, and the field autocorrelation function $\langle E^{+}(\tau)E^{-}(0)\rangle$, which we shall designate by $G^{(1)}(\tau)$. If the spectrum $W(\omega)$ is broad, $G^{(1)}(\tau)$ is narrow (i.e. the decorrelation time - or coherence time in this context - is short), and vice versa. Equation 1.40 is therefore the mathematical formulation of the well-known phenomenon that the narrower the spectral line, the longer is the coherence time.

In order to resolve two spectral frequencies $\Delta\omega$ apart, we have the choice of two approaches. We can measure $\Delta\omega$ directly, as a beat frequency between the two frequencies ω and $\omega+\Delta\omega$; this involves a time resolution $T = 2\pi/\Delta\omega$. Or we can measure the distance over which the two frequencies ω and $\omega+\Delta\omega$ get out of step by one wavelength; this requires an instrument of length $L = \omega\lambda/\Delta\omega$. For optical wavelengths ($\lambda \sim 0.5\mu m$), the requirements are summarised in Table 1.2 (Pike 1973).

TABLE 1.2

Instrument Lengths and Time Resolutions Required for Given Frequency Resolutions in Spectroscopy

Resolution $\frac{\Delta\omega}{\omega}$	$L = \frac{\omega\lambda}{\Delta\omega}$	$T = \frac{2\pi}{\Delta\omega}$
10^{-5}	50mm	200ps
10^{-6}	500mm	2ns
10^{-7}	5m	20ns
10^{-8}	50m	200ns

It can now be seen why intensity fluctuation spectroscopy was not developed earlier. Prior to the invention of the laser, the nature of conventional sources was such that, due to the Doppler broadening mentioned above, resolutions in excess of 10^{-6} were rarely feasible. At these relatively low resolutions, the fluctuation times (T) are too short to be easily measurable, while the length of interferometric instrument required is perfectly realisable. However, the increased resolution offered by laser sources swung the balance in favour of correlation techniques, since the fluctuation times became more reasonable, while interferometer lengths became unmanageable. The dotted line in Table 1.2 represents the break-even point between interferometric and correlation techniques. A detailed comparison of interferometric and correlation techniques has been given by Vaughan (1973).

Although, as we have said, fluctuation measurements did not come into prominence until the laser was available as a light source, it should perhaps be mentioned that the first correlation experiments were, in fact, carried out a few years earlier using non-laser sources. These were the beating together of the two components of the Zeeman-split mercury line at 546nm, observed by Forrester et al. (1955), and the intensity interferometry experiments of Hanbury Brown and Twiss (1956a,b). It was the laser, however, which created a real need for higher resolution, and hence stimulated the rapid development of the techniques of intensity fluctuation spectroscopy.

The idea of investigating spectral content in the temporal rather than the spatial domain was not, of course, an entirely new concept. Techniques had been developed for use at longer wavelengths - those of the microwave and radio bands - and were in regular use as methods of information extraction from carrier waves. However, such techniques cannot be taken over directly to optical frequencies because in the latter case square law detection is involved - there is no way of continuously monitoring the field itself, but only the quantity $|E^{+}(t)|^{2}$.

Two methods have been developed to extend the resolution beyond the practical limit of classical techniques, and thus to take advantage of the narrow spectral lines of laser sources. The first of these involves the use of analogue spectrum analysers to measure the Wiener spectrum, $W(\omega)$, directly. Although this technique has now been virtually superseded by the second one, it is still claimed to have the advantage in some circumstances, such as work at very high frequencies ($>$20MHz) and with narrow spectral lines (Cummins 1973a). A review of the spectral analysis approach to intensity fluctuation spectroscopy has been given by Cummins and Swinney (1970). The second technique, that of digital autocorrelation to give the Fourier transform of the Wiener spectrum, will be discussed in some detail in Section 1.2.3.

1.2.2 - The Mathematics of Intensity Fluctuation Spectroscopy

For the purposes of the present discussion, we shall assume that the intensity fluctuations to be investigated comprise a stationary, ergodic system. (A stationary process is one in which the statistics do not change with time, i.e. $P(V(t)) = P(V(t+\tau))$, where V is any real time-dependent variable; this also implies that $\langle V(t)V(t+\tau)\rangle = \langle V(0)V(\tau)\rangle$. An ergodic system is one in which the ensemble average is equal to the time average, i.e. $\langle f(V)\rangle = \int_{-\infty}^{\infty} f(V)P(V)\,dV = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{+T/2} f(V(t))\,dt$). For such a system, we can define the temporal autocorrelation function as follows:

$$G(\tau) = \langle V(0)V(\tau)\rangle = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{+T/2} V(t)V(t+\tau)\,dt \qquad 1.42$$

The Wiener-Khintchine theorem relates this quantity with the Wiener spectrum, $W(\omega)$:

$$W(\omega) = \int_{-\infty}^{\infty} G(\tau)e^{-i\omega\tau}\,d\tau \qquad 1.43$$

i.e. the Wiener spectrum is the Fourier transform of the autocorrelation function.

In general, a complete statistical description of $V(t)$ can only be provided by a knowledge of the entire set of multiple joint probability distributions $P(V(t_i))$, or, equivalently, by a knowledge of the entire set of moments $\langle V^n(t)\rangle = \int_{-\infty}^{\infty} V^n p(V)\, dV$ and correlation functions $\left\langle \prod_i V^{n_i}(t_i)\right\rangle = \int_{-\infty}^{\infty} p(V_i) \prod_i V_i^{n_i}\, dV_i$

(Jakeman 1973). However, it can be shown that for Gaussian fields, all higher order spectral properties can be expressed in terms of the first and second order statistics. Hence, for such a system it is only necessary to measure the Wiener spectrum, or, by Equation 1.43, the autocorrelation function.

Turning to optical fields, we immediately encounter a problem in that the response time of practical detectors is much too short to follow the oscillations of the field, and the quantity actually measured is the square of the envelope of the field, usually referred to as the intensity. Thus optical detectors are invariably square-law detectors, and neither the phase nor the frequency of the field are measured. Hence we cannot measure the statistics of the field itself, but only those of the intensity. Referring to Equation 1.40, we have the Wiener-Khintchine theorem applied to optical fields:

$$W(\omega) = \int_{-\infty}^{\infty} G^{(1)}(\tau) e^{-i\omega\tau}\, d\tau$$

where $G^{(1)}(\tau) = \langle E^+(\tau)E^-(0)\rangle$ is the temporal autocorrelation function of the optical field $E(t)$.

Due to the limitation of square-law detectors, however, we are unable to measure $G^{(1)}(\tau)$ directly, but are restricted to the measurement of the autocorrelation function of the intensity, defined as follows:

$$G^{(2)}(\tau) = \langle I(0)I(\tau)\rangle$$

It is again useful at this stage to define a normalised autocorrelation function:

$$\text{if} \quad G(\tau) = \langle V(0)V(\tau)\rangle$$

$$\text{then define} \quad g(\tau) = \frac{\langle V(0)V(\tau)\rangle}{\langle V^2\rangle} \qquad 1.44$$

Adopting this device, we have (Jakeman 1973):

$$g^{(1)}(\tau) = \frac{\langle E^+(0)E^-(\tau)\rangle}{\langle I\rangle} \qquad 1.45$$

$$\text{and} \quad g^{(2)}(\tau) = \frac{\langle I(0)I(\tau)\rangle}{\langle I\rangle^2} \qquad 1.46$$

These two normalised functions can be related by the so-called Siegert relation (Siegert 1943, Glauber 1963, Jakeman 1973):

$$g^{(2)}(\tau) = 1 + \left|g^{(1)}(\tau)\right|^2 \qquad 1.47$$

providing Gaussian fields are involved; this implies that a large number of scatterers is involved in the scattering volume.

As an aside at this point, it should be mentioned that for non-Gaussian fields the actual relationship between $g^{(1)}(\tau)$ and $g^{(2)}(\tau)$ might provide useful information about the scattering process. Hence it would be useful if a direct measurement of $g^{(1)}(\tau)$ could be achieved. Notwithstanding the remarks made about square-law detectors earlier in this section, such direct measurements <u>can</u> be made using a heterodyne approach, in which the unscattered laser beam is used as a local oscillator. However, these techniques, which are widely used in the specific application of laser Doppler anemometry - and which are directly analogous to the use of a reference beam in holography - lie outside the scope of this thesis. The usual approach in intensity fluctuation spectroscopy is to measure $g^{(2)}(\tau)$ and use the Siegert relation (Equation 1.47) to find $g^{(1)}(\tau)$.

Returning to Equation 1.47, the following properties of the normalised autocorrelation functions $g^{(1)}(\tau)$ and $g^{(2)}(\tau)$ should be noted:

(i) at zero lag, $g^{(1)}(0) = \frac{G^{(1)}(0)}{\langle I \rangle} = \frac{\langle E^+(0)E^-(0)\rangle}{\langle I \rangle} = 1$

$$g^{(2)}(0) = \frac{G^{(2)}(0)}{\langle I \rangle^2} = \frac{\langle I^2 \rangle}{\langle I \rangle^2} = 2$$

(ii) at long lags (full decorrelation),

$$g^{(1)}(\tau) = \frac{\langle E^+(0)E^-(\tau)\rangle}{\langle I \rangle} = \frac{\langle E^+(0)\rangle\langle E^-(\tau)\rangle}{\langle I \rangle} = 0$$

$$g^{(2)}(\tau) = \frac{\langle I(0)I(\tau)\rangle}{\langle I \rangle^2} = \frac{\langle I(0)\rangle\langle I(\tau)\rangle}{\langle I \rangle^2} = 1$$

These properties, which are identical to those given in Table 1.1 in Section 1.1.3 for spatial autocorrelations, are obviously consistent with the Siegert relation (Equation 1.47), which has been quoted in the form in which it is usually employed by workers in intensity fluctuation spectroscopy. However, if we replace the autocorrelation function of the intensity fluctuations by the autocovariance, the relation becomes even simpler. For a stationary process, we have, by analogy with Equation 1.8:

$$\begin{aligned} C^{(2)}(\tau) &= \langle I(0)I(\tau)\rangle - \langle I \rangle^2 \\ &= G^{(2)}(\tau) - \langle I \rangle^2 \end{aligned} \qquad 1.48$$

Hence the normalised autocovariance, $c^{(2)}(\tau)$, is given by:

$$c^{(2)}(\tau) = g^{(2)}(\tau) - 1 \qquad 1.49$$

and the Siegert relation becomes simply:

$$\begin{aligned} c^{(2)}(\tau) &= \left|g^{(1)}(\tau)\right|^2 \\ &= \left|c^{(1)}(\tau)\right|^2 \end{aligned} \qquad 1.50$$

(since $\langle A \rangle = 0$)

Both $c^{(1)}(\tau)$ and $c^{(2)}(\tau)$ take the value 1 at zero lag, and fall off to zero (though at different rates) at large lags. (Equation 1.50 is the temporal analogue of Equation 1.11 of Section 1.1.3 - see also Equations A26 and A27 of the Appendix). Workers in intensity fluctuation spectroscopy and photon counting, however, usually use the autocorrelation function (Jakeman 1973), and hence use the Siegert relation in the original form of Equation 1.47. It is interesting to note, however, that they usually plot $g^{(2)}(\tau) - 1$ against τ in order to have the curve fall from unity at zero lag to zero at large lags!

Having obtained $g^{(1)}(\tau)$, via $g^{(2)}(\tau)$, we can now find the Wiener spectrum, $W(\omega)$, by applying the Wiener-Khintchine theorem of Equation 1.40. If we are dealing with moving particles, however, it is of more interest to go directly to the velocity distribution $P(\underline{v})$ - or, more likely, the speed distribution $P(|v|)$. It can be shown (Nossal et al. 1971) that $\frac{P(|v|)}{|v|}$ and $\xi\left|g^{(1)}(\xi)\right|$ are a Fourier transform pair, where $\xi = q\tau$ and $q = \frac{4\pi}{\lambda}\sin\frac{\alpha}{2}$, α being the scattering angle. In other words, the Wiener-Khintchine theorem takes the form:

$$P(|v|) = \frac{2|v|}{\pi}\int_0^\infty \xi\left|g^{(1)}(\xi)\right| \sin(\xi|v|)\, d\xi \qquad 1.51$$

Since the autocorrelation function must be truncated at some point, spurious oscillations will be introduced into $P(|v|)$. Some a priori knowledge of $P(|v|)$ must be assumed so that these oscillations can be suppressed; one simple technique that has been proposed by some authors is to take the envelope of the distribution as approximating the real distribution.

If we are interested only in some measure of the average particle velocity rather than the velocity distribution of all the particles, we can avoid the computation of the Fourier transform in Equation 1.51. If we assume that the Wiener spectrum is a Lorentzian distribution of half-width Γ, then the autocorrelation function $g^{(1)}(\tau)$ will be a negative exponential with time constant $\tau_c = 1/\Gamma$,

i.e. if the Wiener spectrum is given by $W(\omega) = \dfrac{\Gamma}{\omega^2 + \Gamma^2}$

then the autocorrelation function is $g^{(1)}(\tau) = \exp(-|\tau|/\tau_c)$.

Even if the assumption of a Lorentzian Wiener spectrum is not completely valid, the autocorrelation function will often be found to approximate to a negative exponential curve. If we take the value of τ at the 1/e point on the autocorrelation curve, we shall obtain a typical "decorrelation time" τ_c, from which we can obtain a typical "half-width velocity" as follows:

$$v_c = \frac{\lambda}{2\pi\tau_c} \qquad 1.52$$

1.2.3 - Digital Autocorrelation Techniques

As we have already mentioned, early experiments in the field of intensity fluctuation spectroscopy were carried out using spectrum analysers to obtain the Wiener spectrum directly. Although these instruments do have advantages in certain circumstances (see Section 1.2.1), they have now been largely superseded by digital autocorrelators such as those developed at the Royal Radar Establishment, Malvern (Pike and Jakeman 1973).

The advent of the laser, with its very narrow spectral lines, prompted the need for the extension of the frequency range of optical spectroscopy below the 1MHz limit of the Fabry-Perot etalon. This led to an intensive programme at RRE which culminated in the development of a digital autocorrelator capable of covering the frequency range from 1 to 10^8Hz. The original aim in developing such an instrument was the investigation of the statistical properties of various light sources (Johnson et al. 1965). Hence the requirement was to measure photon statistics by the detection of individual photons. Although the original application of the digital autocorrelator was as a tool in photon-counting experiments, its use at higher light levels, where continuous intensities are being measured, is equally valid, and it was not long before its potential as a tool in intensity fluctuation spectroscopy was realised and exploited. (It should be noted, however, that an autocorrelator working in the photon-counting mode displays a higher accuracy than one operating in a capacitative integration mode such as is usually adopted when analogue (continuous) signals are being monitored (Oliver 1973)).

A digital autocorrelator can be constructed, in principle, from a shift register and a digital store (see Figure 1.6). The signal at a given time, either an instantaneous value of the intensity or the integrated intensity (or total photon count) over a chosen integration time, is multiplied by previous signals in turn and each product is fed into the appropriate store corresponding to that lag time. Continuous cycling allows the accumulation of data in these stores, and in this way the autocorrelation function of the intensity is built up. The resolution of the instrument is determined by the shift time, and the frequency range by the total sample length (cycle time).

Unfortunately, the attainment of even a modest resolution involves the multiplying together of many pairs of numbers in very short times, and this would involve very expensive computing facilities. In order to circumvent this practical objection to digital autocorrelators, the technique of "clipping" the signals has been introduced. The clipping procedure is a one-bit quantisation in which the analogue signal $V(t)$ is replaced by a two-level scheme $V_c(t)$. In an extreme form of clipping known as "hard limiting", $V_c(t)$ takes the value -1 if V is less than its mean value, and $+1$ if it is greater than this,

$$\text{i.e.} \quad V_c = -1 \qquad \text{if } V < \langle V \rangle$$

$$V_c = +1 \qquad \text{if } V \geq \langle V \rangle$$

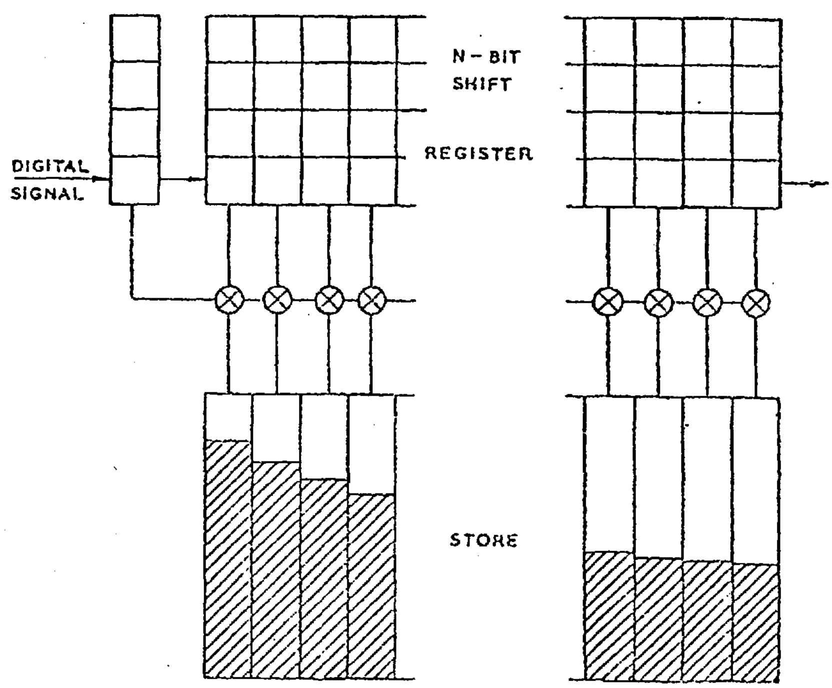


Figure 1.6 - Block diagram of a many-bit correlator (after Jakeman 1973 and Oliver 1973).

If we can assume that $V(t)$ is normally distributed about $\langle V \rangle$, it can be shown that the clipped normalised autocorrelation function, $g_{dc}(\tau) = \langle V_c(t) V_c(t+\tau) \rangle$, is given by the following:

$$g_{dc}(\tau) = \frac{2}{\pi} \arc \sin g(\tau) \qquad 1.53$$

This is known as the Van Vleck theorem (Van Vleck and Middleton 1966). $g_{dc}(\tau)$ is much more easily obtained than $g(\tau)$ since it involves only the multiplication of a series of +1's and -1's.

The above form of clipping, in which <u>all</u> the signals are clipped according to the above scheme, is known as double-clipping (hence the use of the subscript "dc" in the above expressions). It is also possible to define a single-clipped version, in which only one channel is hard-limited, according to the following scheme:

$$V_c = 0 \qquad \text{if } V < \langle V \rangle$$

$$V_c = 1 \qquad \text{if } V \geqslant \langle V \rangle$$

In this case the function measured is called the single-clipped autocorrelation function $g_{sc}(\tau) = \langle V_c(t) V(t+\tau) \rangle$. This is related to the unclipped autocorrelation function, $g(\tau)$, by an even simpler formula:

$$g_{sc}(\tau) = \frac{1}{\sqrt{2\pi}} g(\tau) \qquad 1.54$$

If we try to apply the Van Vleck theorem (Equation 1.53) to the optical case, we come up once more against the problem of square-law detectors, and the theorem is not valid. However, a similar relationship <u>can</u> be derived for the case of <u>single-clipped</u> autocorrelation functions and Gaussian signals (Jakeman and Pike 1969):

$$g^{(2)}_{sc}(\tau) = 1 + \frac{1+k}{1+\langle I \rangle} \left| g^{(1)}(\tau) \right|^2 \qquad 1.55$$

where k = clipping level:

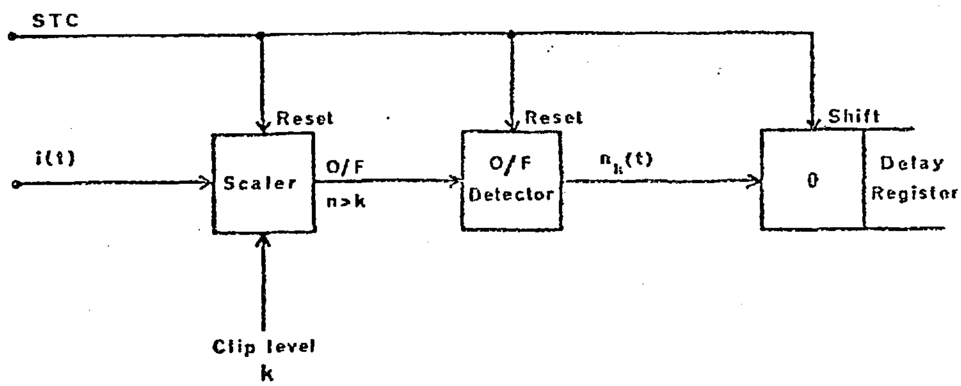
$$V_c(t) = 0 \quad \text{if } V < k$$
$$= 1 \quad \text{if } V \geqslant k$$

No such simple relationship can be derived for the case of double clipping, although some reduction is possible for the special case of clipping at zero (Jakeman and Pike 1969). However, double clipping introduces distortion of the time-dependence of the intensity autocorrelation function, whereas no such distortion is introduced by the single clipping approach - indeed, if $k = \langle I \rangle$, the relationship expressed by Equation 1.55 becomes identical to the Siegert relation of Equation 1.47.

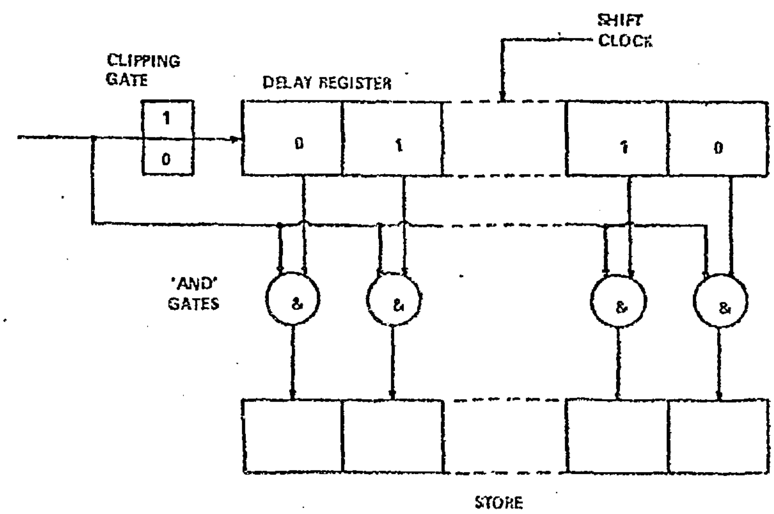
The design of a digital autocorrelator to operate in the single-clipping mode is relatively straightforward and is illustrated in Figure 1.7. The shift register remains as in the original unclipped version, but input to the stores is now controlled by simple AND gates. The clipped signals are fed into the shift register as a set of 0's and 1's every shift time and the value of the unclipped signal is fed simultaneously to each of the gates. Storage is only accepted in channels where the shift register holds a 1. Further details of this autocorrelator, including the effects of spatial and temporal integration and of non-Gaussian fields, can be found in the review paper by Jakeman (1973). The practical applications of the autocorrelator, together with potential sources of error such as source and detector imperfections (frequency and amplitude instabilities), have been discussed by Oliver (1973).

1.2.4 - Applications of Intensity Fluctuation Spectroscopy

In general, two effects are apparent when a laser beam is scattered by an agglomeration of particles. First, any random velocity distribution of the scattering particles results in a Doppler broadening of the spectral line, i.e. in a reduction of coherence of the laser light. Representing the incident laser light as an idealised δ-function spectrum of frequency ω (i.e. zero bandwidth, or infinite coherence length), the spectrum of the scattered light will have a finite bandwidth $1/\tau_c$, where τ_c is



(a)



(b)

Figure 1.7 – Schematic design of a single-clipping autocorrelator:
(a) block diagram of the clipping gate;
(b) block diagram of the complete autocorrelator (after Jakeman 1973 and Oliver 1973).

is the decorrelation time (or the coherence time). Both τ_c and the velocity distribution of the particles can be derived from the autocorrelation curve, and this can be obtained by means of an autocorrelator such as was described in the preceding section. The second effect is that any overall motion of the scatterers leads to a Doppler shift $\Delta\omega$ in the frequency of the scattered light. This is usually detected by heterodyne techniques, using the unscattered laser neam as a local oscillator, and provides the basis of the techniques of laser Doppler anemometry. This is a very active field (Abbiss et al. 1974) but lies outside the scope of this thesis, in which we shall concentrate mainly on the spectral broadening of the scattered light.

This may be an appropriate point at which to emphasise the equivalence between the Doppler broadening, mentioned above, and the phenomenon of the fluctuating speckle patterns produced by groups of moving scatterers. If we consider just two scatterers, one stationary with respect to the observer and one moving with a velocity component v along the line of sight (positive towards the observer), then a classical interferometry approach tells us that the light scattered from the two particles and reaching an observer will be in phase when the separation of the two particles is a whole number of wavelengths and out of phase when it is an odd number of half-wavelengths. Thus a fluctuation rate of 1 Hz implies that the separation of the two particles is changing at the rate of one wavelength per second, and hence that $v = \lambda$. Using the Doppler approach, if the frequency of the light from the stationary particle is ν, then the frequency of the light from the moving particle is given by $\nu' = \nu(1 + v/c)$, where c is the velocity of light. Hence the difference in frequency between the light from the two scatterers (and hence the beat frequency measured by the observer) is given by $\Delta\nu = \nu' - \nu = \nu . \frac{v}{c} = \frac{v}{\lambda}$ (since $\frac{c}{\nu} = \lambda$), and for $\Delta\nu = 1$ Hz, we again have $v = \lambda$. (In order to keep this discussion simple, we have neglected the effects due to the fact that the incident light is propagating towards a moving particle, and have considered what happens only after the light has left the scattering particles; it is a trivial matter to generalise the argument).

Historically, the first application of intensity fluctuation spectroscopy was reported twenty years ago by Forrester et al. (1955), who used a spectrum analyser to measure the beat frequency between the two Zeeman components of the mercury 546nm line. Similar techniques have been used to investigate the stability of light sources, and it was this application that led to the development of the RRE digital autocorrelator. The large and important field of laser Doppler anemometry has already been mentioned, and other applications have included the study of critical phenomena in fluids (Swinney 1973) and of liquid crystals.

One of the main applications of intensity fluctuation spectroscopy has been in the study of Brownian motion by the measurement of diffusion constants. Spectral broadening of laser light by diffusion was first observed by Cummins et al. (1964), who used a suspension of polystyrene latex spheres. For spherical scatterers, it can be shown (Cummins 1973b) that:

$$g^{(2)}(\tau) = 1 + \exp\left(-2D_T q^2 \tau\right) \qquad 1.56$$

where D_T = diffusion constant
$q = \frac{4\pi n}{\lambda} \sin\frac{\alpha}{2}$
n = refractive index
α = scattering angle.

In practice, the quantity actually measured in digital autocorrelation experiments is:

$$\hat{g}^{(2)}(\tau) = 1 + a.\exp\left(-2D_T q^2 \tau\right) \qquad 1.57$$

where a = a constant depending on spatial coherence, etc.

D_T can be found from the slope of a plot of $\log_e(\hat{g}^{(2)}(\tau) - 1)$ against τ. The above applies strictly only to ideal spherical scattering particles. The situation is more complicated for non-spherical scatterers, and this, together with other effects of non-ideal scatterers, has been dealt with by Cummins (1973b), and in more detail by Pusey (1973).

Diffusion measurements are very important in biological work, the first reported experiments being carried out by Dubin et al. (1967) on the translational diffusion constants of biological macromolecules. A detailed review of the biological applications of intensity fluctuation spectroscopy has been given by Cummins (1973b), and we shall content ourselves here by mentioning only the main developments. Spectral broadening by motile micro-organisms was first reported by Bergé et al. (1967), who showed, in experiments with rabbit spermatazoa, that motile organisms show more spectral broadening than do passive ones. Similar experiments on bacteria have been carried out by Nossal et al. (Nossal 1971, Nossal et al. 1971, 1972a,b), who have deduced the speed distribution $P(|v|)$ of the organisms from autocorrelation measurements, and by Schaeffer (1973). Experiments on the motion of isolated bacterial flagella have been reported by Fujime et al. (1972a,b), and electrophoresis measurements have been described by Ware and Flygare (1971, 1972) and Uzgiris (1972).

Turning to larger biological structures, it is appropriate at this point to quote a passage from Cummins (1973b) which is of relevance to the experiments on fluctuating speckle from biological specimens to be described in Chapter 1.3 of this thesis:

> "Living cells are known to be characterised by many internal dynamical processes, and it should be possible to analyse many of these processes by light scattering if the changes occurring during the process can couple to the light".

Experiments have been performed on contractile muscle tissue (Carlson and Fraser 1973), blood flow within the eye (Riva et al. 1972), and motions inside a single cell, using a microscope to ensure that a small scattering volume was involved (Maeda and Fujime 1972). The transport of ions across a cell membrane in blood cells was reported by Bargeron et al. (1972), but the spectral broadening that they had observed later turned out to be a spurious effect caused by convection set up by the incident laser beam (Bargeron et al. 1973) - a dire warning to workers in this field! Finally, work on protoplasmic streaming has been reported by Piddington (1973) (see also recent work by Piddington (1975) and Langley et al. (1976)). Piddington and his colleagues claim to have separated streaming from the Brownian motion of passive particles. This, of course, is of direct relevance to the work reported in Chapter 1.3 of this thesis, and provides further support that intensity fluctuation spectroscopy (or speckle fluctuations) should be a useful tool in the monitoring of biological activity.

CHAPTER 1.3

SPECKLE FLUCTUATIONS WITH BOTANICAL SPECIMENS

1.3.1 - Qualitative Observations of Speckle Fluctuations

After a remark by Abramson (1974a) that the speckle pattern observed when an apple is illuminated with laser light appears to scintillate, and the realisation that this effect may be the source of some of the problems that we were encountering in applying holographic interferometry to living plants, we decided to investigate the properties of these speckle fluctuations.

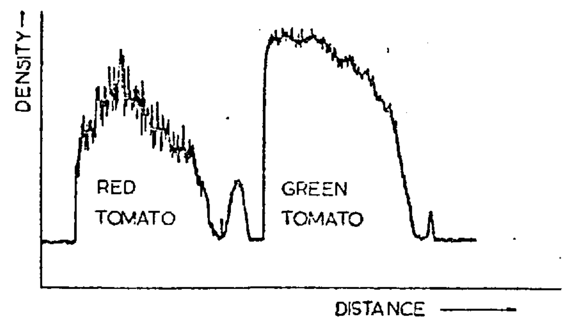
The first simple experiment consisted of placing a red tomato and a red chess piece side by side on a table and illuminating them with light from a helium-neon laser. (These objects were chosen to represent living and inanimate objects because they were approximately the same colour, and had comparable curvatures). It was observed that the speckle pattern on the tomato appeared to be of lower contrast than that on the chess piece. When the scene was observed through a pinhole in order to increase the size of the speckles (see Section 1.1.1), so that they were more easily resolvable by the eye, it could be seen that the contrast in the speckle pattern on the tomato was actually quite high, but that the whole pattern was scintillating, with an apparent time constant of a few seconds. (It should be noted here that the fact that the speckle fluctuations could be followed by the eye in real time immediately places us on a much different time scale from that usually encountered in intensity fluctuation spectroscopy - the speckle fluctuations resulting from Brownian motion, for example, are so rapid that no speckle pattern is visible to the eye when a suspension such as milk is illuminated with laser light).

A hologram taken of the tomato and the chess piece, with an exposure time comparable with the observed time constant of the speckle fluctuations, produced the interesting result of a very good reconstruction of the chess piece but only a hint of an image of the tomato. This was strong supporting evidence that the speckle fluctuations were at the root of some of the problems encountered in applying holographic interferometry to living plants (see Chapter 2.2 of this thesis).

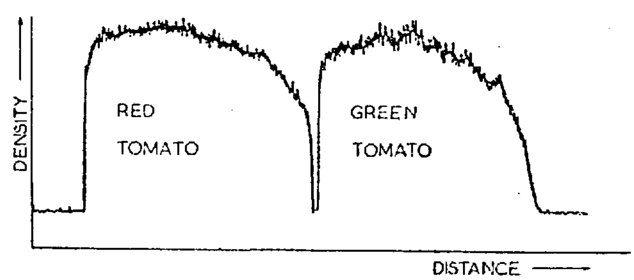
Observations of speckle fluctuations when other botanical specimens were placed in the helium-neon laser beam soon showed that the effect varied from subject to subject, and even from point to point on the same specimen.

1.3.2 - Wavelength Dependence of the Speckle Fluctuations

When the green line of an argon ion laser (wavelength 514nm) was used to illuminate a tomato, it was apparent that the speckle fluctuations were not so pronounced as they were with the helium-neon laser (wavelength 633nm). (In order to eliminate the possibility that the laser light itself was acting as a stimulus for the effect, the observations were made with the specimen illuminated by both lasers at once, the viewing being carried out through appropriate filters - in any case, if energy absorption were to be the cause of the wavelength dependence we should have expected the more violent fluctuations to occur with the green light, since this is absorbed by a red tomato). In order to investigate further this apparent wavelength dependence, a green (i.e. unripe) tomato was obtained and the observations repeated. This time the effect was reversed - the speckle fluctuations were more pronounced in the green light from the argon ion laser. Figure 1.8 shows typical microdensitometer traces taken across photographs of a red and a green tomato placed side by side and illuminated with (a) argon ion laser light and (b) helium-neon laser light (Briers 1975a). The modulation of the traces reflects the contrast of the speckle pattern recorded on the film, and this contrast is in turn a measure of the fluctuation that has occurred



(a)



(b)

Figure 1.8 - Microdensitometer scans of photographs of a red and a green tomato illuminated with (a) green light ($\lambda = 514$nm) from an argon ion laser, and (b) red light ($\lambda = 633$nm) from a helium-neon laser (from Briers 1975a)

in the speckle pattern during the exposure time of the photographs (16 seconds for (a) and 32 seconds for (b)). The smoother the trace in Figure 1.8, the more pronounced is the fluctuation of the speckle pattern.

This wavelength dependence was also observed in far-field speckle patterns, and with other specimens. For example, Figure 1.9 shows typical microdensitometer traces across photographs of far-field speckle patterns (obtained by simply placing the photographic film in the far field, with no camera lens) produced by a single pepper (capsicum) illuminated with laser light. The pepper chosen was partially ripe, and had both green and red areas. The traces of Figure 1.9 are listed in Table 1.3, which also gives values of $\sigma/\langle I\rangle$ calculated from the traces.

TABLE 1.3

Details of the Microdensitometer Traces of Figure 1.9

Trace	Colour of portion of pepper illuminated	Colour and nature of laser light	Exposure time	$\frac{\sigma}{\langle I\rangle}$
(a)	red	red (He-Ne, 633nm)	$\frac{1}{2}$ s	0.07
(b)	red	red (" ")	2 s	0.05
(c)	red	green (Ar, 514nm)	$\frac{1}{15}$ s	0.48
(d)	red	green (" ")	4 s	0.39
(e)	green	green (" ")	$\frac{1}{15}$ s	0.18
(f)	green	green (" ")	4 s	0.14
(g)	green	red (He-Ne, 633nm)	$\frac{1}{2}$ s	0.21
(h)	green	red (" ")	2 s	0.12

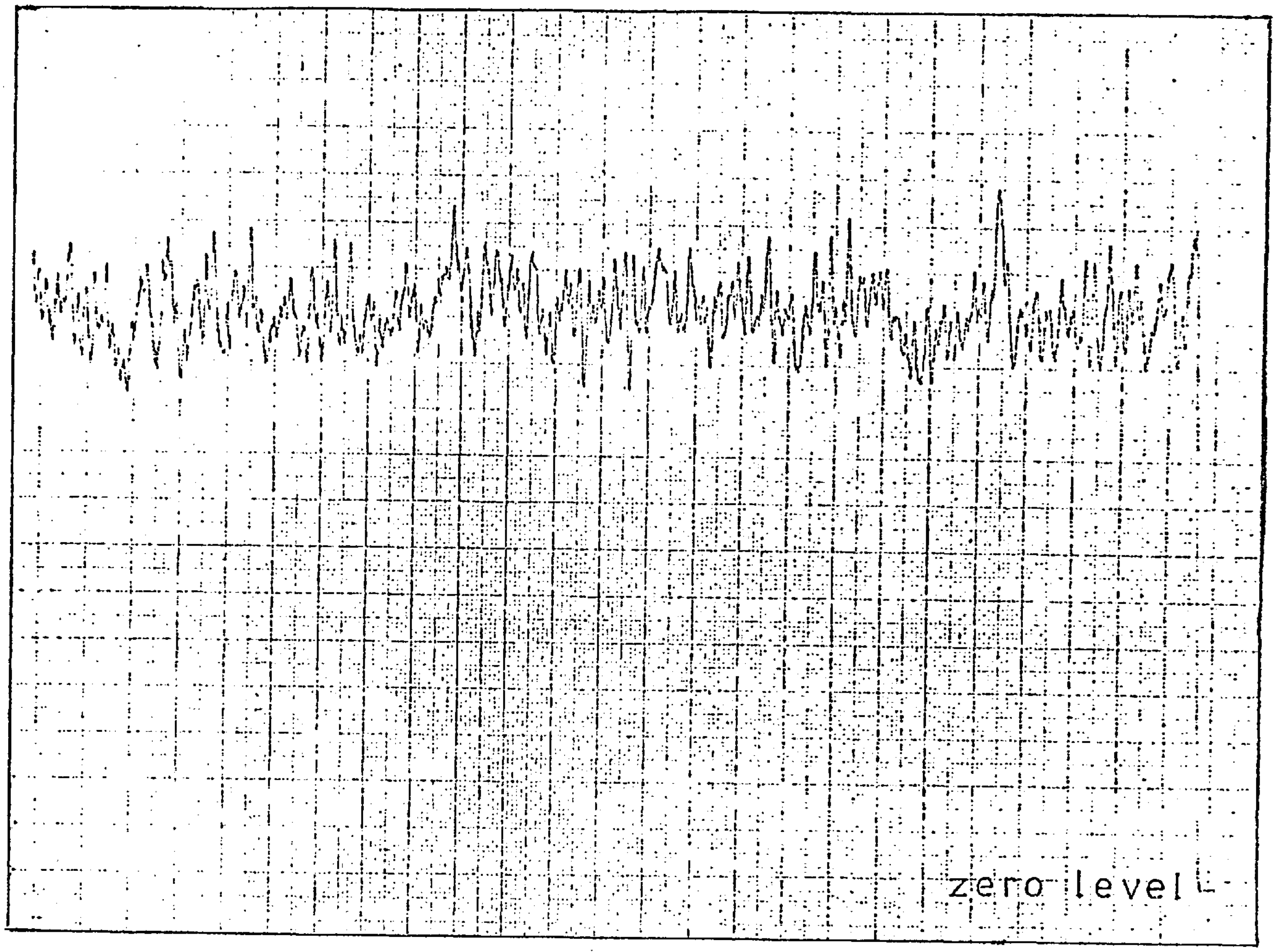


Figure 1.9 - Microdensitometer scans across photographs of far-field speckle patterns produced by a pepper (capsicum) :

(a) Colour of illuminated area = red
Source = helium-neon laser, 633nm
Exposure time = 0.5 seconds.

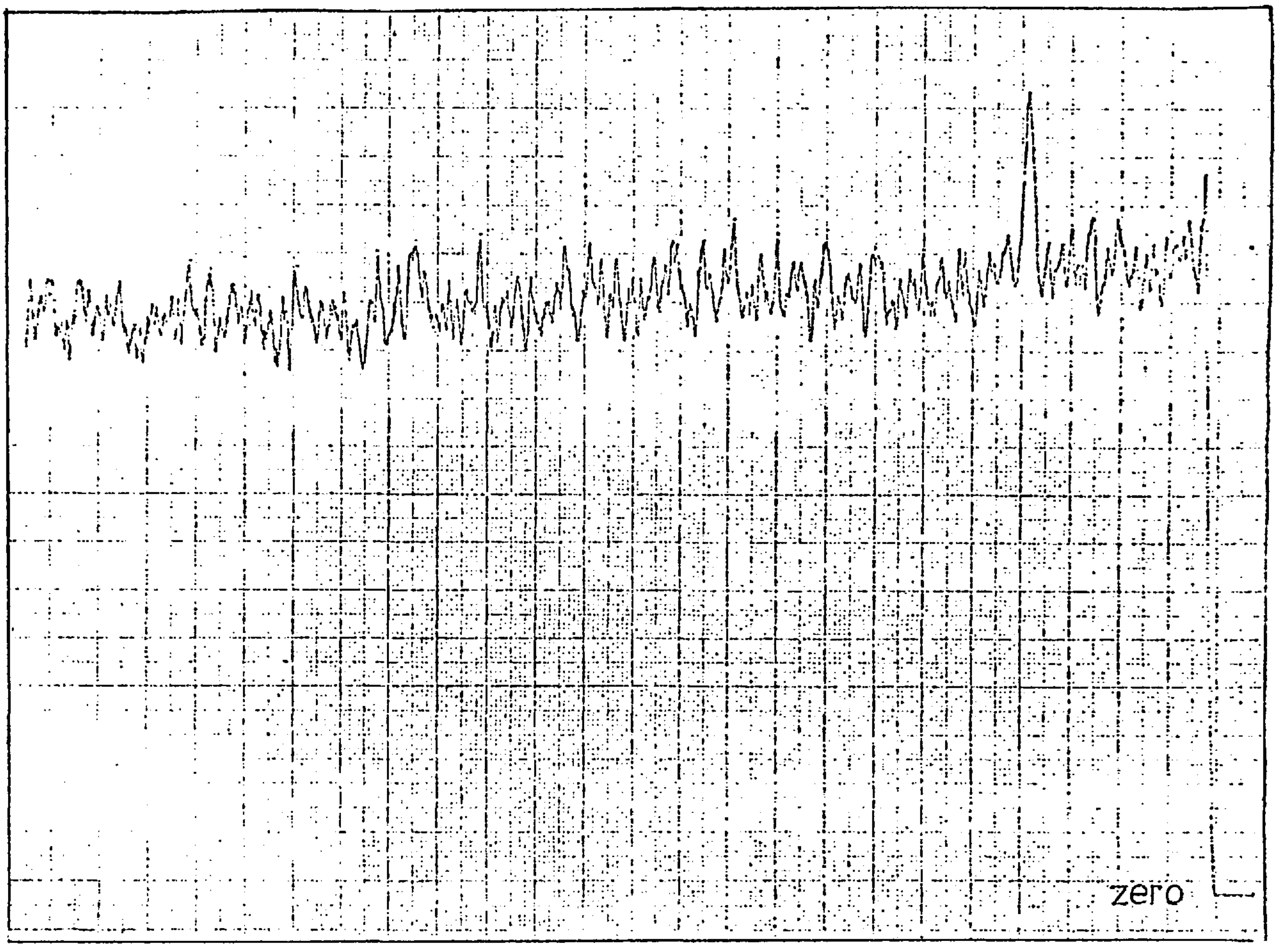


Figure 1.9 (b) - Colour = red
Source = He-Ne, 633nm
Exposure = 2 seconds.

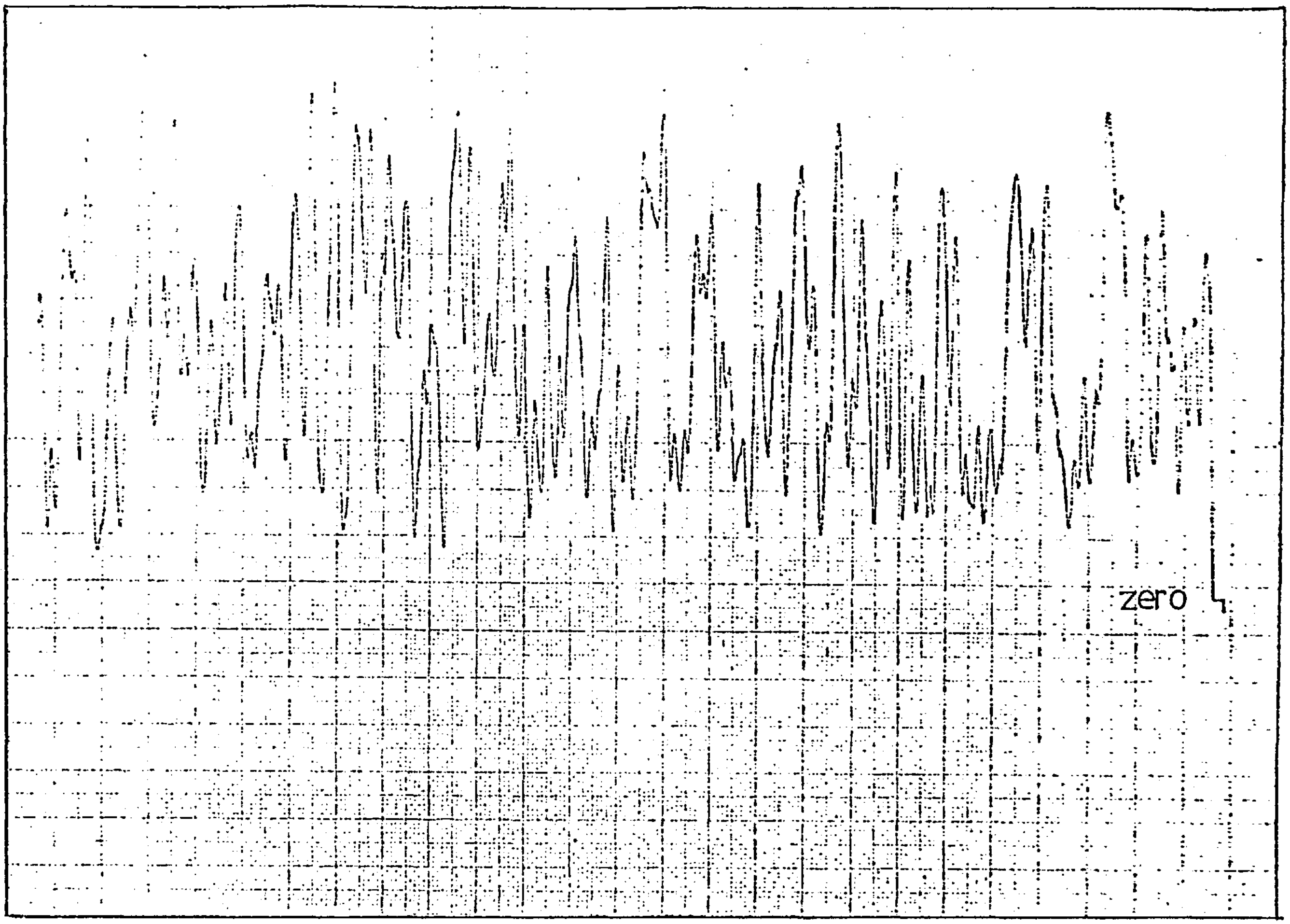


Figure 1.9 (c) - Colour = red
Source = argon ion laser, 514nm
Exposure = 0.07 seconds.

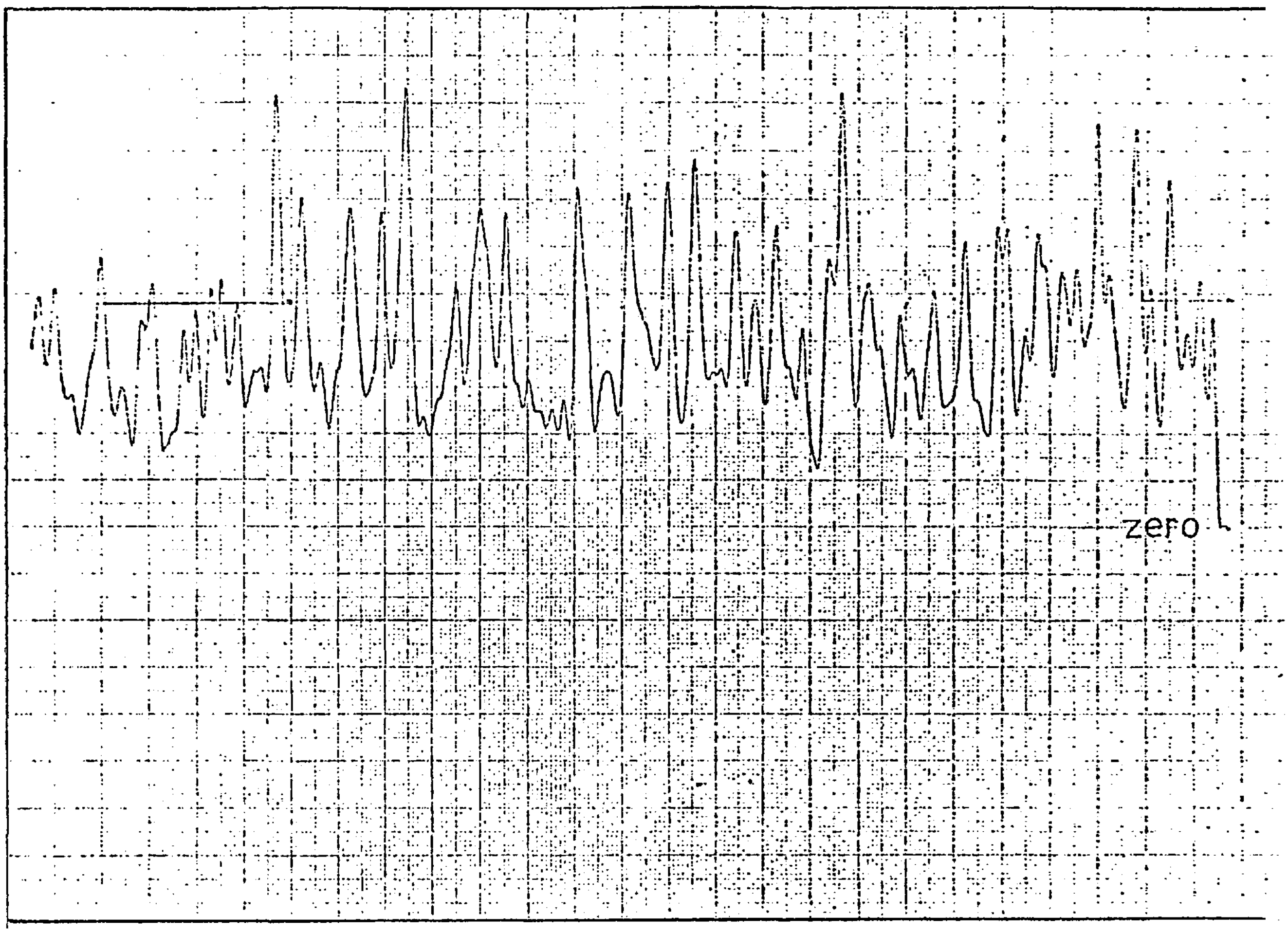


Figure 1.9 (d) - Colour = red
Source = Ar, 514nm
Exposure = 4 seconds.

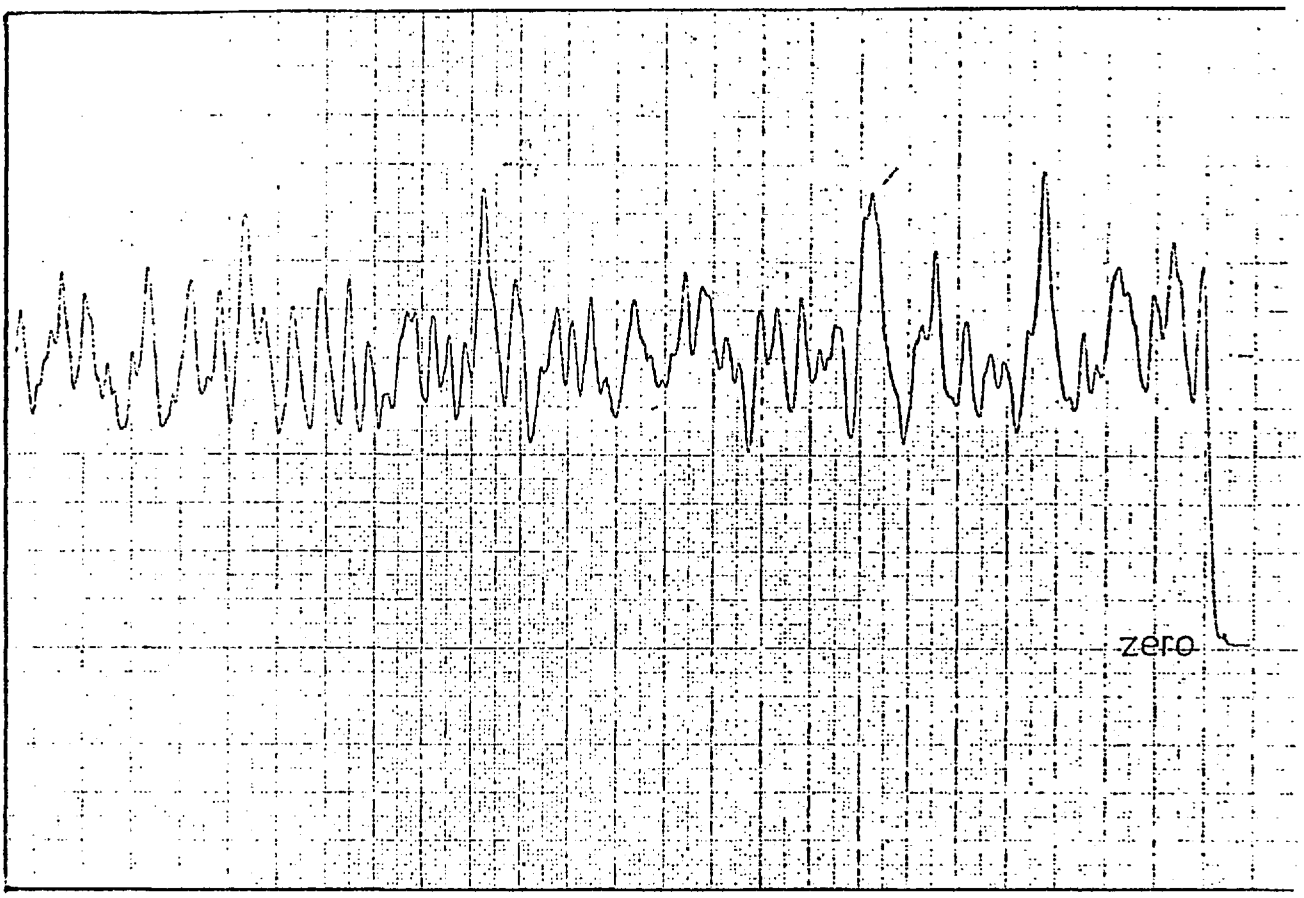


Figure 1.9 (e) - Colour = green
Source = Ar, 514nm
Exposure = 0.07 seconds.

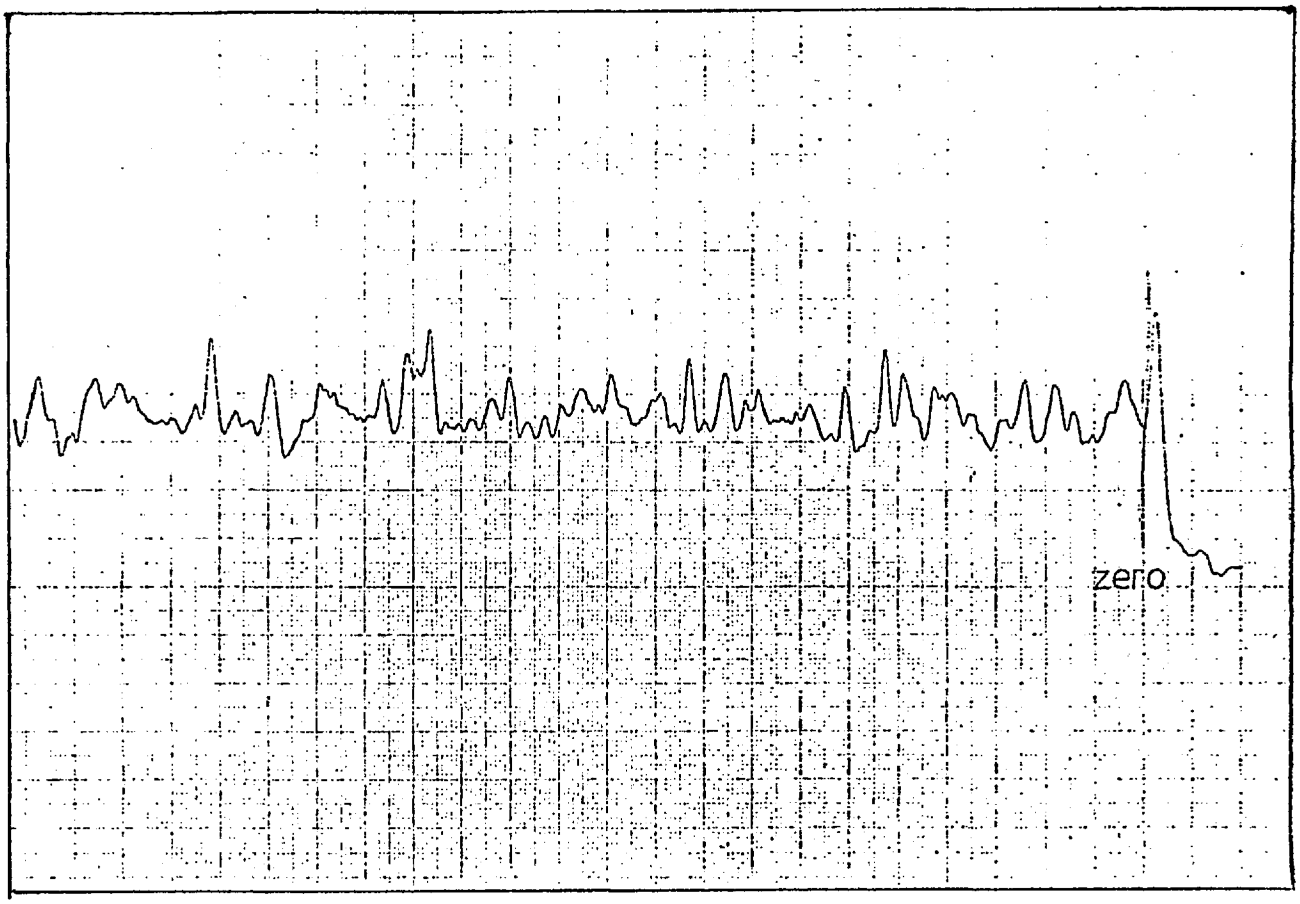


Figure 1.9 (f) - Colour = green
Source = Ar, 514nm
Exposure = 4 seconds.

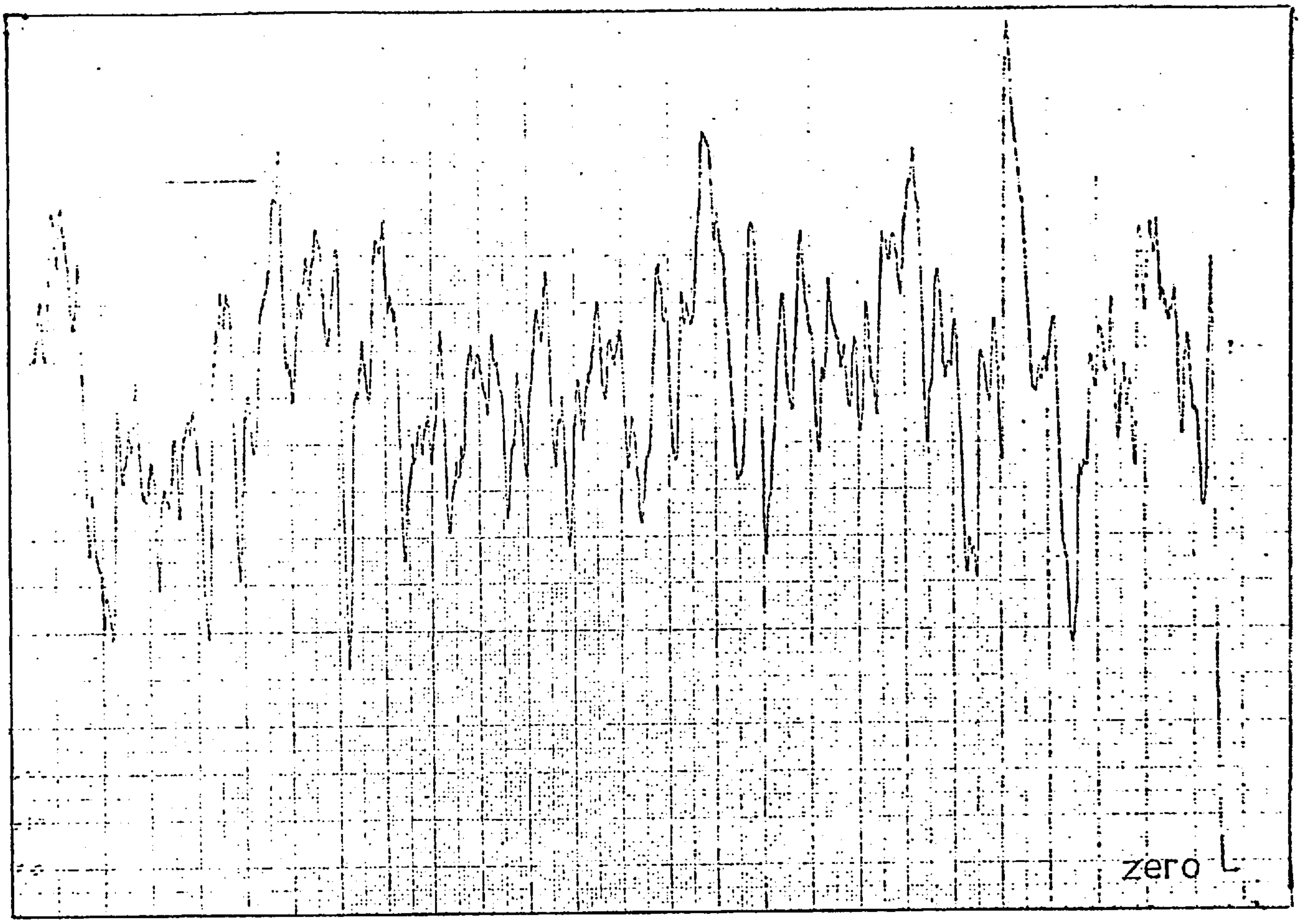


Figure 1.9 (g) - Colour = green
Source = He-Ne, 633nm
Exposure = 0.5 seconds.

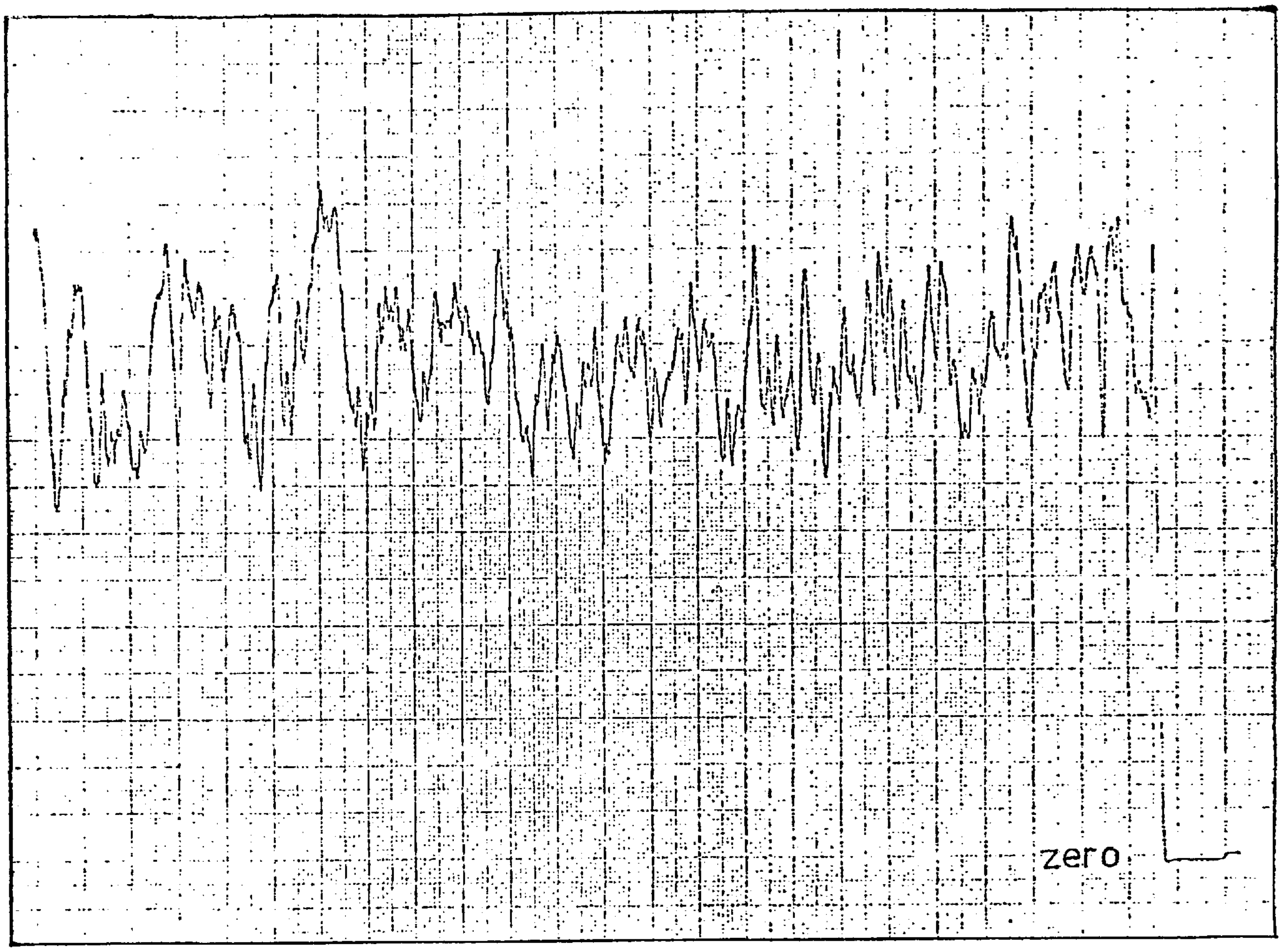


Figure 1.9 (h) - Colour = green
Source = He-Ne, 633nm
Exposure = 2 seconds.

It should be pointed out that the values of σ / I in Table 1.3 are all on the low side - a result of the finite scanning aperture of the microdensitometer and (perhaps) of the smoothing effect of the photographic process. This was emphasised by measurements taken on a photograph of a stationary, inanimate object (trace not illustrated). The theoretical value of σ / I in this case should be 1.0 (see Section 1.1.3), but was, in fact, 0.38. However, we are not interested at present in absolute measurements of speckle statistics, but merely in comparing the statistics for the different combinations of conditions listed in Table 1.3. From this table the following trends are apparent:

(i) the contrast, given by $\sigma / \langle I \rangle$, decreases as the exposure time is increased;

(ii) $\sigma / \langle I \rangle$ is higher when the colour of the laser light is complementary to that of the specimen, and lower lower when the colour of the light is the same as that of the specimen.

(The second property above is very apparent for the red portion of the pepper, but less so for the green; also, we have the anomalous result of $\sigma / \langle I \rangle$ for two of the traces ((c) and (d)) being higher than that obtained for the inanimate object. These discrepancies are probably due partly to the relatively low number of sampling points used in the calculation of the $\sigma / \langle I \rangle$ ratios (250), and partly to the crude nature of the experiment; more carefully controlled experiments are required in order to confirm the trends recorded above).

1.3.3 - Angular Dependence of the Speckle Fluctuations

Superimposed on the wavelength dependence of the speckle fluctuations was an angular dependence. The fluctuations were observed to be much less pronounced in the neighbourhood of the "highlights" caused by specular reflection. This is illustrated in Figure 1.10, in which the microdensitometer trace has been

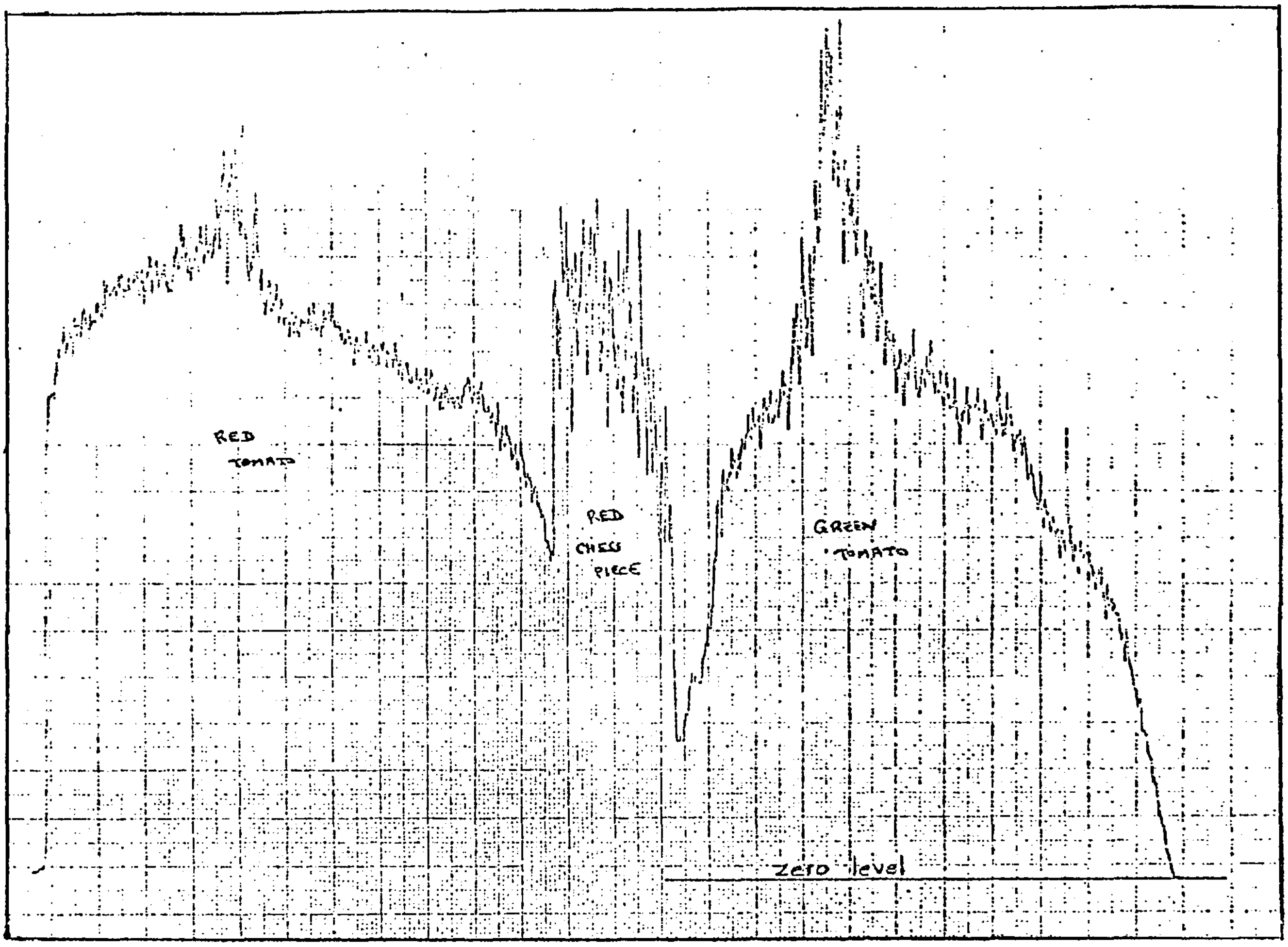


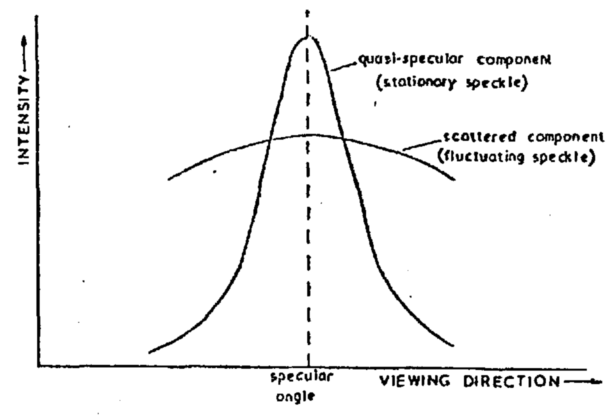
Figure 1.10 - Microdensitometer scan across a photograph of a red and a green tomato, including specular highlights; a red chess piece is also shown, for comparison (source = helium-neon laser, 633nm exposure = 64 seconds at f/22).

arranged to pass across these highlights in a photograph of a red and a green tomato illuminated with red light from a helium-neon laser. (The exposure time of the photograph was 64 seconds, and a red chess piece was included in the photograph for comparison). It can be seen that the modulation of the trace is more pronounced in the neighbourhood of the highlights (indicated by peaks in the trace), which in turn means that the time-averaged speckle in these areas is of higher contrast than that in the remainder of the image, and hence that the temporal fluctuations in these areas are less pronounced.

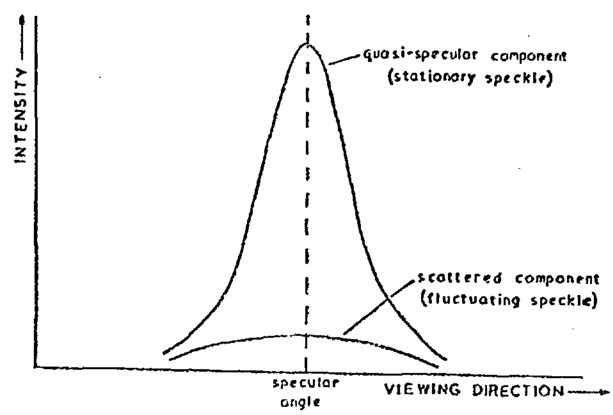
1.3.4 - A Proposed Model to explain the Wavelength and Angular Dependence of the Speckle Fluctuations

A tentative explanation of the phenomena described in the preceding sections has been formulated (Briers 1975a), based on the fact that the colour of a tomato, for example, is due to the presence of discrete, pigmented bodies (plastids) that selectively scatter light of that colour. When a red tomato is illuminated with red light, most of the light is scattered from red chromoplasts inside the cells of the tomato. When green light is used, this is absorbed by the chromoplasts, and most of the light reaching an observer or a detector arises from specular reflection at the skin of the tomato. For a green tomato the situation is reversed: the green chloroplasts within the cells preferentially absorb red light and scatter green. Thus it is suggested that light from an illuminated tomato consists of two components, a component scattered from the plastids, and a quasi-specular component reflected from the skin. The specular component will show much more angular dependence than the scattered component, as illustrated in Figure 1.11.

This model can be used to explain all the phenomena described in the preceding sections, if it is further assumed that the plastids are in motion and that the skin of the tomato is stationary. This would result in fluctuating speckle in the light scattered by the plastids (due to time-variations in the optical path lengths from plastid to detector), and stationary speckle in



(a)



(b)

Figure 1.11 – Suggested explanation of the wavelength and angular dependence of the speckle fluctuations:
(a) colour of specimen = colour of laser light;
(b) colour of specimen complementary to colour of laser light

(from Briers 1975a).

the specular component. When the colour of the tomato is the same as the colour of the light, most of the observed speckle pattern will be due to the scattered light from the plastids, and will hence show fluctuations; in one particular direction, however, the specular component will be predominant (Figure 1.11(a)) and the fluctuations less pronounced. When the colour of the fruit is complementary to that of the laser light, the specular component, and hence stationary speckle, will always predominate (Figure 1.11(b)).

No movement of plastids has been observed during microscopic examination of tomato sections, but it can be argued that the taking of a section interferes with the life processes and might arrest any motion. In fact, motion of chloroplasts (cyclosis) has long been established for thin biological specimens which can be examined in vivo under a microscope, the classic example being Canadian pondweed (Elodea canadensis). When a leaf of this plant, mounted in water on a microscope slide, was illuminated with the argon ion laser beam, the far-field speckle pattern showed fluctuations whose intensity varied as the beam was scanned across the leaf, the fluctuations being most pronounced near the central vein of the leaf. Microscopic examination of the same leaf revealed that the chloroplasts in the cells near the central vascular tissue were much more active than those nearer the edge of the leaf. The same leaf was examined some time later, when it was found that all cyclosis had ceased; the far-field speckle pattern was again observed, and this time showed no signs of any fluctuations.

Order-of-magnitude measurements also tend to support the validity of the model. The chloroplasts of Elodea canadensis were observed under the microscope to be moving with velocities varying between zero and $4\mu m\ s^{-1}$, with a heavy bias towards the lower end of this range. A rough estimate put the mean velocity between 0.1 and $0.4\mu m\ s^{-1}$. The intensity fluctuations of a single speckle were observed, and the time constant of the fluctuations was of the order of five seconds. This can be associated, to an order-of-magnitude approximation, with an average movement of the scattering particles of one wavelength, and hence with an average velocity of about $0.1\mu m\ s^{-1}$. (It should be pointed out, however, that the microscope observations give an estimate of the transverse component of the velocity, while the speckle fluctuations are caused by the line-of-sight component). Further, the intensity envelope of the speckle pattern is determined by the size of the scattering particles, and diffraction theory gave the size in the present case as approximately 8μm. Microscopic measurements gave the diameters of the chloroplasts as 4 to 6μm.

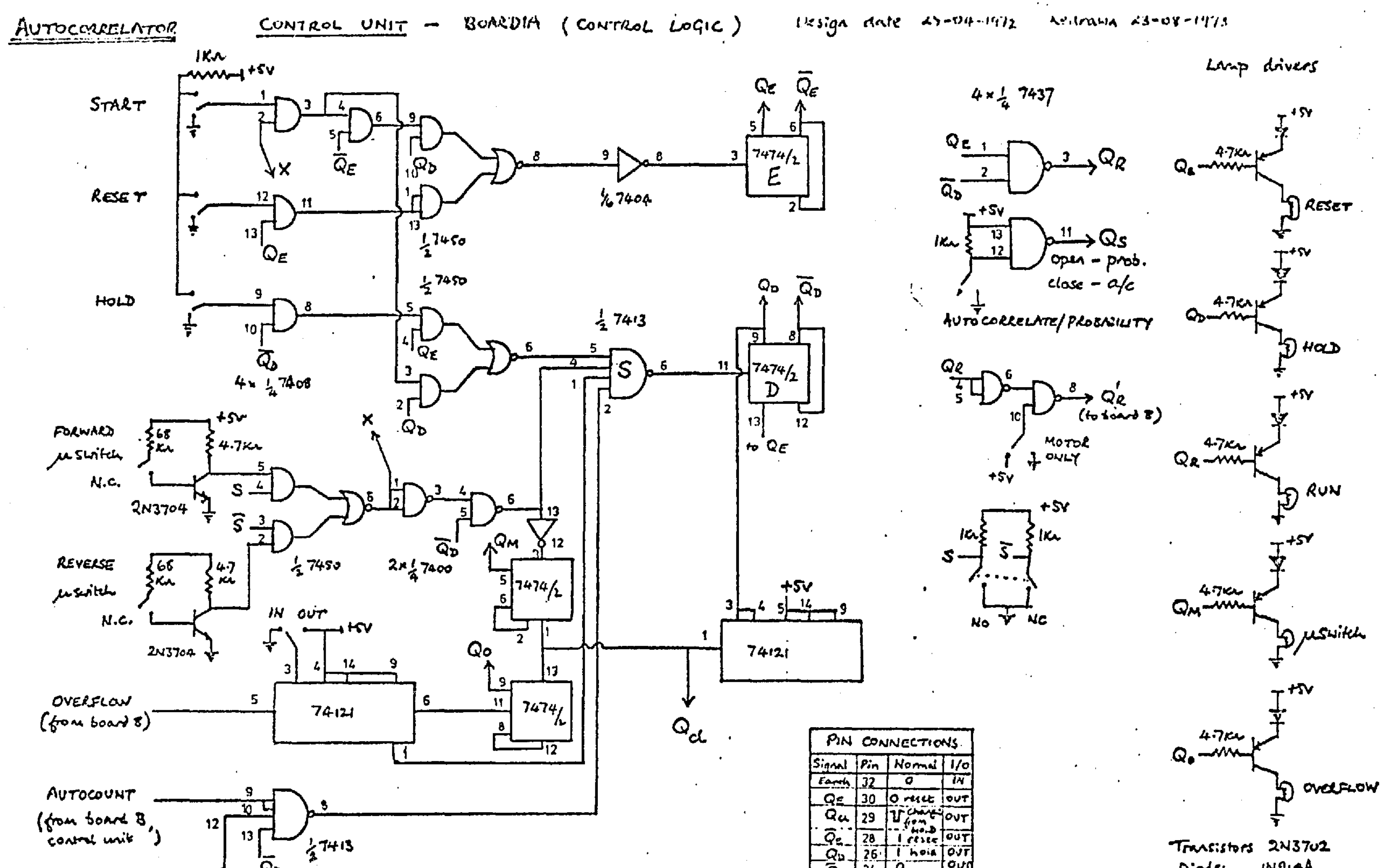
If the above interpretation of the phenomena is correct, an analysis of the statistics of the speckle fluctuations should yield information about the motion of particles inside the cells of living organisms. Such analysis could be carried out, at increasing levels of sophistication, by (i) a study of the time-averaged speckle patterns discussed in Sections 1.1.5, 1.3.2 and 1.3.3; (ii) by a study of the first order statistics of the fluctuations, as described in Section 1.1.6 and to be mentioned again in Section 1.3.6; (iii) by a study of the second order statistics of the fluctuations such as will be discussed in Sections 1.3.7 and 1.3.8; and (iv) by the use of the intensity fluctuation spectroscopy techniques described in Chapter 1.2.

1.3.5 - Recording and Analysing the Speckle Fluctuations using a Double-Clipped Autocorrelator

It will be obvious* from Chapter 1.2 and Section 1.3.4 that the ideal method of analysing the speckle fluctuations observed when laser light is scattered from biological specimens would be to use a single-clipped autocorrelator of the type described by Pike and Jakeman (1973) and the techniques of intensity fluctuation spectroscopy (Cummins and Pike 1973). Unfortunately, such an instrument was not available, and the nearest approach to it was a double-clipped autocorrelator which had been designed by Dr J C Dainty. The circuit diagram of this instrument is shown in Figure 1.12. It will be remembered from Section 1.2.3 that the double-clipped autocorrelator is not a very suitable tool for

---

* It would appear from recent discussions with Dr M Corti (1975), of the Centro Informazioni Studi Esperienze Segrate (Milan), that perhaps this is not so "obvious" after all - Dr Corti warned us of difficulties in using fast autocorrelators to analyse slow fluctuations.



STATUS	Q_E	$\overline{Q_E}$	Q_D	$\overline{Q_D}$
RUN	1	0	0	1
HOLD	1	0	1	0
RESET	0	1	0	1

PIN CONNECTIONS			
Signal	Pin	Normal	I/O
Earth	32	0	IN
Q_E	30	0 reset	OUT
Q_a	29	⊔ change from hold	OUT
$\overline{Q_E}$	28	1 reset	OUT
Q_D	26	1 hold	OUT
$\overline{Q_D}$	24	0	OUT
Q_R'	22	0 run	OUT
Pr/Ac	20	1 A/C	OUT
Awrec	18	⊓	IN
Overflow	3	⊓	IN
+5V	1	1	IN

Figure 1.12 - Circuit diagram for double-clipping autocorrelator:

(a) Board 1A - control logic.

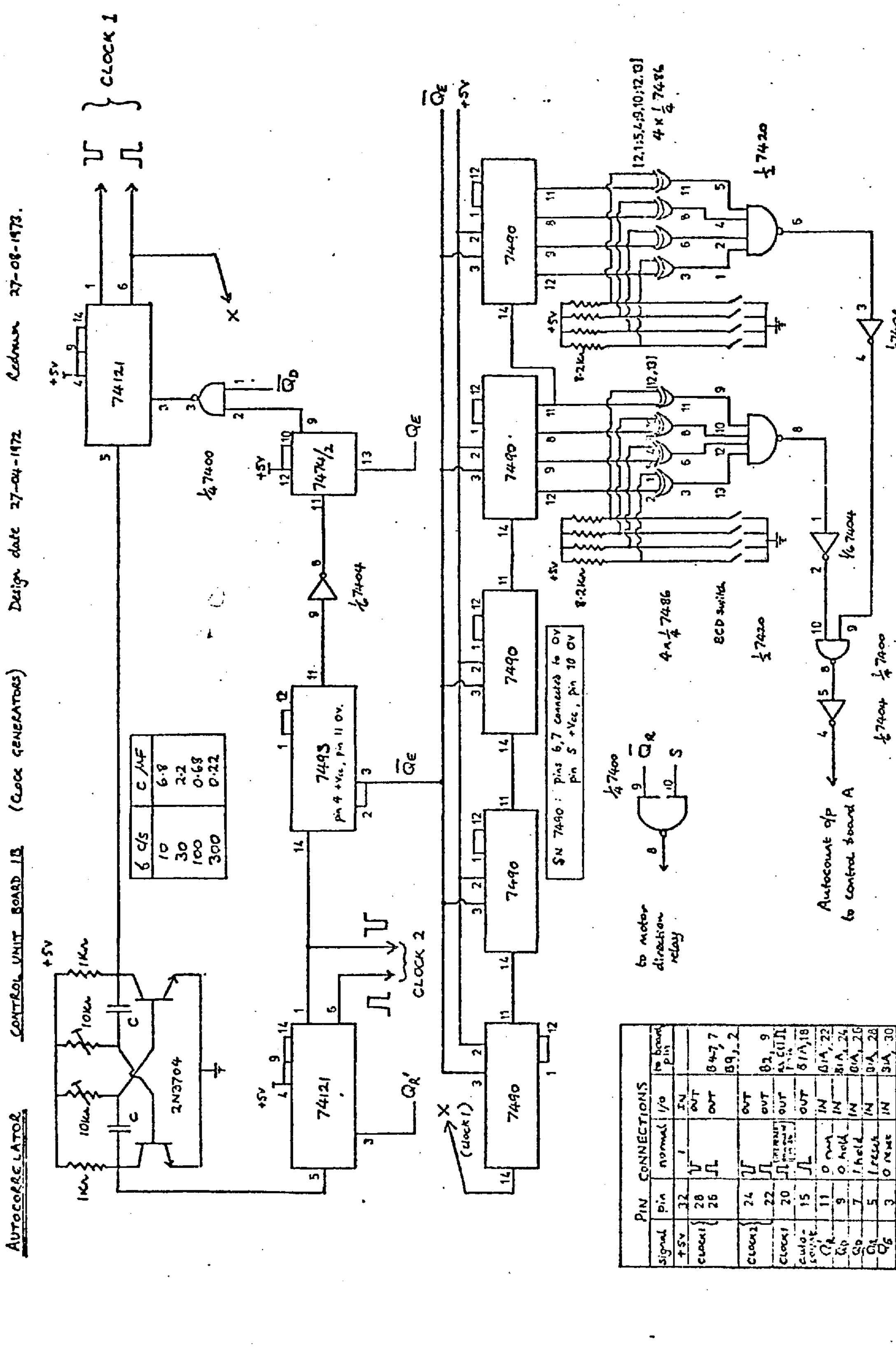


Figure 1.12 (b) - Board 1B (clock generator) of autocorrelator.

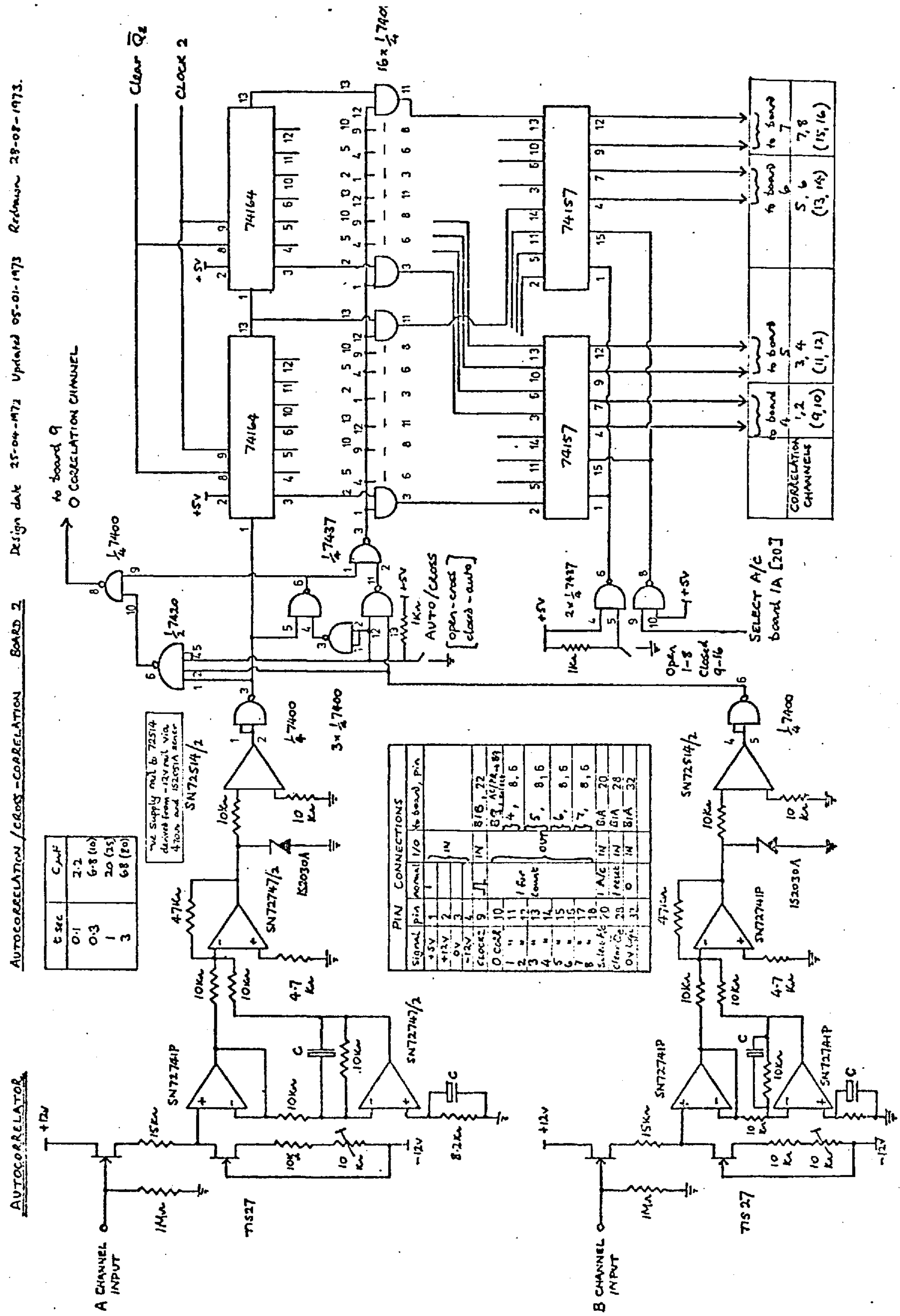


Figure 1.12 (c) - Board 2 (autocorrelation/cross-correlation) of autocorrelator.

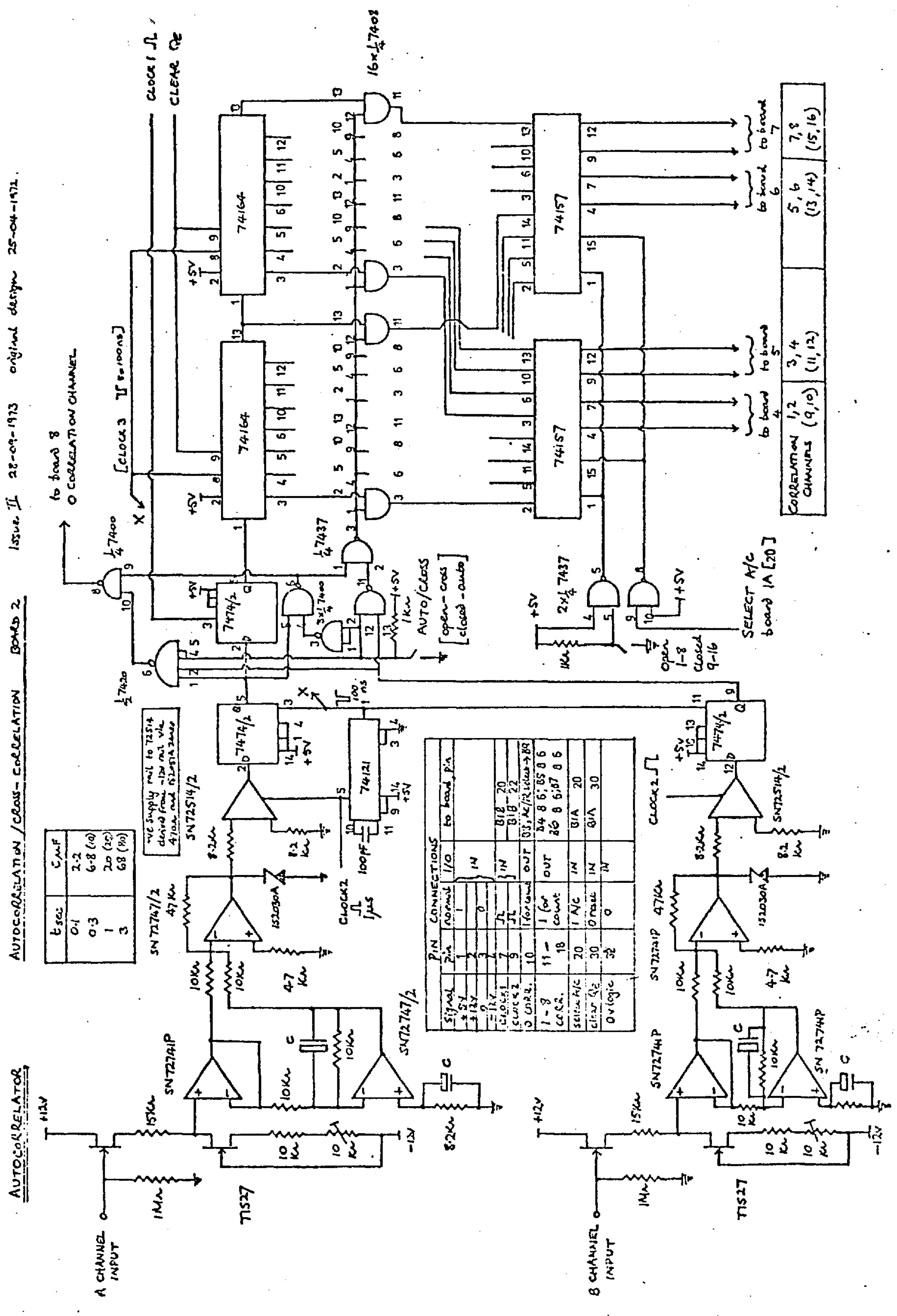


Figure 1.12 (d) - Board 2 (revised version) of autocorrelator.

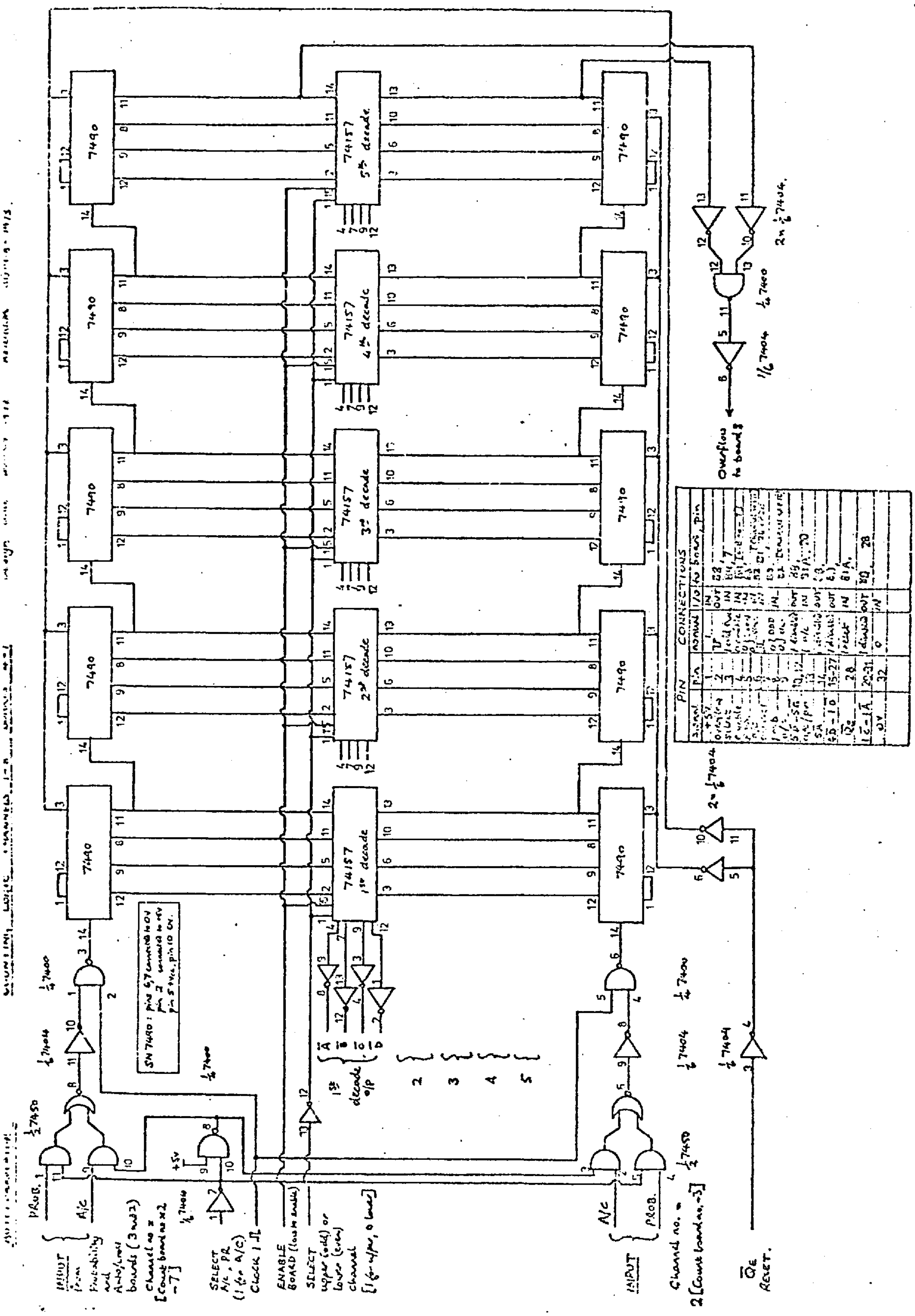


Figure 1.12 (e) - Boards 4 - 7 (counting logic) of autocorrelator.

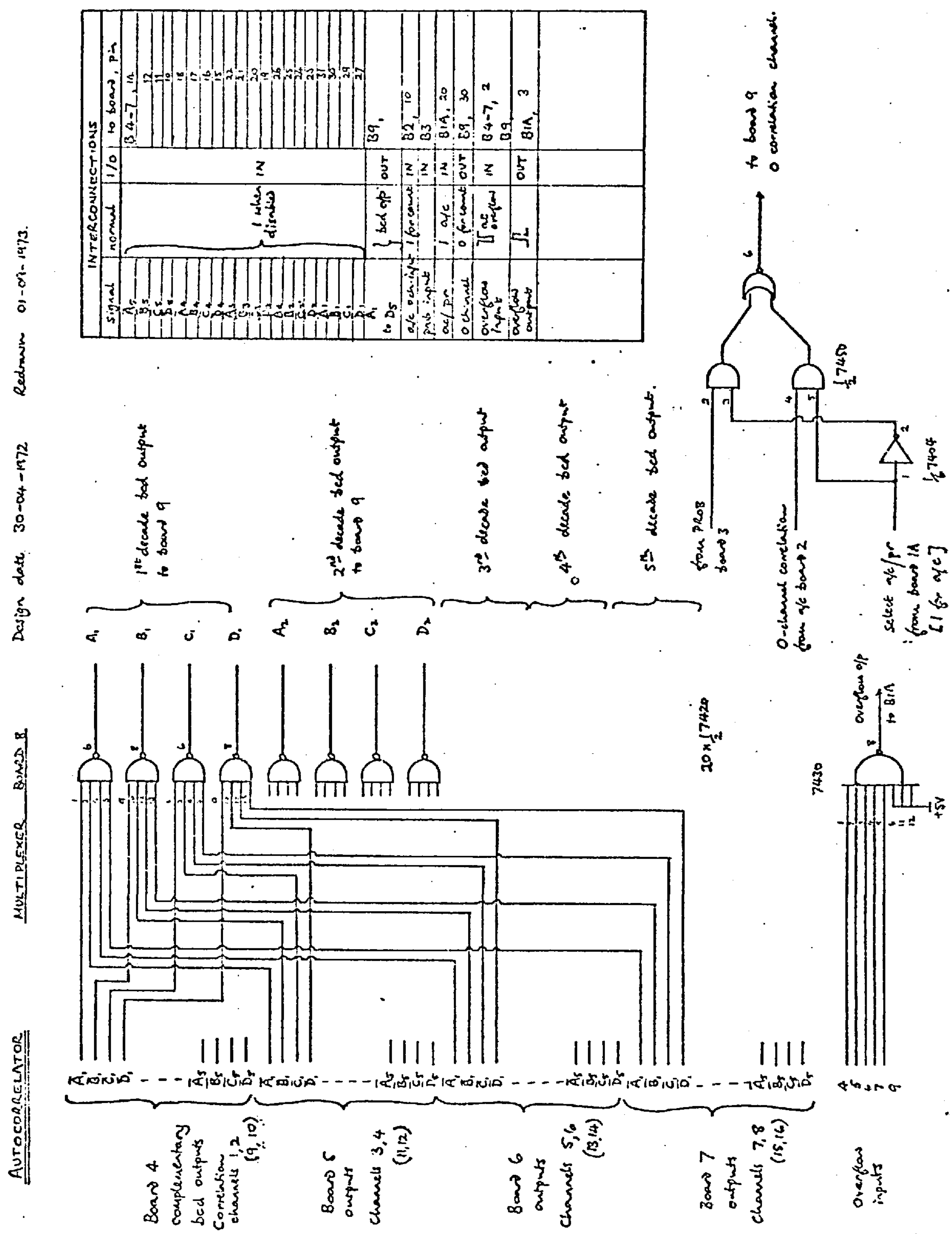


Figure 1.12 (f) - Board 8 (multiplexer) of autocorrelator.

analysing intensity fluctuations, but we decided to carry out some experiments with the instrument in order to show the general nature of results obtained with fast autocorrelators.

The experimental arrangement consisted simply of a photomultiplier placed in the far field of a specimen illuminated by an undiverged laser beam, with its output fed directly to the autocorrelator. A pinhole, whose aperture was much smaller than the average speckle size, was placed in front of the photomultiplier to ensure that the intensity of a single speckle was being monitored.

Figure 1.13 shows the averaged results of several experiments carried out using a green tomato as the specimen and the green line (514nm) of an argon ion laser as the illuminating source. The readings from the autocorrelator have been normalised to give $g_{dc}^{(2)}(\tau)$, and we have plotted the function $g_{dc}^{(2)}(\tau) - 1$ against the lag τ, following the usual practice in intensity fluctuation spectroscopy. (It will be remembered from Section 1.1.3 that this function is the one that we have called the normalised autocovariance, $c_{dc}^{(2)}(\tau)$. It can also be seen from Table 1.1 in that section that this function should tend to zero at infinite lag. The slight discrepancy between this theoretical limit and the experimental curve can probably be attributed to an error in setting the clipping level of the autocorrelator; this clipping level should be set accurately to the mean signal level if the correct limiting value at infinite lag is to be obtained).

The decorrelation time, taken as the 1/e point on the curve of Figure 1.13, is approximately 0.07 seconds. This is very much on the low side, compared with both the visual estimates given at the beginning of this chapter (Section 1.3.1) and the experiments to be described later, and the discrepancy is probably due to the distortions introduced by taking the double-clipped autocorrelation function (see Section 1.2.3). Figure 1.13 is included in this thesis merely to show the type of curve that is obtained from digital autocorrelators.

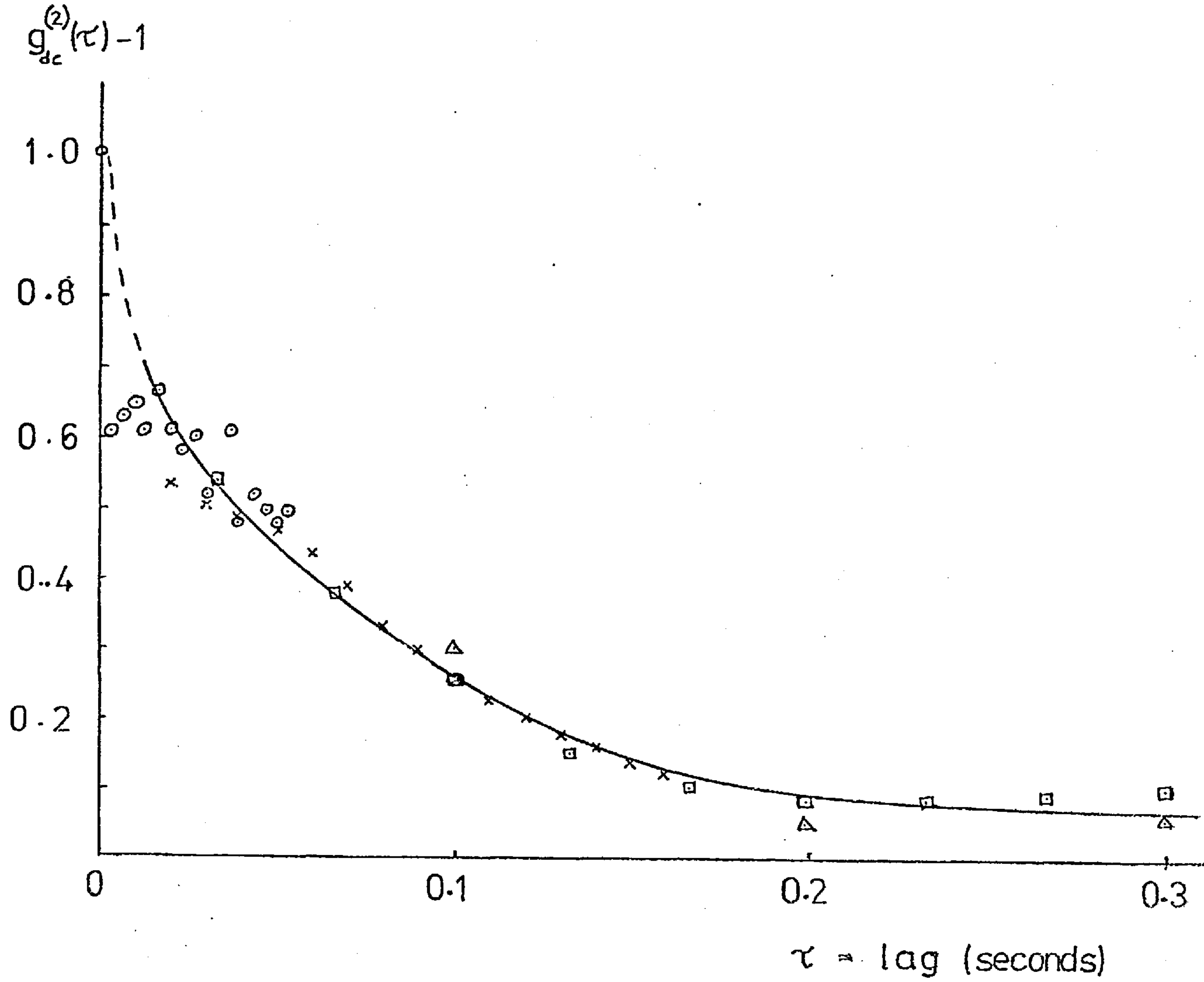


Figure 1.13 - Double-clipped normalised autocovariance curve for a green tomato: averaged results of several runs with different sample lengths (source = argon ion laser, 514nm).

1.3.6 - Recording and Analysing the Speckle Fluctuations using an Oscilloscope: $\sigma/\langle I\rangle$ Measurements

Since the double-clipped autocorrelator does not give full information about the autocorrelation function of the unclipped signal, and since the preferred single-clipped autocorrelator was not available, we decided to carry out some experiments on far-field speckle fluctuations by recording the time-varying signal from a photomultiplier by means of an oscilloscope and a camera. The arrangement is illustrated in Figure 1.14. The output from the photomultiplier, placed at a point in the far-field speckle pattern produced by a specimen was fed via a variable filter to the oscilloscope. The capacitance of the filter could be adjusted to filter out fluctuations above a predetermined frequency. The oscilloscope traces were recorded photographically by means of a 35mm camera. Typical traces are illustrated in Figures 1.15 to 1.19.

Once the traces are recorded, it is possible, in principle, to obtain both the first and the second order statistics of the fluctuations. This, however, requires some method of extracting the data from the oscillographs. Since the only method that we had available for doing this was the very tedious and time-consuming one of reading off the data point by point, a full analysis, leading to autocorrelation curves, was only carried out for a few of the traces. It is possible, however, to obtain an estimate of the standard deviation, σ, and the mean, $\langle I\rangle$, of the traces by adopting an approximate method used in industry for quality control purposes. The method is based on the property of a Gaussian distribution which states that 68% of the members of such a population will lie within one standard deviation of the mean. Applying this property to a batch of data, assumed to follow a Gaussian distribution, the method consists of discarding the upper 16% and the lower 16% of the data, and taking the difference between the highest and the lowest values remaining as twice the standard deviation, and the mean of these two numbers as the mean of the population. In our case, the method is applied by laying a straight-edge across the trace, parallel to the time axis, and adjusting its

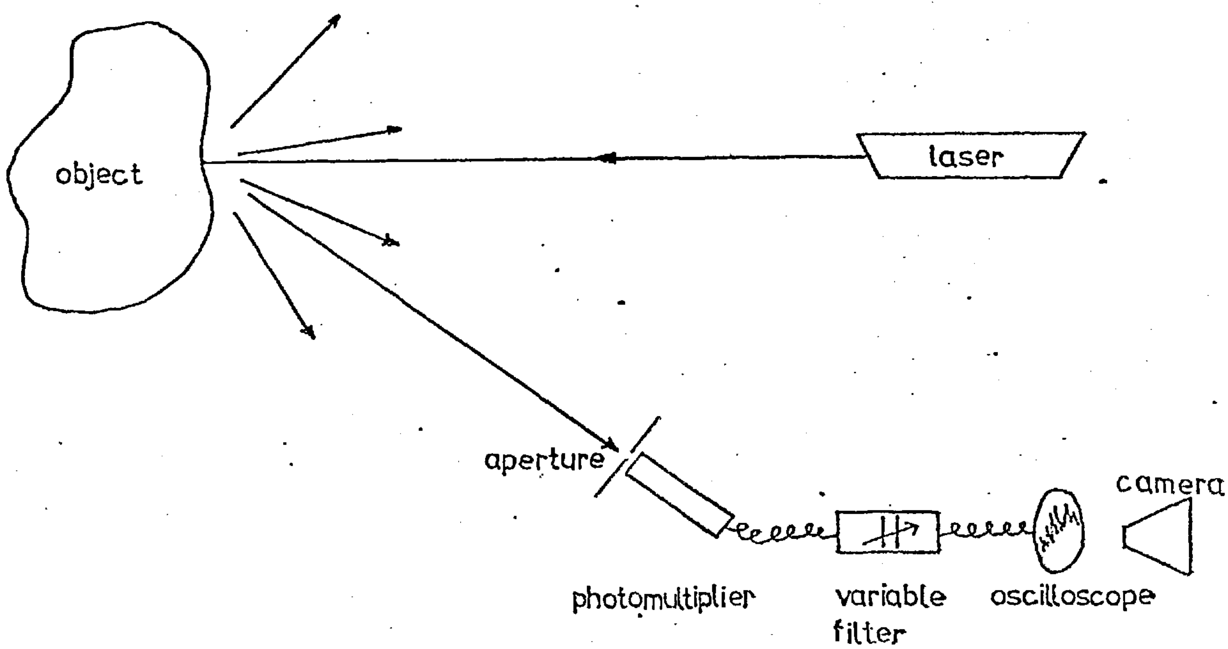


Figure 1.14 - Experimental arrangement for monitoring intensity fluctuations in the far-field speckle pattern.

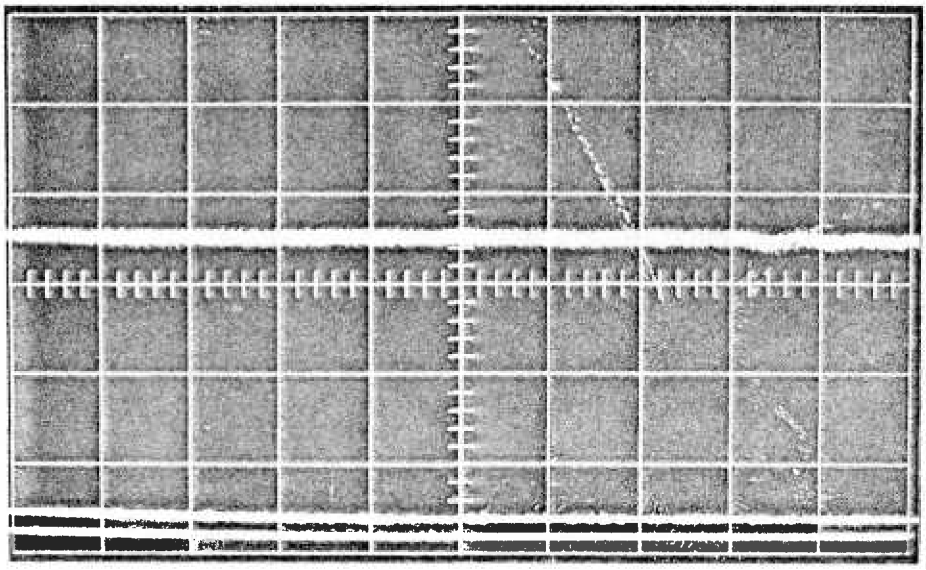
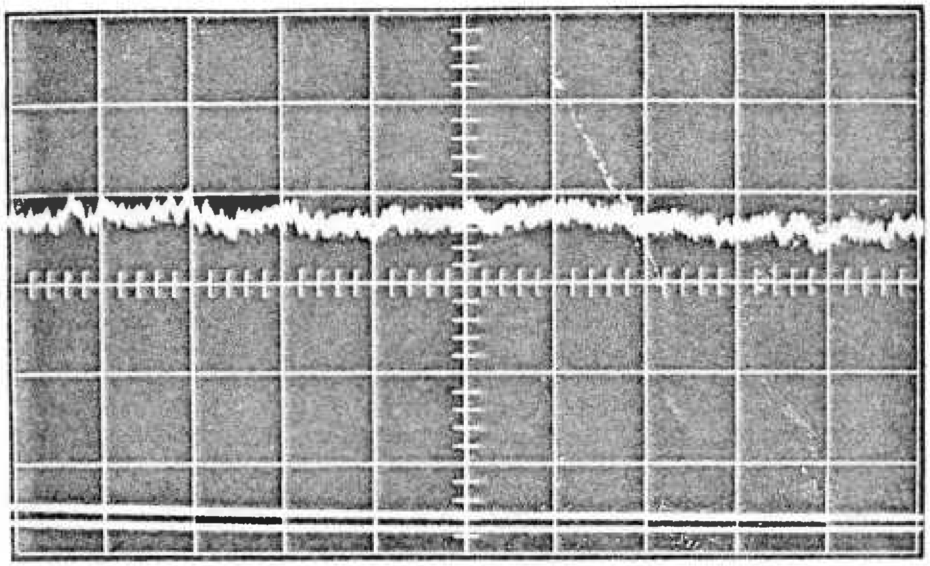
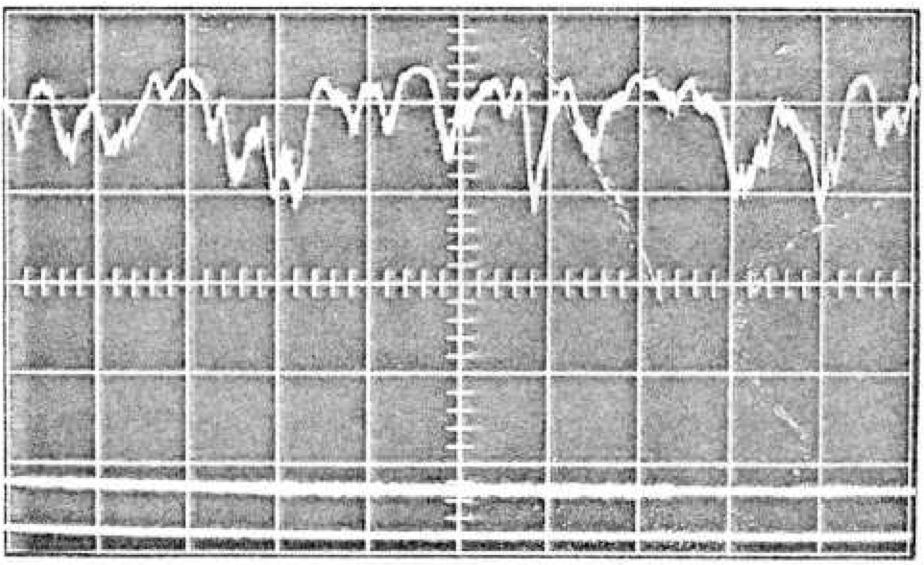


Figure 1.15 - Oscilloscope trace of the intensity at a point in the far-field speckle pattern produced by a stationary metal block (for comparison purposes)

(time base = 8 seconds, full scale; source = argon ion laser, 514nm).



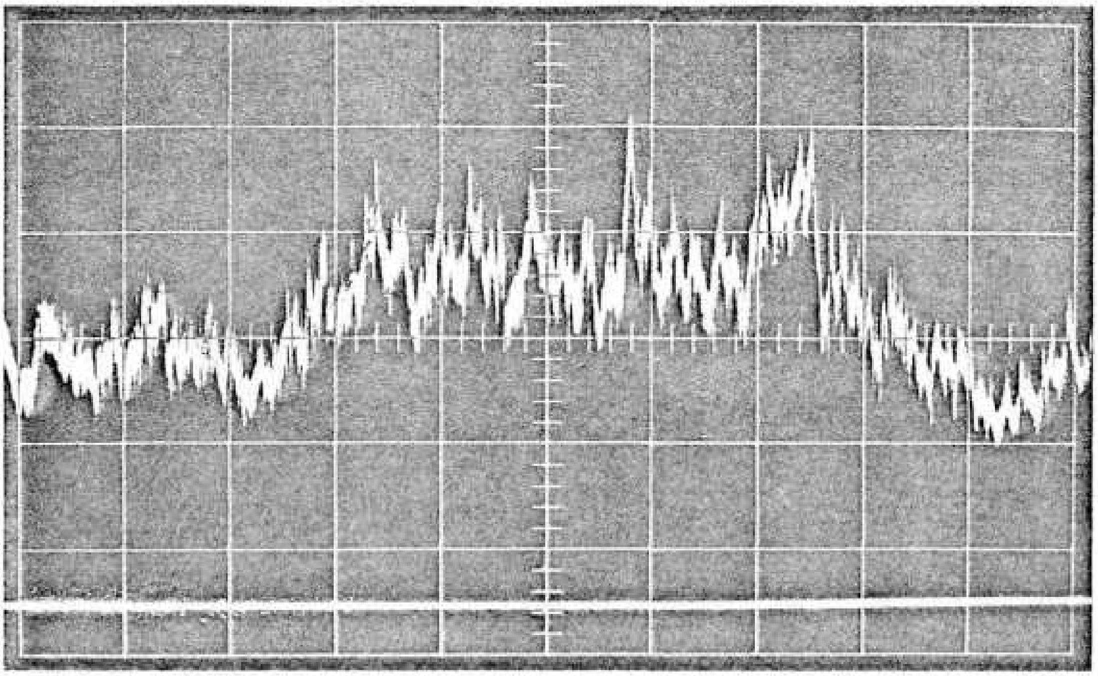
(a)



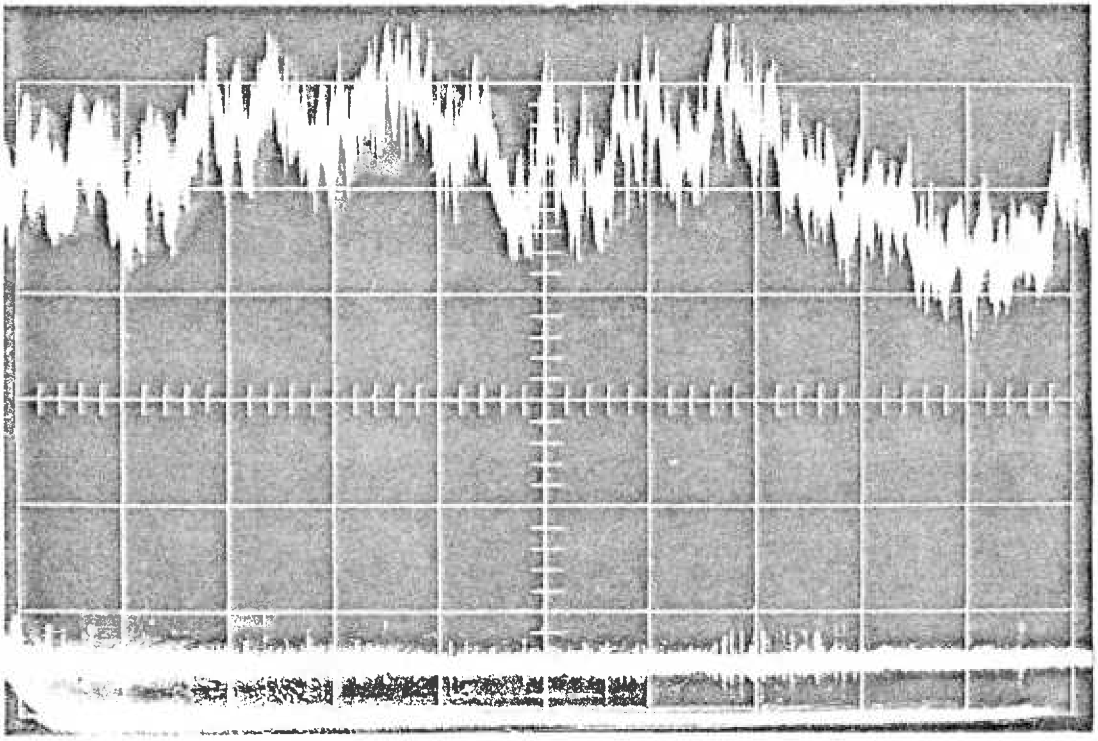
(b)

Figure 1.16 - Oscilloscope traces from tobacco, showing differences between different parts of the same specimen: (a) stem; (b) leaf

(time base = 8 seconds, full scale; source = argon ion laser, 514nm).



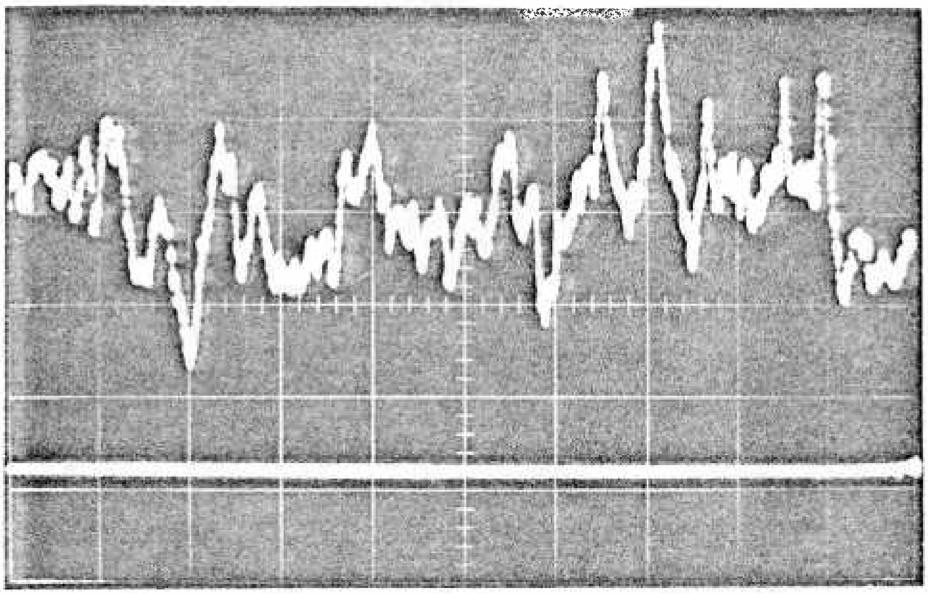
(a)



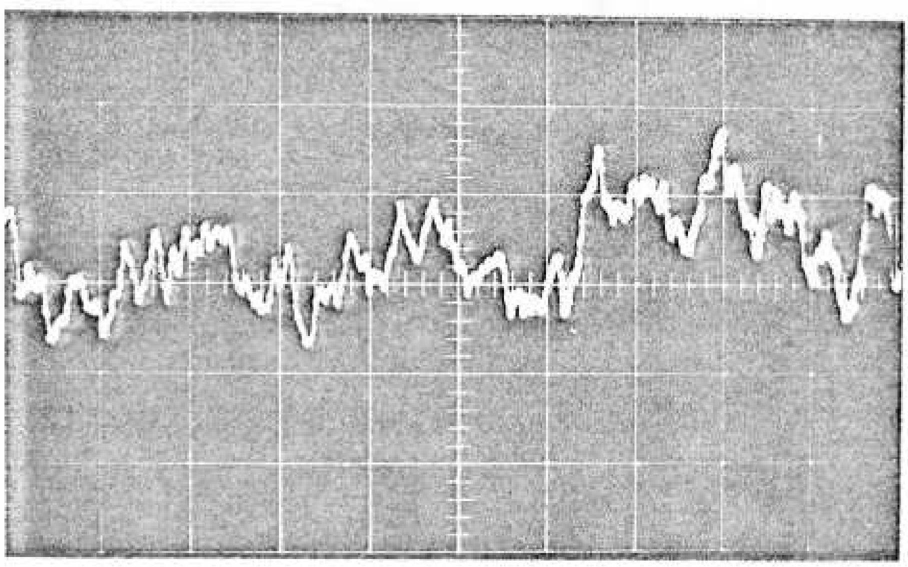
(b)

Figure 1.17 - Oscilloscope traces from Elodea canadensis: (a) near the central vein; (b) mid-way between the vein and the edge of the leaf

(time base = 8 seconds, full scale; source = argon ion laser, 514nm).



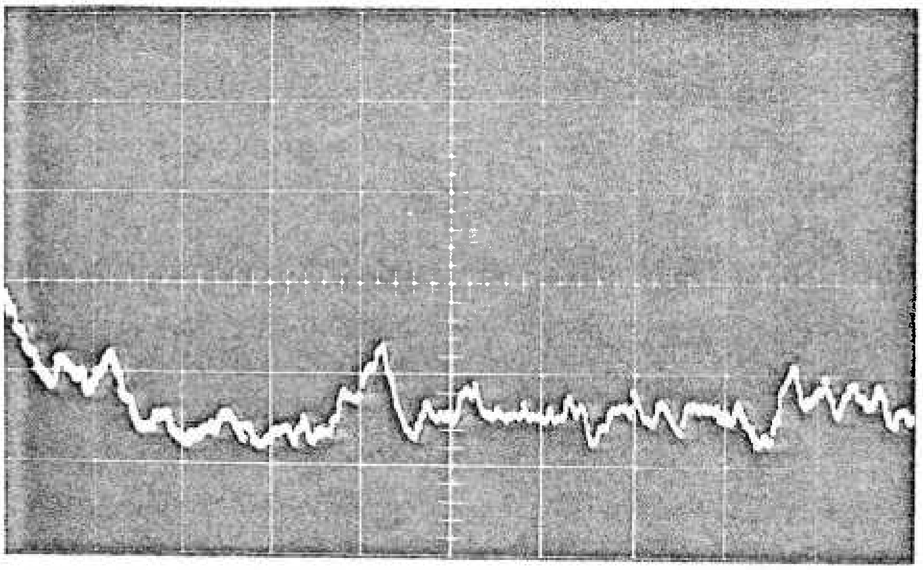
(a)



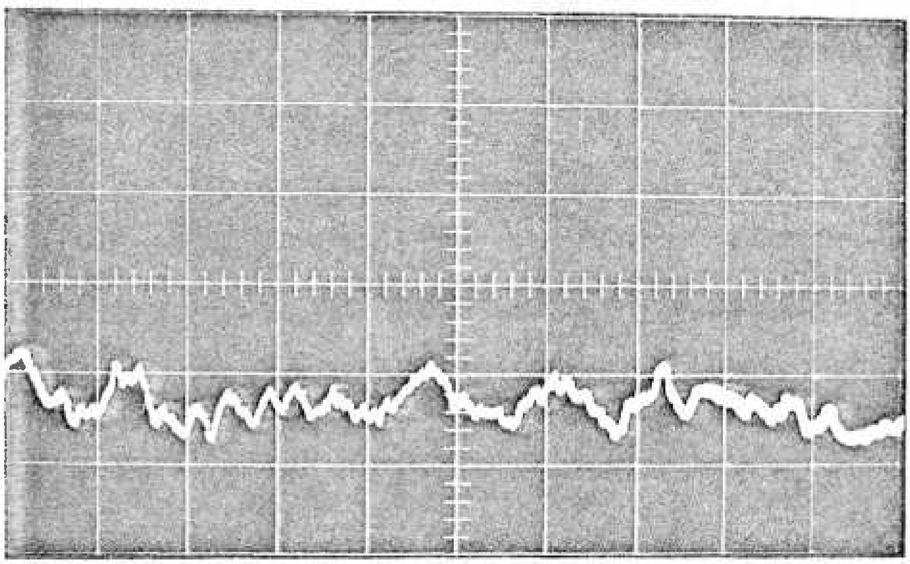
(b)

Figure 1.18 – Oscilloscope traces from green tomato:
(a) 20^{o} from specular direction;
(b) in specular direction

(time base = 10 seconds for (a), 8 seconds for (b), both full scale;
source = argon ion laser, 514nm).



(a)



(b)

Figure 1.19 - Oscilloscope traces from red tomato:
(a) 20^{o} from specular direction;
(b) in specular direction

(time base = 8 seconds, full scale;
source = argon ion laser, 514nm).

position so that the trace is above it for 16% of its range. This determines the value which is one standard deviation above the mean, and a similar procedure is followed to find the lower limiting value. The assumption of Gaussian statistics implies, in our case, that the method is only valid, even as an approximation, for traces with a low value of $\sigma/\langle I \rangle$.

The traces of Figures 1.15 to 1.19, and others (not illustrated), were analysed in this way, and the results are summarised in Table 1.4. Some values of $\sigma/\langle I \rangle$ calculated subsequently by a computer program (see Section 1.3.7) are also included for comparison; it can be seen that, in general, there is quite good agreement between these computed values and those calculated by the approximation method outlined above.

The following trends can be deduced from Table 1.4:

(a) Tobacco

These results show the differences obtained from different parts of the same specimen - see also Figure 1.16.

(b) Elodea

The qualitative observation reported in Section 1.3.4, that the fluctuations are more pronounced nearer the central vein of the leaf, is supported by these results.

(c) Tomato

The wavelength/colour dependence of the speckle fluctuations is confirmed by the measurements on green and red tomatoes, the value of $\sigma/\langle I \rangle$ being significantly higher for the green tomato in the green argon ion laser light used for the experiments. (It was not possible to reinforce these results with experiments in red light, since we had no helium-neon laser available which was powerful enough for the photomultiplier used).

TABLE 1.4

Approximate Values of $\sigma/\langle I\rangle$ from Oscilloscope Traces

Subject	Example of trace (if any)	Time base (full scale)	$\frac{\sigma}{\langle I\rangle}$	No. of traces	Computed $\sigma/\langle I\rangle$	No. of traces
Metal block	-	5 s	0.03	1	-	-
	1.15	8 s	0.03	2	-	-
	-	50 s	0.03	1	-	-
White paper	-	8 s	0.03	1	-	-
Tobacco: stem	1.16(a)	8 s	0.05	1	-	-
" : leaf	1.16(b)	8 s	0.11	3	-	-
Elodea: near vein	-	400ms	0.11	1	-	-
	-	2 s	0.37	1	-	-
	1.17(a)	8 s	0.46	3	0.29	1
" : off vein	-	400ms	0.09	2	-	-
	-	2 s	0.15	2	-	-
	1.17(b)	8 s	0.16	2	0.13	1
Green tomato: 20^o from specular	-	5 s	0.21	3	-	-
	1.18(a)	8 s	0.20	4	0.22	2
	-	10 s	0.24	3	-	-
	-	50 s	0.23	3	0.22	1
	-	3 m	0.27	8	0.24	1
Green tomato: specular	1.18(b)	8 s	0.19	5	0.24	1
	-	40 s	0.25	1	-	-
Red tomato: 20^o	1.19(a)	8 s	0.15	2	0.15	1
	-	32 s	0.18	1	-	-
Red tomato: specular	1.19(b)	8 s	0.16	5	0.13	1
Cucumber: paper*	-	2 s	0.08	3	-	-
	-	8 s	0.08	3	-	-
Cucumber: lower stem	-	2 s	0.38	2	-	-
	-	8 s	0.26	3	-	-
Cucumber: mid-stem	-	2 s	0.18	1	-	-
	-	8 s	0.43	1	-	-
Cucumber: plumular hook	-	2 s	0.29	2	-	-
	-	8 s	0.34	3	-	-
Wheat: paper*	-	8 s	0.04	1	0.06	1
" : lower coleoptile	-	8 s	0.34	2	0.41	1
" : 5mm below tip	-	8 s	0.44	3	-	-
" : tip	-	8 s	0.49	3	0.42	1

* "paper" = the paper supporting the seedling in the specimen tube.(included for comparison).

The angular dependence of the fluctuations, observed qualitatively (see Section 1.3.3) is not supported by these measurements, the values of $\sigma/\langle I\rangle$ for the specular region being for both colours of tomato of the same order as those for 20^{o} off axis. There are two possible explanations for this discrepancy. One is simply that the experiments were not carried out with sufficient care, or enough times, to obtain a true spatial average for $\sigma/\langle I\rangle$; further experiments should be carried out to check the results of Table 1.4. The second possibility is that the observed differences in the degree of fluctuation between the specular and off-axis zones are due, not to differences in $\sigma/\langle I\rangle$ (i.e. the amplitude of the fluctuations, related to the number of moving scatterers), but to differences in the frequency of fluctuation (related to the velocity distribution of the moving scatterers); detailed autocorrelation experiments would be required to check this possibility.

(d) Cucumber

The difference in the measured value of $\sigma/\langle I\rangle$ at different points on the stem of the seedling is not significant compared with the expected experimental error. Hence the qualitative observation that the fluctuations are most pronounced in the neighbourhood of the plumular hook (see Section 2.2.6 in Part 2 of this thesis) is not borne out by these measurements. Further work on this aspect of the speckle fluctuations is called for.

(e) Wheat

The results suggest that the fluctuations are more pronounced nearer the tip of the coleoptile, an observation which agrees with the biological fact that the active growth region of a cereal coleoptile is in this area; however, the differences in the values of $\sigma/\langle I\rangle$ are only slight, and may not be significant.

1.3.7 - Recording and Analysing the Speckle Fluctuations using an Oscilloscope: Computed Autocorrelations

A computer program was written (see Figure 1.20) to calculate the autocorrelation function (or, more precisely, the normalised autocovariance as defined in Section 1.1.3) from data entered via the keyboard of the computer terminal. The program also calculates the value of $\sigma/\langle I\rangle$ for the data. During the testing of the original program, in which the autocovariance was calculated from Equation 1.48:

$$C^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle - \langle I\rangle^2,$$

we found that for several sets of data the results of the program were far from being well-behaved autocovariances. Instead of tending to zero at large lag-values, the autocovariance would often start to increase again after an initial fall, and continue to increase to values well in excess of unity! This behaviour was eventually attributed to the non-stationarity of the traces involved: due either to genuine long-term fluctuations or to spurious external effects such as laser instabilities, the mean signal level was not steady over the duration of the experiment. It will be remembered from Section 1.1.3 that the condition for Equation 1.48 to represent the true autocovariance is that of stationarity, i.e.:

$$\langle I(t)\rangle = \langle I(t+\tau)\rangle$$

This condition was not satisfied for several of the oscilloscope traces, and hence the use of Equation 1.48 was invalid. (That this factor could be the cause of the anomalous autocovariances obtained is demonstrated by the fact that for one particular trace it was found that a 1% error in the mean could lead to a 100% error in the autocorrelation at large lag-values!).

Another factor which may have been affecting the computation was a result of using a truncated signal as the input data. The use of a truncated signal involves averaging over a progressively smaller sample (the region of overlap) as the lag is

```
 75/09/10. 15.41.15.
PROGRAM   NORMA

      DIMENSION S(250),G(50),GN(50)
      WRITE (6,1000)
 1000 FORMAT (//52H NORMALISED AUTOCORRELATION PROGRAM NORMA (JDB 7509)//)
    5 WRITE (6,1001)
 1001 FORMAT (///33H ENTER N AND ZERO POINT, I3, F5.0 )
      READ(5,2000) N,S0
 2000 FORMAT (I3/F5.0)
      ST = S0
      WRITE (6,2003)
 2003 FORMAT (19H ENTER DATA, 12F5.0)
      READ (5,2001) (S(I),I=1,N)
      WRITE (6,2004) S0
 2004 FORMAT (//11H INPUT DATA  //F6.1/)
      WRITE (6,2005) (S(I),I=1,N)
 2005 FORMAT (1X,12F6.1)
 2001 FORMAT (12F5.0)
      DO 100 J = 1,N
      ST = ST+S(J)
  100 CONTINUE
      N = N+1
      B = N
      SM = ST/B
      S0 = S0-SM
      G0 = S0*S0
      DO 101 J = 1,N-1
      S(J) = S(J) - SM
      G0 = G0 + S(J)*S(J)
  101 CONTINUE
      G0 = G0/B
      SD = SQRT(G0)
      SDN = SD/SM
      WRITE (6,1002) SM,SD,SDN
      WRITE (6,1003)
 1002 FORMAT (9H IMEAN = ,F7.2//9H SIGMA = ,F7.2//14H SIGMA/IMEAN =,F6.3//

 1003 FORMAT (13H LAG       G ,/)
      K = N/5
      DO 3 L = 1,K
      G(L) = S0*S(L)
      DO 4 I = 1,N-L-1
      G(L) = G(L) + S(I)*S(I+L)
    4 CONTINUE
      C = N + L
      G(L) = G(L)/C
      GN(L) = G(L)/G0
      WRITE (6,1004) L,GN(L)
 1004 FORMAT (I3,5X,F7.4)
    3 CONTINUE
      WRITE (6,1005)
 1005 FORMAT (23H TYPE 1 FOR ANOTHER RUN )
      READ (5,2002) J
 2002 FORMAT (I1)
      IF (J.EQ.1) GO TO 5
      STOP
      END
```

Figure 1.20 – Computer program for calculation of normalised covariance.

increased. Several methods of correcting this were tried, including (i) weighted means (weighted towards the area of overlap, since the data in this area were used twice in the computation while those at the ends of the batch were only used once); (ii) constant sample length (using the end of the data batch only et large lag-values, and re-calculating the mean for each lag-value); and (iii) including a string of additional values, each equal to the mean of the whole sample, at the beginning and end of the sample (these being brought into use progressively as the lag was increased, thus ensuring that all of the original data were used twice at all lag-values). The third of these methods, it was found, also had the effect of removing the discrepancy caused by non-stationarity (see above), and restoring the validity of Equation 1.48 as a representation of the autocovariance: it also resulted, whem tested with several sets of data, in autocovariance curves which were relatively well-behaved - tending to zero at large lag-values. We decided, therefore, to adopt this approach and to incorporate the device into the computer program. It must be remembered, however, that this method of avoiding end-effects and non-stationarity effects in the autocorrelation of a truncated sample is somewhat ad hoc and may have a distorting effect on the computed autocorrelations (or autocovariances).

Since the extraction of data from the oscilloscope traces is a very tedious business - the only method available for doing this was simply to read off the data point by point - only a few of the many traces obtained were treated in this way. Figures 1.21 to 1.26 show typical autocorrelation (strictly, normalised autocovariance) curves computed using the program of Figure 1.20. The large oscillations at large lag-values are statistical fluctuations caused by the relatively small number of data used (250 data points from each trace; it can be shown (McKechnie 1974) that the statistical error is proportional to $N^{-\frac{1}{2}}$, where N is the number of data points). The following trends can be detected from the curves:

(a) Elodea

There is a slight, but probably insignificant, tendency for the fluctuations in light scattered from near the vein of the leaf to decorrelate more slowly than those in light

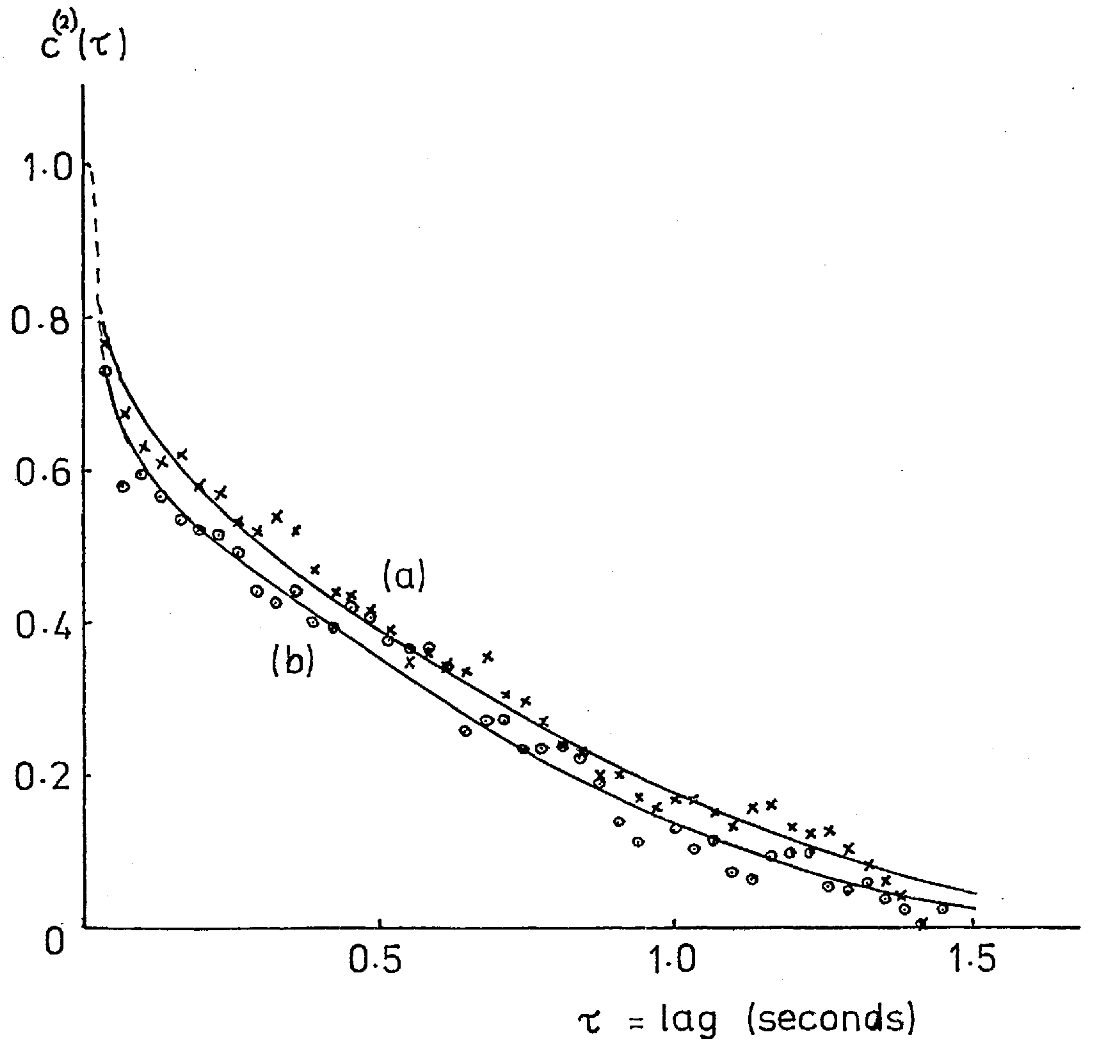


Figure 1.21 - Normalised autocovariance curves for Elodea: (a) near the central vein; (b) mid-way between the vein and the edge of the leaf

(sample length = 8 seconds
source = argon ion laser, 514nm).

scattered from other parts of the leaf. The decorrelation time, τ_c (taken at the 1/e points on the curve), for both curves of Figure 1.21 is approximately 0.6 seconds, and the time taken for complete decorrelation is about 1.5 seconds. This is a little on the low side compared with the time constant of 5 seconds reported earlier for Elodea and based on visual observations (see Section 1.3.4, also Briers 1975a); the discrepancy could be due to any of a number of possible causes, such as a tendency for the eye to integrate out the more rapid fluctuations in the visual observations, distortion of the autocovariance by the truncated signal and the small number of data points, or simply the fact that the observations were not carried out on the same specimen of Elodea! The most likely explanation, however, is that the short sample time of the traces (8 seconds) prevented the recording of the low-frequency fluctuations which had been apparent in the visual observations. According to Equation 1.52 of Section 1.2.2, a decorrelation time of 0.6 seconds would correspond to a half-width velocity of approximately 0.15μm s^{-1}.

(b) Tomato

Figure 1.22 shows normalised autocovariance curves for a green and a red tomato in green light (514nm) from an argon ion laser. The scattering was approximately 20^o backscatter and was also about 20^o from the specular direction. The difference between the two curves is probably significant, and indicates that there are slower fluctuations in the light scattered from the green tomato. If the model proposed in Section 1.3.4 is correct, these slower fluctuations would be due to chloroplast motion. The decorrelation times, τ_c (at the 1/e points) for the curves are 0.3 seconds for the green tomato and 0.15 seconds for the red tomato. These would correspond to half-width velocities of approximately 0.3 and 0.6μm s^{-1} respectively.

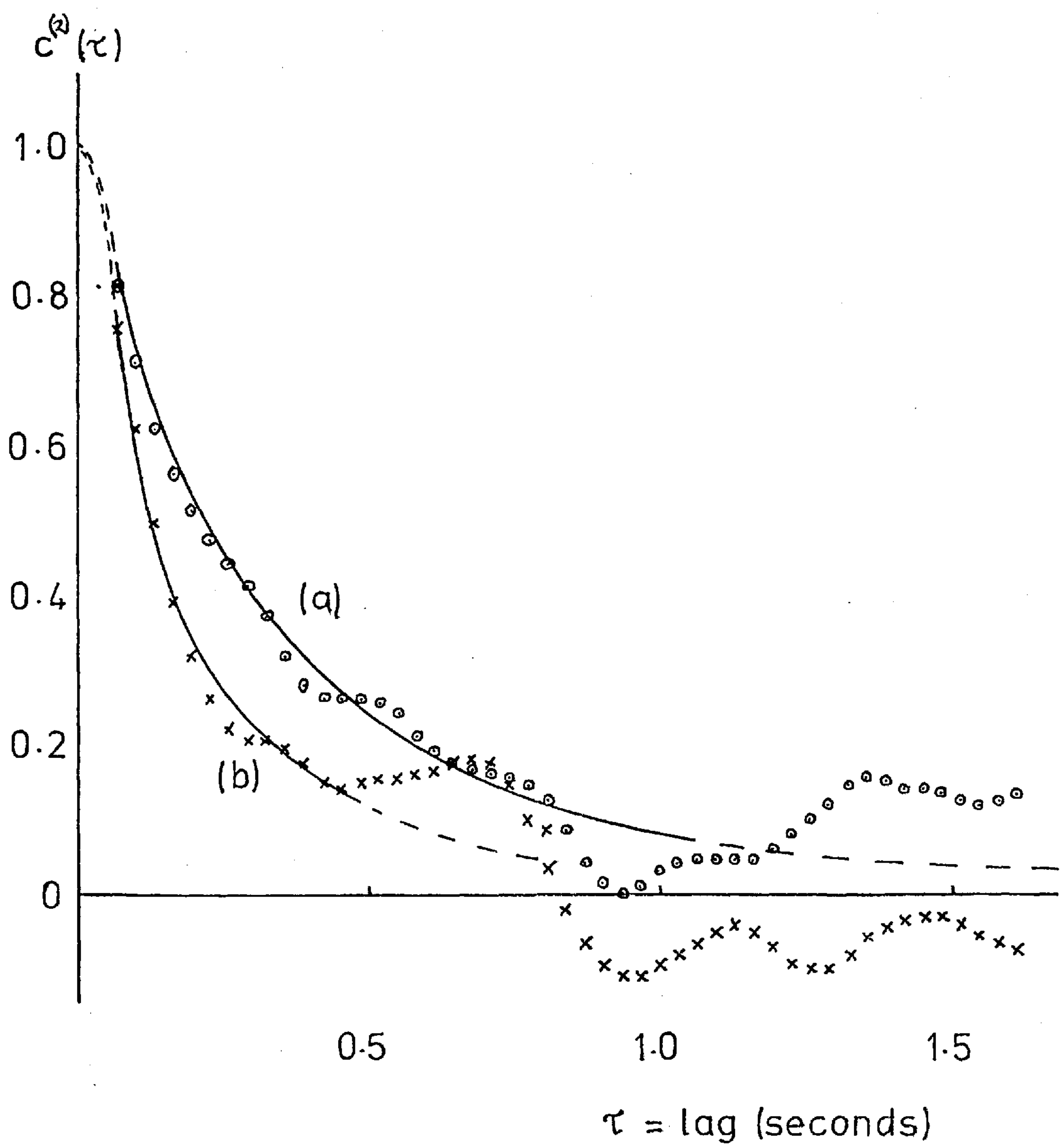


Figure 1.22 - Normalised autocovariance curves for tomatoes: (a) green tomato; (b) red tomato

(sample length = 8 seconds
source = argon ion laser, 514nm).

Figures 1.23 and 1.24 also show autocovariance curves for a green tomato in green light, but with longer sampling times - 50 seconds and 3 minutes respectively. The decorrelation times (1/e points) for the two curves are approximately 1.3 seconds and 3.3 seconds respectively, indicating that the longer sampling times have allowed the inclusion of lower-frequency components of the fluctuations. The corresponding half-width velocities for the two cases are 0.06 and 0.025μm s^{-1}.

(c) Wheat

The two curves in Figure 1.25, one for the tip and one for the lower part of a wheat coleoptile, can be seen to be virtually coincident, with a decorrelation time of about 0.3 seconds, corresponding to a half-width velocity of approximately 0.3μm s^{-1}. (These values are very close to those obtained for the green tomato with the same sampling time of 8 seconds).

(d) Cucumber

Figure 1.26 shows an autocovariance curve for a cucumber seedling with a sampling time of 2 seconds. The decorrelation time for this curve is approximately 0.4 seconds, corresponding to a half-width velocity of about 0.2μm s^{-1}.

1.3.8 - Computation of Autocorrelations from Directly Recorded Data

An alternative to the use of an oscilloscope and camera for recording the fluctuations is the direct logging of the output from the photomultiplier. A system built by Dr T S McKechnie (1974) for investigating the statistics of stationary speckle patterns can also be used to monitor fluctuations at one point in the far-field speckle pattern. This has the very important advantage of avoiding the need for manual extraction of the data from the oscilloscope traces, since the instantaneous values of intensity measured by the photomultiplier are punched directly on to paper tape in a form which enables the tape to be fed directly into a mini-computer for

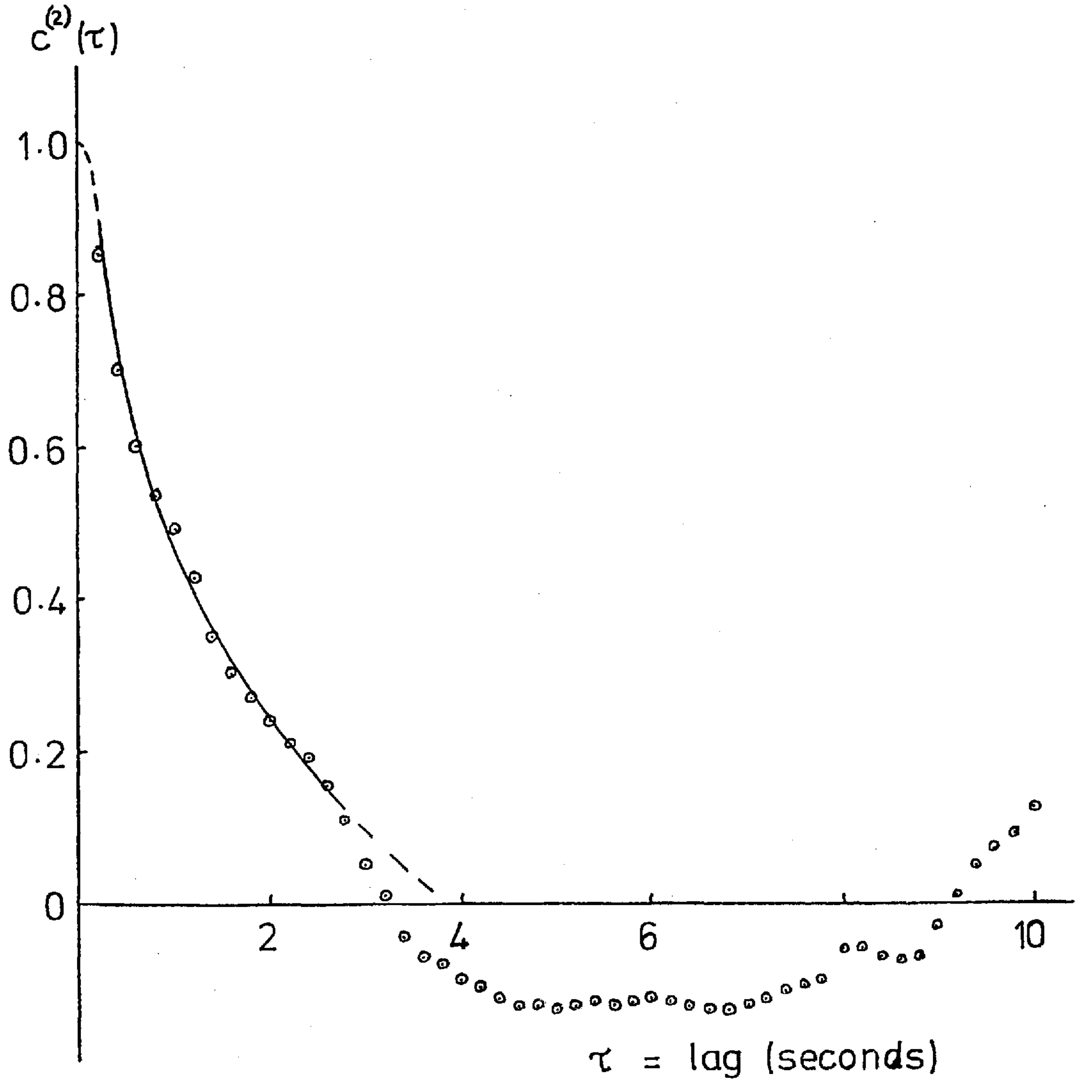


Figure 1.23 - Normalised autocovariance curve for green tomato (sample length = 50 seconds source = argon ion laser, 514nm).

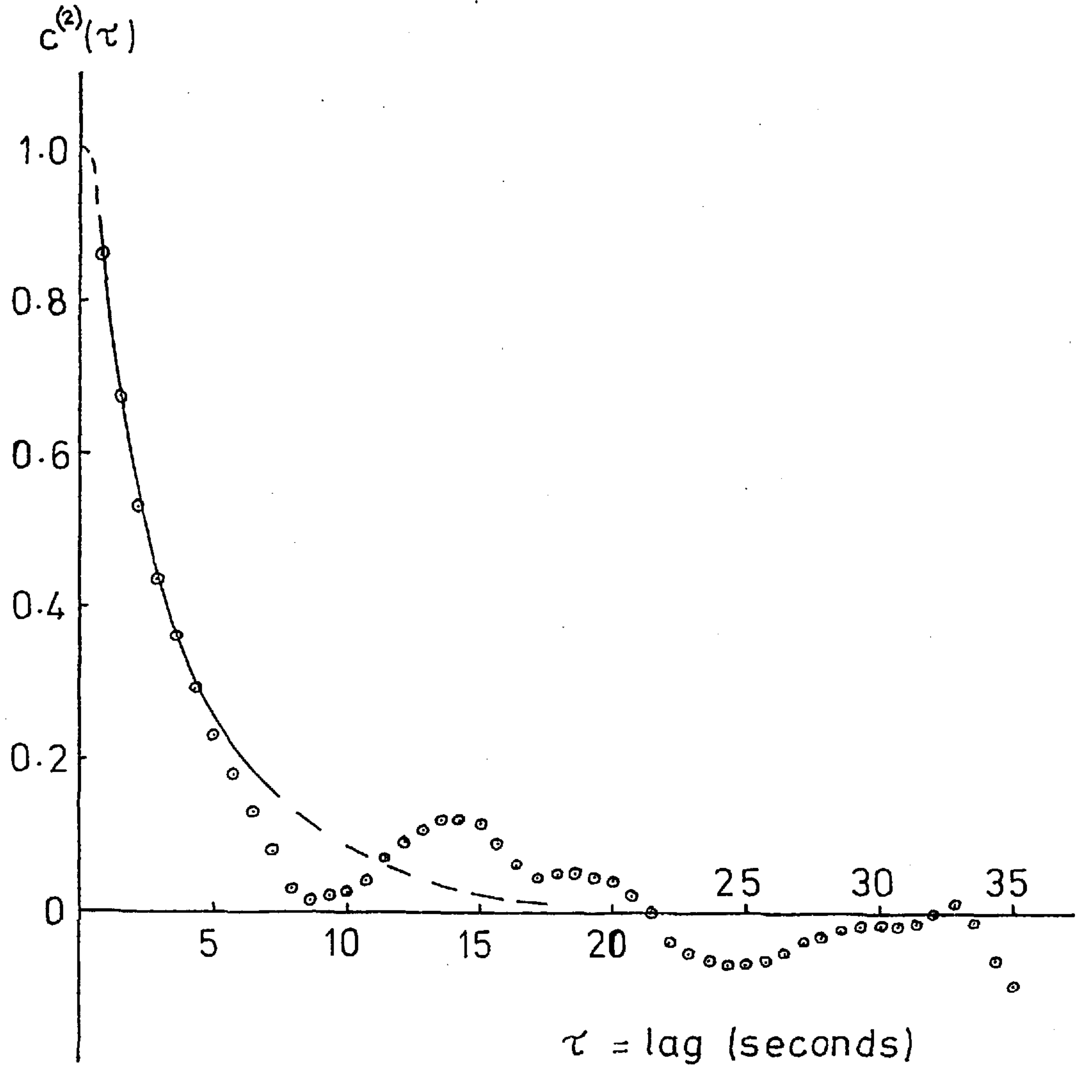


Figure 1.24 - Normalised autocovariance curve for green tomato

(sample length = 3 minutes

source = argon ion laser, 514nm).

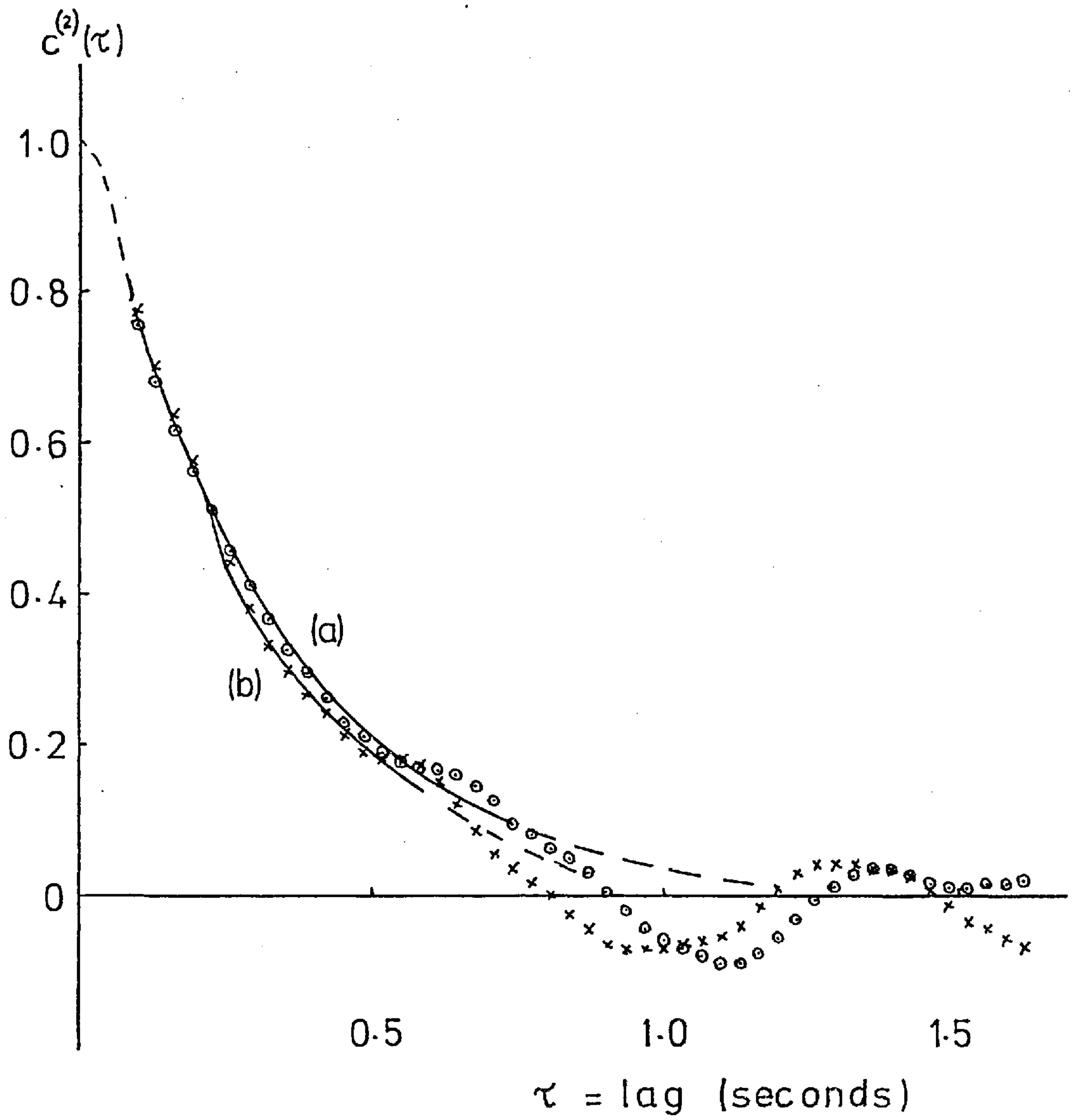


Figure 1.25 - Normalised autocovariance curves for wheat coleoptile: (a) lower coleoptile; (b) tip

(sample length = 8 seconds
source = argon ion laser, 514nm).

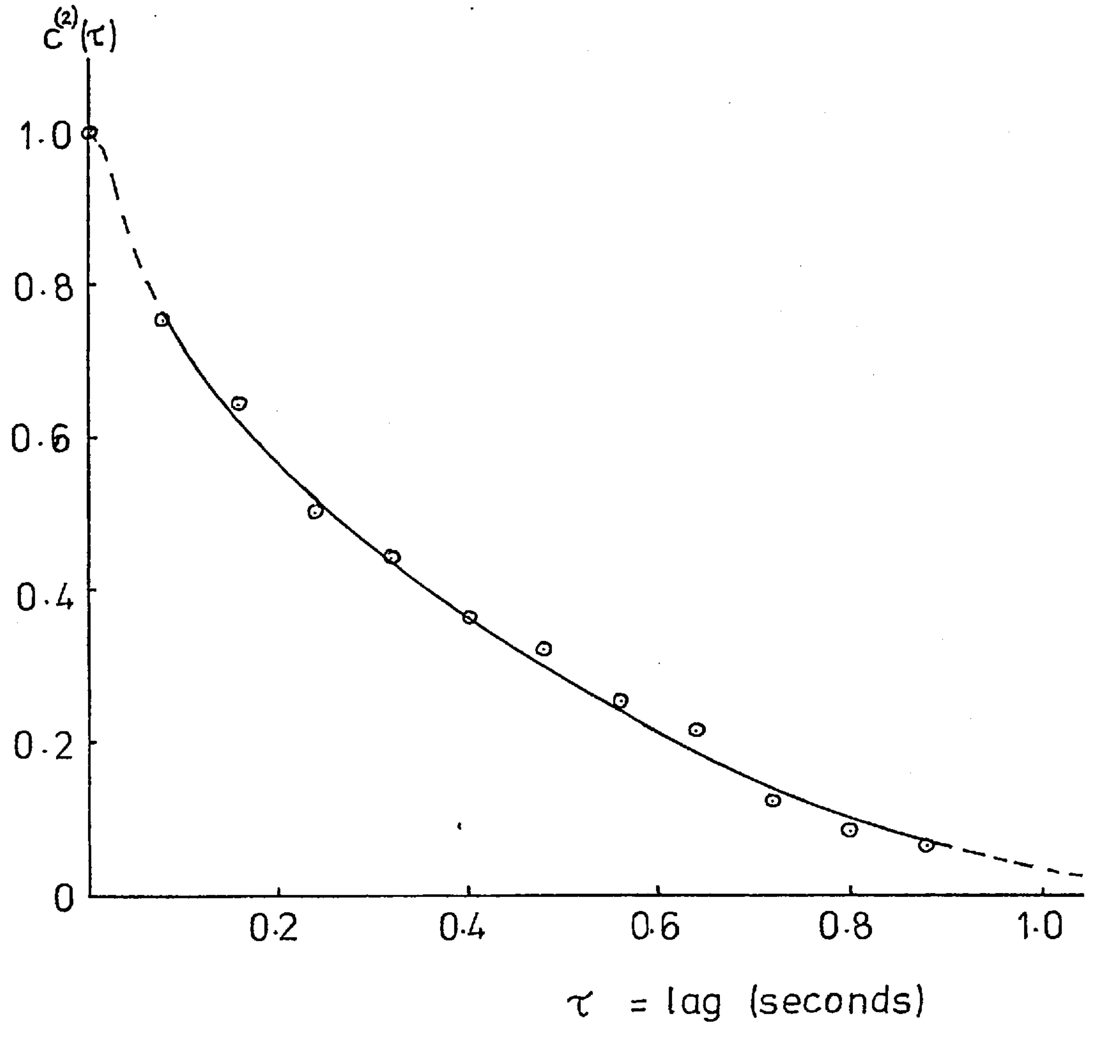


Figure 1.26 - Normalised autocovariance curve for cucumber (sample length = 2 seconds source = argon ion laser, 514nm).

processing. Unfortunately. the system is limited (in the present application) by the slow speed of the tape punch to sampling rates of less than 3 per second. Hence high-frequency fluctuations are lost when using this equipment. (According to the sampling theorem, only frequencies lower than 50% of the sampling rate can be detected, which limits the method to the analysis of fluctuations with a frequency below 1.5Hz).

Despite this limitation, we decided to carry out one experiment with the equipment, and chose to measure the autocovariance of the fluctuations from a green tomato in the 514nm line of the argon ion laser. Several runs were taken, a typical one being illustrated in Figure 1.27. The decorrelation time (1/e point) for this curve is approximately 0.5 seconds, corresponding to a half-width velocity of about 0.16μm s^{-1}. (The total sampling time for the run was about 15 minutes, much longer than was feasible using the oscilloscope).

1.3.9 - Discussion of the Experimental Results, and Suggestions for Further Work

The results of the experiments described in this chapter go some way towards confirming the validity of the model proposed as an explanation for the wavelength dependence of the speckle fluctuations (see Section 1.3.4, also Briers 1975a). There are, however, some discrepancies - for example, the failure to confirm the angular dependence of the fluctuations. Also, it is not clear at this stage whether the differences observed, for example in the statistics of the fluctuations from red and green tomatoes, are sufficiently large for them to lie sufficiently far outside the range of experimental error and statistical variation to be useful. On the other hand, the measured statistics are compatible, at least to an order of magnitude, with the proposed model, and suggest that the fluctuations could indeed be due, at least in part, to the motion of chloroplasts or other bodies inside the cells of the specimens. Much more work needs to be done before

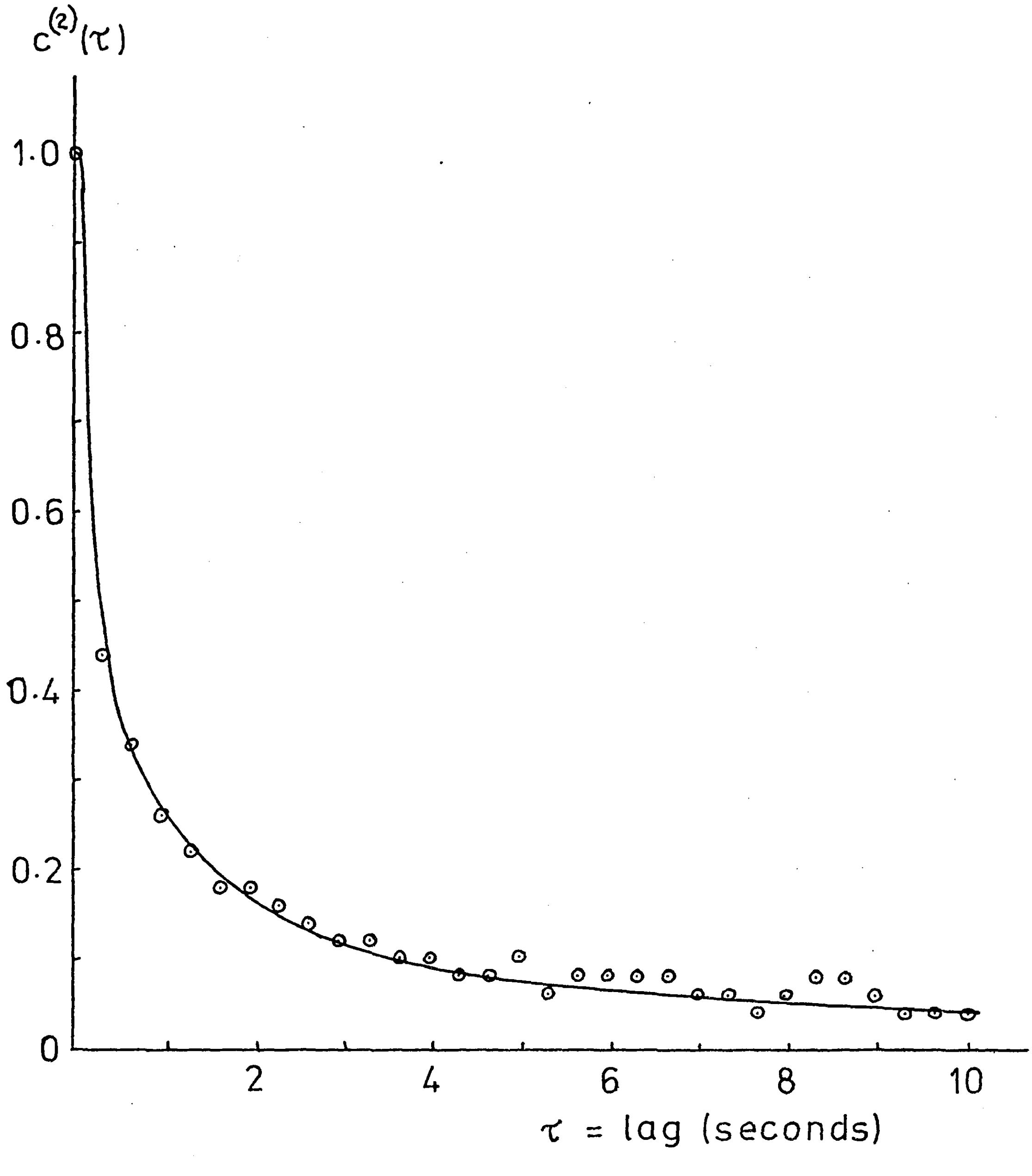


Figure 1.27 - Normalised autocovariance curve for a green tomato, using direct data-logging system

(sample length = 15 minutes, approx.
source = argon ion laser, 514nm).

we can decide whether or not the statistics of the fluctuations can be related to processes occurring inside the specimens, and we strongly recommend that if such work be undertaken, it should be as a joint project between a physicist and a biologist. We would make the following suggestions for further experiments:

(i) the use of equipment which will allow the monitoring of the fluctuations to be carried out over a wide range of sampling rates and sampling times, and which will give both the first order statistics ($\sigma / \langle I \rangle$) and the second order statistics (autocorrelations) of the fluctuations; some provision should be included to allow the spatial averaging of the former;

(ii) detailed measurements of the fluctuations from differently coloured specimens, and using different wavelengths of laser light, in order to investigate further the validity of the proposed model;

(iii) an investigation of the effects, if any, on the statistics, of changes in the environment of the specimens and of different treatments of the specimens - preliminary experiments along these lines, using tomatoes, were inconclusive.

If the results of these experiments were the confirmation of the validity of the proposed model, and the analysis of the speckle fluctuations were to provide a means of monitoring cell activity in living organisms, then it is possible that the technique could be used as a diagnostic tool in biological experiments.

Finally, we suggest that the wavelength dependence of the speckle fluctuations might prove to be a useful tool in other applications of intensity fluctuation spectroscopy, where the motions of coloured particles are involved (Briers 1975a,b).

PART 2 - HOLOGRAPHIC INTERFEROMETRY APPLIED TO THE MEASUREMENT OF PLANT GROWTH

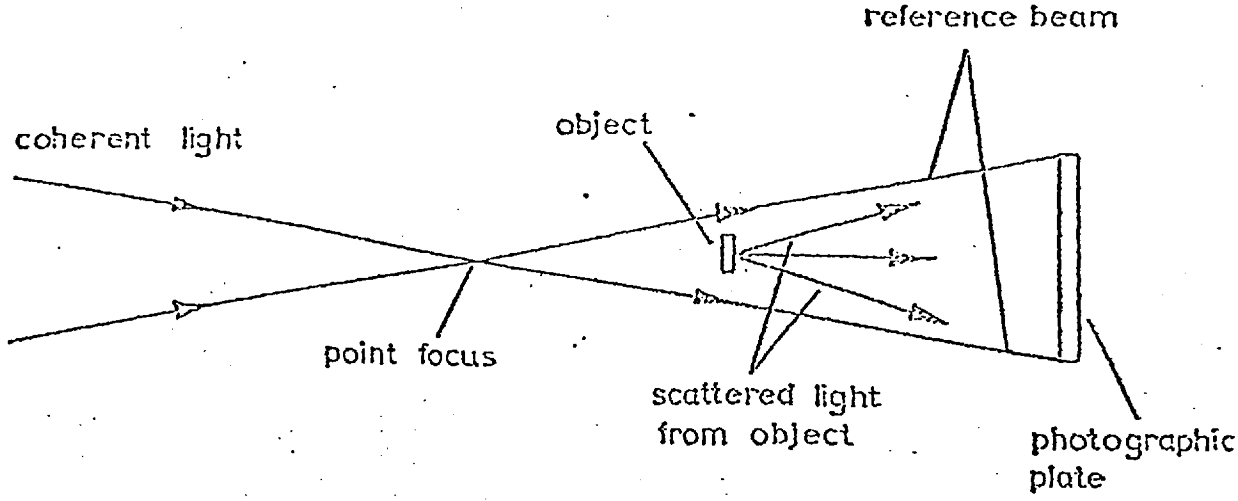
CHAPTER 2.1

THE PRINCIPLES OF HOLOGRAPHY AND HOLOGRAPHIC INTERFEROMETRY

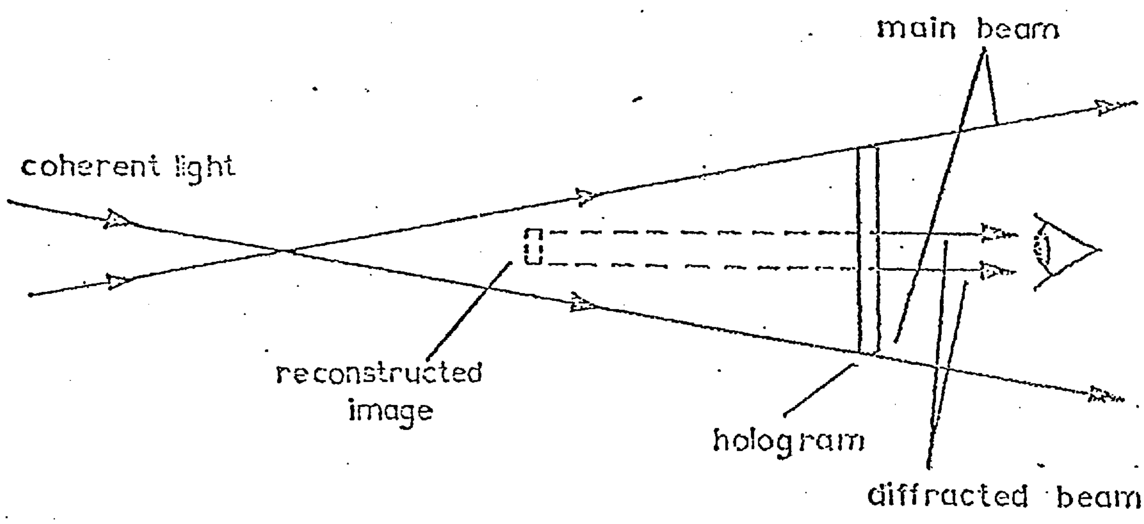
2.1.1 - The Origins of Holography

In 1947 Dennis Gabor, at that time employed by the British Thomson-Houston Company, was working on the possibility of improving the resolution of the electron microscope. The performance of that instrument was limited by aberrations of the electron lenses, and Gabor had the idea of recording an electron image produced by such an instrument, but in a manner which would retain all the information in the electron beam, both amplitude and phase, and correcting the defects later by optical means. Two short steps from this starting point led Gabor to the realisation that in order to retain the phase information he would need both coherent illumination and a reference beam with which to compare the phase of the wavefront from the object being recorded. Thus was holography born (Gabor 1948), and the first crude holograms were produced using a mercury lamp, the best compromise available at the time between intensity and coherence. Gabor's technique for recording and reconstructing holographic images is illustrated in Figure 2.1.

During the next few years the mathematical basis of holography was formulated (Gabor 1949, 1951) and physical interpretations were devised (Rogers 1950, 1951). However, apart from a small amount of work carried out in the USA in 1952, little or no further progress was made for over a decade, and holography appeared to die a natural death for lack of a good coherent source of light. The world had to wait until the invention of the laser in 1960 before the real potential of holography could be realised.



(a)



(b)

Figure 2.1 – Principle of Gabor ("in-line") holography: (a) recording; (b) reconstruction.

In the fifteen years since the invention of the laser, the science and art of holography have been investigated in great depth. It is not the purpose of this thesis to present a detailed review of the subject, and the reader is referred to the standard text books on holography (Stroke 1966, 1969, DeVelis and Reynolds 1967, Smith 1969, Collier et al. 1971), and to a DSIR (New Zealand) report by the present author (Briers 1973a). An introduction to holography and its applications is also included in a recent review of laser applications (Briers 1973b). We shall content ourselves here with a brief historical account of holography, and of its important sub-branch, holographic interferometry, together with a few words about the basic principles involved in the techniques.

2.1.2 - The Relationship between Holography and Photography

The technique of recording an image on light-sensitive materials by the process known as photography has been with us for well over a century. Unfortunately, the photographic process records only the intensity distribution across a light beam - all the phase information is lost. This loss of information results in the loss of a dimension, so that a photograph is a flat, two-dimensional picture of the scene being recorded. Even so-called three-dimensional or stereoscopic photographs, which achieve their effect by presenting slightly different views of the scene to each eye, only give the view from one angle - it is not possible to look round an object by moving one's head.

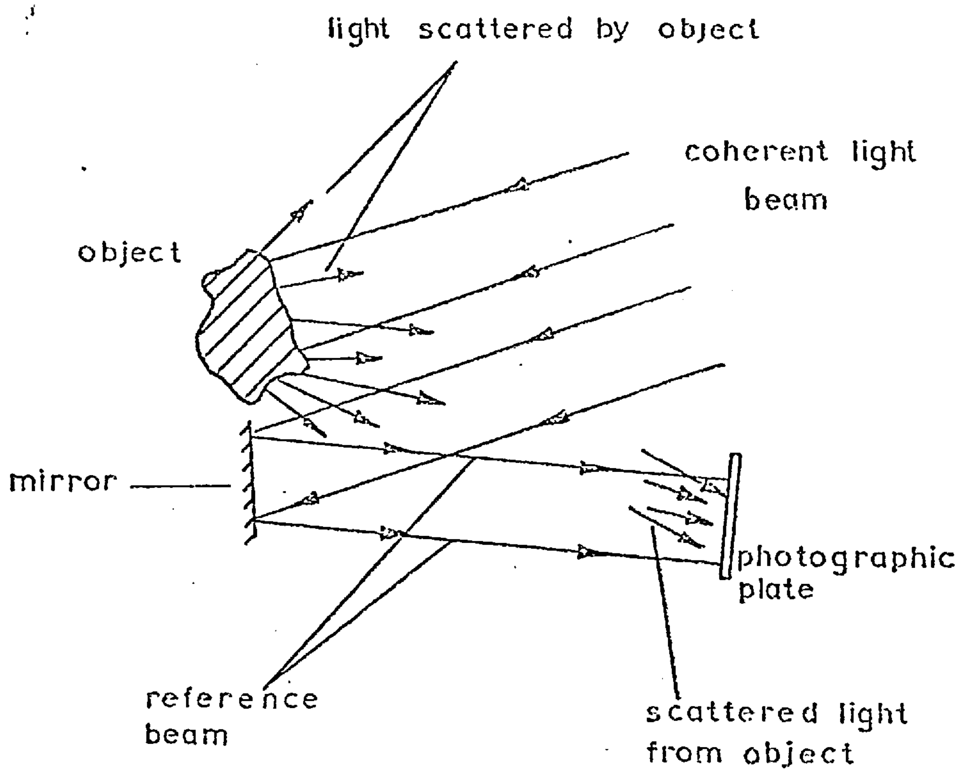
If the third dimension of a scene is to be recorded, some method must be found of recording the phase information in a light beam. Holography solves the problem by using a reference beam with which to compare the phase of the light beam from the scene to be recorded. In practice, coherent light must be used if the phase information is to be stored in a recoverable form. Since all the information in the light beam is recorded, phase as well as amplitude, the resulting "hologram" contains all the information necessary to reconstruct a full-parallax, three-dimensional image of the original scene. The reconstruction is carried out by "decoding" the hologram with a reference beam similar to the one used for the recording of the hologram.

In simple terms, the main differences between holography and conventional photography may be expressed in the following way. In photography a lens, placed between the object to be recorded and the film, focuses all the light from one point on the object to one definite point on the film. Thus each point on the film receives information from one point on the object. The result is a two-dimensional perspective of the object as seen from a point at the centre of the camera lens. In holography, on the other hand, no lens is used and light from all parts of the object is allowed to reach all parts of the film. Each point on the film receives information from all parts of the object visible from that point. The information is coded for recording on the film by using coherent light and arranging for a reference beam to illuminate the film at the same time as the light from the object. The reference beam and the object beam must be mutually coherent, which in practice means that they must originate from the same coherent source. The result is that each point (or, more precisely, a small area around each point) on the hologram reconstructs a perspective image of the object as seen from that point, so that changing the viewpoint changes the perspective - in other words, a three-dimensional image is produced. For an account of the physical and mathematical explanation of the mechanisms involved in holography, the reader is referred to the CSIR report already mentioned (Briers 1973a).

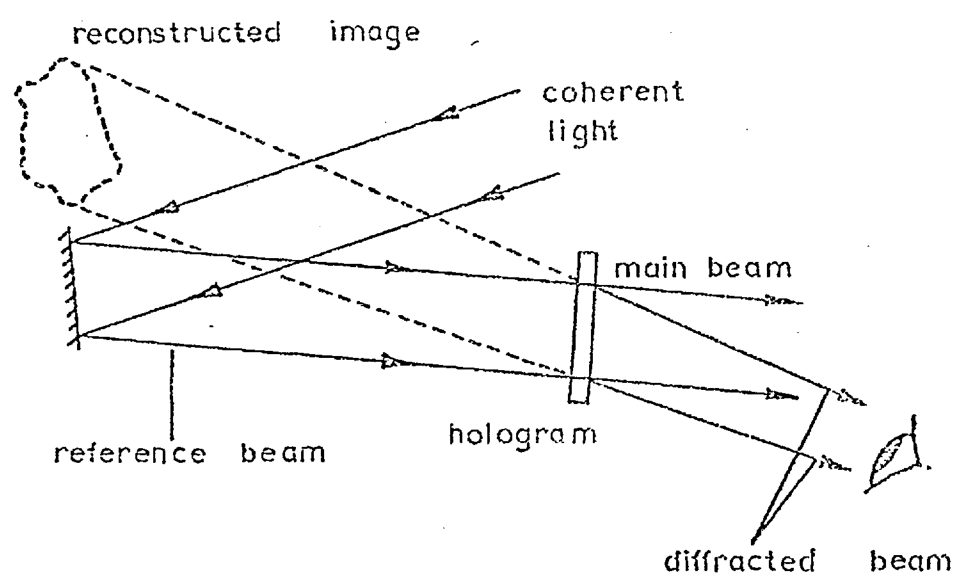
2.1.3 - A Brief History of Holography

The invention of holography by Gabor (1948, 1949, 1951) was described in Section 2.1.1, where we left the new technique languishing for want of a good source of coherent light. The great leap forward came in 1960 with the invention of the laser. For the first time the world had a source of really coherent light, and the whole field of optics was revitalised. Workers at the University of Michigan had been applying holographic techniques to side-looking radar, and when the laser became available they were able to translate their ideas into the visible realm. By combining holography with the new coherent source, and by applying the ideas of communication theory to the problem, they produced the first laser hologram (Leith and Upatnieks 1962, 1963). These early Michigan holograms were already vastly superior to the original efforts of Gabor. Not only did the coherence of the laser allow a greater latitude in path length, but it also permitted the concept of "off axis" holography to be developed (see Figure 2.2). This resulted in the complete separation of the reconstructed image from both the reference beam and the so-called conjugate image (a major problem in Gabor holography was the superposition of the zero-order beam and both diffracted beams from the hologram), and allowed the application of holography to opaque, three-dimensional objects, instead of merely to transparencies as in Gabor's original "in-line" arrangement.

Following their initial successes, the Michigan team intensified their efforts in the field and the science of holography progressed rapidly. It was not long before other universities and laboratories, and industrial organisations, were taking an interest, and today there is hardly a major firm or laboratory in the world that is not involved in some aspect of holography. During the middle 1960's refinements and extensions to the techniques of holography were being announced almost daily. The original holograms of Gabor and of Leith and Upatnieks, now known as Fresnel holograms, were joined by Fraunhofer holograms (DeVelis 1964) and Fourier transform holograms (Stroke 1965). The introduction of bleaching processes to remove the silver from the exposed and developed hologram resulted in the use of "phase holograms" with brighter reconstructed images (Cathey 1965). Colour holography was introduced by using two laser beams of different wavelengths (Pennington and Lin 1965), and later improved by the use of three wavelengths (Friesem and Fedorowicz 1966).



(a)



(b)

Figure 2.2 - Principle of "off axis" holography: (a) recording; (b) reconstruction.

An idea, pioneered by Denisyuk (1962) in the USSR, which used the full thickness of the emulsion by setting up standing waves in it between the object beam and the reference beam travelling in opposite directions, was developed to produce "volume holograms" (Stroke and Labeyrie 1966) which could be reconstructed in white (incoherent) light. This interest in having holograms which could be viewed in ordinary light also led to the development of "image-plane holograms" (Brandt 1969), in which the three-dimensional image is produced approximately in the plane of the hologram rather than at some distance behind it.

The emphasis in recent years has been on the applications of holography (Briers 1973a,b) rather than on its techniques, although work is continuing on some aspects, such as extending the range of holography beyond the visible spectrum and into the infrared (Sakusabe and Kobayashi 1971) and X-ray regions (Kikuta et al. 1972), and the development of alternative recording media such as photochromic glass (Kirk 1966), photopolymers (Colburn and Haines 1971, Booth 1975), photoresist (Sheridon 1968, Beesley and Castledine 1970), and thermoplastics (Lee 1974, Gray and Barnett 1974, Colburn and Tompkins 1974).

2.1.4 - The Origins of Holographic Interferometry

Although holography has found application in a wide variety of fields, among which we might mention three-dimensional display, data processing, pattern recognition, and image enhancement, it is generally conceded that one of its most dramatic and important applications has come from its combination with interferometry.

Holographic interferometry was born - by accident - at the University of Michigan in 1964. R L Powell and K A Stetson had been given the problem of investigating and eliminating dark bands which tended to appear superimposed on reconstructed holographic images. Instead, they deduced that the bands were really interference fringes formed as a result of slight motion of the object during the recording of the hologram. They then went on to show how these fringes could be used to analyse the vibrational modes of an object, leaving it to their colleagues at Michigan, K A Haines and B P Hildebrand, to extend the technique to the analysis of simple translations and rotations, and of general deformations.

Powell and Stetson (1965) reported their work to the Spring Meeting of the Optical Society of America in April 1965, and are thus credited with the first announcement of holographic interferometry. (A paper by Horman (1965), published in Applied Optics in March 1965 and received by that journal in October 1964, although combining the techniques of holography and interferometry in wind tunnel work, did not go so far as to realise that interference could actually occur within the holographic reconstruction itself).

However, since the work of Haines and Hildebrand (1966) on the application of the technique to unidirectional motion was not announced until January 1966 (letter received by Applied Optics in July 1965) (Hildebrand and Haines 1966a), equal credit must also go to the National Physical Laboratory (UK) team of Burch, Gates and Wilton, who developed the technique independently in May 1965 and announced it at a meeting of the Institution of Production Engineers in June of that year (Burch 1965). (In the published version of this paper, dated September 1965, Burch refers to the independent discovery of holographic interferometry by Reid and Wall of the Atomic Weapons Research Establishment, Aldermaston). The NPL work was fully reported in a subsequent paper by Burch et al. (1966).

Most of the large number of papers on holographic interferometry published since 1966 have been concerned either with particular applications of the technique, or with methods of interpreting holographic interferograms; the former mainly lie outside the scope of this thesis, and the latter will be considered in detail in Part 3.

2.1.5 - The Basic Techniques of Holographic Interferometry

As mentioned in the previous section, the first problem to be attacked by means of holographic interferometry was vibration analysis (Powell and Stetson 1965). If a hologram is recorded of a vibrating object, and the exposure time of the hologram is long compared with the period of vibration, the resulting holographic image will exhibit an interference fringe pattern. The fringes indicate loci of equal amplitudes of vibration, with the mechanical nodes showing up as exceptionally bright fringes. This technique is now known as time-averaged holographic interferometry.

The phenomenon of time-averaged holographic interferometry can be regarded as the interference of many different holographic wavefronts produced continuously during the exposure of the hologram as the object vibrates. A particular case of this general situation might be, say, the interference of just two such holographic wavefronts representing two different states of the same (stationary) object. This is the basis of the "before-and-after" type of holographic interferometry developed independently at the NPL (Burch 1965) and at the University of Michigan (Hildebrand and Haines 1966a, Haines and Hildebrand 1966). If two nearly identical holographic wavefronts are superimposed, they will reinforce where they are in phase and cancel where they are in antiphase. In practice the superposition can occur either at the time of recording the hologram or at the reconstruction stage. In the first case, if a hologram is recorded of an object, the object is then moved or deformed, and a second hologram is recorded on the same photographic plate, the reconstructed image from the final composite hologram will be crossed by interference fringes indicating the amount of displacement or deformation applied. This technique is now known as double-exposure (or frozen-fringe) holographic interferometry. Alternatively, the hologram of an object can be replaced in its original position so that the reconstructed image is superimposed on the actual object, which is viewed through the hologram. If the object is now displaced or deformed an interference pattern will appear which can be followed dynamically. This technique is now known as real-time (or live-fringe) holographic interferometry. Yet another technique, that of superimposing the reconstructed wavefronts from two separate holograms, can also be used (Gates 1968, Havener and Radley 1972, Gori and Mallamace 1973, Hariharan and Hegedus 1973). Havener and Radley (1972) referred to this method as dual hologram interferometry. A similar technique, in which the two holograms are sandwiched together and moved, either together or separately, to assist in fringe interpretation, has recently been developed by Abramson (1974b).

This summary of the basic techniques of holographic interferometry has been intentionally very brief and cursory. A more detailed review of the subject will be given in Part 3 of this thesis.

CHAPTER 2.2

HOLOGRAPHIC INTERFEROMETRY APPLIED TO THE MEASUREMENT OF PLANT GROWTH

2.2.1 - The Role of Growth Rate Monitoring in Botany

In many botanical experiments it is necessary to monitor the rate of growth of the plant, and in particular to detect changes in the rate of growth which occur in response to applied stimuli or changes in environment. Typical experiments might include the effect of light stimuli on growth rates and on the direction of growth (for example, the study of phototropism, the growth of a plant towards, or away from, a light source), the effect of the local environment on growth patterns, and the response of plants to hormones or other chemical stimuli. The actual chemical and physical mechanisms of differential growth (as in the phototropic response) are also topics of great interest to biologists.

Traditional methods of measuring the growth rates of plants have been purely visual, usually either by a shadowgraph technique or by physically marking the specimens, and it has generally been possible only to measure the average rate of growth over a period of several hours. These rather crude techniques have proved adequate for most work, but it would be useful if the botanist could have available an instantaneous method of measuring the rate of growth. This would also facilitate the measurement of the response times of plants to the application of stimuli.

Some attempts have been made in the past to devise methods of measuring instantaneous growth rates, but they have usually failed because they have entailed physically attaching something to the plant, and this can interfere with the growth rate that the experimenter is trying to measure (C Mer, private communication, 1974). What is required is a non-contact technique that does not interfere with the plant in any way.

2.2.2 - The Attractions of Holographic Interferometry as a Method of Measuring Plant Growth Rates

We can summarise the requirements of a suitable technique for measuring the growth rate of plants as follows:

(i) the method should be capable of measuring growth rates which are typically of the order of 1mm per hour for cereal coleoptiles (much used by botanists in experiments on plant growth), but which may be much slower for other types of specimen;

(ii) the method should be a non-contact technique, and should not influence the growth of the plant in any way.

The first of these requirements immediately suggests that holographic interferometry might be a suitable technique, since we are dealing with growth rates which are of the order of one wavelength of light in a few seconds. Depending on the type of specimen, the motions to be measured might be in-plane extensions or translations, line-of-sight deformations or translations, rotations (for example in phototropism), or any combination of these. Holographic interferometry can handle any of these types of motion (see Part 3 of this thesis), and is ideally suited to the time scales involved.

The second requirement can also be met, at least partially, by holographic interferometry. Although the technique is certainly a non-contact one, it is known, of course, that plants do respond to even small amounts of light, and there is a real danger that the laser used for the holography will affect the growth rate of the plant. Fortunately, we can avoid this problem to some extent by making use of the same property of laser light that makes holography possible - its narrow spectral bandwidth. Plants respond in different ways to light of different wavelengths, and there are some wavelength bands to which the response is negligible, or even zero. It is known that plant growth is stimulated by red light, and that a phototropic response is triggered by blue light. Most plants are, however, relatively insensitive to green light. By a careful choice of laser, and taking into account the known spectral response curve of the particular plant under investigation, it will usually be possible to devise a holographic system that will not influence the growth rate being measured. Light of other wavelengths could then be used as controlled stimuli in the experiments.

It would appear, then, that holographic interferometry should be an ideal tool for measuring the growth rates of plants. Yet this application of the technique has been noticeably absent from the literature, and this suggests that there may be hidden problems in store for the experimenter. We shall see later that this is indeed the case.

2.2.3 - Summary of Previous Work

The first reference to the possibility of applying holographic interferometry to the measurement of plant growth occurred in the discussion following the presentation of Burch's paper on applications of lasers, given at the meeting of the Institution of Production Engineers in June 1965. It was in this paper (Burch 1965) that the new technique of holographic interferometry was announced. In reply to a question from the floor, regarding potential applications of the technique, Burch said that "there might be biological applications" and that he "had even been tempted to try recording the growth of mushrooms!".

The only other published references that we have been able to find are a reconstruction of a double-exposure hologram of an opening tulip bud, published in two papers by Martienssen (1968, 1970), and a hint of interest in the subject in a paper by Ashton et al. (1971).

Martienssen's photograph showed fringes on the reconstruction of the tulip bud, which was taken from a double-exposure hologram with a time interval of 15 seconds between the two exposures. The photograph was included merely as an example of a possible application of holographic interferometry in a general discussion of the subject, and it is believed that the matter was not taken any further.

The paper by Ashton et al. described the application of holographic interferometry to the stretching of a rubber band and to the corrosion of metal surfaces by acid attack. The experiments were intended to be a preliminary to work on plants - the stretching of the rubber band simulated growth, and the corrosion experiment was used to investigate the effect of a change of surface texture (such as might occur in plant growth) on the visibility of holographic fringes. The result of the first experiment was that the stretching of the rubber band produced fringes which could be interpreted by standard techniques. (Ashton et al. expressed surprise that the fringes tended to be localised in front of the image rather than behind it, as is usually the case in holographic interferometry (see, for example, Haines and Hildebrand 1966); however, it is easy to show that localisation in front of the image is to be expected for in-plane deformation by stretching). The corrosion experiments showed that a change in surface texture had a very deleterious effect on fringe visibility, and the authors warned that this might be a serious problem in experiments on plant growth. No sequel to this paper was published, but in response to an enquiry by the present writer one of the authors of the paper, Professor Gerritsen (private communication, 1973), said that some experiments had, in fact, been carried out on the growth of cactus plants, but had been abandoned because of uncertainties about the interpretation of the fringes obtained.

The only other investigation that we have been able to trace is some unpublished work by Brooks and Knox. We were informed of this work by J W Goodman (verbal communication, 1974), and in response to our enquiry Dr Brooks (private communication, 1974) kindly provided photographs of reconstructions of double-exposure holograms of growing mushrooms. Although these showed definite changes on the cap of the mushroom compared with a reconstruction from a single-exposure hologram (also supplied by Dr Brooks), there were no really distinct fringes and it would appear that surface texture changes (or other factors) were affecting the experiments.

2.2.4 - Experiments with Growing Plants: Qualitative Descriptions

(a) Opening rose bud

As a preliminary feasibility study we repeated Martienssen's experiment (Martienssen 1968, 1970), but using a rose instead of a tulip. Figure 2.3 shows a reconstruction from a double-exposure hologram with a time interval of approximately 20 seconds between the two exposures. The fringes indicate movement of the petals between the two exposures, and their form and localisation properties are consistent with their being caused by the opening of the bud. The visibility of the fringes is not as high as would be expected with inanimate objects, but the fringes could certainly be used to give a measure of the rate of movement of the petals.

(b) Growth of cereal coleoptiles

Cereals such as oats, wheat and maize are often used by botanists in experiments on plant growth, and this would be a very useful area for the application of a rapid method of measuring growth rates. Experiments have been carried out in an attempt to measure the growth rates of some cereals, and particularly of wheat coleoptiles, by means of holographic interferometry. (The coleoptile is the sheath which grows from the seed of a cereal and which protects the young leaf of the plant. If the seedling is grown in the dark, the coleoptile will grow to a length of several centimetres; if the tip is stimulated by light, however, the coleoptile will stop growing and the leaf will grow from it. This is the plant's mechanism for determining when the coleoptile has reached the surface of the soil and is no longer required to protect the tender leaf).

Initial results of experiments on wheat and maize coleoptiles were encouraging, if somewhat inconsistent. Using a low-powered helium-neon laser, exposure times of the order of a few seconds, and time intervals between exposures of up to one minute, clear fringes were obtained on some double-exposure holograms,

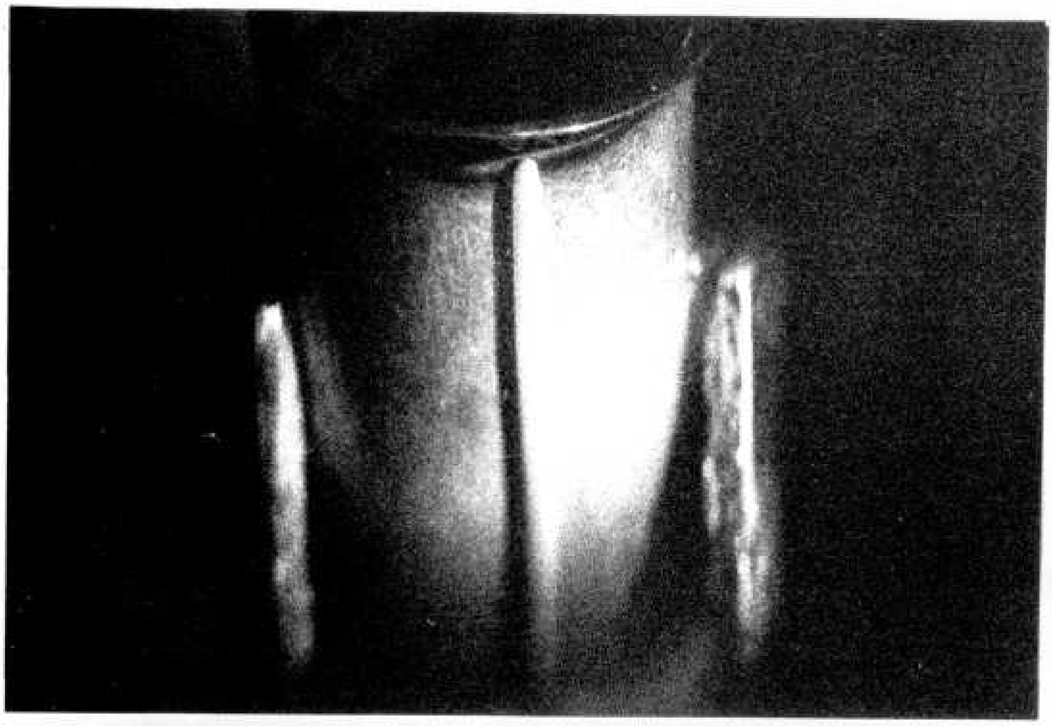
Figure 2.3 - Reconstruction from a double-exposure hologram of a rose bud, showing fringes caused by the opening of the petals

(source = helium-neon laser, 633nm
time interval = 20 seconds, approx.).

but not on others. It was soon realised, however, that these fringes were being caused, not by growth of the coleoptile between the two exposures, but by gross movement during the actual exposures. (This was confirmed when fringes were produced on single-exposure holograms). When the laser power was increased, and the coleoptiles grown between filter paper and glass to prevent gross movement, it became much more difficult to obtain fringes on the coleoptiles (although faint fringes were sometimes obtained). What did happen, however, was the appearance of clear fringes on the filter paper, indicating the movement of the paper caused by the growth of the plant. This is illustrated in Figure 2.4, which shows two reconstructions from double-exposure holograms of a wheat coleoptile grown (in the dark) in a specimen tube, and supported by a cylinder of filter paper inside the tube. Figure 2.4(a) shows the start of the formation of a fringe pattern when the time interval between the two exposures (each of 50 milliseconds) was 5 seconds, and Figure 2.4(b) shows the fringe pattern when the time interval was 30 seconds. It is possible that this in itself might be a useful technique for botanists to use, but we repeat that the production of high-visibility fringes on the coleoptile itself is both difficult and unpredictable. We also found that holograms of coleoptiles, with an exposure time of the order of a few seconds, often gave very dark reconstructions of the plant, and especially of the area near the tip of the coleoptile.

<u>(c) Growth of cereal leaves</u>

By exposing a growing coleoptile to light its growth can be arrested and the production of the leaf stimulated. Holograms of the growing leaf were usually successful in producing fringes of reasonable visibility, and they certainly did not cause the same degree of difficulty as we experienced with the coleoptiles. When the fringes were analysed by means of the fringe counting technique (see Section 3.2.3 of this thesis) they gave growth rates that were consistent with the long-term growth rates measured over a period of several hours by traditional techniques.



(a)

(b)

Figure 2.4 - Reconstructions from double-exposure holograms of a wheat coleoptile grown between filter paper and the glass wall of a specimen tube:
(a) time interval 5 seconds; (b) time interval 30 seconds
(source = argon ion laser, 514nm; exposure = 50ms).

(d) Experiments with other plants

Experiments with pea, bean and cucumber seedlings have shown that it is sometimes possible to see faint fringes on reconstructions from double-exposure holograms of these plants (see Figure 2.5). The presence or absence of fringes is rather unpredictable, but they are more likely to be observed on the stem of the plant well below the plumular hook. If fringes are obtained, they can be interpreted by means of the fringe counting technique and give growth rates which are consistent with measured long-term growth rates.

Dark reconstructions of holograms were also observed with these plants, and especially with the plumular hooks.

2.2.5 - A Possible Explanation of the Difficulties Encountered

Gross movement of the plant can be discounted as a cause of the unpredictable appearance of fringes, their low visibility when they are present, and the dark reconstructions, since the phenomena occur even when the plant is prevented from moving laterally. We must therefore seek another solution to the problem.

When a botanical specimen is illuminated with laser light and observed through a small aperture, the speckle pattern produced on the surface of the specimen is observed to fluctuate. This phenomenon has already been reported in the literature (Briers 1975a), and forms the basis of the work described in Part 1 of this thesis. We suggest that these speckle fluctuations are the source of the difficulties encountered when holograms are recorded of plants, since it is essentially the speckle pattern that is recorded in the holographic process. (The phenomenon was also noticed by Brooks (private communication, 1974) and blamed by him for his difficulties with holographic interferograms of mushrooms - see Section 2.2.3).

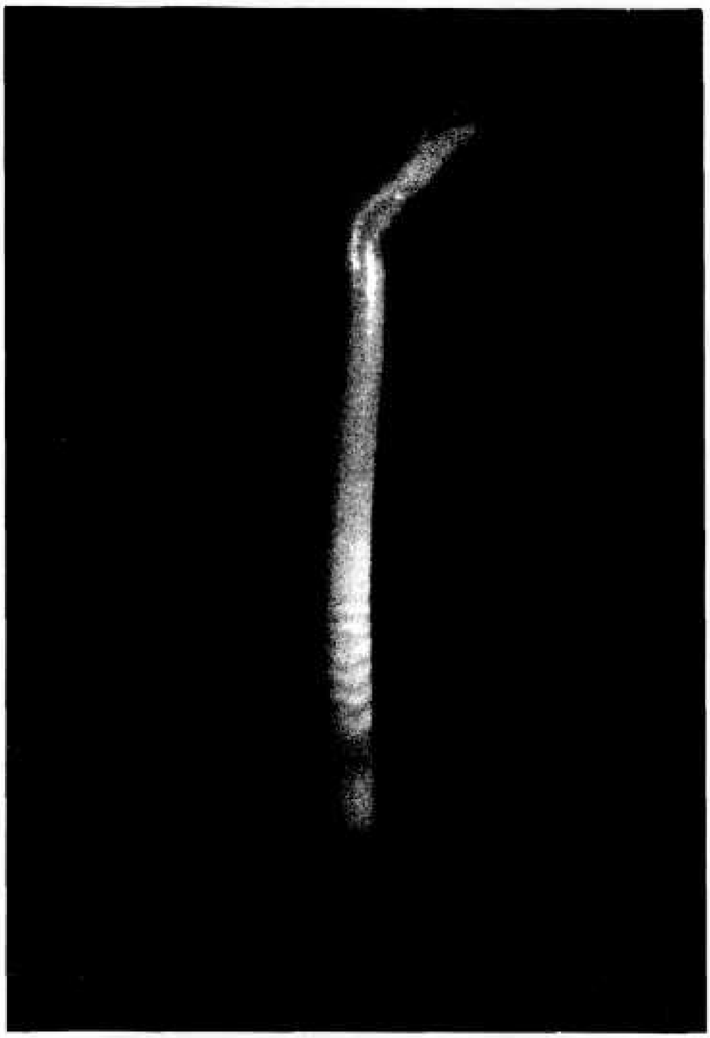


Figure 2.5 - Reconstruction from a double-exposure hologram of a broad bean seedling, showing fringes

(source = helium-neon laser, 633nm
time interval = 2 minutes
exposure time = 10 seconds).

2.2.6 - Correlation between Speckle Fluctuations and Holographic Problems

The first experiment that we carried out to establish a link between speckle fluctuations and the problems encountered when using plants as holographic subjects was the attempt described in Section 1.3.1 to record a hologram of a red tomato. As we saw, when the hologram was subsequently reconstructed a chess piece included in the scene for comparison was clearly visible, while the image of the tomato was extremely faint. The hologram was recorded in the red light from a helium-neon laser, and it will be remembered from Section 1.3.2 that a red tomato exhibits marked speckle fluctuations when illuminated with red laser light (see also Briers 1975a,b).

We reported in Part 1 that the speckle fluctuations are often more pronounced in some parts of a specimen than in others. There appears to be a definite correlation between the degree of speckle fluctuation and the quality of holographic reconstruction. For example, investigation of a cucumber seedling grown in a specimen tube showed that the fluctuations were more pronounced in the neighbourhood of the plumular hook than they were lower down the stem: a single-exposure hologram, with an exposure time of eight seconds, produced a reconstruction in which the lower part of the stem was clearly visible, but whose brightness diminished towards the top of the plant, the plumular hook being hardly visible.

We suggest that these speckle fluctuations are the cause of the difficulties experienced with holograms of growing plants. If the time constant of the fluctuations is of the same order as the time interval between the two exposures of a double-exposure hologram, the effect will be a deterioration in the visibility of any fringes, and ultimately their complete loss; if the time constant is of the same order as the actual exposure time, the result will be a reduction in the brightness of the reconstruction. The latter problem can be overcome by increasing the laser power and reducing the exposure time (see Figure 2.4), but little can be done about the time interval since this is dictated by the growth rate itself.

2.2.7 - Quantitative Analysis of some Holographic Interferograms

Although the appearance of fringes on double-exposure holograms of plants is a rather unpredictable phenomenon, and appears to be affected adversely by the fluctuating speckle effect, we have already indicated that when fringes are obtained they can be interpreted by standard techniques and tend to give answers of the right order of magnitude. We now present some examples of the analysis of holographic interferograms, using the examples of double-exposure holograms given in Figures 2.3 to 2.5.

(a) Rose

The fringes of Figure 2.3 are seen to be, in general, sets of approximately parallel lines, one set to each petal. Observation of the actual hologram shows that there is virtually no parallax between the fringes and the specimen as the eye is scanned across the hologram. We shall see in Part 3 of this thesis (Section 3.2.2) that parallel fringes localised at the image in this way are caused by rotation of the object about an axis in the plane of the object, or in other words by tilt of the object. The axis of rotation is parallel to the fringes, and the angle of tilt is given by Equation 3.4:

$$\psi = \frac{\lambda}{h(\cos\theta_2 + \cos\theta_1)}$$

where λ = wavelength of the light
h = fringe spacing
θ_1 = angle of incidence of the illuminating beam
θ_2 = angle between the direction of view and the normal to the surface of the object.

The geometry of the experiment was such that we can approximate $\cos\theta_2 + \cos\theta_1$ as being 1.5, and taking the value of the fringe spacing for the petals at the extreme left and extreme right of the photograph as 1mm, we find from the above equation that this corresponds to a tilt of the petals of approximately 0.4 milliradians, or 1.4 minutes of arc. The fact that the fringes are not exactly parallel, especially for the petal at the centre top, means that some deformation (change of shape) of the petals has also occurred, as would be expected. These tilts and deformations have, of course, occurred during the twenty seconds between the two exposures of Figure 2.3.

(b) Wheat

The fringes of Figure 2.4, produced on the paper used for supporting the specimen against the wall of a glass tube, can be interpreted to give a measure of the swelling (or possibly the lateral movement along the line of sight) of the coleoptile during the time interval between the two exposures of each hologram. The single fringe on Figure 2.4(a) indicate that the paper has moved approximately half a wavelength, or 0.25μm, in the five seconds time interval, and the three fringes on Figure 2.4(b) represent a movement of about 0.75μm in thirty seconds.

(c) Bean

The fringes of Figure 2.5 are localised by parallax in front of the image, a property typical of fringes caused by extension or growth (see Section 2.2.3). The region over which fringes are visible extends over about 10mm of the stem of the broad bean seedling, and the parallax between the fringes and the image varies from zero at the bottom of this zone to three fringes at the top, as the eye is scanned a distance of 35mm vertically across the hologram. The distance from the reconstructed image to the hologram when these measurements were made was 170mm, and the fringe counting technique (see Section 3.2.3 of Part 3 of this thesis) gives the vertical motion of the object at this point as approximately fifteen wavelengths, or 10μm. Thus the differential growth between the two extreme ends of this 10mm zone is 10μm in the two minutes between the two exposures of the hologram. If the same differential growth rate were to occur over the rest of the seedling above this zone, then by extrapolation the tip of the plant would have increased its height by about 40μm in the same two minutes. If the same growth rate can be assumed to persist over twenty-four hours, this would be equivalent to a growth of the seedling of just under 30mm per day, which is of the right order of magnitude for a seedling at this stage of its development, and agrees with the long-term growth rates measured over a period of twenty-four hours.

Other experiments (not illustrated) have also given growth rates which agree with the measured long-term growth rates. For example, when faint fringes were obtained on a wheat coleoptile - one of the few occasions when they were not completely destroyed by the speckle fluctuations - analysis by the fringe counting technique gave the growth during the ten seconds interval between the two exposures as 2.3μm. This is equivalent to a growth rate of about 20mm per day, which again agrees with the observed daily growth.

2.2.8 - Possible Applications of Holographic Interferometry in Botany

Success with holographic interferograms of growing plants is more likely to be achieved if the areas of pronounced speckle fluctuations can be avoided. Not surprisingly, we find that the known areas of high activity in a plant produce the more marked speckle fluctuations. For example, growth in a coleoptile occurs mainly in an area a few millimetres below the tip; growth of a cereal leaf, on the other hand, occurs at its base, the upper portion of the leaf being pushed upwards by the new material being produced below. It will be remembered that it was considerably easier to produce fringes on a leaf than it was on a coleoptile.

It would appear, therefore, that the application of holographic interferometry to the measurement of plant growth is likely to be more successful if the areas of high activity can be avoided. Unfortunately, it is just those areas that are more interesting to the botanists, but if they can accept this limitation we believe that the technique might prove to be of value.

If fringes can be obtained in holographic interferometry experiments they can usually be interpreted, and we have seen that when growth fringes are obtained on holographic interferograms of plants they yield growth rates which are consistent with those measured by more traditional techniques. Since these traditional techniques can only give an average growth rate over a period of minutes, or even hours, we suggest that holographic interferometry, which can give an almost instantaneous measure of the growth rate, might, despite its limitations caused by speckle fluctuations, be a useful tool for botanists. The following precautions should, however, be taken:

(i) gross movement of the specimen should, if possible, be prevented;

(ii) speckle fluctuations should be used as a guide as to whether a particular specimen, or a particular part of a specimen, is a suitable subject for holographic interferometry - the speckle fluctuations can be observed by viewing the specimen, illuminated with laser light, through a pinhole (say 1mm aperture), and areas of high activity should be avoided.

Subject to the restrictions outlined above, the instantaneous measurement of growth rate provided by holographic interferometry should provide a method of measuring the response times of plants to external stimuli or to changes in conditions. An obvious application would be an investigation of the phototropic response of plants.

2.2.9 - Suggestions for Further Work

We have shown that in certain cases holographic interferometry can be a viable tool for the monitoring of plant growth, though there are certainly severe restrictions on its usefulness. If the biologists feel that, despite these limitations, the technique is worth investigating further, we strongly recommend that any such development should be carried out in close consultation with a botanist, and preferably as a joint project between a physicist and a botanist.

It is likely that the provision of some degree of magnification in the holographic arrangement would make the technique more attractive to botanists. A low-powered holographic interference microscope such as that described by Magill and Wilson (1968) might be a useful instrument, and it is also possible to use a full-scale, high-powered holographic microscope (Williams 1973) for holographic interferometry of microscopic specimens.

Finally, it would be useful to have a real-time process for measuring the growth rates so that a truly instantaneous measurement could be made and the development of the fringes followed dynamically. Unfortunately, this is not practicable with photographic emulsions, since the growth of the plant cannot be halted while the hologram is processed, and too much growth is likely to have occurred before the processed hologram is returned to the apparatus. What is required is a recording material which can be processed in situ virtually instantaneously. A suitable material, for example, might be a thermoplastic medium such as that used by Gray (1975) (see also Gray and Barnett 1974), and we suggest that the use of this material be investigated for use in the application of real-time holographic interferometry to the measurement of plant growth.

PART 3 - THE INTERPRETATION OF HOLOGRAPHIC INTERFEROGRAMS

CHAPTER 3.1

A HISTORICAL REVIEW OF HOLOGRAPHIC INTERFEROMETRY

(Note: The dates in the headings to the sections of this chapter refer to the dates of verbal presentation of a paper, or of its receipt by a journal, rather than to the actual publication dates; this has been done in order to establish chronological precedence.)

3.1.1 - Laying the Foundations: 1965-1966

The story of the discovery of holographic interferometry has already been told in Section 2.1.3, where due credit was given to the pioneering Michigan teams of Powell and Stetson (1965a,b) and Hildebrand and Haines (1966a), and to the NPL team led by Burch (1965).

The work of Powell and Stetson concentrated on "time-averaged" holographic interferometry, in which a hologram is recorded of a vibrating object, and this paved the way for the development of a whole sub-branch of holographic interferometry, its application to vibration analysis (Stetson and Powell 1965a,b, 1966).

Haines and Hildebrand were more concerned with the application of the technique to displacements, rotations and deformations by analysing the fringes produced by interference between the holographic images of an object before and after the motion or deformation had occurred. This interference could occur either between the object itself and a holographic reconstruction of it ("real-time" or "live-fringe" holographic interferometry), or between two holographic images recorded on one photographic plate, the displacement or deformation having been applied to the object between the two exposures ("double-exposure" or "frozen-fringe" holographic interferometry). Haines and Hildebrand (1966a,b) gave a method of interpreting the fringe patterns for rotation about an axis in the plane of the object, when the fringes were localised on the surface of the object, and for in-plane translations, when the distance from the object to the plane of localisation of the fringes was used as a parameter. (This fringe localisation (FL) technique of fringe interpretation will be discussed in detail in Section 3.2.2). They also gave a set of very complex equations for interpreting fringes due to a combination of rotations and translations (Haines and Hildebrand 1966a).

The NPL team also published further details of their work in holographic interferometry. Burch et al. (1966) introduced the concept of multiple-exposure holography with equal increments of strain between each exposure and the next in order to sharpen the fringes - and to indicate failure of Hooke's law when a lack of sharpness followed. Burch and Ennos (1966), and later Archbold et al. (1967), gave an account of the application of holographic interferometry to the comparison of cylinder bores.

Work on holographic interferometry had also been in progress, independently, at Bell Telephone Laboratories, and an account of that work was published by Collier et al. (1965). They described both double-exposure and real-time techniques, and explained the fringes as a moiré effect similar to that obtained by the superposition of two diffraction gratings, though they went on to suggest that "interference" might be a better word than "moiré". Prompted by this work, Nassenstein (1966), of Farbenfabriken, Leverkusen, Germany, published some examples of holographic interferograms of subjects such as cogwheels, but attempted no analysis of the fringe patterns.

The interpretation scheme put forward by Haines and Hildebrand (1966a) suffered from several major disadvantages. First, it relied on the determination of the plane of localisation of the fringes. It will be seen later (Chapter 3.3) that fringe localisation in holographic interferometry is not so straightforward as might at first be thought, and Haines and Hildebrand admitted that uncertainties in the localisation of the fringes often led to errors of up to 60% in the calculated translations! Secondly, the method gave only the "in-plane" components of any translation (i.e. the components normal to the line of sight). If the third component of the motion was required, this could only be found by changing the angle of view through the hologram, repeating the measurements, and combining this new set of data with the first set in order to deduce the component along the line of sight. Since the geometry usually allows only a slight change in the viewing angle to be made in this way, the line-of-sight component of motion could not be measured with any great accuracy. Thirdly, the equations to be solved when both translations and rotations were present together were so complex that analysis by this technique was not really a practicable proposition. It was not surprising, therefore, that alternative methods of analysis began to appear in the literature.

The first of these alternative methods of interpreting holographic interferograms came from the USSR, when Aleksandrov and Bonch-Bruevich (1967) announced a "dynamic" method of interpretation which we shall call the fringe counting (FC) technique. Details of the technique will be given in Section 3.2.3, but the principle of the method is that the observer counts the number of fringes which pass the point under consideration on the holographic reconstruction of the object as he moves his head (and hence the area of hologram through which he is looking) in a linear direction. Thus one of the objections to the method of Haines and Hildebrand, that of locating the fringes, was overcome - the parallax between the fringes and the object was used, rather than the actual position of the fringes in space. On the other hand, this method also gave only the in-plane components of the displacement of the point under consideration, and relied on a change of viewing angle to determine the third component.

The analysis of fringe patterns obtained with holograms of phase objects presents little difficulty, since the geometry of the apparatus can be made much simpler than when dealing with diffusely reflecting objects. Heflinger et al. (1966) - see also

Brooks et al. (1965) - of TRW published holographic interferograms of shock waves caused by bullets, and of electric lamps. A qualitative explanation of the fringes in terms of moiré effects at the hologram was given, and interpretation of the fringes was carried out simply in terms of changes in optical path length along the line of sight.

Another application of holographic interferometry was developed in these early days of the subject. By using two wavelengths of laser light for recording the hologram, but only one for reconstruction, contour fringes could be superimposed on a holographic image (Haines and Hildebrand 1965). The effect relies on the magnification which occurs when a hologram is reconstructed in light of a wavelength different from that of the recording light: interference takes place between the two different-sized reconstructed images. Haines and Hildebrand showed that for a contour spacing of 1mm depth, the wavelength difference should be about 0.4nm. It was later shown that the same effect could be achieved by using two separated sources of the same wavelength at the recording stage (Hildebrand and Haines 1966b). Both techniques were fully described in a later paper by Hildebrand and Haines (1967).

3.1.2 - The Years of Consolidation: 1967-1969

Following the initial enthusiasm for holographic interferometry in the previous two years, 1967 was actually a lean year for the subject. No major advances were announced, and the only significant contribution was a review by Ennos (1967) of holography as a whole, and of the NPL work on holographic interferometry in particular. He also mentioned the decrease in fringe contrast which occurs if the surface texture of the object changes during the experiment.The possibility of using holography to measure velocity was reported independently by Redman (1967) and Lurie (1967), the latter dealing in general terms with holograms of moving objects. This topic was also discussed by Neumann (1968). A new method of contouring, by immersion of the object in liquids of different refractive indices for the two exposures, was introduced by Tsuruta et al. (1967). A progress report on holographic techniques of vibration analysis was presented by Barnett (1967), and the effect of speckle on fringe visibility was discussed in a paper by Tanner (1968).

In contrast, 1968 and 1969 were very busy years for holographic interferometry. A new method of interpreting holographic interferograms was introduced by Ennos (1968). Based on first principles, the technique assumed that the zero-order fringe could be identified. For this reason we shall refer to it as the "zero-order fringe" (ZF) technique, and we shall discuss it in detail in Section 3.2.6. The method gave the component of motion in a direction bisecting the illuminating and viewing directions, which in most cases is in a direction close to the line of sight. The other components of the motion were obtained by recording three separate holograms with three different lines of sight and operating numerically on the fringe order number derived from the three patterns in order to obtain the resolved part of the motion in any direction. The application of the technique to a particular example, the stretching of a metal foil strip under tension, was described in the same paper. Papers by Lurie (1968a) and Lurie and Zambuto (1968) gave a generalised theory of holographic interferometry for any type of motion. Using coherence theory, they produced a general formula for the coherence between the incident and reflected light and showed that it reduced to the formula of Powell and Stetson (1965b) for the case of sinusoidal vibration. A generalised theory of fringe formation was also given by Brown et al. (1969), and applied to the specific cases of time-averaged sinusoidal vibrations (J_0^2 fringes), uniform motion at constant velocity (sinc^2 fringes), double-exposure or real-time case in which the object has two discrete states (cosine fringes), and real-time vibration analysis ($(1+J_0)$ fringes). An interesting result given in this paper is that the maximum allowable velocity of an object if a good hologram is to be recorded in an exposure time T is given by the expression $v_{max} \leqslant \lambda/5T$ (for motion towards the hologram). Other work reported before the Strathclyde Symposium of 1968 included the use of stroboscopic techniques as an aid to real-time holographic interferometry of vibrating surfaces (Archbold and Ennos 1968), the extension of Hildebrand and

Haines' (1967) contouring method to larger surfaces, using an auxiliary imaging system (Zelenka and Varner 1968), the application of holographic interferometry to the monitoring of the bending of a bar and the propagation of a stress wave through a bar (Gottenberg 1968), and a description of a holographic interference microscope, operating at magnifications between x6 and x15, for time-averaged holography of small vibrating objects and real-time holography of stresses applied to small transparent objects (Magill and Wilson 1968).

At the 1968 Strathclyde Symposium on "Engineering Uses of Holography" (Robertson and Harvey 1970), there were several papers on the applications of holographic interferometry. Most of these described specific applications, such as the measurement of crystal vibrations and of creep (Bradford 1968), the testing of diaphragms (Butters 1968), the measurement of large-amplitude vibrations (Wall 1968), and contouring by the two-source and the two-wavelength techniques (Hildebrand 1968). The possibility of combining moiré techniques with holographic interferometry was raised by Der Hovanesian and Varner (1968). Among the more general papers, Abramson (1968, 1969a) introduced his "holo-diagram"; this algorithm, based on confocal ellipses, was claimed to make the task of the experimenter much simpler. The holo-diagram, which its inventor suggested should be drawn directly on the holography table, could, it was claimed, be used to ensure good holograms even of long objects (by positioning the object carefully in relation to the ellipses of the holo-diagram), to pre-select the required fringe spacing (sensitivity), to optimise the position of the reference mirror, to minimise the sensitivity of the holographic apparatus to unwanted movements, and finally to evaluate the hologram (in conjunction with the ZF technique - see Section 3.2.6). Abramson also extended the principle of the holo-diagram to classical interferometry to produce his so-called "interferoscope", a grazing-incidence interferometer for non-optical surfaces (see also Abramson 1969b, Briers 1971). The problem of a "unified theory" of holographic interferometry, to explain all types of fringe pattern, was attacked separately by Lurie (1968b) and by Stetson (1968). Lurie produced a formula of which the previously reported formulae for vibrations (Powell and Stetson 1965b) and for constant velocity (Redman 1967) were special cases. Stetson derived a similar formula and applied it to the specific cases of pure translation, pure rotation, pivot motion (rotation about an axis normal to the line of sight, but not necessarily in the plane of the object), screw motion (translation along the axis of rotation), and in-plane rotation. He also considered the problem of fringe localisation, defining the surface of localisation in terms of maximum visibility of the fringes. Both authors later developed these general formulations further (Zambuto and Lurie 1970, Stetson 1969, Molin and Stetson 1970).

Still at the 1968 Strathclyde Symposium, Viénot et al. (1968) described an investigation of the fringe patterns of holographic interferometry. The object that they used was a rectangular metal plate oriented perpendicular to the plane containing the source, the hologram, and the observation point (defined as the centre of the aperture of the camera used to photograph the resulting fringe patterns). Photographs of typical fringe patterns were presented for the cases of in-plane translation, line-of-sight translation, rotation about an axis in the plane of the object, rotation about an axis parallel with the line of sight, and a so-called "coplanar" combination of translation and rotation. These results were then used to analyse fringe patterns according to whether the fringes were (i) equally spaced straight lines localised on the reconstructed image (caused by rotation about an axis normal to the line of sight), (ii) equally spaced straight lines localised behind or in front of the image (coplanar displacement), (iii) closed rings (translation along the line of sight). The authors introduced the concept of "homologous rays" - rays from the same point on the object, which may be travelling in different directions after the object is displaced. It was claimed that the contrast of the interference pattern would be highest, and hence the fringes localised, at the intersection of these homologous rays. The fringe localisation (FL) method was used to calculate in-plane translations from the fringe patterns. During the discussion of the paper, a question from the floor by Denby, about the method used to determine the plane of localisation of the fringes, was not answered very satisfactorily by the authors. (It will be remembered from the previous section that the main disadvantage of the FL technique is the difficulty of locating the fringe plane). An extended version of this paper was later published by Viénot (1970).

A later paper by the authors of the paper discussed above (Froehly et al. 1969) introduced a so-called "second class" of straight interference fringes which were visible at the "image of the source" (i.e. in the undiffracted, or zero-order, beam after passing through the hologram). These fringes were claimed only to exist if the object had undergone pure displacement without deformation. It was suggested, therefore, that this "second class" of fringes could be used to differentiate between deformation and pure displacement. We shall return to the subject of these fringes in Section 3.2.8, but we might perhaps point out at this stage that the fringes appear to be the same fringes that are observed in speckle interferometry (Archbold et al. 1970, Leendertz 1970).

In 1969, Boone and Verbiest (1969) applied the fringe localisation technique to the cases of bending and in-plane translations. For the bending experiments they used the simple rotation formula of Haines and Hildebrand (1966a) (see Section 3.2.2), and they improved on the latter authors' technique for measuring in-plane translation by using the real (conjugate) image from the hologram; in this way the fringes could be projected directly on to a screen and localised (by the criterion of maximum visibility) much more accurately. The homologous ray concept of localisation introduced by Viénot et al. (1968) was also invoked.

A detailed description of the fringe patterns produced by different types of displacement was given by Tsujiuchi et al. (1969) (see also Tsujiuchi and Matsuda 1970). Photographs were given of fringe patterns due to displacement along the line of sight (concentric Haidinger fringes, located at infinity in collimated light), in-plane displacement (straight, equally spaced Brewster fringes, located at infinity in collimated light), and rotations (straight, equally spaced fringes localised on the image). Formulae were given in each case connecting the fringe spacing with the displacement or rotation; the problem of the fringes being at infinity in some cases was overcome by using the <u>angular</u> fringe spacing, though it was not stated how this was measured. The paper went on to explain how deformations along the line of sight could be regarded as rotations about an axis in the plane of the surface, and how in-plane deformations could be treated as in-plane rotations. A method of differentiating between deformations and rigid-body rotations was then proposed, using information about the fringe orders at three points on the surface of the image to calculate the deformation separately. Rigid-body translations, it was suggested, could also be obtained by measuring the angular fringe spacing and off-set of the centre of the Haidinger fringe system mentioned above. Hence, it was claimed, it was possible to measure and differentiate between rotations, displacements and deformations from a single hologram, merely by separating the fringes localised on the surface of the image from those localised at infinity - and this separation could be achieved simply by using an aperture to vary the depth of field of the observing system. We shall return to this possibility of obtaining full three-dimensional information about the motion of an object from a single hologram and the use of the off-axis Haidinger rings in the next chapter.

An interesting paper by Sollid (1969) reviewed the FC interpretation method used by Aleksandrov and Bonch-Brüevich (1967) and the ZF technique of Ennos (1968), pointing out the disadvantages of each. Sollid then gave a "general theory" of fringe formation for two-dimensional motion, and showed that the resulting very simple formula was consistent with both the aforementioned techniques. The generalisation of the formula to three dimensions, however, led to a very complex set of equations. (A fundamental error in the derivation of these equations was later pointed out and corrected, though without any resultant simplification of the equations, by Vest (1973)).

Ennos' (1968) zero-order fringe (ZF) technique was reviewed in an important paper by Gates (1969), who pointed out the need to use three separate holograms in order to measure all three components of a motion. He then showed that this restriction only applies if a small portion of the hologram is used, as when the fringe pattern is photographed. If the full area of the hologram is used, three-dimensional information can be obtained from a single hologram. If this is done all at once, for example by using a large aperture for photographing the fringe pattern, the result is merely to blur out the fringes, since the patterns due to different illumination and viewing directions are all superimposed. Gates described four methods of using the full area of the hologram without losing the information in this way. The first technique consisted of placing a small aperture stop at the point on the object under consideration. The interference pattern then seen on the surface of the hologram "is a series of curved bands whose spacing and orientation give the transverse displacement (of the point)..... ....in two orthogonal directions and whose curvature shows the displacement in the direction of the centre of the ring pattern" (Gates 1969). Each point on the object was treated in turn, and "a contour map of the three-dimensional distortion.... ....built up" (Gates 1969). This was the first account of a technique which we shall call the hologram fringe (HF) technique, and which we shall describe in more detail in Section 3.2.7. Gates' second method of obtaining three-dimensional motion from a hologram involved the recording of two separate holograms, one of each state of the object, and the manipulation of one of them until the fringe pattern observed on the superimposed reconstructed images was "fluffed out". The movement of the hologram then indicated how the object had moved between the two exposures. (The technique of using two separate holograms had been introduced earlier by Gates (1968) for use with phase objects; the method was subsequently used, in various forms, by Havener and Radley (1972), Gori and Mallamace (1973), Hariharan and Hegedus (1973), and Abramson (1974b,1975)). The third technique described by Gates was a "real-time" version of the second, in which the distorted object was superimposed on the undistorted reconstructed image from a hologram and the object (or the hologram) moved until the fringes observed at the point on the object being investigated did not change shape or position when the angle of view was changed; the applied motion of the object (or the hologram) had then compensated for the original displacement, and the process could be repeated for other points on the object. This and the previous technique we shall call the "null" method of interpretation. The fourth and final method described in this paper was the fringe counting (FC) technique of Aleksandrov and Bonch-Bruevich (1967) - though Gates did not give a reference to the latter paper. The method is only applicable to in-plane motion, and Gates suggested the use of one of the null techniques for measuring the line-of-sight displacement of an object.

A detailed description of a practical application of holographic interferometry was given by Sampson (1970). He used Ennos' (1968) zero-order fringe (ZF) technique, but chose three carefully selected and closely controlled positions for the three holograms used. This was done in order to facilitate the analysis, and the resulting formulae are certainly easy to apply, though the technique assumes that the object can be positioned and manipulated at will. The paper also described a multiple-exposure method of fringe sharpening, a moiré technique for measuring the <u>differences</u> between two loads, using <u>two</u> double-exposure holograms on the same plate, the two-liquid method of contouring, and flaw detection in an aluminium honeycomb structure vibrated at its resonant frequency. Another specific application of holographic interferometry was described by Wilson (1970a), who used the technique to investigate the rotation of a solid cylinder and the torsion of a flexible shaft. The analysis of the fringe patterns was from first principles.

The problem of fringe localisation was attacked by several writers in the late 1960's. Tsuruta et al. (1969) and Walles (1970a) used fringe visibility to define the plane of localisation, while Welford (1969, 1970a) used zero parallax as the criterion. This problem of fringe localisation will be discussed in detail in Chapter 3.3.

Other papers on holographic interferometry written during the three years 1967 to 1969 included a moiré interpretation of the fringes, using the Ronchi test as an analogy (Pastor et al. 1970), a further contribution on the two-liquid technique of contouring (Zelenka and Varner 1969), the measurement of combination mode patterns in vibration analysis (Molin and Stetson 1969), the application by Abramson (1970a) of his holo-diagram to the fringe counting (FC) method of interpreting holographic interferograms, some methods of eliminating gross object motion when small deformations are being investigated (Champagne and Kersch 1969), and a multiple-wavelength technique for reducing the sensitivity of holographic interferometry (Varner 1970).

3.1.3 - Contributions of the 'Seventies

The past six years have produced a large number of papers on holographic interferometry, many dealing with the problems of interpretation. The zero-order fringe (ZF) technique, first used by Ennos (1968), has received the most attention, probably because it is conceptually the most pleasing, being based on the first-principles approach of classical interferometry and needing no information about fringe localisation. Although most writers concede that three separate holograms must be used if three-dimensional displacements are to be measured, a paper by Aplin et al. (1970), presented at the Besançon Symposium on "Applications of Holography" (Viénot et al. 1970) described a technique using only two holograms, at 90^{o} to each other, to deduce the three components of linear translation and the rotations about the three axes. Plane-wave illumination was used, and the illumination and viewing directions were coincident. A paper by Bijl and Jones (1970), delivered at the same conference, described the more usual technique of using three mutually orthogonal holograms. A detailed analysis technique was later published by the same authors (Bijl and Jones 1974, Jones 1974). A modification, using two illumination beams in order to separate the in-plane and the normal components of the displacement, was described at the same meeting by Luxmoore and House (1970), and this idea of multiple illumination directions was taken further by Wall (1970). The three-hologram technique was later described in detail by Shibayama and Uchiyama (1971), their analysis being in terms of direction cosines, and by Hecht et al. (1973), who used a vector analysis approach. A vector approach was also used by Macáková (1974) to obtain the direction and magnitude of the displacement from a _single_ hologram. The problem of establishing a common datum point between the separate holograms in the multiple-hologram method was overcome by Abramson (1972), who used a length of elastic rubber strip fixed at one end to the object and at the other end to the (stationary) holographic bench. Using the same technique, Sciammarella and Gilbert (1973) claimed an accuracy of $\lambda/8$ for the ZF technnique, after measuring fringe positions to an accuracy of $\lambda/100$. The rubber strip idea was also used by Hung et al. (1973), who suggested that an alternative to using three separate holograms was to use a single hologram plate and to record three double-exposure holograms on it using three different illumination directions. This technique avoids the need for fringe projection inherent in the original three-hologram method, since all the fringe patterns can be photographed from the same viewing direction. The need for some simplification of the analysis if holographic interferometry were to become a widely used engineering tool was stressed by Hansche and Murphy (1972, 1974), who showed that if the geometry of the object and the expected direction of the displacement are known a priori, then the analysis from first principles can be very straightforward. Specific applications of the ZF technique were described by Wilson (1971a), Burchett and Irwin (1971), Michael (1973), and Matsumoto et al. (1974). Comparisons of the ZF technique with the fringe counting (FC) and fringe localisation (FL) techniques were made in papers by Sollid (1970b) and Hansche and Murphy (1972). The combination of the ZF technique (to measure line-of-sight displacements) with speckle interferometry (to measure in-plane displacements) has been suggested by Adams and Maddux (1974). (This approach is also favoured by the NPL, and will be discussed again in Chapter 3.4).

The fringe counting (FC) technique, introduced originally by Aleksandrov and Bonch-Bruevich (1967), has also been developed in recent years. Kohler (1974) proposed the use of cut-out masks in the hologram plane to dictate the way in which the viewing direction is changed. Landry and Wise (1973) proposed a semi-automatic method of data analysis for the FC technique, and a similar scanning method, but using the real image from the hologram, was suggested by Bellani and Sona (1974). A paper by King (1974) combined the advantages of the two-hologram methods (the accurate measurement of all the components of motion) with the simplicity of the FC technique. He used two holograms mounted perpendicularly to each other. The FC technique has been compared with the ZF (zero-order fringe) and FL (fringe localisation) techniques by several authors, including Sollid (1970a,b) and Hansche and Murphy (1972).

The fringe localisation (FL) technique, pioneered by Haines and Hildebrand (1966a), has also had its adherents. Monneret (1970) used the homologous ray concept of Viénot et al. (1968) to define the plane of localisation of the fringes, while Dubas and Schumann (1974) used the criterion of maximum fringe visibility, The method was also used by Ashton et al. (1971), and was discussed in two papers by Stetson (Molin and Stetson 1971, Stetson 1974a). Hansche and Murphy (1972) compared the technique with the fringe counting (FC) method.

The hologram fringe (HF) technique, first proposed by Gates (1969, 1970) has been taken up and developed by other workers, notably by Boone and De Backer (Boone 1972, Boone and De Backer 1973) and by Ewers et al. (1974). A similar technique was used by Marom and Mueller (1970).

Fringe localisation has remained a focus of attention, with contributions by Stetson (1970b,1974a), Molin and Stetson (1971), Walles (1970b), Welford (1970b), Steel (1970), Machado Gama (1973), Dubas and Schumann (1974), and Přikryl (1974). This subject will be discussed in more detail in Chapter 3.3.

The application of moiré techniques to holographic interferometry has generated some considerable interest. Boone (1970a,b) used a double-illumination method similar to that proposed by Butters (1968) in order to measure in-plane deformations by means of a moiré technique. Twin apertures in the hologram plane were used by Velzel (1973) to produce moiré fringes which were contours of equal in-plane displacement, while Der et al. (1973) used a four-exposure technique to obtain moiré fringes which gave the difference between two displacements of the object. Moiré fringes were also used by Hariharan and Hegedus (1974) in a method designed to eliminate the effects of spurious movements of the object and to record only symmetrical deflections. Moiré fringe theory has also been used to explain and interpret holographic interferograms - the fringes can be regarded as moiré fringes between the two complex gratings making up the two holograms. This line of approach has been developed by such authors as Abramson (1971a,b 1973) and Gori and Mallamace (1973).

Abramson has developed his holo-diagram technique in a series of papers (Abramson 1970b,c, 1972, 1973), and has also introduced an "analogue computer", based on the holo-diagram and stretched strings, to facilitate the interpretation of holographic interferograms (Abramson 1972, 1973). The use of the holo-diagram has been extended to time-averaged holography and the study of vibrations by Bjelkhagen (1973). An interesting set of photographs of holographic interferograms was also published by this team (Abramson and Bjelkhagen 1973).

A detailed account of vibration analysis by time-averaged holography is outside the scope of this thesis, but it should perhaps be recorded that a considerable amount of work has been put into this branch of holography. The reader is referred to the literature for further details of this topic, and particularly to the contributions of Wilson and Strope (1970) and Wilson (1970b,1971b) on vibration mode patterns, Hazell and Liem (1970), Moffat and Watrasiewicz (1970), Mottier (1970), and Waddell et al. (1970) on stroboscopic techniques, Stetson (1971, 1972a,b,c) on non-sinusoidal vibrations and higher-order fringes, Bjelkhagen (1973) on the use of Abramson's holo-diagram, and Vikram on vibrating objects undergoing constant acceleration (Vikram 1973a) and on a technique for extending the range of time-averaged holography (Vikram 1973b).

Several techniques using two or more separate holograms, instead of either one double-exposure hologram or a single-exposure hologram and the original object, have been proposed. Havener and Radley (1972) called their method, which was used with phase objects, "dual-hologram interferometry". A similar approach, also using phase objects, was described by Gori and Mallamace (1973). A multiplex technique, in which several images of the object in different states were recorded either on separate hologram plates or on different parts of the same plate, was described in a paper by Hariharan and Hegedus (1973). Finally, a novel approach by Abramson (1974b,1975) uses two holographic plates sandwiched together; the holograms can be manipulated either separately, or together as a single "sandwich hologram", as an aid to fringe interpretation.

Other contributions to the field of holographic interferometry during the past few years have included the further development by Stetson (1970a, 1974~~a~~, 1975) of his generalised mathematical theory of fringe formation, the application of holographic interferometry to moving objects (Lohmann 1970), the use of a double-pulsed laser for work with impact-loaded objects (Gates et al. 1972), a high-speed technique for phase objects (Pasteur 1970), the application of the technique to photoelastic stress analysis (Bazelaire and Prade 1970, Nicolas 1970), flaw detection (Leadbetter 1970), and the stretching of rubber (Ashton et al. 1971), a further paper on the two-wavelength method of contouring (Robertson and Elliott 1970), the use of projected fringes to reduce the sensitivity of the technique (Rowe 1971), the use of temporally modulated illumination (Aleksoff 1971), considerations of errors and the accuracy attainable in holographic interferometry (Matsumoto et al. 1973, Sciammarella and Gilbert 1973), the use of a small frequency shift to achieve fringe interpolation down to better than $\lambda/100$ (Dändliker et al. 1973), and a Fourier transform method of analysis (Tribillon and Miles 1974).

This survey of the literature of holographic interferometry has, of necessity, been very cursory. A detailed review of the subject is outside the scope of this thesis, and would be a major undertaking. We believe, however, that all the major contributions - and quite a few of lesser importance - have been mentioned in this chapter, and can hence be identified from the bibliography listed at the end of this thesis. We now turn to the branch of holographic interferometry which is of most direct concern to us in this part of the thesis, the question of the interpretation of holographic interferograms.

CHAPTER 3.2

A CLASSIFICATION OF METHODS FOR THE INTERPRETATION OF HOLOGRAPHIC INTERFEROGRAMS

3.2.1 - The Confusion in Holographic Interferometry

Since the introduction of holographic interferometry in 1965 (Powell and Stetson 1965a, Burch 1965, Hildebrand and Haines 1966a) the field has become somewhat confused by the very large number of papers that have been published on the problem of interpreting the fringe patterns obtained. Most authors tend to derive their own methods of interpretation, often designed around their own particular application, but sometimes of more general interest. It is with these more general interpretation techniques that we are mainly concerned in this thesis, but even here there is some considerable confusion, if not conflict. Although most of the proposed techniques have been derived from first principles and are mathematically sound, there has been little attempt in the literature either to classify the various methods, or to compare them and demonstrate their equivalence (or contradictions), or even to suggest which technique should be used in a particular application. It is the aim of this part of the thesis to attempt to fill this gap. A thorough search of the literature of holographic interferometry has been carried out, and the various interpretation techniques have been analysed and classified.

In general, the interpretation schemes can be divided into four main techniques, which we shall call the fringe localisation (FL), fringe counting (FC), zero-order fringe (ZF) and hologram fringe (HF) techniques. There are also a few other methods, of lesser interest, which do not fall into any of these classes. We should stress that, in general, all of the proposed interpretation techniques encountered in the literature are equally valid and sound, and the confusion about the subject lies not so much in the question "Which approach is correct?" as in the question "Which approach is more appropriate to my particular application?". There is, however, one aspect of the subject where some real confusion has arisen, and this is the problem of fringe localisation. Since the localisation of the fringes plays a direct part in the FL technique, and is used implicitly in the FC technique, the question would appear to be of some importance, and we shall devote a separate chapter to the topic (see Chapter 3.3). Meanwhile, the present chapter will be devoted to a detailed description and discussion of the main interpretation techniques.

3.2.2 - The Fringe Localisation (FL) Technique

The fringe localisation method was historically the first reported technique for interpreting holographic interferograms, it being the method used in the original paper of Haines and Hildebrand (1966a). The technique relies on the fact that, in general, fringes caused by translation of the object are localised some distance behind the image. This plane of localisation is identified and its distance from the object is measured (see Figure 3.1). The following formulae are then used to determine the in-plane components of the translation (i.e. the components in the plane perpendicular to the line of sight):

$$d_x = \frac{\lambda R}{h_x} \quad ; \qquad d_y = \frac{\lambda R}{h_y} \qquad\qquad 3.1$$

where d_x and d_y = translation components parallel to the x and y directions;

λ = wavelength of the light;

R = distance from the object to the plane of localisation of the fringes;

h_x and h_y = fringe spacings parallel to the x and y directions, measured in the localisation plane.

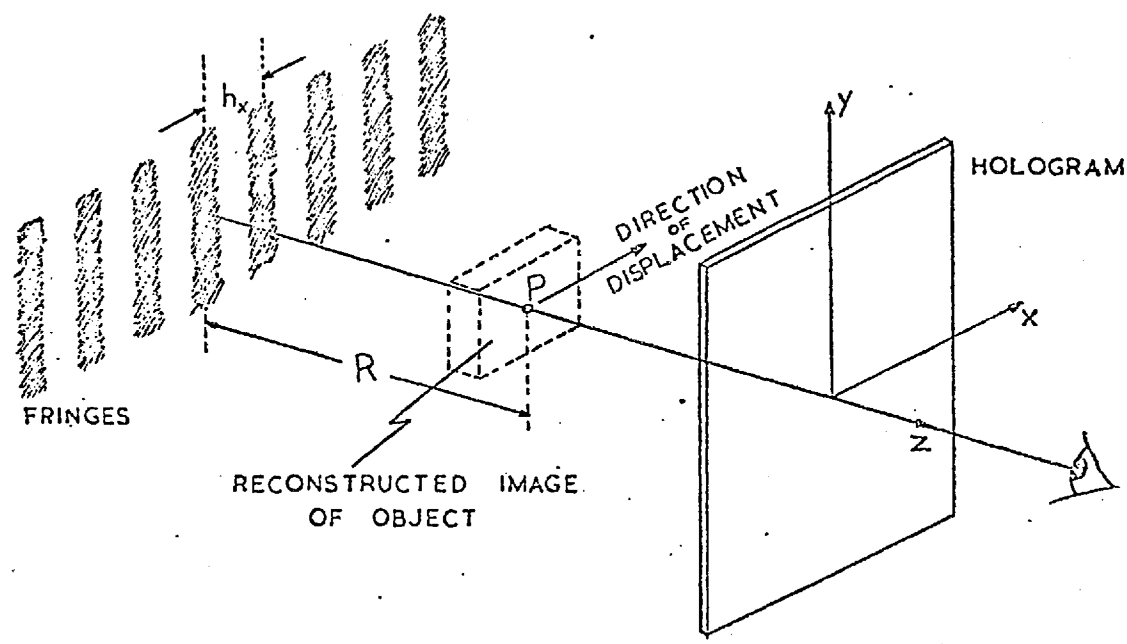


Figure 3.1 - Parameters used in the fringe localisation (FL) technique.

Formulae equivalent to Equation 3.1 have also been given by Viénot et al. (1968), Froehly et al. (1969), Monneret (1970), Boone and Verbiest (1969), and Stetson (1970b, 1974a).

If collimated illumination of the object is used, an arrangement often adopted by workers in holography, it is found that the fringes produced by in-plane translation are localised at infinity. In this case the formulae of Equation 3.1 cannot be used; however, an equivalent expression for use in this case has been given by Tsujiuchi et al. (1969):

$$d_x = \frac{\lambda}{\chi_x} ; \qquad d_y = \frac{\lambda}{\chi_y} \qquad 3.2$$

where χ = angular fringe spacing.

It can be seen by inspection that Equation 3.2 is equivalent to the Haines and Hildebrand formulae of Equation 3.1, providing $h \ll R$.

The line-of-sight component (z-component) of the translation can be found, in principle, by repeating the measurements of fringe spacing and localisation from a different viewpoint (i.e. by using a different portion of the hologram). The "in-plane" components will now be in a different plane from those measured in the first instance, and by resolving the two sets of measurements the z-component of the motion can be deduced. Unfortunately, the size of the hologram is usually too small to allow any significant change in viewing direction, and the accuracy with which the z-component can be measured is very poor.

The technique is only applicable in cases where the fringes are localised in a plane remote from the object (or image) surface. Haines and Hildebrand (1966a) interpreted fringes that formed on the surface itself as being due to rotation about an axis in the plane of the surface (i.e. to a tilt of the surface).

The angle of rotation in this case can be calculated from the following formula:

$$\psi = \frac{\lambda}{h(1 + \cos\theta_1)} \qquad 3.3$$

where ψ = angle of rotation,(tilt), assumed to be small;
λ = wavelength of the light;
h = fringe spacing;
θ_1 = angle of incidence of the illuminating beam.

The axis of rotation is parallel to the fringes, which in the case of simple rotation are parallel straight lines. Equation 3.3 assumes that the object (or image) surface is viewed normally; if this is not the case, the formula should be modified as follows (Stetson and Powell 1966):

$$\psi = \frac{\lambda}{h(\cos\theta_2 + \cos\theta_1)} \qquad 3.4$$

where θ_2 = angle between the direction of view and the normal to the surface.

Formulae equivalent to Equation 3.4 have also been given by Viénot et al. (1968), Tsujiuchi et al. (1969), Welford (1970a), Boone and Verbiest (1969), Wilson (1970a), and Hecht et al. (1973).

If a combination of translations and rotations is present in the motion of the object, the interpretation of the fringes by the FL technique becomes very difficult. Haines and Hildebrand (1966a) did consider the general case of rigid-body motion and derived a set of simultaneous equations, the solution of which would give the various components of the motion. The equations are very complex, however, and difficult to apply in practice.

The main advantages of the FL technique are that only one hologram is needed, in principle, to obtain all three components of translation, and that it is not necessary to know either the absolute order of the fringes or the geometry of the recording process.

On the other hand, the technique suffers from several disadvantages. The actual hologram is required for measurement purposes - it is not possible to use a photograph of the reconstruction since some method must be available for locating the fringe plane. The technique cannot be used when the fringes localise on the surface of the image (such fringes are interpreted as being due to rotation, as described above), and is very difficult to apply when the localisation plane is close to the surface. The line-of-sight component is very difficult to measure with any accuracy, since only a limited change in viewing direction is possible with a typical hologram. (This problem could be overcome by using a separate hologram, recorded at right angles to the first, to measure the line-of-sight component of the movement, as is the usual practice with the ZF technique (see Section 3.2.5); this possibility does not appear to have been reported in the literature, though the adoption of such a procedure would, of course, result in the loss of one of the few advantages of the technique - the use of only one hologram). However, the major disadvantage of the FL technique is the difficulty of locating the plane of localisation of the fringes. This leads to very poor accuracy for the technique, and indeed Haines and Hildebrand (1966a), using the virtual image from the hologram, reported errors of up to ±60% due to this cause. The problem can be overcome to some extent by using the real reconstructed image from the hologram (Boone and Verbiest 1969), since the fringes can then be projected on to a screen which can be moved until the fringe visibility is maximised; however, it proves to be still difficult to localise the fringes with any degree of accuracy, a point which will be discussed further in Chapter 3.3. Finally, the complexities of analysing motions which are a combination of translations and rotations are another barrier to the widespread acceptance of the FL technique.

The only major improvements in the FL technique since its introduction by Haines and Hildebrand (1966a) have been the use of the real image, allowing more accurate location of the fringe plane, by Boone and Verbiest (1969), and the use of a slit aperture to improve the localisation of the fringes (Welford 1969, Stetson 1970b, 1974, Molin and Stetson 1971). The latter point will be discussed on more detail in Chapter 3.3. However, the technique has been largely superseded by the fringe counting (FC) technique (see the next section), which provides a simpler and more accurate method of obtaining the same information.

3.2.3 - The Fringe Counting (FC) Technique

The fringe counting technique was first proposed by Aleksandrov and Bonch-Bruevich (1967), and was later adopted by Gates (1969). This method also makes use of the fact that the fringes are, in general, localised some distance from the surface of the reconstructed image, but uses the parallax between the fringes and the image rather than the actual location of the fringe plane. The optical system used for viewing the fringes (this may, of course, be the unaided eye) is focused on the reconstructed image and is stopped down until the fringes have good visibility. The direction of view is then changed progressively by scanning the line of sight linearly across the hologram. The number of fringes which pass across the image point under consideration is counted as the viewing direction is changed by a known amount. The technique is illustrated in Figure 3.2. The method gives the component of translation of the point in a direction perpendicular to the bisector of the two extreme lines of sight and in the plane containing these lines of sight. This can be seen from Figure 3.2 to be approximately the in-plane component parallel to the direction in which the line of sight is scanned across the hologram. The

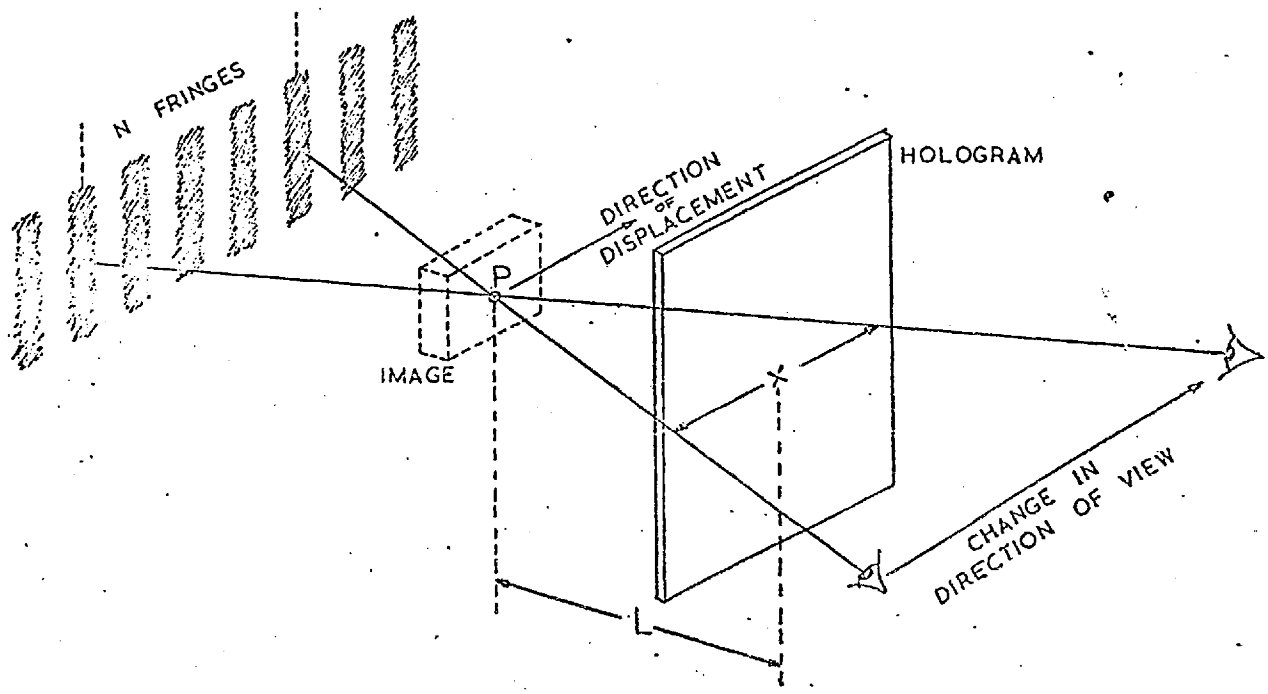


Figure 3.2 - Parameters used in the fringe counting (FC) technique.

magnitude of the component is calculated from the following formula:

$$d_x = \frac{N\lambda L}{x} \qquad 3.5$$

where d_x = component of translation as defined above;

N = number of fringes passing the point under consideration;

λ = wavelength of the light;

L = distance from the hologram to the reconstructed image;

x = distance scanned across the hologram.

This formula, or its equivalent, was used by Aleksandrov and Bonch-Bruevich (1967), Gates (1969), Sollid (1970a,b), Landry and Wise (1973), Kohler (1974), King (1974), and Bellani and Sona (1974).

Since the plane of localisation of the fringes is used only by implication (using the parallax between the fringes and the holographic image) and not by direct measurement, the actual physical location of the fringes is irrelevant in the FC technique. Thus Equation 3.5 is valid for all cases in which the fringes are localised sufficiently far from the image, even if the localisation plane is at infinity, as is the case when collimated light is used.

As stated above, the FC technique gives the components of translation of a point in the plane normal to the bisector of the extremes of the range of lines of sight used (see Figure 3.2), and hence the same "in-plane" components that would be measured by means of the FL technique using this bisector as the line of sight. As with the FL method, the z-component of the translation (along the bisector of the two extreme lines of sight) can be found by repeating the observation using a different portion of the hologram and combining the two sets of results. The same limitations on the accuracy with which the z-component can be measured apply equally to the FC technique as to the FL method.

Since the FC approach depends on parallax between the fringes and the holographic image, it obviously cannot be used when the fringes are localised on the surface of the image. As stated in Section 3.2.2, such fringes are caused by rotation (or tilt) of the object about an axis in the plane of the object surface, and are interpreted by means of Equation 3.4. However, if there is any parallax between the fringes and the image, the FC technique can always be relied upon to give the (approximately) in-plane components of the motion of each point on the image surface, and by measuring the motions of each point in turn a picture of the motion of the whole object can be built up, without the need to resort to the complex equations of the FL technique.

The advantages of the FC technique are largely those of the FL technique: the use of a single hologram and the irrelevance of the absolute fringe order. There is, however, one big advantage that the FC method enjoys over the FL approach - it is much easier to apply. This is partly because there is no need physically to find the plane of localisation of the fringes, but mainly because the displacement components of each point on the object can be found, even if the total motion of the object is very complex, without recourse to complicated formulae.

The main disadvantages of the FC technique are also shared with the FL method: the actual hologram is required rather than a photograph of the reconstruction, the technique cannot be used when the fringes are localised at or very close to the image, and it is very inaccurate for measuring the line-of-sight component. (The latter objection can be overcome, at the expense of losing the "single-hologram" advantage, by using the two-hologram technique usually associated with the ZF (zero-order fringe) technique; this approach has recently been used by King (1974)). Finally, although the FC technique does not suffer from the crippling disadvantage of the FL method, namely the problem of locating the fringe-plane with any degree of accuracy, it is nevertheless capable of only limited accuracy, even for the in-plane components of motion, unless very large holograms are used. This is because of the limit imposed on the change in viewing angle by the limited size of the hologram. Typical accuracies quoted in the literature are the $\pm 3\lambda$ of Sollid (1970b) and the ±5% of Landry and Wise (1973), the latter using automatic data reduction and fractional fringe-order measurement.

Since its announcement by Aleksandrov and Bonch-Bruevich in 1967, the FC technique has attracted the attention of many workers in holography, and several improvements and developments have been proposed. Among these we might mention the combination of the FC technique with Abramson's holo-diagram (Abramson 1970a,b,c), the introduction of automatic data reduction by Landry and Wise (1973), the use of a cut-out mask in the hologram plane by Kohler (1974) to dictate the direction and the extent of the change in viewing direction, the application to the FC technique of the two-hologram approach (King 1974), and the use of the real image in a scanning arrangement described by Bellani and Sona (1974). The FC techique remains a very useful holographic method, especially for rapid, semi-quantitative assessments of double-exposure holograms, or for detailed measurements of rigid-body in-plane translations.

3.2.4 - The Equivalence of the FL and FC Techniques

The reader will have noticed from Sections 3.2.2 and 3.2.3 that the FL and FC techniques are basically very similar. They both rely on the fact that the fringes caused by translation of an object are localised at some distance from the holographic image, and in fact the localisation of the fringes is used in calculating the translation. In the FL technique, this information is used directly, the distance between the image and the plane of localisation of the fringes being used as a parameter in the calculation; in the FC method the fringe localisation is used implicitly as the parallax between the fringes and the image. In addition, both techniques give basically the same components of the displacement - the in-plane translation. (More strictly, the FL technique gives the components of translation normal to the line of sight, and the FC technique gives the components normal to the bisector of the two extreme lines of sight). We shall now show that the two techniques are equivalent - an essential requirement if we are to have any confidence in either method. The proof is very elementary, though the only account of it that we have been able to find is in a recent article by Stetson (1974b).

Figure 3.3 represents a typical holographic interferometry experiment in which fringes, reconstructed image and hologram are as shown. The image is formed at a distance L from the hologram, and the fringes are localised at a distance R behind the image. Consider first the analysis of the fringes by the FL technique. The plane of localisation of the fringes is found (by the criterion of maximum visibility), and its distance R from the image point under consideration (P on the diagram) is measured. R and the fringe spacing, h_x, are used to calculate the transverse translation of point P, using Equation 3.1:

$$d_x = \frac{\lambda R}{h_x} \qquad 3.6$$

Now let us consider the FC technique, in which the observing system is scanned horizontally across the hologram and the number of fringes passing the point P is counted. As Figure 3.3 is drawn, it can be seen that just one fringe will pass the point P as the line of sight is scanned through the distance x, measured at the hologram. Hence N = 1 in Equation 3.5, which becomes:

$$d_x = \frac{\lambda L}{x} \qquad 3.7$$

By comparing Equations 3.6 and 3.7 it will be seen that the FL and FC techniques give the same answer for the translation component d_x, providing:

$$\frac{R}{h_x} = \frac{L}{x}$$

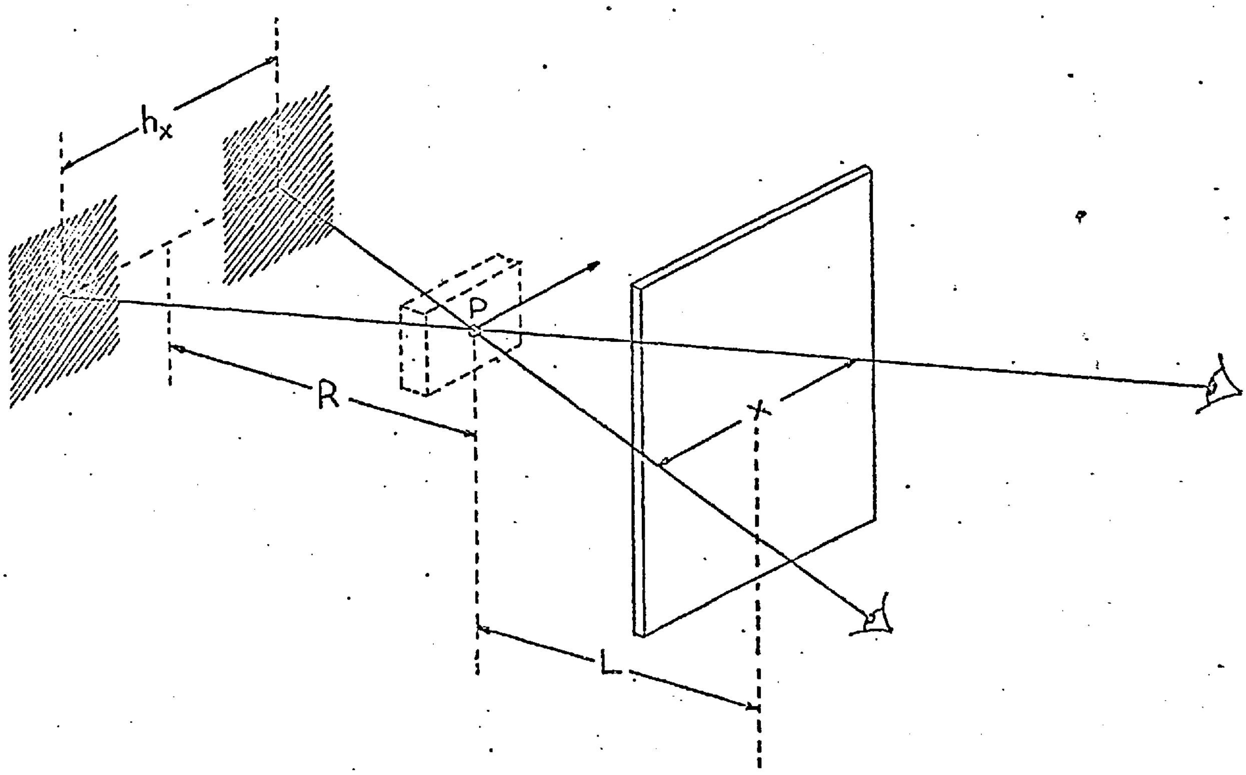


Figure 3.3 - Equivalence of the FL and FC techniques.

Inspection of Figure 3.3 shows that this equality is, in fact, true (by similar triangles), providing the localisation of the fringes by maximum visibility (as used in the FL technique) is equivalent to their localisation by parallax (as in the FC case). For the present we shall accept that this is the case, and hence that the FL and FC methods are exactly equivalent. The question of fringe localisation is, however, more complex than it might appear, and the problem will be discussed in more detail in Chapter 3.3.

3.2.5 - A Note on Haidinger Fringes

The FL and FC techniques both use the fact that the fringes are localised in a plane remote from the image to measure the in-plane components of translation of the object. They interpret any fringes localised at a distance from the image as being due to a displacement with an in-plane component. However, it has been reported by several authors that pure line-of-sight translation (i.e. displacement along the z-axis) also gives rise to fringe patterns with similar localisation properties. Gates described such fringes as "curved bands", while other authors were more specific and reported them as concentric circles centred on the line of sight (Viénot et al. 1968, Froehly et al. 1969, Tsujiuchi et al. 1969, Viénot 1970). These fringes can be used to measure the line-of-sight translation that produced them. Thus Tsujiuchi et al. (1969) used the angular fringe spacing of the pattern in a manner analogous to their formulae for in-plane translations (see Equation 3.2):

$$d_z = \frac{2m\lambda}{\chi_m^2} \qquad 3.8$$

where m = fringe-number, counting from the centre of the pattern;

$2\chi_m$ = angular subtense of the m'th fringe.

Viénot et al. (1968), on the other hand, used a fringe counting technique in which they counted the number of fringes crossing the field as the direction of view was changed:

$$d_z = \frac{N\lambda}{\Delta\gamma . \sin\gamma} \qquad 3.9$$

where N = number of fringes crossing the field of view as the viewing direction is changed by $\Delta\gamma$;

γ = angle between the viewing direction and the displacement vector.

The explanation of these fringes is that they are Haidinger fringes (fringes of equal inclination) and are due entirely to the variation in path difference across the object caused by the change in angle of view. They are identical to the fringes seen in the Michelson interferometer in diverging light. Although their origin is easy to explain, and they can be used to measure rigid-body translations along the line of sight, it is possible that they might be a source of error when using the FL or FC techniques to measure in-plane displacements. This is unlikely to be the case, however, since the holographic system is much less sensitive to rigid-body translations along the line of sight than it is to in-plane displacements. (This can be seen by inspection of Equations 3.2 and 3.8). Also, the concentric Haidinger fringes can usually be recognised for what they are. If a large line-of-sight component of translation is combined with a small transverse displacement, then the Haidinger fringes will modify the expected interference pattern. The resultant pattern will be an off-centre set of concentric circles and can be used to measure both the in-plane and the line-of-sight components of the motion (Tsujiuchi et al. 1969). Finally, if it is still feared that the Haidinger fringes might be a source of error in using the FL and FC techniques, they can always be eliminated by arranging for the directions of illumination and viewing to be constant over the whole object. This can be achieved by using collimated illumination and viewing: rigid-body translation along the line of sight now introduces the same change in optical path for all points on the object, and no fringes will be observed.

3.2.6 - The Zero-order Fringe (ZF) Technique

The zero-order fringe method was first described by Ennos (1968), although it is really the direct application of classical interferometry techniques to holographic interferometry. It is based on the first-principles argument that the change in optical path from the source to the observer (or the detector) is related to the fringe order by the following equation:

$$\Delta = m\lambda \qquad 3.10$$

where Δ = change in optical path;
m = fringe order;
λ = wavelength of the light.

The fringe order, m, can only be determined absolutely if a known zer-order fringe is in the field of view. In many cases it will be known that a particular point on the object has remained stationary between the two holographic exposures, and the fringe orders can be determined by counting fringes from such a point. A technique which has been used by several authors (Abramson 1972, Sciammarella and Gilbert 1973, Hung et al. 1973, Matsumoto et al. 1974) consists of fixing one end of a strip of rubber to a point on the object, and fixing the other end firmly to the holographic bench: fringe counting can then be commenced from the latter point, which is known to have undergone zero displacement. Ennos (1968) and Sollid (1969) suggested that a zero-order fringe could be identified from the fact that there would be no parallax between it and the image; this approach must be adopted with caution, however, since although it is true if it is known that the object has undergone pure translation, we have already seen that rotations about an axis in the plane of the object produce a system of parallel fringes, all of which display no parallax with the image.

Once the fringe order at all points on the image is known, Equation 3.10 gives the change in optical path for each point. The problem then becomes one of pure geometry - relating the change in optical path for each point to the actual displacement of that point. There are two schools of thought regarding this problem. One approach is to consider the general case of arbitrary illumination and viewing conditions and to derive a set of equations from which the three components of the displacement can be calculated. Other authors have taken the view that the resulting expressions are so complex for the general case that it is better to treat each case on its own merit and to devise an experimental arrangement which will simplify the calculation of the displacement from first principles.

Considering the generalised approach first, it should be pointed out that, in general, three separate holograms will be required if all three components of the displacement are to be measured. Each hologram gives the component of the displacement parallel to the bisector of the illumination and viewing directions (Ennos 1968). Most authors have used vector analysis methods to derive equations connecting the displacement components of each point on the object with the path differences calculated from the three holograms. In his original paper on the technique, Ennos (1968) predicted that this would be a "formidable task" in the general case, and this has certainly proved to be so. The problem has been attacked by several authors, among whom we might mention Sollid (1969), Stetson (1969), Shibayama and Uchiyama (1971), Hecht et al. (1973), and Bijl and Jones (1974 - see also Jones 1974), but the resulting equations are usually very complex, and either difficult or tedious to apply in practice. The equations become more manageable, at the expense of some loss of generality, if the viewing direction (i.e. the positions of the three holograms) are specified in certain ways. For example, Sampson (1970), in a paper which perhaps deserves to be more widely known than it is, described a technique in which the object under test is rotated about specific axes for the second and third holograms, the illumination and viewing conditions remaining fixed. The full procedure for each point on the object is as follows:

(i) Take double-exposure hologram (1) with any convenient experimental arrangement; measure the fringe order m_o for the point under consideration.

(ii) Rotate the object through an arbitrary angle ψ_1 about the x-axis (the horizontal axis normal to the line of sight), and take double-exposure hologram (2); measure the fringe order m_1.

(iii) Calculate the angle ε between the displacement vector and the x-z plane (the horizontal plane) from the following equation:

$$\tan\varepsilon = \frac{m_o(\cos\psi_1 + \cos(\psi_1-\theta)) - m_1(1+\cos\theta)}{m_o(\sin\psi_1 + \sin(\psi_1-\theta)) + m_1\sin\theta} \qquad 3.11$$

where θ = angle between the illumination direction and the line of sight (z-axis).

(iv) Rotate the object through this angle ε about the x-axis, so that the displacement vector now lies in the (horizontal) x-z plane.

(v) Rotate the object through an arbitrary angle ψ_2 about the (vertical) y-axis, and take double-exposure hologram (3); measure the fringe order m_2.

(vi) Calculate the angle γ_x between the displacement vector and the x-axis from the following equation:

$$\tan\gamma_x = \frac{m_o\sin\psi_2}{m_2 - m_o\cos\psi_2} \qquad 3.12$$

(vii) Calculate the magnitude d of the displacement vector from the following equation:

$$d = \frac{m_o\lambda}{\sin\gamma_x(\cos\varepsilon + \cos\varepsilon\cos\theta + \sin\varepsilon\sin\theta)} \qquad 3.13$$

We have changed the notation in the above equations from that used by Sampson in order to make the equations compatible with the notation used in this thesis. In addition, we have replaced the factor $\frac{2m - 1}{2}$ which occurs in the original version of the equations by the simple factor m; the discrepancy arises from the fact that Sampson was assigning fringe-order numbers to dark fringes in a double-exposure hologram. (This may be an opportune moment to point out that in double-exposure (frozen-fringe) holographic interferometry the zero-order fringe is a bright fringe, while in the real-time (live-fringe) technique it is a dark fringe (Ennos 1968); it is important to remember this when interpreting fringe patterns and applying the equations quoted in this part of the thesis).

Of course, it is not always convenient, or even feasible, to rotate the object in the way required by Sampson's method, but we have described the technique in full in order to show how the complexities of fringe interpretation can be simplified by a careful choice of the viewing directions for the three holograms. In Sampson's method, the three views are obtained by rotating the object about selected axes, but other methods are equally valid and may often be more appropriate. Hung et al. (1973), for example, suggest using a single position for the object and a single hologram plate, and recording three separate double-exposures on the one plate, using three different and carefully chosen directions of illumination. Collimated light is used to illuminate the object, and the three incident-beam directions chosen are:

(i) along the z-axis (line of sight), using a beam splitter;

(ii) at an arbitrary angle θ_1' to the z-axis, and in the y-z plane;

(iii) at an arbitrary angle θ_2'' to the z-axis, and in the x-z plane.

If the fringe-order numbers associated with each hologram are m_o, m_1 and m_2, then the three components of the displacement are given by the following set of equations:

$$\left.\begin{aligned} d_x &= \left(\frac{2m_2 - (1 + \cos\theta_1'')m_o}{2\sin\theta_1''}\right)\lambda \\ d_y &= \left(\frac{2m_2 - (1 + \cos\theta_1')m_o}{2\sin\theta_1'}\right)\lambda \\ d_z &= \frac{m_o\lambda}{2} \end{aligned}\right\} \qquad 3.14$$

Another way of avoiding complex formulae is to use Abramson's "holo-diagram" (Abramson 1968, 1969a). The direction of the component of displacement measured by a particular configuration is given by the normal to the holo-diagram ellipse which passes through the point on the object under consideration; the holo-diagram also gives the sensitivity, k, for the configuration, from which the magnitude of the displacement component can be calculated from the simple equation:

$$d = \frac{km\lambda}{2} \qquad 3.15$$

where m = fringe order.

Before leaving the generalised interpretation schemes for the ZF technique, we should mention three more significant contributions to the subject. Shibayama and Uchiyama (1971) gave a matrix formulation for calculating the three components of the displacement in terms of the direction cosines of the incident light and of the three viewing directions used. An interesting approach by Bijl and Jones (1974), followed up later by Jones (1974), combined the vector equations into a single tensor equation expressing the displacement in terms of fringe spacing rather than absolute fringe order; using this approach there is no need to identify a zero-order fringe, and the method represents a link between the ZF technique and the FC and FL methods, which also use the fringe spacing as a parameter. A similar technique, but using vector rather than tensor notation and restricting the treatment to plane-wave illumination, was suggested by Macáková (1974).

The main simplification of the general method is to use collimated illumination and viewing, to view the object surface (if flat) normally, and to arrange for the illumination and viewing directions to coincide. This arrangement, illustrated in Figure 3.4, greatly simplifies the problem of interpreting holographic interferograms by means of the ZF technique, and has been adopted by most workers in the field as the standard technique. Since the displacement component measured by the ZF technique is

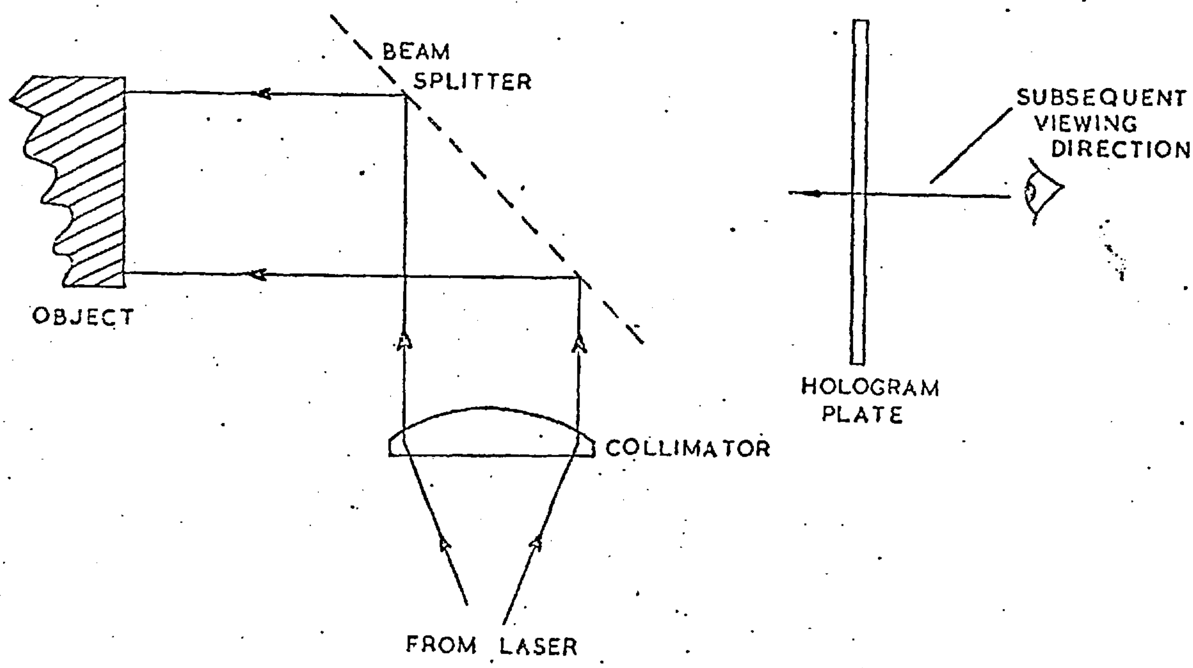


Figure 3.4 - Simplified geometry for the zero-order fringe (ZF) technique: collimated, normal and coincident illumination and viewing.

that in the direction of the bisector of the illumination and viewing directions (Ennos 1968), it follows that in the arrangement described above, and illustrated in Figure 3.4, it is the line-of-sight component that is measured, the necessary equation being simply the following (Hung et al. 1973):

$$d_z = \frac{m\lambda}{2} \qquad 3.16$$

If the direction of the actual displacement is known (as is often the case), and the illumination and viewing direction can be arranged to be along this direction, then Equation 3.16 gives the value of the actual displacement of the point under consideration. This is not always feasible, however, and the resolved component along the line of sight is the quantity usually measured by Equation 3.16. The magnitude of the actual displacement vector is then given by:

$$d = \frac{m\lambda}{2\cos\gamma} \qquad 3.17$$

where γ = the angle between the line of sight (and hence also the illumination direction) and the (known) direction of the displacement vector.

Equation 3.17, which is particularly useful in the case of in-plane displacements, is equivalent to the forms used by Ennos (1968), Wilson (1971a), Burchett and Irwin (1971), Hansche and Murphy (1972), and Michael (1973). If the illumination and viewing directions are not coincident, but are at an angle θ to each other (see Figure 3.5), then the ZF technique gives the component along the bisector of that angle by means of the following equation:

$$d_{\theta/2} = \frac{m\lambda}{2\cos\theta/2} \qquad 3.18$$

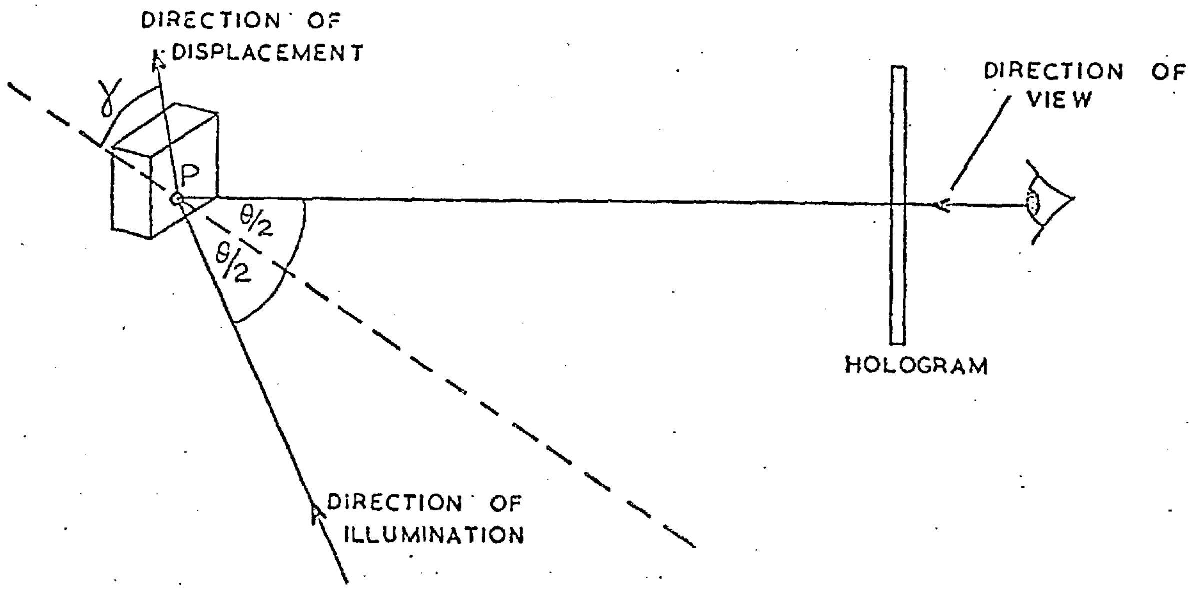


Figure 3.5 - General geometry for the ZF technique.

Again, if the direction of the displacement vector is known, its magnitude can be calculated directly:

$$d = \frac{m\lambda}{2\cos\theta/2 \cos\gamma} \qquad 3.19$$

where γ = the angle between the displacement vector and the bisector of the illumination direction and the line of sight.

We have assumed so far that we are interested in measuring the absolute displacement of points on the surface of the object. It is often the case, however, that it is only the <u>differential</u> displacements of the different points on an object that are important. This is particularly so when investigating deformations of objects. If this is the case, the requirement of the identification of the zero-order fringe no longer applies, and the ZF technique becomes a very simple fringe-counting method, especially if one of the simplified geometries mentioned above is adopted. Consider, for example, the case of normal viewing, collimated illumination, and coincident viewing and illumination directions, illustrated in Figure 3.4. If we are interested only in the difference between the line-of-sight displacements of two points, P and Q, on the object surface (as we might be, for example, in the case of line-of-sight deformation of the object caused by an applied load), then we can write two equations for the separate displacements of the two points (from Equation 3.16):

$$d_P = \frac{m_P\lambda}{2} \qquad 3.20$$

$$d_Q = \frac{m_Q\lambda}{2} \qquad 3.12$$

where m_P and m_Q = fringe-orders at P and Q.

Subtracting Equations 3.20 and 3.21 gives the difference between the two displacements:

$$\triangle d_z = \frac{(m_P - m_Q)\lambda}{2} \qquad 3.22$$

$(m_P - m_Q)$, the difference in fringe order, is, of course, simply the number of fringes that appear on the surface of the object between points P and Q, and can be obtained directly from the hologram (or from a photograph of a reconstruction). Hence the ZF technique provides a very simple method of differential strain analysis. (There is, however, a problem in determining whether the fringe order is increasing or decreasing as one goes from P to Q, or even whether there is a change of sign between the two points; several methods are available for overcoming this problem - see for example Přikryl 1975).

The main advantages of the ZF technique are that it can be used with photographs of reconstructions of the holographic images instead of with the actual holograms (providing a zero-order fringe can be identified if absolute displacements are to be measured), that the physical interpretation is simple (providing the geometry of the experimental arrangement is simplified), and that it is capable of high accuracy - Sciammarella and Gilbert (1973) reported an accuracy of $\lambda/8$, and Dändliker et al. (1973), using a double-frequency technique to interpolate between the fringes, claimed an accuracy of $\lambda/100$.

The major disadvantages of the method are the need for three separate holograms if a complete three-dimensional picture of the displacement is required, the need, in general, to identify a zero-order fringe (unless differential strains are being measured), and the complex equations that must be resorted to when general illumination and viewing conditions are used. Even when simplified geometries are used, the ZF technique always takes more time to carry out than does the FC method, which can be very quick if only an order-of-magnitude measurement of the displacement is required.

The main developments in the ZF technique since it was first used by Ennos (1968) have been on the one hand the attempts to produce generalised equations for interpreting the fringe patterns, and on the other hand the simplification of experimental arrangements so that such complex equations are not required. The method of Hung et al. (1973), who used a single hologram plate and three different illumination directions for three separate double-exposure holograms, and thus avoided the need for fringe projection, deserves special mention. Finally, the increase in accuracy to $\lambda/100$ by using double-frequency illumination, described by Dändliker et al. (1973), should also be noted.

The ZF technique can be used for any type of displacement or deformation of an object, though it is most useful when the fringes are localised on the image. This occurs when the object has undergone rotation about an axis in its own plane, or, more importantly, when it has been subjected to out-of-plane deformation (Tsujiuchi et al. 1969, Sollid 1970b). The technique is thus to some extent complementary to the FC and FL technqiues, which can only be used when the fringes are localised at some distance from the image, and which give the in-plane components of the displacement. However, the ZF technqiue is basically a more powerful and more accurate technique than either of the other two, and is the most widely used method of interpreting holographic interferograms.

3.2.7 - The Hologram Fringe (HF) Technique

The interpretation methods described so far have used a limiting aperture in the hologram plane through which to view the fringes. (Unless the fringes are localised on the image, the use of too large an aperture at the hologram has a detrimental effect on the visibility of the fringes). The complementary approach of placing a small limiting aperture at the image (or at the object in real-time holographic interferometry), and viewing fringes which appear in the hologram plane, can also be adopted. This technique was first reported by Gates (1969, 1970), and was later developed by Boone and De Backer (Boone 1972, Boone and De Backer 1973).

The basic experimental arrangement is illustrated in Figure 3.6. The real (conjugate) image is used for convenience, and a small aperture (typically less than 1mm in diameter) is placed at this image to isolate the point under investigation. A camera (or the unaided eye) is placed behind this aperture and focused on the hologram plane. Fringes are seen on the hologram; these fringes are straight and evenly spaced if the point selected by the aperture has undergone an in-plane displacement (the fringes are perpendicular to the direction of displacement, as in the FC and FL techniques), and they consist of concentric circles if the point has been displaced along the line of sight. Any combination of in-plane and out-of-plane displacement results in an off-centre pattern of curved fringes. Since the method investigates the absolute displacement of a single point on the image (selected by the aperture), there is no need to worry about whether the motion is due to translation, rotation or deformation; each point is analysed in turn, and a picture of the motion of the whole object is gradually built up. Similarly, we are not concerned in the HF technique with the intricacies of fringe localisation - it is the fringes in the plane of the hologram which are of interest.

In the case of in-plane displacement, the straight-line fringes are interpreted by means of the following equations (Gates 1970):

$$d_x = \frac{\lambda L}{h_x} \quad ; \qquad d_y = \frac{\lambda L}{h_y} \qquad\qquad 3.23$$

where d_x and d_y = displacement components in the x and y directions;

h_x and h_y = fringe spacings in the x and y directions;

L = distance from the hologram to the image (or aperture).

It will be noted that these equations have a very similar form to those used in the FL technique (Equation 3.1); however, the HF method

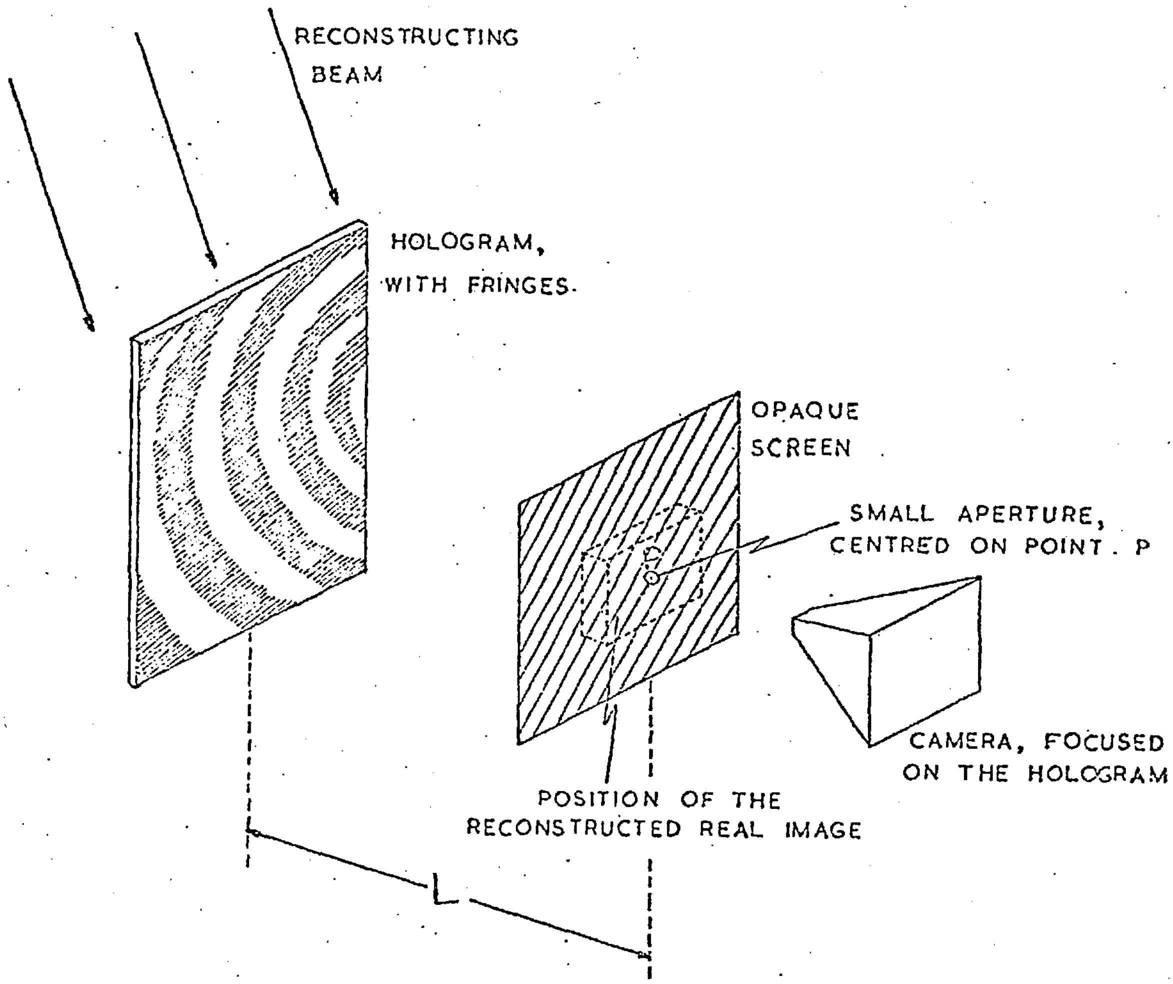


Figure 3.6 - Basic arrangement for the hologram fringe (HF) technique.

is much easier to apply since the parameter L is much easier to measure than the distance R in Equation 3.1, which involves locating the plane of localisation of the fringes.

The circular fringes caused by out-of-plane displacements are analysed as follows (Gates 1970):

$$d_z = \frac{2m\lambda L^2}{s_m^2} \qquad 3.24$$

where d_z = line-of-sight component of displacement;

m = fringe number (counting from the centre of the pattern;

s_m = radius of the m'th fringe.

Gates did not give formulae for the general case, but merely stated that the fringes would be "a series of curved bands whose spacing and orientation give the transverse displacements......and whose curvature shows the displacement in the direction of the centre of the ring pattern" (Gates 1970). It was left to Boone and De Backer (1973) to produce a general interpretation scheme. They showed that the fringe patterns are conic sections, particular cases being as follows:

(i) concentric circles indicate line-of-sight displacement (d_z);

(ii) off-centre ellipses indicate that the direction of $\underline{d}$ is within 45° of the z-axis (line of sight);

(iii) a central parabolic fringe indicates that $\underline{d}$ is at 45° to the z-axis;

(iv) hyperbolic fringes indicate that $\underline{d}$ is within 45° of the plane normal to the z-axis;

(v) a central straight fringe indicates in-plane displacement.

(The slight discrepancy between (v) above and Gates (1969), who stated that <u>all</u> the fringes are straight for in-plane displacements, is presumably a function of the geometry of the experiment, and particularly the question of whether or not collimated illumination is used).

Hence the shape and the orientation of the fringes (and particularly of the central fringe) can be used to determine the direction of the displacement vector $\underline{d}$ in relation to the line of sight (z-axis). Boone and De Backer (1973) suggested that this could be done either by computer, or by visual comparison of the fringe with a set of computed fringe shapes. (Their paper actually includes such a set of computed fringe shapes for this purpose). Denoting the angle between the displacement vector and the line of sight, obtained in this way, by γ, the following formula is used to determine the magnitude of the displacement (Boone and De Backer 1973):

$$d = \frac{N\lambda}{2\sin\gamma\sin\mu} \qquad 3.25$$

where N = number of fringes between two points on the hologram, the two points being collinear with the centre of the fringe system;

γ = angle between the displacement vector, $\underline{d}$, and the z-axis (line of sight), found from the shape of the central fringe;

$\tan\mu = \frac{x}{L}$;

$2x$ = the separation of the two points on the hologram;

L = distance from the hologram to the image (and hence to the aperture).

For the special case of in-plane displacement, $\gamma = 90^{\circ}$, and if we assume that $x \ll L$, so that $\tan\mu \sim \sin\mu$, Equation 3.25 reduces to:

$$d = \frac{N\lambda L}{2x} \qquad 3.26$$

Since in this case the fringes are straight and equally spaced, assuming collimated illumination (Gates 1969), $2x/N$ is equal to the fringe spacing, h, and Equation 3.26 reduces to the form used by Gates (Equation 3.23):

$$d = \frac{\lambda L}{h} \qquad 3.27$$

For the case of line-of-sight displacement, $\sin\gamma \to \infty$, and Equation 3.25 cannot be used. Boone and De Backer (1973) showed that in this case the appropriate formula is:

$$d_z = \frac{m\lambda}{1 - \cos\mu} \qquad 3.28$$

where m = number of fringes between the centre of the ring pattern (on the z-axis) and a point on the hologram a distance s_m from the centre (i.e. s_m = radius of the m'th fringe);

$\tan\mu = \frac{s_m}{L}$.

Again, if $s_m \ll L$, this reduces to the form used by Gates (Equation 3.24):

$$d_z = \frac{2m\lambda L^2}{s_m^2} \qquad 3.29$$

The main advantages of the HF technique lie in the fact that the absolute value and the direction of the displacement of each point on the object can be obtained from a single hologram by observing a two-dimensional fringe pattern situated in the hologram plane, without the issue being complicated by considerations of fringe localisation or of absolute fringe order. The technique is equally valid for both in-plane and line-of-sight displacements, and for any combination of these, and since the displacements are measured point by point the question of whether the overall motion of the object is one of translation, rotation or deformation is irrelevant to the calculation. A point-by-point survey of the whole image gives eventually a composite picture of the motion of the whole object.

There are, however, two important disadvantages of the HF method, both due to the small aperture that is used. It is found that in order to have fringes of reasonable visibility appear on the hologram the aperture in the image plane must be very small - typically 0.5 to 1.0 mm. Since the analysis of each point, selected by this aperture, is rather involved, requiring an assessment of fringe shape in order to determine the direction in which the point has moved, followed by a calculation to find how much it has moved, the analysis of the complete image is a very time-consuming affair. (This objection could be overcome to some extent by the use of automatic data extraction and a computer). The other problem is the purely practical one of shortage of light. The reconstructed image from a hologram is not usually very bright, and the small aperture required for good fringe visibility has an adverse effect in reducing the total amount of light available. In practice, we have found it almost impossible to observe the fringes with small general-purpose laboratory lasers (2mW helium-neon), and it would appear that the technique really requires the use of much higher powers. A proposal by Boone and De Backer (1973) to use image-plane holography (Brandt 1969) would also help to alleviate this problem. A third problem with the HF technique derives from the small size of hologram usually used in experiments. It is often found that this restricts the number of fringes observed to such an extent that the accuracy of the technique is severely limited. (This is really the same problem that occurs with the FC technique, of course, where the small size of the hologram restricts the angle through which the direction of view can be changed).

In spite of the disadvantages outlined in the preceding paragraph, we feel that the HF technique could be a very useful tool for interpreting holographic interferograms, and that it has been neglected by workers in the field. Apart from the works already cited (Gates 1969,1970, Boone 1972, Boone and De Backer 1973), the only other reference that we have found which uses the technique is a paper by Ewers et al. (1974) describing a three-hologram approach aimed at determining the absolute fringe order at each point on the image. We feel that the HF technique deserves a wider popularity than it appears to enjoy at the moment, giving as it does full three-dimensional information about the displacement of an object point by point, and from a single double-exposure hologram. We admit, however, that it is not the technique to use if speed is a criterion, or if lack of laser power is a problem.

3.2.8 - Other Interpretation Techniques

In Sections 3.2.2 to 3.2.7 we have described what we believe are the four main classes of interpretation techniques for holographic interferometry. Most of the interpretation techniques proposed in the literature can be assigned to one or other of these classes, and we hope that this classification of techniques will help to rationalise the subject and remove some of the confusion that exists. For the sake of completeness, however, we shall mention now several analysis methods that do not fall into these four classes.

The first approach that we shall consider might be called the "null technique". First proposed by Gates (1969), in its simplest form this involves the physical manipulation of the object in a real-time holographic interferometry experiment in such a way as to "fluff out" the interference fringes. The argument is that the object has been restored to its original position and, if the amount of movement has been monitored, this gives the displacement suffered by the object in the first place. The technique can also be used to cancel out unwanted rigid-body motion and leave only the deformation fringes which may be the object of the experiment (Champagne and Kersch 1969). In either case, the hologram itself may be manipulated in place of the object - a technique which may be much more practicable. The method cannot be used, of course, with the conventional double-exposure hologram, since the fringes are frozen into the hologram and cannot be changed. The use of two separate holograms, however, does allow the method to be used (Havener and Radley 1972), as does the new "sandwich-hologram" technique recently developed by Abramson (1974b, 1975).

A "second class" of interference effect was introduced by Froehly et al. (1969). They showed that in addition to the "first class" of fringes formed either in the vicinity of the holographic image or longitudinally displaced with respect to it, there was a second set of fringes observable in the direction of the source, i.e. in the undiffracted or zero-order beam. These "second class" fringes were straight and evenly spaced, and oriented in a direction perpendicular to the lateral component of the object displacement. The fringes were only present if the object had been subjected to a rigid-body translation with a component in the plane perpendicular to the line of sight. If there was any deformation of the object, or if the motion of the object was along the line of sight, no fringes were observed. Froehly et al. suggested that this "second class" of interference effect could be used to determine whether or not deformation of the object had occurred, and also to measure the in-plane components of rigid-body translation or the amount of rotation about an axis parallel to the plane of the hologram. Formulae were given for these calculations. It would appear, however, that this "second class" of interference phenomenon is none other than the fringes observed in speckle interferometry (Leendertz 1970, Archbold et al. 1970, Butters and Leendertz 1971). The use of speckle interferometry as a complementary technique to holographic interferometry, using the former for the measurement of the in-plane components of displacement and the latter for the line-of-sight component (using the ZF technique) has, in fact, been adopted by Adams and Maddux (1974). A similar approach of obtaining speckle interferometry fringes from a hologram was proposed by Velzel (1973), who used twin apertures in the hologram plane to obtain contours of equal in-plane displacement, and the ZF technique to measure the line-of-sight component from the same hologram. The subject of speckle interferometry lies outside the terms of reference of this thesis, but this complementary technique cannot be completely ignored in a discussion of holographic interferometry, and we shall return to the subject briefly in Chapter 3.4.

Finally, we should mention an interesting paper by Tribillon and Miles (1974), who have proposed a novel method of interpreting double-exposure holographic interferograms. A photograph (transparency) is taken of the reconstruction from the hologram (complete with interference fringes modulating the holographic image). The transparency is then illuminated with a plane coherent wavefront, and a lens is placed behind the transparency. In the focal plane of this lens is formed the Fourier transform of the image, including the spectrum of the interference fringes, which gived the autocorrelation of the displacement.

To conclude this chapter, however, we repeat the claim that we have devised a classification scheme into which most reported interpretation methods can be fitted. The four techniques - fringe localisation (FL), fringe counting (FC), zero-order fringe (ZF), and hologram fringe (HF) - will be summarised and compared in Chapter 3.4, and guidelines will be offered for selecting the most appropriate technique to use in different situations. Meanwhile, however, we shall turn our attention to the vexed question of fringe localisation.

CHAPTER 3.3

FRINGE LOCALISATION IN HOLOGRAPHIC INTERFEROMETRY

3.3.1 - The Role of Fringe Localisation in Holographic Interferometry

Holography is basically a technique of recording a three-dimensional image on a two-dimensional photographic plate; in the general case, then, the fringes of holographic interferometry should represent the three-dimensional motion of an object between the two states recorded (or between the recorded state and the actual object in real-time holographic interferometry). However, if these fringes are viewed from one position, only a two-dimensional pattern will be seen, and there will inevitably be some ambiguity in its interpretation. It is found, though, that there is a third dimension in the fringes, since they are not necessarily localised on the image. This additional parameter, the position of the plane of localisation of the fringes, can be used to advantage in the interpretation of the fringes.

Fringe localisation was used as a parameter in the original papers of Haines and Hildebrand (1966a,b), who used the plane of maximum visibility as the criterion of localisation in the FL technique (see Section 3.2.2). Fringe localisation is also used in the FC technique of Aleksandrov and Bonch-Bruevich (1967), but in this case it appears only implicitly as the parallax between the fringes and the image (see Section 3.2.3). In Section 3.2.4 we showed that the FL and FC techniques, which both purport to give the in-plane components of translation of the object, are mathematically equivalent, providing the definitions of fringe localisation by visibility and by parallax are equivalent. This, however, is not obvious, and there have been some conflicting views expressed in the literature on this subject. It is the aim of this chapter to try to resolve this problem.

Before discussing the question of fringe localisation in detail, we should perhaps point out that the issue is not so relevant to the other two major interpretation techniques. The ZF method calculates the line-of-sight component of the displacement (strictly the component parallel to the bisector of the line of sight and the illumination direction) from first principles, and uses several (usually three) holograms in order to build up a three-dimensional picture of the displacement of the object. The HF technique achieves the same objective from a single hologram by using the two-dimensional fringe pattern observed in the hologram plane when a very small area of the image is isolated; the convenience of needing only one hologram and no knowledge of the recording geometry is gained at the expense of having to repeat the rather complex measurement procedure for every point on the image. However, for the FL and FC techniques the question of fringe localisation is very relevant, and has received the attention of several workers in the field.

3.3.2 - Localisation Properties of Holographic Fringes

It was observed in the early days of holographic interferometry that fringe localisation depended on the type of motion undergone by the object. Haines and Hildebrand (1966a) reported that in-plane translation gave rise to fringes localised some distance behind the image, while rotation (tilt) of the object about an axis in its own plane produced fringes localised at the image. These observations were confirmed by other workers, who also added the localisation properties of fringes caused by other types of motion. These properties are summarised for convenience in Table 3.1.

TABLE 3.1

Localisation Properties of Holographic Interferometry Fringes

Type of Motion	Illumination	Localisation	References
in-plane translation (normal to line of sight)	divergent	behind image	1-13
"	collimated	infinity	9,10,14-18
longitudinal translation (along line of sight)	divergent	ill-defined	3,4
"	collimated	infinity	3,4,14,15, 17-19
tilt (rotation about an axis in the plane of the object)	any	on the image	1,3-5,9,11,14, 15,18,20-22
in-plane rotation (about line of sight as axis)	any	(a) at the image	14,15
		(b) on inclined plane	23
		(c) on line in space	10,16
		(d) at the stationary point only	24
in-plane deformation (e.g. stretching)	any	(a) at the image	14
		(b) on inclined plane	23,25
deformation along the line of sight	any	at the image	11,14,18, 20-22

Key to references:

1. Haines and Hildebrand (1966a)
2. Haines and Hildebrand (1966b)
3. Viénot et al. (1968)
4. Viénot (1970)
5. Froehly et al. (1969)
6. Monneret (1970)
7. Boone and Verbiest (1969)
8. Stetson (1970b)
9. Aleksandrov and Bonch-Bruevich (1967)
10. Molin and Stetson (1970)
11. Tsujiuchi and Matsuda (1970)
12. Molin and Stetson (1971)
13. Stetson (1974a)
14. Tsujiuchi et al. (1969)
15. Sollid (1969)
16. Stetson (1968)
17. Macáková (1974)
18. Přikryl (1974)
19. Tsuruta et al. (1969)
20. Welford (1969)
21. Welford (1970b)
22. Sampson (1970)
23. Walles (1970a,b)
24. Steel (1970)
25. Ashton et al. (1971)

3.3.3 - The Confusion about Fringe Localisation

Table 3.1 represents the consensus of opinion among authors as to the localisation properties of fringes caused by various types of object motion. There are, however, some points of disagreement, and some dissenting voices. One area of uncertainty is the case of in-plane rotation (rotation about the line of sight as axis), where opinion is divided between a plane of localisation (the surface itself), a line of localisation, and a single point! This discrepancy can almost certainly be resolved in terms of the illumination and viewing conditions used by the different authors.

A more fundamental disagreement with the scheme summarised in Table 3.1 has arisen with the suggestion by Welford (1969, 1970a,b) that, in general, there is "no unique region of localisation" of the fringes (Welford 1969). Welford bases his argument on the observed fact that when fringes are produced which exhibit parallax with the image, this parallax depends on the direction in which the line of sight is changed; thus it is possible (in fact it is always the case!) for the straight fringes produced by in-plane translation to show considerable parallax with the image when the line of sight is changed in a direction perpendicular to the fringes, and to show zero parallax when the line of sight is changed in a direction parallel to the fringes. Using this parallax definition of localisation, Welford concludes that this implies that there is no unique plane of localisation, and distinguishes between localisation (defined by parallax) and the "plane of maximum visibility" used by most authors to locate the fringe plane. (This distinction between parallax and visibility has also been stressed by Machado Gama (1973), who extended the argument to include broad-source classical interferometry). Welford further shows that if the fringes <u>are</u> localised on the image (i.e. exhibit no parallax with the image for any change in viewing direction), then the displacement of the object is in a direction parallel to the viewing direction. At first sight this appears to imply that line-of-sight translations lead to fringes localised on the image, a situation that would be completely at variance with Table 3.1.

Further consideration, however, reveals that this is not the case, and that Welford's criterion only applies to path differences introduced by the translation, and hence to differential displacements along the line of sight. This covers the cases of tilt (rotation about an axis in the plane of the object) and deformation along the line of sight; in both these cases the displacement of any point on the object is, to a first approximation, along the line of sight, and the fact that these two types of motion yield fringes which are localised at the image is strongly endorsed by other workers (see Table 3.1). For rigid-body translations along the line of sight, the only path differences (and hence fringes) that are introduced result from purely geometrical considerations (the fringes are fringes of equal inclination - Haidinger fringes), and these are not considered in Welford's treatment.

Hence the only controversial point to remain is the question of whether fringe localisation by the criterion of maximum visibility, used by most authors, is the same as fringe localisation by the criterion of parallax. It will be remembered from Section 3.2.4 that they must be the same if the FC and FL techniques are to be equivalent (and hence both valid!). On the other hand, Welford's argument that the parallax between the fringes and the image varies with the direction in which the angle of view is changed (and hence that there is no unique localisation for the fringes) would appear to be directly at variance with the practice of many authors of defining the localisation plane as the plane of maximum visibility, and even using this definition in the interpretation of holographic interferograms by means of the FL technique! Steel (1970) dismisses the problem merely by referring to Welford's criterion of localisation (parallax) as being "not equivalent to the condition used by others". This, however, would appear to be avoiding the issue, and certainly does not explain the approach of several authors who have explicitly equated the localisation criteria of visibility and parallax (Aleksandrov and Bonch-Bruevich 1967, Stetson 1968, Tsuruta et al. 1969, Molin and Stetson 1970, Walles 1970b, Přikryl 1974). Walles (1970b) has also shown that the parallax criterion, in which the localisation of the fringes is defined as occurring where the change in optical path difference (and hence the fringe intensity) with change in viewing direction is zero, is equivalent to the definition of fringe localisation as the crossing-point of homologous rays (Viénot et al. 1968, Boone and Verbiest 1969, Viénot 1970, Monneret 1970). (The criterion of ray intersections, based on classical interferometry ideas, was also used by Steel (1970) and by Machado Gama (1973)). This link brings most authors into full agreement, that fringe localisation by maximum visibility is equivalent to fringe localisation by parallax. We shall now add further fuel to the fire by mentioning a paper by Machado Gama (1973), in which it is shown that the position of the plane of maximum visibility - and therefore of the plane of localisation, if visibility be the criterion - varies with the size of the viewing aperture! This would appear to jeopardise the whole practice of localising the fringes by visibility, since the plane of localisation - and hence the interpretation of the fringes by the FL technique - would vary according to the size of the aperture used!

3.3.4 - Resolving the Localisation Paradox

It would appear, then, that we have a real dilemma in the question of fringe localisation, and there has certainly been considerable confusion and misunderstanding about the subject. The answer to the problem is found in two papers by Stetson (1970b,1974a). Stetson showed that the fringes are localised, in general, not on a plane but on a line in space. (Line localisation had already been introduced by Stetson for fringes due to in-plane rotation (Stetson 1968, Molin and Stetson 1970, 1971); the concept of line localisation was also used by Steel (1970) and by Dubas and Schumann (1974)). Stetson also showed that a plane of localisation can be found if a slit aperture is used to view the fringes in place of the usual circular aperture, and that the position of this plane of localisation varies with the orientation of the slit (Stetson 1974a). This piece of theory from Stetson resolves all the problems about fringe localisation:

(i) If a slit aperture is used, there is a plane of maximum visibility of the fringes which coincides with the plane of localisation found by parallax when the line of sight is changed in the direction parallel to the slit.

(ii) Hence the FL and FC techniques of fringe interpretation are equivalent, providing a slit aperture is used for the former, and the slit is oriented in a direction parallel to that of observer motion in the latter.

(iii) If a circular aperture is used, the fringes are localised on a <u>line</u> rather than on a plane: this leads to very ill-defined localisation of the plane of maximum visibility, whose position may vary with the size of the aperture.

(iv) Using a circular aperture, fringes of high visibility can only be observed if the aperture is below a certain critical size (see also Welford 1969, 1970b); the use of such a small aperture leads to a large depth of focus, and further uncertainty in locating the plane of maximum visibility. (This dilemma explains the large errors reported by Haines and Hildebrand for the FL technique, and the difficulties of locating the plane of localisation of the fringes, reported by many authors).

(v) In the case of the straight fringes produced by in-plane translation, if the slit aperture is oriented to be parallel to the fringes, the plane of maximum visibility is found to be at the image: there is also found to be no parallax between the fringes and the image in this direction. (This leads to a method of identifying the direction of any in-plane component of translation - see Chapter 3.4).

(We should perhaps point out that Welford (1969) had, in fact, already suggested that a slit aperture could help to resolve the localisation problem; in his case, however, he used the slit oriented in a direction perpendicular to the displacement vector, and hence parallel to the fringes, in order to localise the fringes at the image - see (v) above. Stetson, on the other hand, suggested using the slit oriented in a direction perpendicular to the fringes, in order to localise the fringes in a plane behind the image so that the FL technique could be used in a valid way - see (ii) above.)

3.3.5 - Removing Fringe Localisation from the Scene of Holographic Interferometry

Having disposed of the problem of fringe localisation, and reconciled the apparently conflicting views expressed in the literature, we now suggest that the whole question of fringe localisation is of purely academic interest, and should not have been allowed to cloud the issue of holographic interferogram interpretation. The discussion of fringe localisation has, of course, been important insofar as, until the apparent contradictions were resolved, workers could have little faith in those techniques (FL and FC) which used fringe localisation as a parameter. Now that the visibility and parallax criteria have been shown to be equivalent (providing a slit aperture is used), we propose that fringe localisation as such should be ignored by workers interested only in the interpretation of holographic interferograms. The FL technique would seem to have no advantages over the FC technique, and we suggest, therefore, that the latter method be used when in-plane components of displacement are under investigation. Since this involves only the parallax between the fringes and the image, the actual physical location of the plane of localisation of the fringes is irrelevant. (If anyone does want to remain loyal to the FL technique, then he need only use a slit aperture, oriented in a direction parallel to the fringes, and the difficulties of fringe localisation are removed for him also). Experimentalists can then concentrate on the problem of interpreting their holograms, instead of worrying about where the fringes are localised!

CHAPTER 3.4

CHOOSING AN INTERPRETATION TECHNIQUE

3.4.1 - The Fringe Patterns of Holographic Interferometry

Before discussing the pros and cons of the different interpretation techniques, we shall summarise the different types of fringe pattern that occur in holographic interferometry, classified according to the types of motion that produce them. Some of the fringe patterns have already been mentioned in passing in previous chapters, but for convenience they are now brought together and listed in Table 3.2.

In addition to the simple types of motion given in Table 3.2, several authors have dealt with specific combinations of such motions. For example, combinations of in-plane translation and tilt about the orthogonal axis were considered by the Besançon group (Viénot et al. 1968, Froehly et al. 1969, Viénot 1970), and pivot motion (rotation about an axis perpendicular to the line of sight but not in the plane of the object) and screw motion (translation along an axis of rotation) were discussed by Stetson (1968). As expected, such hybrid motions give rise to fringe patterns which are compromises between the "pure" patterns described in Table 3.2. It is only fair to say, though, that for complex combinations of motions the fringe patterns - and their localisation properties - can become very involved. This applies particularly to cases of complex deformation, such as might be encountered in stress analysis problems. Nevertheless, it is possible to interpret even the most complex fringe pattern, and we shall now concentrate on the problem of deciding which interpretation techniques are most suitable for the different types of pattern encountered, and for the particular requirements of the experiment.

TABLE 3.2

The Basic Fringe Patterns of Holographic Interferometry

Type of Motion	Description of Fringes	References
in-plane translation (normal to line of sight)	straight, parallel, evenly spaced; perpendicular to displacement direction; spacing inversely proportional to displacement; localised behind image in divergent, at infinity in collimated light	1-8
longitudinal translation (along line of sight)	concentric circles (Haidinger fringes of equal inclination); localised at infinity in collimated light	4-7
tilt (rotation about an axis in the plane of the object)	straight, parallel, evenly spaced; parallel to axis of rotation; spacing inversely proportional to angle of rotation; localised at the image	1,3-7,9
in-plane rotation (about line of sight as axis)	straight, parallel; parallel to direction of illumination; spacing inversely proportional to angle of rotation; localisation variable	4,6,7
in-plane deformation (e.g. stretching)	irregular; spacing inversely proportional to local displacement, but depends on geometry; localisation variable	6,10-14
deformation along the line of sight	irregular; spacing inversely proportional to gradient of displacement; localised at the image	6,8,10,15,16

Key to references:

1. Haines and Hildebrand (1966a)
2. Haines and Hildebrand (1966b)
3. Aleksandrov and Bonch-Bruevich (1967)
4. Viénot et al. (1968)
5. Froehly et al. (1969)
6. Tsujiuchi et al. (1969)
7. Viénot (1970)
8. Boone and Verbiest (1969)
9. Sollid (1969)
10. Ennos (1968)
11. Luxmoore and House (1970)
12. Wilson (1971a)
13. Ashton et al. (1971)
14. Stetson (1974a)
15. Butters (1968)
16. Welford (1970a)

3.4.2 - A Summary and Comparison of the Main Interpretation Techniques

For the moment we shall leave aside the "second class of interference effect" introduced by Froehly et al. (1969) and described in Section 3.2.8 of this thesis. We have already equated it to the technique of speckle interferometry, and we shall mention this subject again in Section 3.4.3 when we discuss the choice of interpretation technique. Four main classes of interpretation technique were isolated and described in detail in Chapter 3.2, and we now propose to summarise and compare the main features of these methods. For convenience, we present this summary in the form of a table (see Table 3.3).

From Table 3.3 it will be seen that there are no real advantages to be gained from using the FL technique in preference to the FC technique. Both methods give the in-plane components of translation, and the FC technique is much easier and quicker to apply. We therefore suggest that the FL technique be relegated to the category "of historical interest only". This step also removes the spectre of fringe localisation from the scene, since the FC technique only uses this parameter in the form of parallax, and fringe localisation is not a factor in the other two techniques. Hence the experimenter can forget the intricacies of fringe localisation - which is something of a "red herring" as far as fringe interpretation is concerned - and leave the problem to the theoreticians. (He should, however, be sufficiently aware of the role of fringe localisation to be able to make some qualitative deductions from fringe parallax).

The "null" technique is not listed in Table 3.2, but perhaps we ought to say something about the possible role of this method also. We feel that the null technique has little to offer the experimenter as a method of actual quantitative interpretation of holographic interferograms. We believe that the practical difficulties involved in manipulating the object (or the hologram) until the original object motion has been cancelled far outweigh the advantage of avoiding any computation. It does, however, have

TABLE 3.3

Summary of the Four Main Interpretation Techniques

Technique	FL	FC	ZF	HF
Originator(s)	Haines and Hildebrand (1966)	Aleksandrov and Bonch-Bruevich (1967)	Ennos (1968)	Gates (1969)
Method	fringe localisation	fringe parallax	calculation of opd's	aperture at image (or object)
Measures	in-plane component	in-plane component	line-of-sight component (approx.)	all components
Formulae	$d_x = \frac{\lambda R}{h_x}$	$d_x = \frac{N\lambda L}{x}$	$\Delta = m\lambda$ $d_{\theta/2} = \frac{m\lambda}{2\cos\theta/2}$ $d_z = m\lambda/2$ (simplified)	$d_x = \frac{\lambda L}{h_x}$ $d_z = \frac{2m\lambda L^2}{s_m^2}$
Views reqd. for one component	1	linear continuum of views	1	1 for each image point
Holograms reqd. for 3D motion	1 (in principle)	1 (in principle)	3	1
Views reqd. for 3D motion	2 (in principle)	2 lines in hologram plane (in principle)	1 per hologram (3 total)	1 for each image point
Actual hologram reqd.?	yes	yes	no	yes
Fringe order required?	no	no	yes (in general)	no
Localisation restrictions	not on image	not on image	none, but most suitable for fringes on image	none
Practical difficulties	1. localisation 2. combinations of motions	limited size of hologram	1. zero fringe identification 2. analysis in general case	1. shortage of light 2. time-consuming 3. limited size of hologram
Accuracy	very limited	limited	good	fair

a valid role in some instances where the experimenter may wish to add some background fringes to the hologram (in the same way that tilt fringes are used in classical interferometry), or to cancel out unwanted motion of the object support structure, or to cancel out rigid-body motion of the object when only deformations are of interest. We suggest that the null technique be used only for qualitative applications such as these, rather than as a quantitative interpretation technique in its own right.

This leaves us with the FC, ZF and HF techniques as basic interpretation methods, and we suggest that each of these main techniques has a role to play in the interpretation of holographic interferograms.

3.4.3 - Guidelines for Fringe Interpretation

The choice of interpretation technique will depend on several factors. Among the most important of these are the accuracy required, the amount of time that can be allowed for the analysis, the nature and accessibility of the object, the apparatus and computational facilities available, the type of displacement or deformation to be measured, and how much a priori knowledge about that displacement is available. Using the facts outlined earlier in this chapter, and in particular the information summarised in Tables 3.2 and 3.3, we offer the following guidelines for choosing an interpretation technique for holographic interferograms.

(a) Maximum accuracy required

Use the ZF technique.

If possible, adopt the simplified geometry of collimated illumination, normal viewing, and coincident illumination and viewing directions (see Figure 3.4). (NB - This arrangement is only suitable for differential displacements such as deformations or tilts: no fringes will be visible for rigid-body translations).

If the direction of the displacement is known a priori, arrange for the viewing and/or illumination direction to be parallel to this direction; if not, or for three-dimensional analysis, use three separate holograms, one for each coordinate direction. (Alternatively, methods such as that described by Sampson (1970) may be used - see Section 3.2.6).

If very high accuracy is required, adopt a fringe interpolation technique such as the double-frequency method described by Dändliker et al. (1973).

(b) General-purpose interpretation: direction of displacement known in advance

Use the ZF technique, with simplified geometry if possible:

(i) for differential displacements, if the displacement direction is at a reasonable angle to the object surface (i.e. not in-plane);

(ii) for out-of-plane deformations.

Use the FC technique for in-plane displacements.

(c) General-purpose interpretation: direction of displacement not known in advance

If the fringes are localised at the image (test by parallax), they are caused either by tilt (if straight and evenly spaced) or by out-of-plane deformation (if otherwise); for the former use Equation 3.4, for the latter use the ZF technique.

If the fringes are not localised at the image, apply one of the following alternatives:

Alternative 1 : Use the HF technique, if time permits and if the problems of low light levels and of limited hologram size are not prohibitive.

Alternative 2 : Adopt the following scheme, based on a proposal by Welford (1970a):

(i) Find, if possible, a direction of view for which the fringes are localised at the image (no parallax between the fringes and the image for any small change in angle of view).

(ii) This viewing direction is parallel to the displacement direction: use the ZF technique for interpretation (only suitable for differential displacements such as tilt or out-of-plane deformation - not for rigid-body translations).

(iii) If this viewing direction for fringe localisation is not available within the field of view of the hologram, find a direction of change of viewing direction for which the fringes at the point under consideration appear to be located at the image (i.e. show no parallax with the image). This direction of change of viewing direction, projected on to the image plane, is perpendicular to the in-plane component of the displacement.

(iv) Use the FC technique, changing the viewing direction in a direction orthogonal to that found in (iii), to calculate the in-plane component of the displacement.

(v) Repeat (iii) and (iv), using a different portion of the hologram (or a separate hologram), to find the out-of-plane component.

Alternative 3 : Use the ZF technique, using three separate holograms, and possibly a "labour-saving" device such as Abramson's holo-diagram (1968, 1969a).

Alternative 4 :

(i) Use the ZF technique for the line-of-sight component.

(ii) Use speckle interferometry for the in-plane components.

This approach will be discussed in Section 3.4.4.

(d) Single hologram available; details of recording geometry not known

Use the HF technique for a full interpretation.

Use the FC technique for a rapid analysis of the in-plane components, and for an estimate of the line-of-sight component.

(e) Photographs of reconstructions available, but not the actual holograms

Use the ZF technique, but we must know:

(i) the recording and the reconstruction geometries;
(ii) the identity of a zero-order fringe (for absolute displacements).

3.4.4 - Combination of Holographic Interferometry with Speckle Interferometry

As one of the recommended alternatives in the previous section, we proposed the combination of holographic interferometry and speckle interferometry. The latter subject is outside the terms of reference of this thesis, and the reader is referred to the literature for details of the technique (Leendertz 1970, Archbold et al. 1970a,b, Butters and Leendertz 1971, Burch 1971).

Basically, though, speckle pattern interferometry uses the correlation between the speckle patterns produced by a diffuse object illuminated with laser light (see Chapter 1.1 of this thesis) in its original position and in its displaced position to measure the in-plane displacement. The fringes of speckle pattern interferometry are analogous to the fringes of the more traditional moiré methods, and are interpreted in a similar way.

Since speckle pattern interferometry provides a simple technique for measuring in-plane displacements, and since holographic interferometry is most sensitive - and most accurate - for line-of-sight deformations, it has been found very useful to combine the two techniques. This approach is, for example, preferred by the National Physical Laboratory. If the method of Adams and Maddux (1974) is adopted, the two measurements can be carried out on the same hologram, without the need for a separate speckle photograph. It is believed that this is, in effect, what Froehly et al. (1969) were doing when they reported their "second class of interference" (see Section 3.2.8). Boone (1975) has recently proposed the use of reflection holograms for this combined technique.

The technique of measuring the in-plane components of displacement by means of speckle interferometry and the line-of-sight component by means of holographic interferometry (using the ZF method) is certainly a powerful one, and should be seriously considered by workers in the field.

In conclusion, we hope that by reducing the vast literature of holographic interferometry to four basic techniques we have removed some of the confusion in this field. We further hope that the guidelines offered in this last chapter will help workers to choose an appropriate interpretation scheme for their own particular applications of holographic interferometry.

IN CONCLUSION

We have been only too aware during the writing of this thesis of the lack of quantitative experimental results. This shortcoming has been due to a number of causes, but primarily to the use of biological specimens in experiments carried out by a physicist inexperienced in such matters. We have suggested, and now wish to emphasise, that any further work in the fields covered by Parts 1 and 2 of this thesis should be carried out as a joint project by a physicist and a biologist.

Let us examine now what we feel has been achieved in this project.

(1). The project was interdisciplinary, not in the sense of being true biophysics (the physics of living organisms), but in the sense that it involved the application of physical techniques (speckle correlation and holographic interferometry) to biological measurements. The author and his colleagues have certainly been made more aware of the world of the botanist, and we hope that botanists are now more aware of how physics might be of service to them in their work.

(2). We have made some small contributions to the statistical theory of laser speckle patterns, and especially to the statistics of speckle patterns added to uniform backgrounds (see Section 1.1.4), the spatial statistics of time-averaged fluctuating speckle patterns (Section 1.1.5), and the temporal statistics of fluctuating speckle patterns (Section 1.1.6).

(3). We have discovered a wavelength/colour dependence of the speckle fluctuations observed when some botanical specimens are illuminated with laser light (Section 1.3.2), and we have used this phenomenon to propose that the fluctuations are caused, at least partially, by the motion of coloured particles inside the specimens. We have been unable to prove the validity of this model, though some supporting evidence has been collected (Chapter 1.3). If the model is correct, we have demonstrated a possible method of monitoring the activity of pigmented particles inside biological specimens, and a technique (the wavelength/colour dependence of the fluctuations) for isolating the contributions of coloured particles to the speckle fluctuations. We have proposed several methods of monitoring and analysing the fluctuations (Chapter 1.3). We have also suggested that the wavelength/colour dependence of speckle fluctuations might be of value in other applications of correlation techniques and intensity fluctuation spectroscopy.

(4). We have established an apparent link between the speckle fluctuations discussed above and the difficulties of obtaining fringes on double-exposure holograms of plants (Section 2.2.5). We have suggested, in fact, that the speckle fluctuations might be a limiting factor in the usefulness of holographic interferometry as a method of measuring plant growth, and that the fluctuations might provide a method of determining in advance whether a particular specimen (or a particular part of a specimen) is a good subject for holographic interferometry (Section 2.2.8). We have indicated ways in which holographic interferometry might be of use to botanists, and have made several proposals for future work in this field (Section 2.2.9).

(5). Finally, we have attempted a major review of the field of holographic interferogram interpretation in an effort to remove some of the confusion that exists in the literature on this subject. We have reduced the many proposed methods for analysing holographic interferograms to four basic techniques (Chapter 3.2), and have offered guidelines for the choice of appropriate techniques for different circumstances (Chapter 3.4).

In conclusion, we hope that this thesis makes some contribution to the fields of laser speckle theory, intensity fluctuation spectroscopy and holographic interferometry - and not least to the idea of interdisciplinary cooperation between physicists and biologists.

NOTE : The three parts of this thesis were recently presented, in abridged form, as three separate papers at the Tenth Congress of the International Commission for Optics, held in Prague (Briers 1975b,c,d). The Proceedings of this conference will be published in 1976.

APPENDIX

THE BASIC THEORY OF LASER SPECKLE STATISTICS

The purpose of this Appendix is the formal derivation of the relationships quoted in Section 1.1, describing the statistics of laser speckle patterns. The proofs are included in this thesis merely in the interests of completeness, so that the work is more or less self-contained. The formulation closely follows the approach of Dainty (1972b) and McKechnie (1974). The following notation is used:

x, y	=	coordinate directions in the far-field speckle pattern;
$\delta x, \delta y$	=	increments in x and y;
ξ, η	=	coordinate directions in the plane of the scattering surface;
u, v	=	spatial frequencies in the x and y directions in the speckle pattern;
D	=	distance from the scattering surface to the plane of observation in the far field;
s	=	radius of the scattering area, when considered circular;
N	=	number of scatterers contributing to the speckle pattern;
A	=	complex amplitude;
A_r, A_i	=	real and imaginary parts of A $(A = A_r + iA_i)$;
A^*	=	complex conjugate of A;
I	=	intensity;
$\langle \rangle$	=	ensemble average;
σ^2	=	variance of intensity;
σ_A^2	=	variance of complex amplitude;
β	=	phase of single component of scattered light;
ϕ	=	resultant phase of scattered light;
p	=	probability density function;
G	=	autocorrelation function;
g	=	normalised autocorrelation function;
C	=	autocovariance;
c	=	normalised autocovariance;

S	=	intensity distribution of light leaving the scattering area;
J_1	=	first-order Bessel function;
W	=	Wiener spectrum (power spectrum);
λ	=	wavelength of the light;
$\delta(f)$	=	delta function (=1 for f=0, =0 for f$\neq$0).

A.1 - First Order Statistics

Consider a scattering surface illuminated with a laser beam of wavelength λ, and an observing screen set up at a distance D from the surface to display the far-field speckle pattern (see Figure 1.3 of Chapter 1.1). Assume that D and the coherence length of the laser light used are both large compared with the roughness of the scattering surface, and that the surface is rough compared with the wavelength λ. Assume also that N, the number of scatterers in the illuminated area, is large.

The complex scalar amplitude of the light scattered from the surface is given by the sum of the contributions from all N scatterers:

$$A(\xi,\eta) = \sum_{n=1}^{N} A_n e^{i\beta_n} \delta(\xi - \xi_n)\, \delta(\eta - \eta_n) \qquad \text{A.1}$$

Assume now that there is a random distribution of scatterers in the scattering area. Hence, A_n will be a random variable, β_n will be a random variable in the range 0 to 2π, and it can be further assumed that β_n will be independent of A_n.

From Chapter 8 of Born and Wolf (1970), the complex amplitude of the scattered light in the far field is given by the Fourier transform of the complex amplitude at the scattering surface:

$$A(x,y) = \iint_{-\infty}^{\infty} A(\xi,\eta) \exp\left\{-\frac{2\pi i}{\lambda D}(\xi x + \eta y)\right\} d\xi\, d\eta \qquad \text{A.2}$$

Combining Equations A.1 and A.2 gives:

$$A(x,y) = \sum_{n=1}^{N} A_n \exp(i\beta_n).\exp\left\{-\frac{2\pi i}{\lambda D}(\xi_n x + \eta_n y)\right\} \qquad A.3$$

If N is large, the Central Limit Theorem (Middleton 1960) tells us that A(x,y) becomes a complex normal process, with its real and imaginary parts identically distributed with zero means and identical variances, and statistically independent at any point. Let the variance of the real and imaginary parts be $\sigma_A^2/2$. Writing A(x,y) in terms of its real and imaginary parts, we have:

$$A(x,y) = A_r(x,y) + iA_i(x,y) \qquad A.4$$

We can now give its joint probability density function (since it is a complex normal process) as:

$$P(A_r,A_i) = \frac{1}{\pi\sigma_A^2} \exp\left\{-\frac{(A_r^2 + A_i^2)}{\sigma_A^2}\right\} \qquad A.5$$

The intensity and phase of the scattered light in the far field can now be given, by definition, as:

$$I = A_r^2 + A_i^2 \qquad A.6$$

$$\phi = \tan^{-1}\frac{A_i}{A_r} \qquad A.7$$

Inverting Equations A.6 and A.7 gives:

$$A_r = \sqrt{I}\cos\phi \qquad A.8$$

$$A_i = \sqrt{I}\sin\phi \qquad A.9$$

The Jacobian of Equations A.8 and A.9 is given by:

$$\begin{vmatrix} \dfrac{\partial A_r}{\partial I} & \dfrac{\partial A_i}{\partial I} \\ \dfrac{\partial A_r}{\partial \phi} & \dfrac{\partial A_i}{\partial \phi} \end{vmatrix} = \tfrac{1}{2} \qquad \text{A.10}$$

Now, the joint probability density function of intensity and phase, $P(I,\phi)$, is given in terms of $P(A_r,A_i)$ by the probability transform (Davenport and Root 1958):

$$P(I,\phi) = \begin{vmatrix} \dfrac{\partial A_r}{\partial I} & \dfrac{\partial A_i}{\partial I} \\ \dfrac{\partial A_r}{\partial \phi} & \dfrac{\partial A_i}{\partial \phi} \end{vmatrix} \cdot P(A_r,A_i)$$

Substituting from Equations A.5, A.6 and A.10, this becomes:

$$P(I,\phi) = \frac{1}{2\pi\sigma_A^2} \exp(-I/\sigma_A^2) \qquad \text{A.11}$$

The marginal probability density functions, $P(I)$ and $P(\phi)$, are given by integrating over the relevant variable:

$$P(\phi) = \int_0^\infty \frac{1}{2\pi\sigma_A^2} \exp(-I/\sigma_A^2)\, dI$$

$$= \frac{1}{2\pi} \quad \text{for } 0 \leqslant \phi \leqslant 2\pi \qquad \text{A.12}$$

or $= 0$ otherwise.

$$P(I) = \int_0^{2\pi} \frac{1}{2\pi\sigma_A^2} \exp(-I/\sigma_A^2)\, d\phi$$

$$= \frac{1}{\sigma_A^2} \exp(-I/\sigma_A^2) \qquad \text{for } I \geqslant 0 \qquad \text{A.13}$$

or $= 0$ for $I < 0$.

Hence there is a uniform distribution of phase over the range 0 to 2π, and a negative exponential distribution of intensity.

The mean intensity, $\langle I \rangle$, is given by the first moment of the probability density function:

$$\begin{aligned}
\langle I \rangle &= \int_{-\infty}^{\infty} P(I).I \, dI \\
&= \int_{0}^{\infty} P(I).I \, dI \qquad \text{(since } P(I)=0 \text{ for } I<0\text{)} \\
&= \frac{1}{\sigma_A^2} \int_{0}^{\infty} I.\exp(-I/\sigma_A^2) \, dI \\
&= \sigma_A^2 \qquad\qquad \text{A.14}
\end{aligned}$$

Substituting Equation A.14 in Equation A.13 gives the following expression for the probability density function:

$$P(I) = \frac{1}{\langle I \rangle} \exp(-I/\langle I \rangle) \qquad\qquad \text{A.15}$$

The second central moment of intensity gives the variance:

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} P(I).I^2 \, dI - \langle I \rangle^2 \\
&= \int_{0}^{\infty} P(I).I^2 \, dI - \langle I \rangle^2 \qquad \text{(since } P(I)=0 \text{ for } I<0\text{)} \\
&= \frac{1}{\langle I \rangle} \int_{0}^{\infty} \exp(-I/\langle I \rangle).I^2 \, dI - \langle I \rangle^2 \qquad \text{(from Equation A.15)} \\
&= 2\langle I \rangle^2 - \langle I \rangle^2
\end{aligned}$$

$$\text{i.e.} \quad \sigma^2 = \langle I \rangle^2 \qquad\qquad \text{A.16}$$

Hence the variance in the intensity of a speckle pattern is equal to the square of the mean intensity.

A.2 - Second Order Statistics

The second order statistics of a speckle pattern which concern us are the autocorrelation function (or the autocovariance), which gives a measure of the size of a typical speckle, and the Wiener spectrum (the power spectrum), which gives the distribution of speckle size in the pattern. These two functions are related by a Fourier transform relationship.

In this thesis we have defined the autocorrelation function of intensity as:

$$G^{(2)}(\delta x,\delta y) = \langle I(x,y).I(x+\delta x,y+\delta y)\rangle \qquad \text{A.17}$$

and the autocovariance as:

$$C^{(2)}(\delta x,\delta y) = \langle (I(x,y)-\langle I\rangle)(I(x+\delta x,y+\delta y)-\langle I\rangle)\rangle \qquad \text{A.18}$$

We have assumed stationarity, i.e. that $G^{(2)}(\delta x,\delta y)$ and $C^{(2)}(\delta x,\delta y)$ are independent of x and y, and hence are constant at all points in the speckle pattern. This implies that $\langle I(x,y)\rangle = \langle I(x+\delta x,y+\delta y)\rangle$, and it then follows that Equation A.18 can be re-written as follows:

$$C^{(2)}(\delta x,\delta y) = \langle I(x,y).I(x+\delta x,y+\delta y)\rangle - \langle I\rangle^2 \qquad \text{A.19}$$

The following properties of the autocorrelation function should be noted:

(i) The maximum value of $G^{(2)}(\delta x,\delta y)$ occurs when $\delta x = \delta y = 0$, and is given by:

$$G^{(2)}(0,0) = \langle I^2\rangle$$

$$= 2\sigma^2$$ (since $\sigma^2 = \langle I^2\rangle - \langle I\rangle^2$ by definition, and $\sigma^2 = \langle I\rangle^2$ for a speckle pattern - see Equation A.16)

(ii) When δx and δy are so large that correlation has been completely lost, then $I(x,y)$ and $I(x+\delta x,y+\delta y)$ are independent of each other.

$$
\begin{aligned}
\text{Thus} \quad G^{(2)}(\delta x,\delta y) &= \langle I(x,y).I(x+\delta x,y+\delta y)\rangle \\
&\longrightarrow \langle I(x,y)\rangle\langle I(x+\delta x,y+\delta y)\rangle \text{ at large lags} \\
&= \langle I\rangle^2 \\
&= \sigma^2
\end{aligned}
$$

These properties lead to a useful normalisation of the autocorrelation function (which has been adopted by workers in photon counting and intensity fluctuation spectroscopy for use with temporal autocorrelations). We define a normalised autocorrelation function as follows:

$$g^{(2)}(\delta x,\delta y) = \frac{G^{(2)}(\delta x,\delta y)}{\sigma^2} \qquad \text{A.20}$$

$g^{(2)}(\delta x,\delta y)$ thus takes the value 2 at zero lag, and the value 1 at large lags.

Turning to the autocovariance, $C^{(2)}(\delta x,\delta y)$, we deduce from Equation A.19 and the properties of $G^{(2)}(\delta x,\delta y)$ that:

(i) At zero lag, $C^{(2)}(0,0) = \langle I^2\rangle - \langle I\rangle^2$

$$= \sigma^2 \qquad \text{(by definition)}$$

(ii) At large lags, $C^{(2)}(\delta x,\delta y) \longrightarrow 0$

Again we can define a normalised form:

$$c^{(2)}(\delta x,\delta y) = \frac{C^{(2)}(\delta x,\delta y)}{\sigma^2} \qquad \text{A.21}$$

which takes the value 1 at zero lag and 0 at large lags.

It should be pointed out that workers in speckle statistics tend to define and use this normalised autocovariance as the autocorrelation function, in place of $g^{(2)}(\delta x,\delta y)$. We note that the two functions are related by the simple relationship:

$$c^{(2)}(\delta x,\delta y) = g^{(2)}(\delta x,\delta y) - 1 \qquad \text{A.22}$$

To find an expression for $c^{(2)}(\delta x,\delta y)$, let us first consider the autocorrelation function of complex amplitude, defined as:

$$G^{(1)}(\delta x,\delta y) = \left\langle A(x,y).A^{*}(x+\delta x,y+\delta y)\right\rangle \qquad \text{A.23}$$

(NB - Since we are now dealing with complex variables, it is necessary to take the complex conjugate of one of the amplitudes in the above expression).

Recalling Equation A.3:

$$A(x,y) = \sum_{n=1}^{N} A_n \exp(i\beta_n) \exp\left\{-\frac{2\pi i}{\lambda D}(\xi_n x + \eta_n y)\right\}$$

and combining this with Equation A.23, we get:

$$G^{(1)}(\delta x,\delta y) = \sum_{n=1}^{N}\left\langle \left|A_n\right|^2 \exp\left\{\frac{2\pi i}{\lambda D}(\xi_n.\delta x + \eta_n.\delta y)\right\}\right\rangle \qquad \text{A.24}$$

(all cross-products go to zero on taking the ensemble average)

If the scatterers are packed sufficiently closely in the scattering area, $\left|A_n\right|^2$ can be represented by a continuous function, say $S(\xi,\eta)$. Hence we have:

$$G^{(1)}(\delta x,\delta y) = \iint_{-\infty}^{\infty} S(\xi,\eta) \exp\left\{\frac{2\pi i}{\lambda D}(\xi.\delta x + \eta.\delta y)\right\} d\xi\, d\eta \qquad \text{A.25}$$

Returning now to the autocovariance of intensity, let us recall Equation A.19:

$$C^{(2)}(\delta x,\delta y) = \langle I(x,y).I(x+\delta x,y+\delta y)\rangle - \langle I\rangle^2$$

Expanding this in terms of complex amplitudes, and invoking Middleton's result that for V_n real, zero-mean, Gaussian random variables (Middleton 1960):

$$\langle V_1V_2V_3V_4\rangle = \langle V_1V_2\rangle\langle V_3V_4\rangle + \langle V_1V_3\rangle\langle V_2V_4\rangle + \langle V_1V_4\rangle\langle V_2V_3\rangle$$

it can easily be shown that:

$$C^{(2)}(\delta x,\delta y) = \left|G^{(1)}(\delta x,\delta y)\right|^2 \qquad \text{A.26}$$

Hence, from Equation A.25:

$$C^{(2)}(\delta x,\delta y) = \left|\iint_{-\infty}^{\infty} S(\xi,\eta)\ \exp\left\{\frac{2\pi i}{\lambda D}(\xi.\delta x + \eta.\delta y)\right\} d\xi d\eta\right|^2 \qquad \text{A.27}$$

For a uniformly illuminated circular scattering area, $S(\xi,\eta)$ reduces to $\langle I\rangle$, and the above equation becomes:

$$C^{(2)}(\delta x,\delta y) = \langle I\rangle^2 \left[\frac{2J_1\left(\frac{2\pi s}{\lambda D}\sqrt{(\delta x)^2 + (\delta y)^2}\right)}{\frac{2\pi s}{\lambda D}\sqrt{(\delta x)^2 + (\delta y)^2}}\right]^2 \qquad \text{A.28}$$

The Wiener spectrum, or power spectrum, of a speckle pattern gives an indication of the distribution of speckles of different sizes. It is given by the self-convolution of the intensity distribution in the scattering area, or, as below, by the Fourier transform of the autocorrelation function:

$$W(u,v) = \iint G^{(2)}(\delta x,\delta y)\ \exp\ (-2\pi i(\delta x.u + \delta y.v))\ d(\delta x)d(\delta y) \qquad \text{A.29}$$

If we wish to remove the delta function at the origin, caused by the d.c. component due to the finite mean $\langle I\rangle$, we simply replace $G^{(2)}(\delta x,\delta y)$ by the autocovariance $C^{(2)}(\delta x,\delta y)$ in Equation A.29.

BIBLIOGRAPHY

ABBISS, J B et al. (1974) - Opt. Laser Technol. 6, 249.

ABRAMSON, N (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P. (1970), p 45.

ABRAMSON, N (1969a) - Appl. Opt. 8, 1235.

ABRAMSON, N (1969b) - Optik 30, 56.

ABRAMSON, N (1970a) - Appl. Opt. 9, 97.

ABRAMSON, N (1970b) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-7.

ABRAMSON, N (1970c) - Appl. Opt. 9, 2311.

ABRAMSON, N (1971a) - Nature - Phys. Sci. 231, 65.

ABRAMSON, N (1971b) - Appl. Opt. 10, 2155.

ABRAMSON, N (1972) - Appl. Opt. 11, 1143.

ABRAMSON, N (1973) - Optik 37, 337.

ABRAMSON, N (1974a) - verbal communication.

ABRAMSON, N (1974b) - Appl. Opt. 13, 2019.

ABRAMSON, N (1975) - Appl. Opt. 14, 981.

ABRAMSON, N and BJELKHAGEN, H (1973) - Appl. Opt. 12, 2792.

ADAMS, F D and MADDUX, G E (1974) - Appl. Opt. 13, 219.

ALEKSANDROV, E B and BONCH-BRUEVICH, A M (1967) - Zh.Tekh.Fiz.37, 360. (Trans.: Sov. Phys.- Tech. Phys. 12, 258).

ALEKSOFF, C C (1971) - Appl. Opt. 10, 1329.

APLIN, C et al. (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-6.

ARCHBOLD, E and ENNOS, A E (1968) - Nature 217, 942.

ARCHBOLD, E et al. (1967) - J. Sci. Instrum. 44, 489.

ARCHBOLD, E et al. (1970a) - in "Optical Instruments and Techniques", Procs. Eighth Cong. ICO, Reading, 1969 (Ed. Home Dickson), Oriel Press (1970), p 265.

ARCHBOLD, E et al. (1970b) - Opt. Acta 17, 883.

ARSENAULT, H and LOWENTHAL, S (1970) - Opt. Commun. 1, 451.

ASHTON, R A et al. (1971) - Appl. Opt. 10, 440.

BARAKAT, R (1973a) - Opt. Acta 20, 729.

BARAKAT, R (1973b) - Opt. Commun. 8, 14.

BARGERON, C et al. (1972) - Phys. Rev. Letters 28, 1105.

BARGERON, C et al. (1973) - Phys. Rev. Letters 30, 205.

BARNETT, N E (1967) - Opt. Soc. Am. Ann. Mtg., Paper WA17.

BAZELAIRE, E and PRADE, B (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 5-6.

BEESLEY, M J and CASTLEDINE, J G (1970) - Appl. Opt. 9, 2720.

BELLANI, V F and SONA, A (1974) - Appl. Opt. 13, 1337.

BERGÉ, P et al. (1967) - C. R. Acad. Sci. Paris D265, 889.

BIJL, D and JONES, R (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 5-3.

BIJL, D and JONES, R (1974) - Opt. Acta 21, 105.

BJELKHAGEN, H (1973) - Opt. Laser Technol. 5, 172.

BOONE, P M (1970a) - Opt. Technol. 2, 94.

BOONE, P M (1970b) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 5-1.

BOONE, P M (1972) - Opt. Laser Technol. 4, 162.

BOONE, P M (1975) - Opt. Acta 22, 579.

BOONE, P M and DE BACKER, L C (1973) - Optik 37, 61.

BOONE, P M amd VERBIEST, R (1969) - Opt. Acta 16, 555.

BOOTH, B L (1975) - Appl. Upt. 14, 593.

BORN, M and WOLF, E (1970) - "Principles of Optics" (4th Edition), Pergamon Press.

BRADFORD, W R (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P (1970), p 57.

BRANDT, G B (1969) - Appl. Opt. 8, 1421.

BRIERS, J D (1971) - Appl. Opt. 10, 519.

BRIERS, J D (1973a) - "Holographic Techniques and Potential Applications", PEL Report No. 404, Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Lower Hutt, New Zealand.

BRIERS, J D (1973b) - New Zealand Engineering 28, 166.

BRIERS, J D (1975a) - Opt. Commun. 13, 324.

BRIERS, J D (1975b) - "Wavelength Dependence as an Additional Degree of Freedom in Biological Applications of Intensity Fluctuation Spectroscopy", paper presented at the Tenth Cong. ICO, Prague; Procs. to be published in 1976.

BRIERS, J D (1975c) - "The Measurement of Plant Growth by means of Holographic Interferometry", paper presented at ICO X.

BRIERS, J D (1975d) - "A Rational Approach to the Interpretation of Holographic Interferograms", paper presented at ICO X.

BRIERS, J D (1975e) - Opt. Quant. Electron. 7, 422.

BROOKS, R E (1974) - private communication.

BROOKS, R E et al. (1965) - Appl. Phys. Letters 7, 248.

BROWN, G M et al. (1969) - J. Acoust. Soc. Am. 45, 1166.

BURCH, J M (1965) - Prodn. Engnr, 44, 431.

BURCH, J M (1970) - in "Optical Instruments and Techniques", Procs. of Eighth Cong. ICO, Reading, 1969 (Ed. Home Dickson), Oriel Press (1970), p 265.

BURCH, J M (1971) - Procs. of SPIE Seminar on "Developments in Holography", p 149.

BURCH, J M and ENNOS, A E (1966) - Opt. Soc. Am. Spring Mtg., PaperWF14

BURCH, J M et al. (1966) - Nature 209, 1015.

BURCHETT, O J and IRWIN, J L (1971) - Mech. Eng. 93 (3), 27.

BUTTERS, J N (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P. (1970), p 151.

BUTTERS, J N and LEENDERTZ, J A (1971) - Opt. Laser Technol. 3, 26.

CARLSON, F D and FRASER, A B (1973) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press(1974), p 519

CATHEY, W T (1965) - J. Opt. Soc. Am. 55, 457.

CHAMPAGNE, E and KERSCH, L (1969) - Opt. Soc. Am. Ann. Mtg., Paper ThF14.

COLBURN, W S and HAINES, K A (1971) - Appl. Opt. 10, 1636.

COLBURN, W S and TOMPKINS, E N (1974) - Appl. Opt. 13, 2934.

COLLIER, R J et al. (1965) - Appl. Phys. Letters 7, 223.

COLLIER, R J et al. (1971) - "Optical Holography", Academic Press.

CONDIE, M A (1966) - Thesis, University of Stanford.

CORTI, M (1975) - verbal communication.

CRANE, R B (1970) - J. Opt. Soc. Am. 60, 1658.

CUMMINS, H Z (1973a) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p 225.

CUMMINS, H Z (1973b) - ibid, p 285.

CUMMINS, H Z and PIKE, E R (1973) - Editors of "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri, Plenum Press (1974).

CUMMINS, H Z and SWINNEY, H L (1970) - Prog. in Optics VIII, p 133.

CUMMINS, H Z et al. (1964) - Phys. Rev. Letters 12, 150.

DAINTY, J C (1970) - Opt. Acta 17, 761.

DAINTY, J C (1971) - Opt. Acta 18, 327.

DAINTY, J C (1972a) - J. Opt. Soc. Am. 62, 595.

DAINTY, J C (1972b) - PhD Thesis, University of London.

DÄNDLIKER, R et al. (1973) - Opt. Commun. 9, 412.

DAVENPORT, W B and ROOT, W L (1958) - "An Introduction to the Theory of Random Signals and Noise", McGraw-Hill.

DENISYUK, Yu N (1962) - Dokl. Akad. Nauk SSSR 144, 1275.

DER, V K et al. (1973) - Appl. Opt. 12, 2552.

DER HOVANESIAN, J and VARNER, J (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P. (1970), p 173.

DE VELIS, J B (1964) - Opt. Soc. Am. Ann. Mtg., Paper FD17.

DE VELIS, J B and REYNOLDS, G O (1967) - "Theory and Applications of Holography", Addison-Wesley.

DUBAS, M and SCHUMANN, W (1974) - Opt. Acta 21, 547.

DUBIN, S B et al. (1967) - Proc. Nat. Acad. Sci. US 57, 1164.

ENNOS, A E (1967) - Contemp. Phys. 8, 153.

ENNOS, A E (1968) - J. Phys. E (J. Sci. Instrum.) 1, 731.

EWERS, W-M et al. (1974) - Optik 40, 57.

FORRESTER, A T et al. (1955) - Phys. Rev. 99, 1691.

FRIESEM, A A and FEDOROWICZ, R J (1966) - Appl. Opt. 5, 1085.

FROEHLY, C et al. (1969) - Opt. Acta 16, 343.

FUJII, H and ASAKURA, T (1974) - Opt. Commun. 11, 35.

FUJIME, S et al. (1972a) - J. Molecular Biol. 68, 347.

FUJIME, S et al. (1972b) - J. Mechanochem. Cell Motility 1, 81.

GABOR, D (1948) - Nature 161, 777.

GABOR, D (1949) - Proc. Roy. Soc. A197, 454.

GABOR, D (1951) - Proc. Phys. Soc. B64, 449.

GABOR, D (1970) - IBM J. Res. Dev. 14, 509.

GATES, J W C (1968) - Nature 220, 473.

GATES, J W C (1969) - Opt. Technol. 1, 247.

GATES, J W C (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-4.

GATES, J W C (1972) - Opt. Laser Technol. 4, 72.

GERRITSEN, H J (1973) - private communication.

GLAUBER, R J (1963) - Phys. Rev. 131, 2766.

GOLDFISCHER, L I (1965) - J. Opt. Soc. Am. 55, 247.

GOODMAN, J W (1963) - Stanford University Electronics Lab. Tech. Rep. No. SEL 63-140 (TR 2303-1).

GOODMAN, J W (1965) - Proc. IEEE 53, 1688.

GOODMAN, J W (1967) - J. Opt. Soc. Am. 57, 493.

GOODMAN, J W (1974) - verbal communication.

GORI, F and MALLAMACE, F (1973) - Opt. Commun. 8, 351.

GOTTENBERG, W G (1968) - Exp. Mech. 8, 405.

GRAY, P F (1975) - PhD Thesis, University of London.

GRAY, P F and BARNETT, M E (1974) - Opt. Commun. 12, 275.

HAINES, K A and HILDEBRAND, B P (1965) - Phys. Letters 19, 10.

HAINES, K A and HILDEBRAND, B P (1966a) - Appl. Opt. 5, 595.

HAINES, K A and HILDEBRAND, B P (1966b) - IEEE Trans.Inst.Meas.15,149.

HANBURY BROWN, R and TWISS, R Q (1956a) - Nature 177, 27.

HANBURY BROWN, R and TWISS, R Q (1956b) - Nature 178, 1046.

HANSCHE, B D and MURPHY, C G (1972) - ISA Trans. 11 (1), 1.

HANSCHE, B D and MURPHY, C G (1974) - Appl. Opt. 13, 630.

HARIHARAN, P and HEGEDUS, Z S (1973) - Opt. Commun. 9, 152.

HARIHARAN, P and HEGEDUS, Z S (1974) - Opt. Commun. 11, 127.

HAVENER, G and RADLEY, R J (1972) - Opto-electron. 4, 349.

HAZELL, C and LIEM, S (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 6-4.

HECHT, N L et al. (1973) - Appl. Opt. 12, 2665.

HEFLINGER, L O et al. (1966) - J. Appl. Phys. 37, 642.

HILDEBRAND, B P (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P. (1970), p 401.

HILDEBRAND, B P and HAINES, K A (1966a) - Appl. Opt. 5, 172.

HILDEBRAND, B P and HAINES, K A (1966b) - Phys. Letters 21, 422.

HILDEBRAND, B P and HAINES, K A (1967) - J. Opt. Soc. Am. 57, 155.

HORMAN, M H (1965) - Appl. Opt. 4, 333.

HUNG, Y Y et al. (1973) - Opt. Commun. 8, 48.

JAKEMAN, E (1973) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p 75.

JAKEMAN, E and PIKE, E R (1969) - J. Phys. A 2, 411.

JOHNSON, F A et al. (1965) - Procs. Int. Conf. on Phys. Quant. Electron. Puerto Rico (Ed. Kelley et al.), McGraw-Hill (1966)

JONES, R (1974) - Opt. Acta 21, 257.

KIKUTA, S et al. (1972) - Opt. Commun. 5, 86.

KING, P W (1974) - Appl. Opt. 13, 231.

KIRK, J P (1966) - Appl. Opt. 5, 1684.

KOHLER, H (1974) - Optik 39, 229.

LANDRY, M J and WISE, C M (1973) - Appl. Opt. 12, 2320.

LANGLEY, K H et al. (1976) - Biochim. Biophys. Acta (in press).

LEADBETTER, I K (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 17-1.

LEE, T C (1974) - Appl. Opt. 13, 888.

LEENDERTZ, J A (1970) - J. Phys. E (J. Sci. Instrum.) 3, 214.

LEITH, E N and UPATNIEKS, J (1962) - J. Opt. Soc. Am. 52, 1123.

LEITH, E N and UPATNIEKS, J (1963) - J. Opt. Soc. Am. 53, 1377.

LOHMANN, A W (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 6-7.

LOWENTHAL, S et al. (1970) - Opt. Commun. 2, 184.

LURIE, M (1967) - Opt. Soc. Am. Spring Mtg., Paper ThF14.

LURIE, M (1968a) - J. Opt. Soc. Am. 58, 614.

LURIE, M (1968b) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P. (1970), p 397.

LURIE, M and ZAMBUTO, M (1968) - Appl. Opt. 7, 2323.

LUXMOORE, A R and HOUSE, C (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 5-2.

MACÁKOVÁ, M (1974) - Opt. Applic. 4 (1), 45.

MACHADO GAMA, M A (1973) - Opt. Commun. 8, 362.

MAEDA, T and FUJIME, S (1972) - Rev. Sci. Instrum. 43, 566.

MAGILL, P J and WILSON, A D (1968) - J. Appl. Phys. 39, 4717.

MAROM, E and MUELLER, R (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 5-5.

MARTIENSSEN, W (1968) - Procs. Eighth Int. Cong. High Speed Phot., 289.

MARTIENSSEN, W (1970) - Math.-Phys. Semesterber. 17, 13.

MATSUMOTO, T et al. (1973) - Appl. Opt. 12, 961.

MATSUMOTO, T et al. (1974) - Appl. Opt. 13, 1080.

McKECHNIE, T S (1974) - PhD Thesis, University of London.

McKECHNIE, T S (1975) - "Speckle Reduction", in "Laser Speckle and Related Topics" (Ed. Dainty), Vol. 9 in series "Topics in Applied Physics", Springer (in press).

MER, C (1974) - verbal communication.

MICHAEL, M (1973) - Appl. Opt. 12, 1111.

MIDDLETON, D (1960) - "An Introduction to Statistical Communication Theory", McGraw-Hill.

MOFFAT, C and WATRASIEWICZ, B M (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 6-2.

MOLIN, N-E and STETSON, K A (1969) - J. Phys. E (J. Sci. Instrum.) 2, 609.

MOLIN, N-E and STETSON, K A (1970) - Optik 31, 157 and 281.

MOLIN, N-E and STETSON, K A (1971) - Optik 33, 399.

MONNERET, M (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-3.

MOTTIER, F M (1970) - ibid, Paper 6-5.

NASSENSTEIN, H (1966) - Phys. Letters 21, 290.

NEUMANN, D B (1968) - J. Opt. Soc. Am. 58, 447.

NICOLAS, J (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 5-4.

NOSSAL, R (1971) - Biophys. J. 11, 341.

NOSSAL, R and CHEN, S H (1972a) - J. de Phys. 33-C1, 171.

NOSSAL, R and CHEN, S H (1972b) - Opt. Commun. 5, 117.

NOSSAL, R et al. (1971) - Opt. Commun. 4, 35.

OLIVER, C J (1973) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p 151.

PARRY, G (1975) - PhD Thesis, University of London.

PASTEUR, J (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 5-7.

PASTOR, J et al. (1970) - Opt. Acta 17, 81.

PEDERSEN, H M (1974) - Opt. Commun. 12, 156.

PENNINGTON, K S and LIN, L H (1965) - Appl. Phys. Letters 7, 56.

PIDDINGTON, R W (1973) - abstract in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p573.

PIDDINGTON, R W (1975) - in "Perspectives in Experimental Biology", Procs. of 50th Anniv. Mtg. Soc. Exp. Biol. London, Oxford (Ed. Spencer-Davies), Pergamon Press (in press).

PIKE, E R (1973) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p 3.

PIKE, E R and JAKEMAN, E (1973) - "Photon Statistics and Photon Correlation Spectroscopy" (Advances in Quantum Electronics, Vol. 2), Academic Press.

POWELL, R L and STETSON, K A (1965a) - Opt. Soc. Am. Spring Mtg., Paper FA16.

POWELL, R L and STETSON, K A (1965b) - J. Opt. Soc. Am. 55, 1593.

PŘIKRYL, I (1974) - Opt. Acta 21, 675.

PŘIKRYL, I (1975) - "Distinguishing and Determination of the Order of Interference Fringes in Holometry", paper presented at the Tenth Cong. ICO, Prague; Procs. to be published.

PUSEY, P N (1973) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p 387.

REDMAN, J D (1967) - J. Sci. Instrum. 44, 1032.

RIGDEN, J D and GORDON, E I (1962) - Proc. IRE 50, 2367.

RIVA, C et al. (1972) - Investigative Ophthal. 11, 936.

ROBERTSON, E R and ELLIOTT, (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 10-6.

ROBERTSON, E R and HARVEY, J M (1970) - Eds. of "Engineering Uses of Holography", Procs. of Symposium, Strathclyde, 1968, C.U.P.

ROGERS, G L (1950) - Nature 166, 237.

ROGERS, G L (1951) - Proc. Roy. Soc. Edin. A63, 193.

ROWE, S H (1971) - J. Opt. Soc. Am. 61, 1599.

SAKUSABE, T and KOBAYASHI, S (1971) - Jap. J. Appl. Phys. 10, 758.

SAMPSON, R C (1970) - Exp. Mech. 10, 313.

SCHAEFFER, D W (1973) - Science 180, 1293.

SCIAMMARELLA, C A and GILBERT, J A (1973) - Appl. Opt. 12, 1951.

SHERIDON, N K (1968) - Appl. Phys. Letters 12, 316.

SHIBAYAMA, K and UCHIYAMA, H (1971) - Appl. Opt. 10, 2150.

SIEGERT, A J F (1943) - MIT Rad. Lab. Rep. No. 465.

SMITH, H M (1969) - "Principles of Holography", Wiley-Interscience.

SOLLID, J E (1969) - Appl. Opt. 8, 1587.

SOLLID, J E (1970a) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-8.

SOLLID, J E (1970b) - Opt. Commun. 2, 282.

STEEL, W H (1970) - Opt. Acta 17, 873.

STETSON, K A (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P. (1970), p 307.

STETSON, K A (1969) - Optik 29, 386.

STETSON, K A (1970a) - J. Opt. Soc. Am. 60, 1378.

STETSON, K A (1970b) - Optik 31, 576.

STETSON, K A (1971) - J. Opt. Soc. Am. 61, 1359.

STETSON, K A (1972a) - J. Opt. Soc. Am. 62, 297.

STETSON, K A (1972b) - J. Opt. Soc. Am. 62, 698.

STETSON, K A (1972c) - Appl. Opt. 11, 1725.

STETSON, K A (1974a) - J. Opt. Soc. Am. 64, 1.

STETSON, K A (1974b) - in "Holographic Nondestructive Testing" (Ed. R K Erf), Academic Press, p 181.

STETSON, K A (1975) - Appl. Opt. 14, 272.

STETSON, K A and POWELL, R L (1965a) - Opt. Soc. Am. Ann. Mtg. Paper WF13.

STETSON, K A and POWELL, R L (1965b) - J. Opt. Soc. Am. 55, 1694.

STETSON, K A and POWELL, R L (1966) - J. Opt. Soc. Am. 56, 1161.

STROKE, G W (1965) - Appl. Phys. Letters 6, 201.

STROKE, G W (1966) - "An Introduction to Coherent Optics and Holography", Academic Press.

STROKE, G W (1969) - idem (2nd Edition), Academic Press.

STROKE, G W and LABEYRIE, A E (1966) - Phys. Letters 20, 368.

SWINNEY, H F (1973) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p 331.

TANNER, L H (1968) - J. Phys. E (J. Sci. Instrum.) 1, 517.

TRIBILLON, G and MILES J F (1974) - Opt. Commun. 11, 123.

TSUJIUCHI, J and MATSUDA, K (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-5.

TSUJIUCHI, J et al. (1969) - Opt. Acta 16, 709.

TSURUTA, T et al. (1967) - Jap. J. Appl Phys. 6, 661.

TSURUTA, T et al. (1969) - Opt. Acta 16, 723. (see also Jap. J. Appl. Phys. 7 (1968), 1092).

UZGIRIS, E E (1972) - Opt. Commun. 6, 55.

VAN VLECK, J H and MIDDLETON, D (1966) - Proc. IEEE 54, 2.

VARNER, J R (1970) - Appl. Opt. 9, 2098.

VAUGHAN, J M (1973) - in "Photon Correlation and Light Beating Spectroscopy", Procs. of NATO Adv. Study Inst., Capri (Ed. Cummins and Pike), Plenum Press (1974), p 429.

VELZEL, C H F (1973) - Opt. Commun. 7, 302.

VEST, C M (1973) - Appl. Opt. 12, 612.

VIÉNOT, J-C (1970) - Nouv. Rev. d'Opt. Appl. 1, 91.

VIÉNOT, J-C et al. (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P.(1970), p 133.

VIÉNOT, J-C et al. (1970) - Eds. of "Applications de l'Holographie", Procs. of Symposium, Besançon, 1970, Université de Besançon.

VIKRAM. C S (1973a) - Opt. Commun. 8, 355.

VIKRAM, C S (1973b) - Appl. Opt. 12, 2808.

WADDELL, P et al. (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 6-6.

WALL, M R (1968) - in "Engineering Uses of Holography", Procs. of Symposium, Strathclyde (Ed. Robertson and Harvey), C.U.P.(1970), p 355.

WALL, M R (1970) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-9.

WALLES, S (1970a) - Arkiv för Fysik 40 (4), 299.

WALLES, S (1970b) - Opt. Acta 17, 899.

WARE, B R and FLYGARE, W H (1971) - Chem. Phys. Letters 12, 81.

WARE, B R and FLYGARE, W H (1972) - J. Coll. Interf. Sci. 39, 670.

WELFORD, W T (1969) - Opt. Commun. 1, 123.

WELFORD, W T (1970a) - Opt. Commun. 1, 311.

WELFORD, W T (1970b) - in "Applications de l'Holographie", Procs. of Symposium, Besançon (Ed. Viénot et al.), Université de Besançon (1970), Paper 4-2.

WILLIAMS, T H (1973) - PhD Thesis, University of London.

WILSON, A D (1970a) - Appl. Opt. 9, 2093.

WILSON, A D (1970b) - J. Opt. Soc. Am. 60, 1068.

WILSON, A D (1971a) - Appl. Opt. 10, 908.

WILSON, A D (1971b) - J. Opt. Soc. Am. 61, 924.

WILSON, A D and STROPE, D H (1970) - J. Opt. Soc. Am. 60, 1162.

ZAMBUTO, M H and LURIE, M (1970) - Appl. Opt. 9, 2066.

ZELENKA, J S and VARNER, J R (1968) - Appl. Opt. 7, 2107.

ZELENKA, J S and VARNER, J R (1969) - Appl. Opt. 8, 1431.

ACKNOWLEDGMENTS

The author would like to acknowledge the supervision of Professor W T Welford during the course of this project, and to thank him for his continuing interest, advice and encouragement. We are also grateful to the other members of the Optics Section at Imperial College for many useful and stimulating discussions, and in particular to Dr T S McKechnie and G Parry for introducing us to the mysteries of speckle statistics. Cooperation from the Department of Botany at Imperial College has been invaluable throughout the project, and we are particularly indebted to Dr J W Hannay for his advice and suggestions, and also for the provision of botanical specimens.

This work has been financed by a Study Award from the Department of Scientific and Industrial Research, New Zealand, and we are especially grateful to Dr M C Probine, Director of the Department's Physics and Engineering Laboratory, for his support during the application for this award, and for granting the necessary study leave.

PUBLISHED PAPERS

Copies of the following publications by the author are submitted as additional material for consideration by the examiners:

1. "Prism Shearing Interferometer" - Opt. Technol. 1, 196 (1969).

2. "Interferometric Flatness Testing of Nonoptical Surfaces" - Appl. Opt. 10, 519 (1971).

3. "Self-Compensation of Errors in a Lateral Shearing Interferometer" - Opt. Commun. 4, 69 (1971).

4. "Interferometric Testing of Optical Systems and Components: a Review" - Opt. Laser Technol. 4, 28 (1972).

5. "Industrial and Engineering Applications of Visible-Light Lasers" - NZ Engng. 28, 166 (1973).

6. "Wavelength Dependence of Intensity Fluctuations in Laser Speckle Patterns from Biological Specimens" - Opt. Commun. 13, 324 (1975).

7. "A Note on the Statistics of Laser Speckle Patterns added to Coherent and Incoherent Uniform Background Fields, and a Possible Application for the Case of Incoherent Addition" - Opt. Quant. Electron. 7, 422 (1975).

PRISM SHEARING INTERFEROMETER

J. D. BRIERS

Pilkington Perkin-Elmer Ltd, St Asaph, Flintshire

A brief account of the theory of the wavefront shearing interferometer is given, followed by a description of Bates' version of the instrument. This has been redesigned in solid glass to form a 'prism shearing interferometer' whose main advantage lies in its robustness. Methods of using the interferometer are described and illustrated, including the use of moiré fringes to magnify the aberration effects in a 'double-pass' technique.

A MAJOR PROBLEM in the interferometric testing of large-aperture optical systems is the generation of a suitable reference wavefront. If a 20cm aperture system, for example, is to be tested with a standard Twyman-Green interferometer, the aperture of the interferometer must also be 20cm. This implies the use of two good quality 20cm lenses, two 20cm reference flats, and a 30 × 20cm optically flat beamsplitter. The cost of such an instrument would be prohibitive.

This problem can be overcome by using wavefront shearing interferometry in which the wavefront from the system under test interferes with a sheared, or slightly displaced, image of itself.

Theory of wavefront shearing interferometer

The principle of the wavefront shearing interferometer (WSI) is the production of an interferogram by superposition of a wavefront on a sheared image of itself (see Fig. 1). By 'shearing' is meant the fractional displacement of the wavefront by rotation about its focus; hence a perfect spherical wavefront is identical to its sheared self, and no fringes would be observed. In star space, shearing of a plane wavefront implies a lateral displacement, the focus being at infinity.

If the wavefront aberration of the original wavefront is $W(x, y)$ and the amount of shear is h, the fringes represent a 'contour map' of the function $h\ \partial W/\partial x$, providing h is small. Because $\partial W/\partial x$ is proportional to the ray aberration, one is effectively obtaining ray aberration data directly from an interferometric method. (In practice, h is not very small, and this conclusion is not completely valid). To obtain the wavefront aberration $W(x, y)$, the fringe function is plotted and integrated with respect to x. It should be noted that the sensitivity of the test increases as h increases, and hence as the validity of the above theory decreases.

A more common method of using shearing interferometry is to 'tilt' one of the wavefronts with respect to the other. By tilt is meant the rotation of one of the wavefronts about a diameter. Such tilt can be resolved into two components—tilt about the axis orthogonal to the shear axis (α-tilt) and tilt about the shear axis, (β-tilt), see Fig. 2.

Tilt is introduced for the same reason as in other interferometers, such as the Twyman-Green,—to

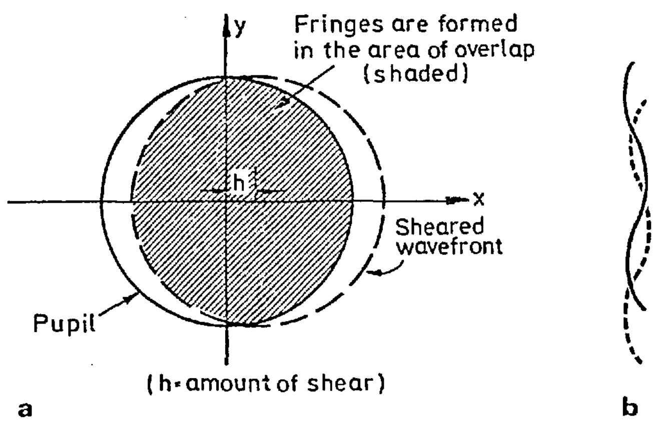


Fig. 1. Principle of shearing interferometry (a) shearing effect, (b) plan view of wavefronts

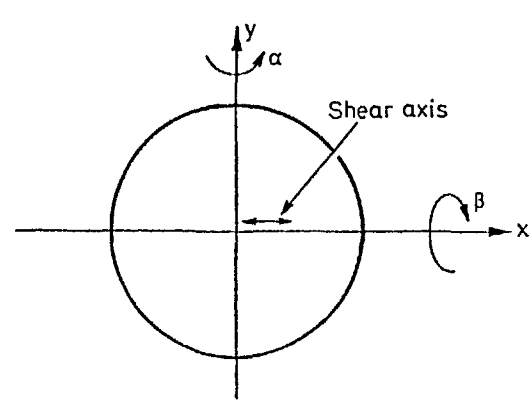


Fig. 2. Tilting the wavefront

provide a reference grid against which fringe displacements are measured. A perfect spherical wavefront produces an interferogram consisting of straight, parallel, evenly-space fringes, parallel to the axis of tilt. Shearing interferometers are usually designed so that the tilt is either pure α or pure β.

In the case of β-tilt, the equation of the central fringe is

$$y = \frac{-h}{\beta} \frac{\partial W}{\partial x}$$

where h is the amount of shear and β the angle of tilt.

Hence the fringe represents a plot of $\partial W/\partial x$, the transverse ray aberration. Further, h can be measured directly, and β is pre-set (or can be obtained from the number of fringes observed). Fig. 3 shows a typical fringe pattern observed with a shearing interferometer having its tilt and shear axes coinciding. The case illustrated is pure spherical aberration, for which the wavefront aberration has a fourth-power dependency on x; hence the cubic fringes shown in the diagram.

α-tilt interferometers are not nearly so useful as those using β-tilt. In the case of α-tilt, the interpretation of the fringe-pattern is more difficult, because the central fringe is now the line $x = 0$. The other fringes represent a function of the longitudinal ray aberration, but in a rather complex form. [redacted] The observed fringe pattern for the case of pure spherical aberration is shown in Fig. 4.

In both the cases discussed above, the fringes are not 'absolute' in that they depend on the amount of shear used. The second method (α-tilt) is less sensitive than the first for the same degree of tilt.

Theory of wavefront shearing interferometers, including analysis of the patterns observed, can be found in the literature.[1-4]

Bates shearing interferometer

Most shearing interferometers, such as those based on Ronchi gratings[5] and Wollaston prisms[6,7] are of the α-tilt type, and hence produce interferograms that are difficult to interpret. A notable exception is the wavefront shearing interferometer developed in 1947 by Bates[8], based on the Mach-Zehnder interferometer. The main advantages of this design are the use of β-tilt, and the fact that both shear and tilt can be varied at will, and independently.

The principle of the interferometer is illustrated in Fig. 5. The beam is focused on the partially reflecting face of mirror 3 (the recombiner). If this mirror is tilted, beam B will be sheared with respect to beam A (the former will be deflected at mirror 3, whereas the latter will be unaffected). If mirrors 2 and 2′ are rotated by equal amounts, tilt will be introduced: the foci will be displaced with respect to each other, and the fact that both mirrors are tilted ensures that the beams are still going in the same direction and will be superimposed. By choosing the axis of rotation of mirrors 2 and 2′ correctly, the tilt can be arranged to be about the shear axis (i.e. β-tilt).

In practical forms of the instrument, compensating plates are included to compensate for the different aberration effects in the two beams.

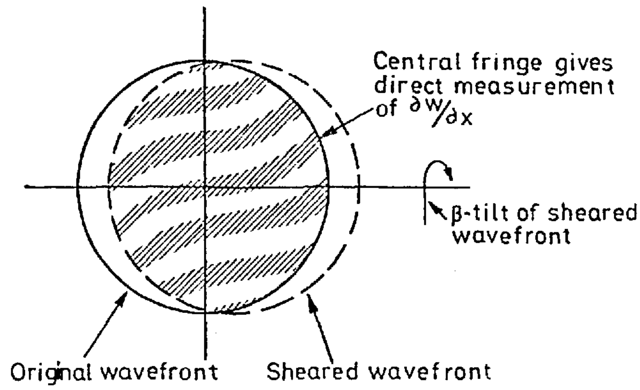


Fig. 3. β-tilt shearing interferogram

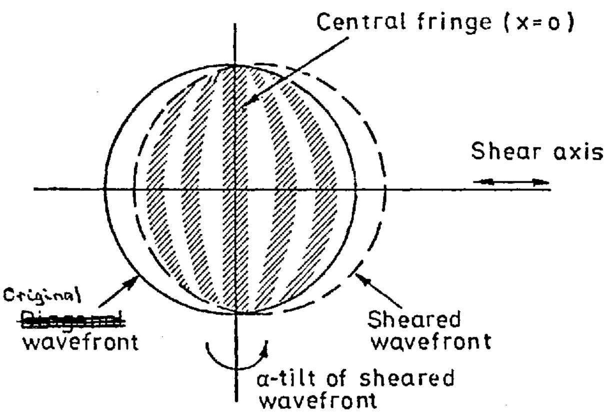


Fig. 4. α-tilt shearing interferogram

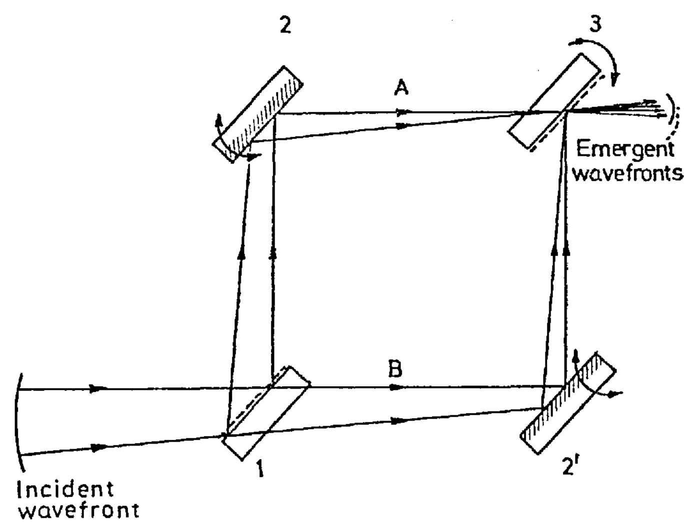


Fig. 5. Bates shearing interferometer

In 1951, Drew[9] simplified the Bates interferometer by replacing the two beam-splitters by a single glass block. Fig. 6 illustrates the basic design of the instrument. By using a wedged block of glass, a fixed shear is introduced, and variable tilt can be introduced by rotating the two mirrors.

Brown[10] simplified the design still further in 1954 by removing all adjustments, and manufacturing the interferometer to give a fixed amount of shear and tilt.

Prism shearing interferometer

A logical development from Brown's fixed shear, fixed-tilt version of the Bates interferometer was the construction of the device in solid glass. This was reported in 1964 by Saunders[11], who suggested various designs according to the particular use for which the instrument was intended. A typical Saunders interferometer is illustrated in Fig. 7.

A slightly different approach has been adopted by the present author, who has designed a prism shearing interferometer based on the classical Bates interferometer.

The interferometer is constructed from four prisms, labelled A, B, C and D in Fig. 8. A and B are ordinary 90°-45°-45° prisms, one of them having its hypotenuse face coated with a semi-reflecting layer. C and D are rhomboids, with the angle θ being slightly greater than 45°. (The deviation of θ from 45° determines the amount of shear introduced by the interferometer—a typical value for θ would be 45° 15′). Faces X and Y are tilted about the axis in the plane of the diagram by a small amount, of the order of, say, 10′, in order to provide the tilt. Faces X and Y are coated with a totally reflecting layer, and the opposite face of one of the rhomboids with a semi-reflecting layer. The prisms are cemented together in the arrangement of Fig. 8, and all faces except the four square faces (marked T in the diagram) can be painted matt black, to eliminate unwanted reflections.

The physical size of the interferometer is immaterial to its performance, because its usefulness is limited only by the f-number of the beam—it will accept any beam of a smaller aperture than f/2.5 in the glass. If the refractive index of the glass used in the interferometer is 1.5, this implies it will accept an f/1.7 beam, and a glass with higher refractive index would enable it to accept an even steeper beam. A convenient size for the instrument is based on 2.5cm-square entrance and exit faces.

In practice, the incidence of a converging beam on plane refracting faces introduces an amount of spherical aberration into the beam. This must be allowed for when the final interferogram is being assessed. The spherical aberration can, however, be eliminated by cementing plano-convex lenses onto the entrance and exit faces of the interferometer, so that the centres of curvature of the convex surfaces are coincident with the focus of the beam, i.e. the centre of the shear plane. This arrangement, also adopted by Saunders[11], ensures that all rays intersect the entrance and exit faces at normal incidence, and hence pass through undeviated. Unfortunately, at the same time as removing the spherical aberration this modification reduces the limiting aperture of the interferometer to f/2.5.

Assembly of the interferometer is simple, providing all the angles (other than the special ones mentioned above) are fairly accurately 90° or 45°. The four prisms are laid on their sides in the correct orientation, and brought together with optical cement on the faces to be joined. Using this technique the zero-order fringe is usually visible, even if not exactly in the centre of the field of view. Thus the interferometer can be used fairly satisfactorily even in white light; if monochromatic light is used, of course, there is no need to identify the zero-order fringe, since it is the centrally-positioned fringe that is used.

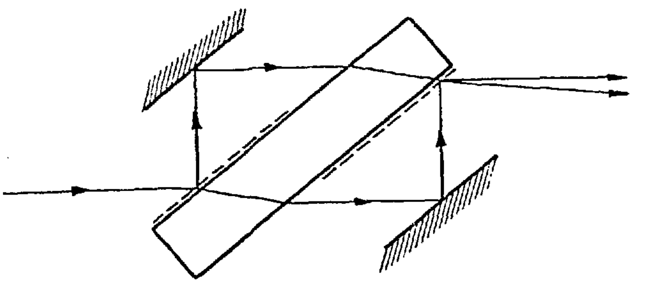


Fig. 6. Drew's interferometer

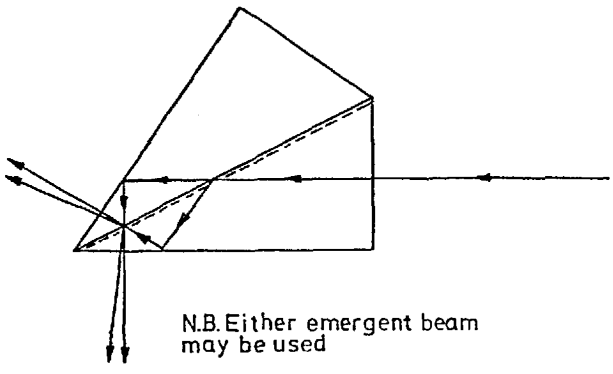


Fig. 7. One of Saunders' prism interferometers

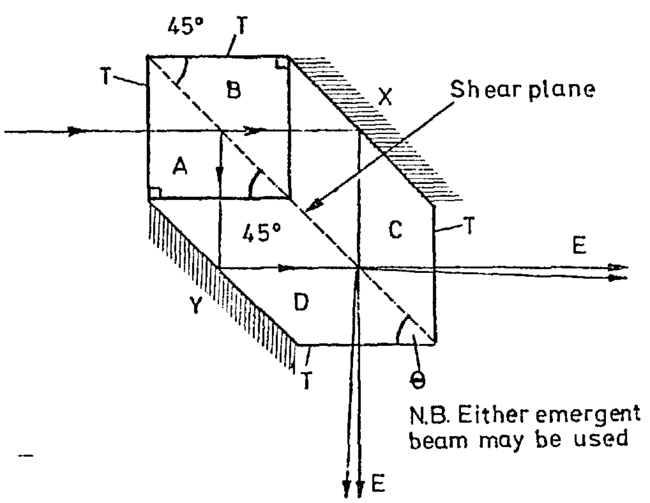


Fig. 8. Prism shearing interferometer

Using the interferometer

Basically, the interferometer is used in exactly the same way as the Bates interferometer, i.e. by placing it in a converging beam so that the beam is focused in the neighbourhood of the second semi-reflecting interface (the shear plane). The interferometer is then moved along the beam until the observed fringes, projected on a screen or viewed through an eyepiece, are horizontal and have maximum spacing. (Either of the two emergent beams labelled E in Fig. 8 can be used.) For a perfectly spherical converging wavefront, the fringes would be horizontal, straight, parallel and evenly spaced. Any deviation from straightness is an indication of aberration in the wavefront, the shape of the fringes giving the derivative of the wavefront aberration, and hence the transverse ray aberration, as described above.

The usual limitations on source size apply to shearing interferometers as to other interferometers such as the Twyman-Green. The usual approach, since the shear is one-dimensional, is to use a slit source, with either white or monochromatic light. An alternative source nowadays is the laser, which produces high-contrast fringes without the need for pinholes or slits. A microscope objective of sufficient power to fill the component under test can be used to diverge the beam.

Testing converging lenses and lens systems

The point or slit source (or focus of the microscope objective if a laser is being used) is placed at one conjugate of the lens under test, and the interferometer at the other conjugate. The lens should be tested at the conjugates for which it was designed. If one of the conjugates is infinite, it will be necessary either to collimate the incident beam at the appropriate beam-width, using a laser with a beam-expander for example, or to place a plane mirror behind the lens and test as though testing a concave mirror (see below). The normal arrangement for testing a lens is shown in Fig. 9.

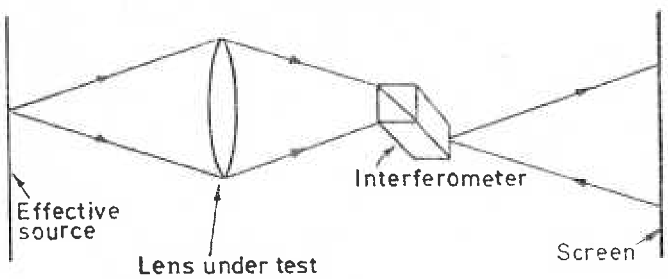


Fig. 9. Testing a converging lens

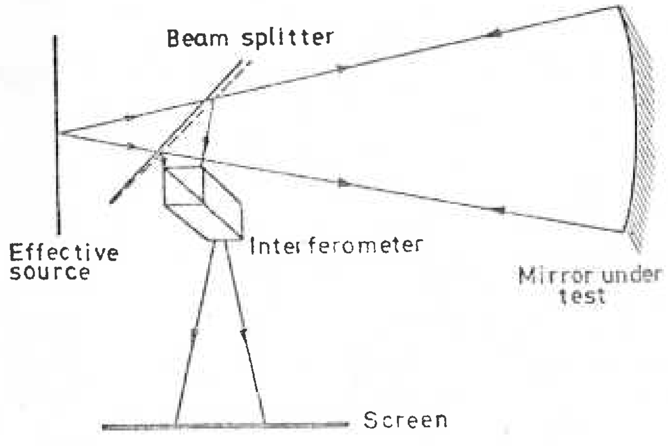


Fig. 10. Testing a concave mirror using a beam-splitter

Testing a concave mirror

A concave spherical mirror is tested at its centre of curvature, where, in theory, both the effective source and the shear plane of the interferometer must be placed. In practice, it is necessary either to use an auxiliary beam-splitter (see Fig. 10), or to test slightly off-axis (see Fig. 11). The latter method, of course, will introduce off-axis aberrations such as coma and astigmatism, and care must be used in this method.

Interpreting results

The interference pattern observed with a perfect spherical wavefront is a series of straight, horizontal, parallel, equally spaced fringes, and this is obviously the optimum pattern to aim for. The transverse ray aberration can be calculated directly from

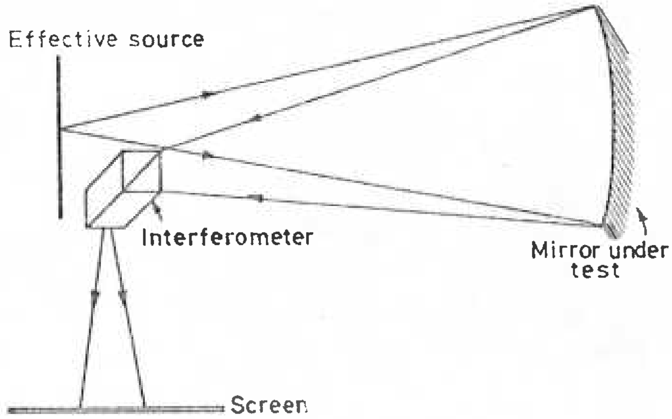


Fig. 11. Testing a concave mirror, off-axis method

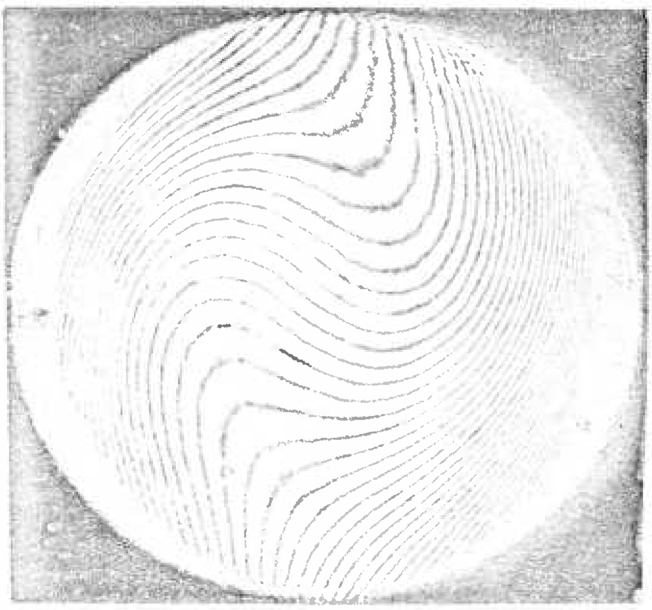


Fig. 12. Interferogram of simple biconvex lens at equal conjugates

the fringe pattern if the shear and tilt of the interferometer are known, or the instrument has been calibrated. Spherical aberration is represented by cubic fringes, and coma by a pattern incorporating elliptical fringes.

Examples of shearing interferograms

Examples of interferograms obtained with the prism shearing interferometer described above are given in Figs 12-15. The source used was a helium-neon laser, with a x40 microscope objective to diverge the beam.

Fig. 12 is the result of testing a simple, symmetrical biconvex lens at equal conjugates using the arrangement of Fig. 9. The cubic nature of the fringes indicates the presence of a large amount of spherical aberration, as expected from a simple, uncorrected lens.

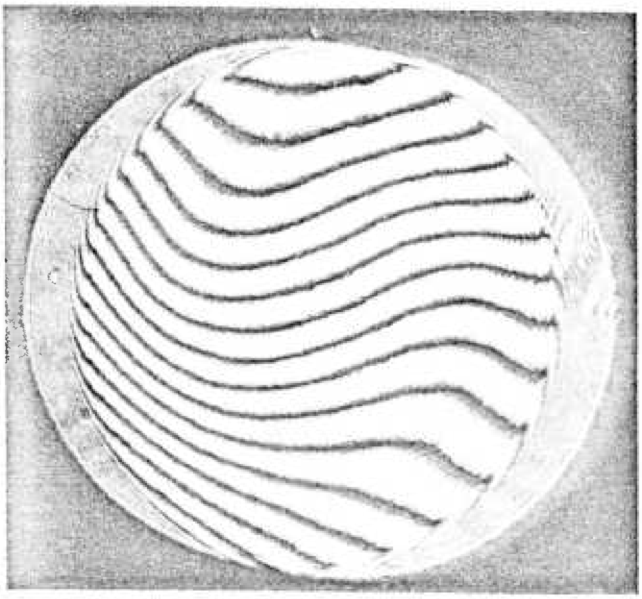


Fig. 13. Interferogram of a complex lens, showing asymmetrical fringes

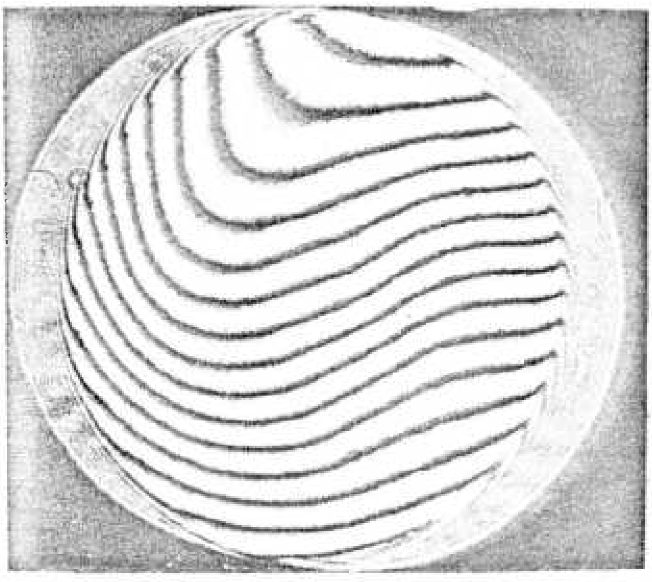


Fig. 14. The same lens, after rotation through approximately 90°

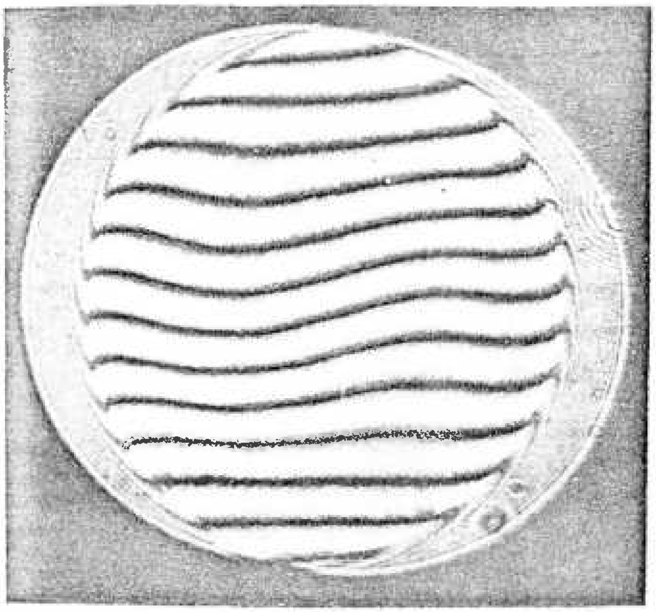


Fig. 15. The same lens after modifying one curve and re-working another—note improvement over Figs 13 & 14

A more complex, eight-element lens was also tested, in its design configuration of equal conjugates. The lens had been rejected in a final-inspection resolution test, and Fig. 13 shows the asymmetrical fringe pattern produced by the interferometer. The asymmetry was even more apparent when the lens was rotated in its mount—Fig. 14. The lens was then stripped down, and each component checked separately against the design tolerances. As a result, the radius of one curve was modified slightly, and another surface was re-worked and re-polished. The lens was re-assembled and inspected again. The resolution requirements were met, and Fig. 15 shows how the interferogram had improved (cf. Figs 13 & 14).

Fig. 16 shows the result of testing a concave spherical mirror on-axis, using a beam-splitter as shown in Fig. 10. There is a slight curvature in the fringes, illustrating a slight amount of spherical aberration together with some asymmetrical effects.

Double-pass technique

Since the interferometer has fixed shear and tilt, the fringe-spacing of Fig. 16 cannot be changed. However, a method has been devised whereby the small aberration effects visible in that photograph can be magnified. The technique involves using the interferometer in a double-pass arrangement, as illustrated in Fig. 17. The interferometer is positioned immediately behind the x40 microscope objective so that the laser beam is focused as close as possible to the shear plane of the device. This causes the beam to be sheared and fringes to be formed in the normal way. The resulting fringe

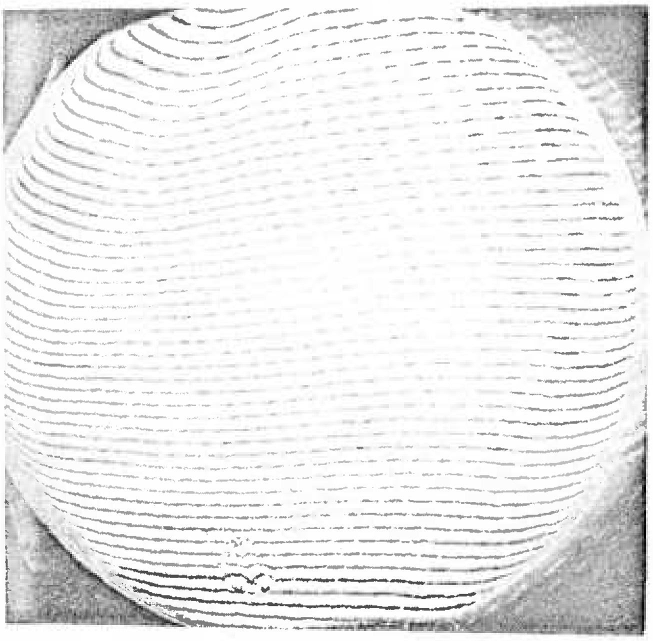


Fig. 16. Interferogram of a concave mirror tested on-axis with a beam-splitter

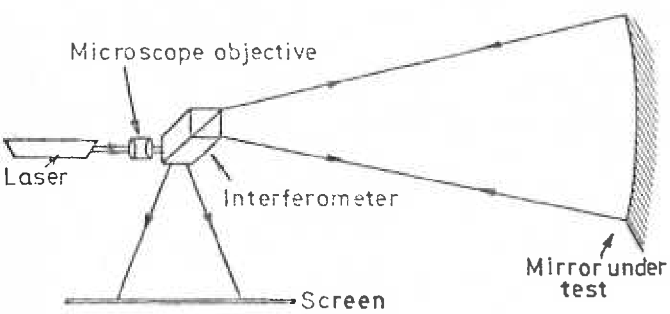


Fig. 17. 'Double-pass' arrangement for magnifying small aberrations

pattern is shown in Fig. 18, the straightness of the fringes indicating the high quality of the objective.

The beam is then reflected from the concave mirror under test, and returns to the interferometer where it is again sheared. This second shearing results in the original fringe pattern of Fig. 16, which is superimposed on the straight fringe pattern of Fig. 18 already 'in the beam'. The combined fringe pattern (Fig. 19) consists of moiré fringes produced by the superposition of the two original patterns. By comparing Figs 16 & 19, it can be seen that the original fringe shape of the former has been magnified by the moiré effect so that the shape is clearly discernible.

Conclusions

The main advantage of the wavefront shearing interferometer (WSI) over other types of interferometric test apparatus is the elimination of the need for a separate reference wavefront. This enables large aperture systems to be tested without the necessity of providing, for example, large reference flats. Additional advantages of the prism shearing interferometer (for which the abbreviation PSI is suggested) include

- —rigid, robust construction, resulting in less chance of breakage or disturbance of the optics
- —lack of adjustment, making it simpler to use
- —relative ease of manufacture and assembly
- —wider acceptance angle (up to f/1.5), providing the associated spherical aberration is calculated and allowed for in the interpretation of interferograms
- —possible removal of inherent aberration effects by the cementing of suitable plano-convex

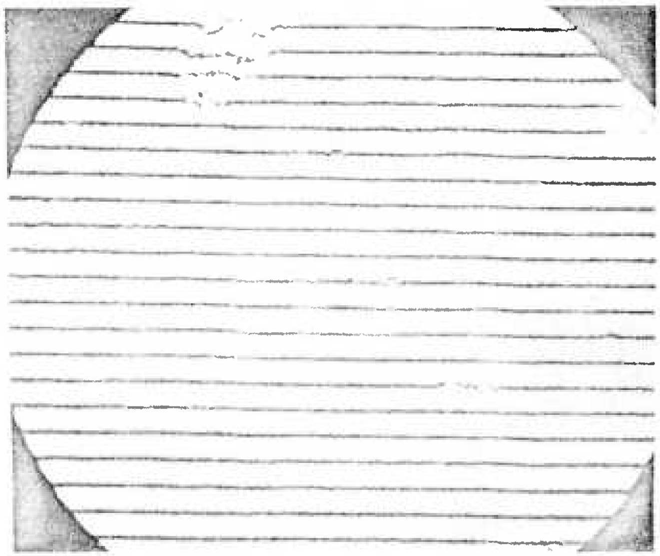


Fig. 18 Interferogram of x40 microscope objective—note straightness of fringes. (N.B. local irregularities are caused by laser diffraction effects and can be ignored)

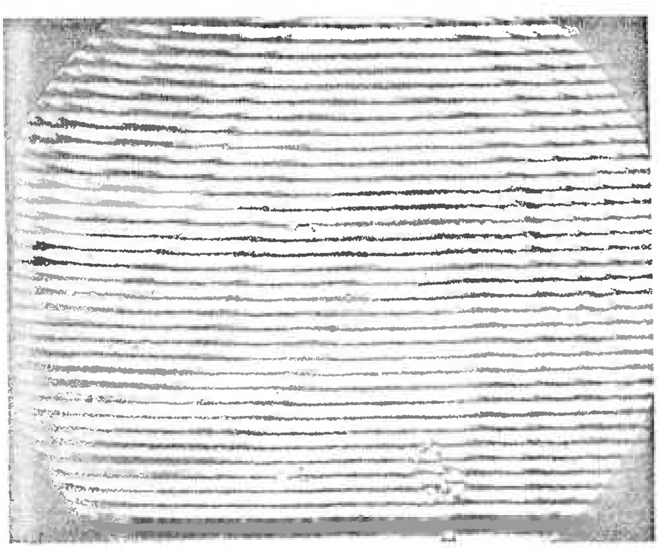


Fig. 19. Interferogram of concave mirror of Fig. 16, tested by 'Double-pass' arrangement of Fig. 17—note moiré fringe pattern

lenses on to the entrance and exit faces
—small physical size, resulting in easier handling.

The major disadvantage of the WSI when compared with devices such as the Twyman-Green interferometer is that the resulting fringe-patterns are far more difficult to interpret quantitatively, especially to an unskilled operator. In addition, the PSI suffers from the following disadvantages

—lack of adjustment limits its use to the range of apertures and defects for which it is specifically designed
—testing of low-aperture beams raises problems because of the long throw needed to produce overlapping of the two emergent beams.

In conclusion, it should be noted that a new design for a PSI has been announced by van Rooyen[12], who has incorporated the facility of variable shear into a solid wavefront shearing interferometer.

Acknowledgement. The author thanks all his colleagues who have contributed to the work described, and Pilkington Perkin-Elmer Ltd for permission to publish.

REFERENCES

1. Brown, D. S., The application of shearing interferometry to routine optical testing, J. Sci. Inst. vol. 32. 1955. pp. 137-9.
2. Saunders, J. B., Measurement of wavefronts without a reference standard. Part 1. The Wave-front-shearing Interferometer, J. Res. Nat. Bur. Stands, vol. 65B. 1961. pp. 239-44.
3. Bryngdahl, O., Applications of shearing interferometry. Progress in Optics, Vol. 4. 1965. pp. 39-83.
4. Malacara, D. & Mendez, M., Lateral shearing interferometry of wavefronts having rotational symmetry, Opt. Acta. vol. 15. 1968. pp. 59-63.
5. Ronchi, V., Forty years of history of a grating interferometer, App. Opt. vol. 3. 1964. pp. 437-51.

   See also:

   Ronchi, V., Forty years of 'gratings', Atti. Fond. G. Ronchi, vol. 17. 1962. pp. 93-143.
6. Lenouvel, L. & Lenouvel, F., Rev. d'Opt. vol. 17. 1938. p. 350 et seq.
7. Bartholomeyczyk, W., Use of the wollaston prism as beam-splitter in shearing interferometers, Instrum. Kde. vol. 68. 1960. pp. 208-13.
8. Bates, W. J., A wavefront shearing interferometer, Proc. Phys. Soc. vol. 59. 1947. pp. 940-50.
9. Drew, R. L., A simplified shearing interferometer, Proc. Phys. Soc. B. vol. 64. 1951. pp. 1005-10.
10. Brown, D., A shearing interferometer with fixed shear and its application to some problems in the testing of astro-optics, Proc. Phys. Soc. vol. 67. 1954. pp. 232-7.
11. Saunders, J. B., Wavefront shearing prism interferometer, J. Res. Nat. Bur. Stands. vol. 68C. 1964. pp. 155-71.
12. Van Rooyen, E., Design for a variable shear prism interferometer. App. Opt. vol. 7. 1968. p. 2433.

Reprinted from APPLIED OPTICS, Vol. 10, page 519, March 1971

Interferometric Flatness Testing of Nonoptical Surfaces

J. D. Briers

An interferometric method of testing the flatness of nonoptical surfaces is described. The method consists essentially of introducing light to the surface at near grazing incidence; this is effected by using an isosceles glass prism, the base of which is the reference surface. The technique can utilize diffuse daylight as a light source and provides a quick and easy method of assessing the flatness of such items as smoothed glass flats prior to polishing, flat metal laps, and other nonoptical surfaces.

Introduction

The interpretation of interference fringes formed between two near perfect optical flats is a well-known and widely practiced method of testing the flatness of a workpiece against a reference flat. In its simplest form, the technique consists simply of putting the two surfaces in contact with each other and observing the fringes produced in the air gap between them. Since white-light fringes are virtually unusable due to chromatic effects, a relatively coherent source such as a sodium or mercury lamp is invariably used. The fringes formed are Fizeau-type contour fringes of the air gap between the flats; if λ is the wavelength of the light used, the contour spacing is $\lambda/2$, providing the fringe pattern is viewed in a direction normal to the interface.

It is apparently not widely known that interference effects can also be observed from nonoptical surfaces, though a suitable technique was described by Herschel[1] over 160 years ago. The technique is to use light incident at very large angles of incidence. Langenbeck[2] mentioned the possibility of testing nonoptical surfaces with the Lloyd's mirror interferometer, but the application of true Fizeau-type fringes to such surfaces appears to have been neglected until a recent paper by Abramson.[3] Using a prism to introduce a collimated light beam onto the surface at near grazing incidence, Abramson described a method of measuring the flatness of nonoptical surfaces such as metal, wood, and even paper. The technique is illustrated in Fig. 1. The fringes observed are Fizeau-type contour fringes but with the contour spacing modified by a magnification $1/\cos\beta$, where β is the angle of incidence in the air gap. The fringes are interpreted in exactly the same way as fringes observed between two optical flats, except that the sensitivity is modified by the $1/\cos\beta$ factor mentioned above.

If the collimated light of Abramson's arrangement is replaced by an extended source, it is possible to observe Haidinger-type fringes (fringes of equal inclination). It is believed that these fringes are the streaks described by Herschel.[1] They can be observed by the naked eye without any auxiliary optics—only the prism is required. The fringes have been in routine use for some years in this laboratory as a qualitative assessment of the flatness of nonoptical surfaces such as ground glass, metal laps, and precision valve seats. The purpose of this paper is to put the technique on a more quantitative basis and to bring it to the attention of other optical laboratories.

Basic Arrangement

Using a simple prism such as is used in the observation of Fizeau-Abramson fringes, and an extended source such as daylight, the observer looks into the exit face of the prism and adjusts his angle of view vertically until he can see the critical angle boundary. Running alongside this boundary, and located more or less at infinity, he sees a system of colored fringes, their number and spacing depending on the roughness of the surface being tested. These boundary fringes are Haidinger-type fringes and are formed by interference between light reflected from the base of the prism and light refracted into the air gap between the base of the prism and the test surface at near grazing incidence and subsequently reflected from the test surface. (The effect is illustrated in Fig. 2, in which the thickness of the air gap has been grossly exaggerated in the interest of clarity.) At these high incident angles even quite rough surfaces reflect the light in a specular manner, and fringes can be observed

The author is with the Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Lower Hutt, New Zealand.

Received 29 June 1970.

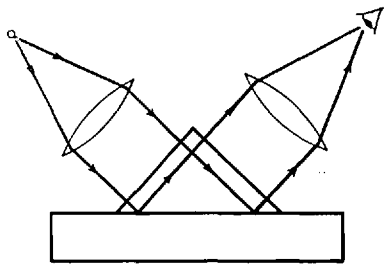


Fig. 1. Abramson's interferoscope.

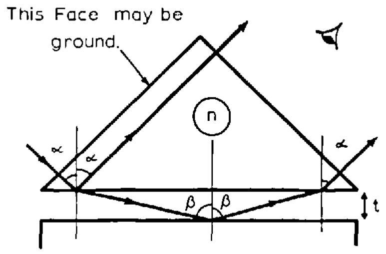


Fig. 2. The interference effect (not to scale).

with most surfaces smoother than those obtained by using 400-mesh grinding powder. If the surface is perfectly flat, the fringes are all parallel to the critical angle boundary, and any deviation from this ideal state indicates a lack of flatness in the test piece. (The base of the prism is the reference standard and is assumed to be optically flat.)

Theory of the Boundary Fringes

The relationship between the boundary fringes and Abramson's fringes is analogous to that between the Haidinger fringes used in the Michelson interferometer and the Fizeau fringes used in the Twyman-Green interferometer.

It is easily shown that the equation of the intensity minima of the boundary fringes is

$$2t \cos\beta = N\lambda, \tag{1}$$

where

t = thickness of the air gap,
β = angle of incidence in the air gap,
N = order of interference (integer),
λ = wavelength considered.

If t is nonzero, then for the zero-order fringe ($N = 0$) $\cos\beta = 0$ and hence $\beta = 90°$. This is the critical angle case, and the zero-order fringe therefore coincides with the critical angle boundary or visual horizon. To the eye this horizon appears curved but approaches a straight line as the eye is drawn back toward infinity.

If t is constant (test surface perfectly flat and parallel to the base of the prism), successive dark fringes occur at regular increments of $\cos\beta$ and are all parallel to the horizon. In white light, one of the fringes is almost free of color. This is due to the achromatizing effect described by Abramson,[3] which will be discussed in more detail later: it does not indicate a zero-order fringe. A further point to note about the boundary fringe system is that the fringe spacing increases gradually as β decreases, i.e., as the order of interference increases. In practice, however, the range of β is usually small enough for this variation to be negligible. The typical appearance of the horizon and the boundary fringes is shown in Fig. 3.

Now let us consider the effect of the test surface not being perfectly flat. Figure 4 represents a convex surface being tested by the boundary fringe technique. Let points A and B be at the same angular distance from the horizon (same β). At point A we have

$$2t_A \cos\beta = N_A\lambda, \tag{2}$$

and at point B,

$$2t_B \cos\beta = N_B\lambda, \tag{3}$$

where t_A and t_B = thickness of the air gap at points A and B; N_A and N_B = order of interference at points A and B. Subtracting these equations we get

$$2\delta t \cos\beta = N'\lambda, \tag{4}$$

where $\delta t = t_A - t_B$ = height difference between A and B; and $N' = N_A - N_B$ = change in the order of interference between A and B.

If the surface is convex, the fringes are concave toward the horizon, and vice versa.

From the above analysis, it is apparent that the boundary fringe pattern can be interpreted in the same way as a normal incidence Fizeau pattern obtained between two flats with an air wedge between them such that the thin end of the wedge is away from the observer, subject to the following provisos:

(1) The sensitivity of the fringe pattern is modified by a factor k:

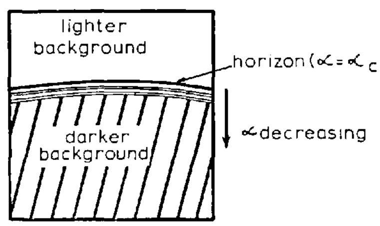


Fig. 3. Typical appearance of the horizon and the boundary fringes.

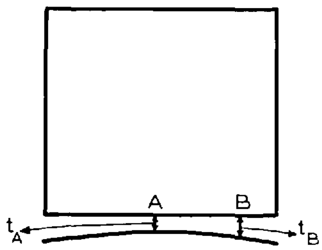


Fig. 4. Testing a convex surface (not to scale).

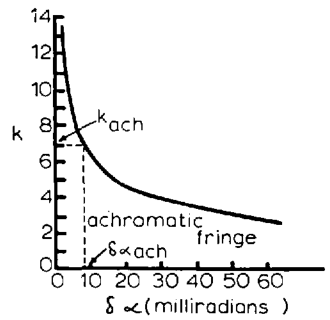


Fig. 5. Variation of sensitivity factor (k) with fringe position ($\delta\alpha$) for zinc crown glass ($n_d = 1.508$).

$$\delta t = kN'\lambda/2, \tag{5}$$

where δt = height difference between two points, N' = change in fringe order between the points, and $k = 1/\cos\beta$ = sensitivity factor;

(2) Fringe displacements should be measured, not with respect to a straight line, but with respect to a line parallel to the critical angle boundary (the horizon), which will be more or less curved according to the angular field of view.

Investigation of the Sensitivity Factor *k*

Referring to Figs. 2 and 3, consider a ray incident at an angle β on the test surface.

Let

α = angle of incidence in the glass prism,
n = refractive index of the glass,
α_c = critical angle for the glass,
$\delta\alpha$ = angular distance between the fringe under consideration and the horizon, as subtended at the eye of the observer.

Then, from above,

$$\delta t = kN'(\lambda/2),$$

where

$$\begin{aligned} k &= 1/\cos\beta \\ &= 1/(1 - \sin^2\beta)^{\frac{1}{2}} \\ &= 1/(1 - n^2\sin^2\alpha)^{\frac{1}{2}} \\ &= 1/[1 - n^2\sin^2(\alpha_c - \delta\alpha)]^{\frac{1}{2}} \\ &= 1/\{1 - n^2\sin^2[\arcsin(1/n) - \delta\alpha)]\}^{\frac{1}{2}}. \end{aligned} \tag{6}$$

For a given prism, n is fixed, and hence if $\delta\alpha$ is fixed (observations always performed at a given angular distance from the horizon), then k is a constant factor. Figure 5 shows how k varies with $\delta\alpha$ for a typical glass of refractive index 1.508 (zinc crown). $\delta\alpha$ would normally lie in the range 4–40 mrad, corresponding to a range of from 3.4 to 10.6 for k—i.e., an order of magnitude change in $\delta\alpha$ results in a change of only a factor of 3 in k.

Possibility of Achromatic Fringe with White Light

As mentioned above, one of the white-light fringes appears to be almost completely color-free. This is due to dispersion effects at the base of the prism. Since the fringe equation is

$$2t\cos\beta = N\lambda,$$

it follows that an achromatic fringe will occur when $\cos\beta \propto \lambda$ for a given α.

Taking the C and F lines as the basis of achromatism, the achromatic condition is:

$$\cos\beta = a\lambda, \qquad (a = \text{arbitrary constant}),$$

$$\text{i.e., } 1 - \sin^2\beta = a^2\lambda^2,$$

$$\text{i.e., } 1 - n_F^2\sin^2\alpha = a^2\lambda_F^2,$$

$$\textit{and } 1 - n_C^2\sin^2\alpha = a^2\lambda_C^2 \text{ (same fringe, therefore same } \alpha).$$

Hence,

$$\sin^2\alpha_{ach} = \frac{\lambda_c^2 - \lambda_F^2}{n_F^2\lambda_C^2 - n_C^2\lambda_F^2}$$

for the achromatic fringe to occur at α_{ach}.

The position of the achromatic fringe thus depends on the choice of glass used for the prism—a low dispersion glass pushes it toward the horizon, and a high dispersion glass brings it nearer to the observer.

Although the fringe pattern is only completely achromatized at one particular value of α, the achromatizing mechanism [the fact that as λ increases, n decreases, and hence β decreases (for the same α) and $\cos\beta$ increases], results in the remaining fringes being visible to a higher order than is normally the case with white-light fringes.

The existence of the achromatic fringe simplifies the use of the technique, since it always appears at a fixed angular distance from the horizon (for a given glass type) and hence provides a calibration for the measurement of $\delta\alpha$. This will be dealt with more fully below, when practical considerations are discussed.

Choice of Glass Type

The glass chosen for the prism has two effects:

(1) *Refractive Index Effect*: This affects the sensitivity, since

$$k = 1/\{1 - n^2\sin^2[\arcsin(1/n) - \delta\alpha]\}^{\frac{1}{2}}.$$

The effect is due to two different phenomena—the variation of critical angle with refractive index and the difference in the variation of β with α (Snell's law).

(2) *Dispersion Effect*: This affects the position of the achromatic fringe, which satisfies the condition:

$$\sin^2\alpha_{ach} = \frac{\lambda_C^2 - \lambda_F^2}{n_F^2\lambda_C^2 - n_C^2\lambda_F^2}.$$

The effect is due to the variation of refractive index with wavelength.

A range of Schott glasses has been investigated for suitability as the prism material, and the results are summarized in Table I. The glasses are arranged in order of increasing separation of the achromatic fringe from the horizon (increasing $\delta\alpha_{ach}$). The

Table I. Investigation of a Range of Schott Glasses for Suitability as the Prism Material

Glass type	Refractive index (n_d)	Dispersion (V_d)	Position of the achromatic fringe ($\delta\alpha_{ach}$) mrad	Sensitivity at the achromatic fringe position (k_{ach})	Sensitivity at $\delta\alpha$ = 10 mrad (k_{10})
FK5	1.487	70.4	7.93	7.56	6.74
PSK3	1.552	63.5	8.86	6.90	6.50
ZKN7	1.508	61.2	9.14	6.96	6.66
SK16	1.620	60.3	9.32	6.50	6.28
K5	1.522	59.5	9.40	6.80	6.60
LaK8	1.713	53.8	10.37	5.90	6.00
LLF6	1.532	48.8	11.48	6.14	6.56
BaF8	1.624	47.0	11.93	5.74	6.26
LaFN2	1.744	44.8	12.42	5.32	5.92
F11	1.621	35.9	15.55	5.04	6.26
SF10	1.728	28.4	19.38	4.30	5.96
SF11	1.785	25.8	21.17	4.02	5.82
SF58	1.918	21.5	24.64	3.54	5.54

following conclusions can be drawn:

(1) The sensitivity k for a given $\delta\alpha$ varies only slightly from glass to glass;

(2) The position of the achromatized fringe $\delta\alpha_{ach}$ depends on the dispersion, increasing as V_d decreases; and

(3) The sensitivity of the achromatized fringe k_{ach} depends on both the dispersion and the refractive index, and varies over the range from 3.5 to 7.6 for the range of glasses considered.

Practical Considerations

It has been suggested that it might be convenient to carry out measurements on the fringe pattern at an apparent distance of about 6 mm from the horizon when viewed from a distance of 500 mm. This corresponds to $\delta\alpha = 12$ mrad. For the achromatic fringe to occur at this position, it can be seen from Table I that a suitable glass would be BaF8 ($\delta\alpha_{ach} = 11.9$ mrad). The sensitivity of the achromatic fringe k_{ach} would then be 5.7. However, from the point of view of scratch resistance, a better choice might be ZKN7 (a zinc crown), which would give the achromatic fringe at $\delta\alpha = 9.1$ mrad (approximately 4.5 mm at 500 mm), with a sensitivity $k_{ach} = 7$.

The ideal system would be to have a range of prisms of different glasses. A suitable selection would be:

ZKN7: $\delta\alpha_{ach} = 9.1$ mrad (4.5 mm at 500 mm), $k_{ach} = 7$;

F11: $\delta\alpha_{ach} = 15.6$ mrad (7.8 mm at 500 mm), $k_{ach} = 5$;

SF11: $\delta\alpha_{ach} = 21.2$ mrad (10.6 mm at 500 mm), $k_{ach} = 4$.

The ZKN7 prism would be used for the rougher surfaces, when only fringes very close to the boundary can be seen, and the SF11 for the smoother surfaces.

If a single prism only is to be used, ZKN7 has proved to be a satisfactory choice in practice.

One of the advantages of using the boundary fringes is the possibility of using diffuse white light; in practice, daylight has been found to be ideal. Because the light can be diffuse, it is convenient to have the entrance face of the prism ground rather than polished (see Fig. 2): this removes unwanted extraneous images from the field of view, without detracting from the quality of the fringes. In order to reduce the effects of dispersion at the entrance and exit faces of the prism, it is recommended that the base angles of the prism be made equal to or slightly less than the critical angle for the glass used; this ensures that the light rays enter and leave the prism more or less normal to these faces. Since, however, the critical angle for most glasses, and certainly for the low index ones, is not much different from 45°, it has proved satisfactory to use a standard 90°–45°–45° prism of ZKN7 glass, with its entrance face ground as mentioned above, and its hypotenuse face polished flat to about $\lambda/4$.

The fact that the achromatized (black) fringe always appears at the same angular distance from the horizon, and hence has the same sensitivity k_{ach}, for a given glass type, is of value when using the technique. If measurements are always carried out on the achromatic fringe position, the sensitivity factor will be constant. If, because of the nature of the surface, it is more convenient to use a different fringe for assessing the quality of the surface, the achromatized fringe can still be used as a calibration to determine the separation $\delta\alpha$ of the chosen fringe from the horizon. The sensitivity k of the chosen fringe can then be calculated from Eq. (6), using n_d as the refractive index of the glass.

It should be noted that the sensitivity factor k has been defined to relate the boundary fringes to ordinary Fizeau fringes observed at normal incidence, and hence gives the number of *half*-wavelengths per fringe.

Use of the Technique

Qualitative Assessments

The curvature of the fringe is compared with that of the horizon and interpreted in the same way as the Fizeau fringes observed between two optical flats with an air wedge between them when the thin end of the wedge is away from the observer. In other words, if the fringes are concave toward the horizon the surface is convex, and vice versa. By observing from a sufficiently great distance, the horizon tends to become straight, and it is then a simple matter of assessing the absolute curvature, convex or concave side toward the observer, of the fringes. An additional advantage of increasing the viewing distance arises from the fact that the fringes occur at constant angular, rather than linear, distances from the horizon. This means that as the eye is drawn back, the vertical scale of the fringe pattern increases with respect to the horizontal scale. This results in an enhancement of any change of slope of a fringe, enabling deviations from flatness to be picked out much more easily; it is particularly valuable in the detection of turned down edges.

Order of Magnitude Measurements

The fringe pattern is viewed from a convenient distance, say 500 mm, and observations are made at a suitable distance from the horizon, preferably that corresponding to the achromatic fringe. The change in fringe order between the two points being compared is estimated, and the difference in height between the two points is then calculated from Eq. (5):

$$\delta t = kN'(\lambda/2).$$

N' is the change in fringe order (not necessarily an integer), and k is the sensitivity factor for the particular fringe position used and for the glass of the prism. For measurements made at the achromatic fringe position using a ZKN7 prism, the sensitivity factor $k_{ach} = 7$.

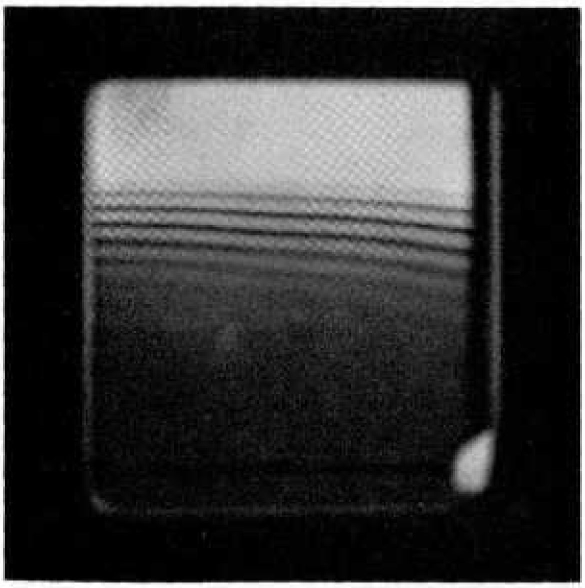


Fig. 6. Fringes obtained from a flat brass lapping tool.

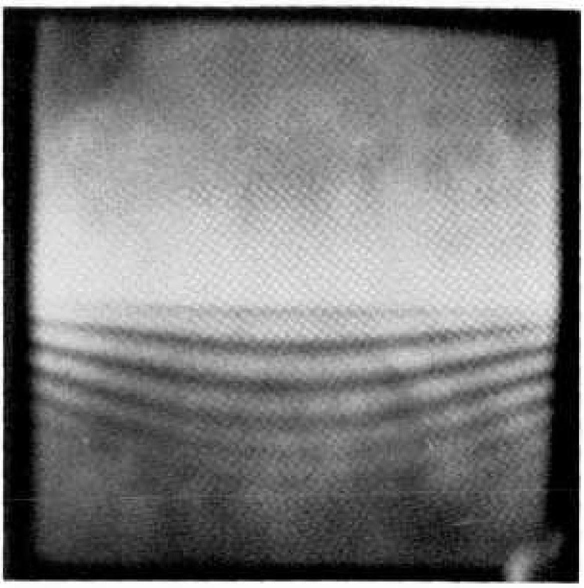


Fig. 7. Fringes obtained from a ground glass plate.

More Exact Measurements

In general, the order of magnitude assessment described above is satisfactory for most purposes, such as the testing of smoothed optical flats before polishing. However, if higher precision is required, this could be achieved by using a more sophisticated arrangement. Some method would be required to locate the observer's eye at a fixed viewing distance from the prism, and a beam splitter and a collimator with a graticule at its focus could be used to superimpose reference lines at given intervals of $\delta\alpha$ on the fringes at infinity. In this way, the value of $\delta\alpha$ for the fringe position used could be known with reasonable precision, and the restriction of using the achromatized fringe would be removed. A filter could then be used to select a given wavelength from the fringe pattern in order to increase the accuracy still more—the graticule, of course, would have to be designed for the particular wavelength chosen, and the value of k would also depend on the wavelength, as well as on the value of $\delta\alpha$. If necessary, a telescope could be used to observe the fringe pattern.

Examples of Fringe Patterns

The photographs of Figs. 6 and 7 illustrate typical applications of the boundary fringe technique. In Fig. 6 a flat brass lapping tool is being tested by means of a prism of ZKN7 glass. Irregularities of the order of one tenth of a fringe (equivalent to about 0.7 wavelength) can be seen. In Fig. 7 the same prism is being used to test a ground glass plate, and the fringe pattern reveals a convexity on the surface of about 3.5 wavelengths. The lack of sharpness in the photographs is due to the fact that the camera was focused on the fringes, approximately at infinity, and not on the test surface or the prism.

Conclusions

The method described in this paper provides a simple and quick way of testing the flatness of a nonoptical surface. The interpretation of the fringes is exactly the same as that of Fizeau fringes observed in the normal contact method of testing optical flats when a wedge is introduced between the flats (thin end away from the observer), except that:

(1) The fringe profile indicating perfect flatness is not straight and horizontal but parallel to the critical angle boundary (the horizon); and

(2) The difference in height between two points is given, not by $\delta t = N(\lambda/2)$, but by $\delta t = kN(\lambda/2)$, where k varies with fringe position and the type of glass used for the prism. Using a ZKN7 prism (useful from the point of view of scratch resistance), the value of k for the achromatic fringe position is 7.

When compared with the similar technique described by Abramson[3] the main advantage of the boundary fringe method is its simplicity—it is capable of producing satisfactory fringes in diffuse daylight: a smoothed glass flat or a metal surface such as a flat lap can be tested in a matter of seconds, without the need for any auxiliary optics or light source. Its main disadvantages lie in the variation of the sensitivity factor (k) with fringe position ($\delta\alpha$) and its inability to produce absolute contour fringes—the fringes are Haidinger-type rather than Fizeau-type and must always be interpreted in a similar way to wedge fringes. It is suggested that when a quick estimate of the flatness of a nonoptical surface is required, as is often the case in an optical workshop, the boundary fringe technique is ideal, while the Abramson arrangement of collimated light and viewing system should be used when higher precision is desired.

The usefulness of both techniques is limited by the size of the prism used, but in practice a prism based on a 60-mm square entrance face has proved perfectly satisfactory for most applications, especially since usually only an order-of-magnitude assessment is required.

The author would like to thank his colleagues at the Physics and Engineering Laboratory for their helpful advice and discussions, Professor Barber of the Victoria University of Wellington for several useful comments, and M. C. Probine, Director of the Physics and Engineering Laboratory, for permission to publish this paper.

References

1. W. Herschel, Phil. Trans. Roy. Soc. London **99**, 274 (1809).
2. P. H. Langenbeck, Appl. Opt. **6**, 1707 (1967).
3. N. Abramson, Optik **30**, 56 (1969).

SELF-COMPENSATION OF ERRORS IN A LATERAL SHEARING INTERFEROMETER

J. D. BRIERS
Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Lower Hutt, New Zealand

Received 22 July 1971

It is shown that small manufacturing errors in a lateral shearing interferometer, leading to an undesirable component in the tilt of the sheared wavefront, can be compensated for merely by using the interferometer out of focus. This self-compensation property suggests a possible relaxation of tolerances in the design of such devices.

The superposition of a wavefront upon a sheared image of itself such that interference effects occur in the area of overlap is a well known and useful method of testing optics. The main advantage is that a standard reference surface, such as is necessary in the Twyman-Green interferometer, is not needed. Lateral shearing is usually defined as the rotation of the wavefront about its focus, so that a perfect spherical wavefront is identical to its sheared image and gives a null result. In the Bates type of wavefront shearing interferometer [1] the converging wavefront is split into two components which travel along equivalent optical paths as in a Mach-Zehnder interferometer (see fig.1). Lateral shearing is achieved by focusing the beams on to the final semi-reflecting plate (labelled 3 in fig.1) and rotating this plate about an axis in its own plane so that one of the component beams is deviated slightly. Some later modified versions of the Bates interferometer [2, 3] introduce shear by rotating one of the two fully reflecting plates (2 or 2' in fig. 1). In all cases, however, the beam to be sheared is focused on to the surface which is rotated to produce the shear. This surface will be referred to in this paper as the 'shear plane'.

As in other interferometers, such as the Fizeau and Twyman-Green, it is often desirable to introduce tilt between the interfering wavefronts in order to provide a reference grid of straight fringes against which to measure fringe displacements. In the lateral shearing interferometer this is achieved by displacing the foci of the two component beams with respect to each other. Care must be taken to ensure that after recombination the two beams are still superimposed and are both travelling in the same direction. This can usually be achieved by rotating two of the four reflecting surfaces of the interferometer together, either about a single axis [1] or through the same angle [3]. The concepts of shear (rotation of a wavefront about its focus) and tilt (lateral displacement of the focus, equivalent to the rotation of a wavefront about a diameter) are illustrated and differentiated in fig.2. The degrees of freedom available in the design of the interferometer are such that it is usually possible to arrange that the tilt is about any desired diameter of the wavefront. There is, however, an important advantage in arranging for the wavefront to be tilted about the shear axis, as shown in fig. 3. Interferograms produced with this arrangement are easier to interpret than those in which the tilt is about some other diameter. In fact it was this feature (tilt about the

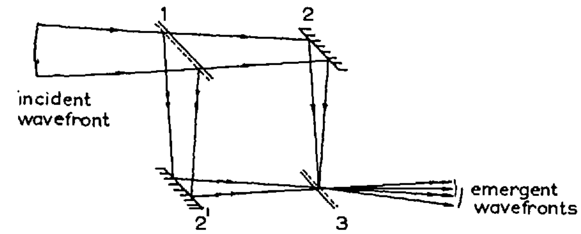


Fig.1. The Bates wavefront shearing interferometer (diagrammatic).

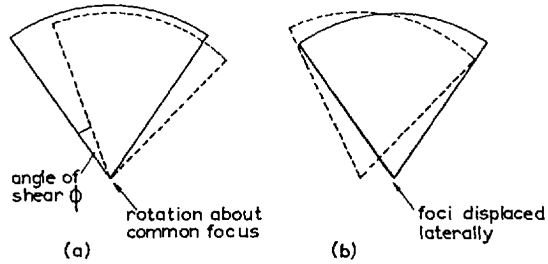


Fig.2. Meaning of (a) lateral shear, and (b) tilt.

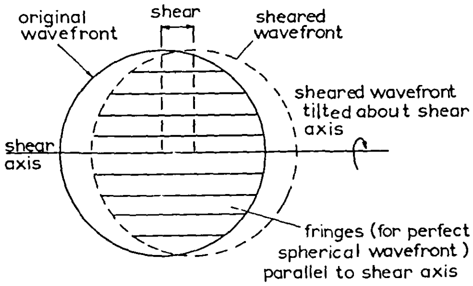


Fig.3. Recommended axis of tilt, identical to shear axis.

shear axis) that made Bates' interferometer [1] so important, since previous shearing interferometers had had an intrinsic tilt about an axis perpendicular to the shear axis.

It is usually assumed that high levels of accuracy are required in the various angles of the components of a shearing interferometer. For example, Brown [3] has stated that the angles must be accurate to about 1 second of arc. On the other hand, a solid glass version of the Bates interferometer designed by the author [4] and illustrated in fig. 4 was assembled without the aid of any special jigs or techniques, simply by placing the four component prisms on a flat surface in the correct orientation, introducing cement on to their adjacent faces, and moving the prisms together. The manufacturing tolerances on the angles of the prisms had been specified as 1 minute of arc. Upon completion, the prism shearing interferometer was found to work successfully, though admittedly in white light the zero-order fringe was not quite central. Since, however, it was the intention to use the device primarily with laser light, this was no inconvenience, and interferometers of this design have been used successfully for the testing of optical components and systems. (As an example, an interferogram of a 250 mm diameter $f/6$ paraboloidal mirror is illustrated in fig.5.)

Since the author's prism shearing interferometer is designed to give tilt about the shear axis, it follows that the fringes for a perfect spherical

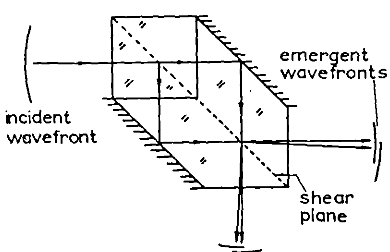


Fig.4. The prism shearing interferometer.

wavefront should be parallel to the shear axis, and of maximum spacing, when the beam under test is focused on the shear plane (the second semi-reflecting interace of fig. 4). It has been the practice to move the interferometer along the optic axis until maximum fringe spacing is achieved, and to take the resulting interferogram as corresponding to the 'in focus' situation. Recently, however, it was observed while testing a high-quality wavefront of high numerical aperture that the fringes were parallel to the shear axis and of maximum spacing when the beam was focused very close to the exit face of the interferometer rather than on the shear plane. When the beam was actually focused on the shear plane, the fringes were inclined, suggesting the presence of a tilt component about some axis other than the shear axis. Further experiments showed that this held for all beam apertures, and an investigation of the phenomenon was obviously called for.

The effect of defocus on the shearing action of the interferometer is illustrated in fig. 6. S-S is the shear plane, inclined at an angle $\phi/2$ to the 45° direction that would ensure exact superposition (without shear) of the two converging component wavefronts W_1 and W_2. W_1 and W_2 are assumed to be perfect spherical wavefronts, and are focused at a distance d beyond the shear plane. The inclination of $\phi/2$ gives an angular shear ϕ to the reflected beam. In practice the value of $\phi/2$ is very small (about 15' in the current version of the prism shearing interferometer), but is exaggerated in fig.6 in the interests of clarity. It can be seen from that diagram that the effect of the inclined shear plane on the defocused beams is to produce an angular shear ϕ (just as in the in-focus case), together with a relative displacement of the foci of the two beams. Inspection of fig. 6, in fact, shows that the focus of the deviated beam is displaced laterally through

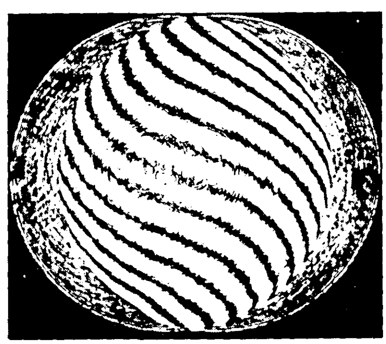


Fig.5. Typical interferogram obtained with the interferometer of fig.4.

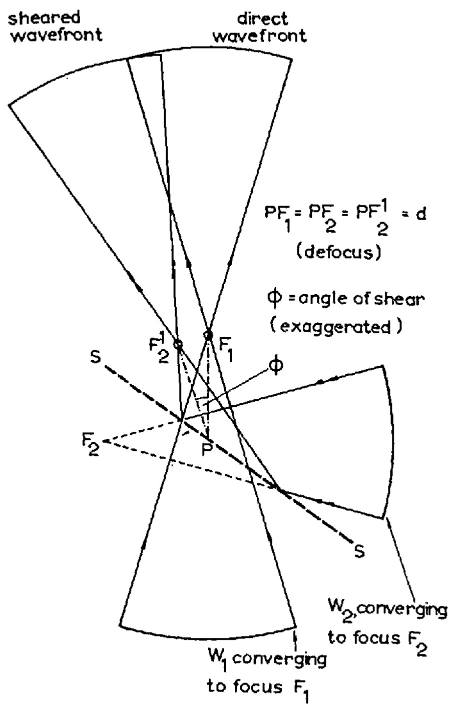


Fig.6. Effect of shear when the interferometer is used out of focus.

a distance $d \sin \emptyset$ and backwards (towards the shear plane) through a distance $d(1 - \cos\emptyset)$. The lateral displacement of the focus introduces tilt about an axis perpendicular to the shear axis, while the backwards displacement leads to a slight defocus of one component wavefront with respect to the other. In general, however, this backwards displacement will be extremely small (about $0.6\,\lambda$ for $d = 1$ cm and $\emptyset = 30'$), and will have a negligible effect on the fringe pattern. Hence the net effect of the defocus is to introduce tilt about an axis perpendicular to the shear axis.

The explanation of the observed phenomenon is now apparent. Due to slight errors arising during the manufacture or assembly of the interferometer, a component of tilt about an axis perpendicular to the shear axis was introduced, in addition to the intended tilt about the shear axis. This results in the fringes being inclined to the shear axis and having a closer spacing. The act of moving the interferometer along the axis of the beam introduces an additional tilt about the axis perpendicular to the shear axis, a tilt that is proportional to the amount of defocus (d) and whose sign depends on the sign of d. Hence there will be some point at which the undesirable component of tilt caused by manufacturing errors is exactly compensated for by the tilt introduced by defocusing, and this point can be found merely by moving the interferometer along the beam until the fringes (from a spherical wavefront) are parallel to the shear axis and have maximum spacing.

It is therefore suggested that manufacturing tolerances on the angles of lateral shearing interferometers might be relaxed somewhat, at least to the level of 1 minute of arc mentioned above. The simple method of assembly described for the prism shearing interferometer also appears satisfactory. This relaxation of tolerances is possible because of the self-compensation effect described in this paper - small manufacturing errors leading to undesirable tilt of the wavefronts can be compensated for by defocusing. The amounts of shear and tilt remaining in the interferometer may now not have exactly the value intended, but the actual shear and tilt can be measured very easily from the interferograms obtained with the instrument. Manufacturing errors may also upset the path-length requirements for white light applications, but with the increased availability of inexpensive lasers this is unlikely to be important - measurements are merely carried out on the central fringe of the interferogram, not necessarily on the zero-order fringe.

The author wishes to thank Dr. M. C. Probine, Director of the Physics and Engineering Laboratory, for permission to publish this note.

REFERENCES

[1] W. J. Bates, Proc. Phys. Soc. (London) 59 (1947) 940.
[2] R. L. Drew, Proc. Phys. Soc. (London) B64 (1951) 1005.
[3] D. Brown, Proc. Phys. Soc. (London) B67 (1954) 232.
[4] J. D. Briers, Opt. Technol. 1 (1969) 196.

Reprinted from **OPTICS AND LASER TECHNOLOGY** Printed in England

Interferometric testing of optical systems and components: a review

J. D. BRIERS
Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Lower Hutt, New Zealand

The tolerances on optical systems and components are usually very tight and the testing of such products is a vital part of their manufacture. This review describes how the interference properties of light have been applied to optical testing and indicates the main fields of application of the various techniques.

The testing of optical systems and components has grown into a very large field; this article is intended as a review of one branch which covers tests based on the interference properties of light. The main advantages of such tests are their high sensitivity and the fact that they are, in general, relatively easy to interpret. The unit of measurement is usually the wavelength of the light used, or a multiple or sub-multiple. Because of the wavelength dependence of most interferometric tests, high precision usually calls for a monochromatic light source and, in some cases, such a source is necessary for the fringe pattern to be observed at all. Suitable sources are sodium lamps, filtered mercury lamps and lasers. In many forms of interferometric testing the effective size of the source must also be limited by a pinhole or slit. Further details of the general theory and practice of interferometry can be obtained from the various excellent textbooks on the subject[1-3].

Test plates

The simplest application of interferometry to optical testing lies in the use of test plates. When two nominally flat pieces of glass are placed in contact, interference fringes can be seen at their interface. The fringes are caused by interference between light reflected from the lower surface of the top glass and light reflected from the upper surface of the bottom glass. Each fringe is a locus of constant optical path difference, and hence of constant thickness of the air gap between the two surfaces. (This explanation is strictly true only if the observer is at an 'infinite' distance from the glasses). If one of the glasses (the test plate) is known to be optically flat, the shape of the fringe pattern is determined by the surface profile of the other. The interpretation of the fringe patterns is identical to that for the Fizeau interferometer except that errors are introduced by viewing the patterns from a finite distance and at an angle to the normal to the surfaces.

If white light is used with test plates, coloured fringes will be observed, and only a few orders of interference will be visible. This disadvantage is overcome by using a quasi-monochromatic source such as a sodium or mercury lamp. The light does not have to be collimated and, in fact, diffuse lighting is commonly used.

The use of test plates is widespread throughout the optical industry, and is extended to curved surfaces—matching test plates are used to test lens surfaces during manufacture. The main advantages of the technique are its simplicity and speed—the interpretation is straightforward and easy to understand, and there is no setting up required. The major disadvantage lies in the high risk of damage to the surfaces as a result of their contact. This danger becomes much more important when very high precision work is being carried out requiring many hours of figuring and repeated testing. Not only is the risk of damage increased every time the testing is carried out, but also if damage does occur it may well destroy the results of several days laborious work. Another disadvantage is that the handling of the test plate and work piece can cause warping and spurious fringe patterns because of heat from the hands of the operator. For these reasons, non-contact methods of testing have been developed, and some will be described below. The use of these more sophisticated techniques also leads to an improvement in precision compared with the test plate method. However, the test plate method is still a very valuable tool in the routine testing of optical surfaces, and further details of the method can be found in the standard glass-working textbooks[4,5].

The Fizeau interferometer

Since the main disadvantage of the test plate method is the danger of damaging the surfaces involved, the obvious answer is to separate the two surfaces. This raises two major problems, one mechanical and one optical. First, the lack of physical contact between the two surfaces means that some form of mechanical supports must be introduced to prevent relative motion. Secondly, the increased air gap introduces restrictions on both the coherence and the size of the light source used. The result is that a reasonably well engineered system is desirable for any non-contact method of interference testing. The use of a collimating lens, first suggested by Fizeau[6] when he devised the arrangement that now bears his name, has the double effect of ensuring normal incidence of light in the air gap and of giving a fully illuminated field of view when the eye is placed at its focus. Fizeau's apparatus was modified and adapted to the testing of plane surfaces by Laurent[7]. The basic arrangement is virtually unchanged in present-day Fizeau interferometers, and is illustrated in Fig. 1. Further details of practical versions of the Fizeau interferometer can be found in an NPL report by Dew[8].

The fringes observed in a Fizeau interferometer represent a 'contour map' of the thickness of the air gap between the test piece and the reference flat. If the latter can be assumed to be perfectly flat, and if the two compared surfaces are parallel, the fringes will represent a contour map of the actual surface topography of the test piece. The contour spacing is $\lambda/2$, where λ is the wavelength of the light used. An alternative approach, and one that is essential when the errors involved are less than one wavelength, is to use tilt fringes. The Fizeau interferometer is invariably fitted with a means of levelling either the test piece or the reference flat, and this adjustment can be used to introduce a wedge of known direction into the air gap. If both glass surfaces are perfectly flat, the observed fringes will be straight, parallel and evenly spaced. The fringe spacing and the sensitivity of the test can be adjusted by varying the amount of tilt. Any deviation from straightness of the fringes indicates a deviation from flatness of the surface under test. If the thin end of the wedge is arranged to be nearer the observer, then the shape of any fringe gives a direct picture of the surface profile along the line of that fringe. A displacement away from the observer ('upwards') of one fringe spacing corresponds to a hill on the surface of height $\lambda/2$.

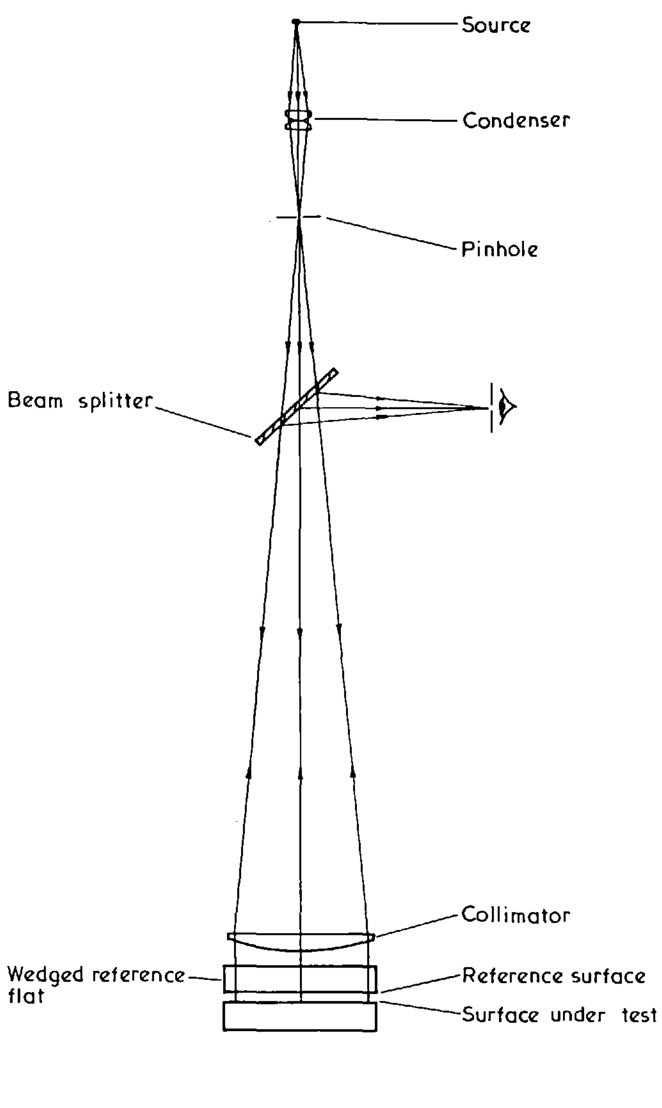


Fig. 1 Principle of the Fizeau interferometer

Several developments of the Fizeau interferometer have been announced in recent years, the most notable being its extension to spherical surfaces. This will be described later. Other modifications have included the use of liquid rather than glass reference surfaces, such as the mercury reference surface used by Bünnagel et al[9]; coated reference flats, such as those described by Clapham & Dew[10] for converting the Fizeau interferometer into a multiple beam device suitable for the flatness testing of highly reflecting samples; and the adoption by Langenbeck[11] of off-axis illumination in order to achieve fringe-sharpening.

The Fizeau interferometer is an essential tool for the testing of high quality optical flats. Its main advantages compared with the test plate method are the avoidance of contact between the reference surface and the test surface, and the increase in precision by the use of a collimating lens and well-engineered mechanics. The main limitation of the instrument is that of size versus cost. A Fizeau interferometer for testing flats up to 150 mm in diameter is fairly easy and cheap to build, but the cost and difficulty of manufacture of the reference flat and the support system increase rapidly if larger versions are considered, and these drawbacks must be weighed against the number of times the larger field of view is necessary. Most optical workshops find that a Fizeau interferometer of 150-200 mm aperture is sufficient for most of their work, and is usually the first piece of standard test equipment acquired.

The Lloyd moiré interferometer (LMI)

The extremely high cost of a very large aperture Fizeau interferometer is not usually justified by the relatively small number of times its full aperture will be used. Alternative methods of testing large optical flats have been investigated. One such method, based on the classical Lloyd's mirror experiment, has been described by Langenbeck[12,13]. The method uses near-grazing incidence and employs a moiré fringe technique to measure the changes in fringe-spacing caused by deviations in flatness of the surface under test. Because of the high angle of incidence, even relatively rough surfaces can be tested by this method. The main limitation of the Lloyd moiré interferometer is its low sensitivity compared with normal-incidence methods. Typical LMI fringes are illustrated in Fig. 2.

The spherical Fizeau interferometer

Adapting the Fizeau interferometer to test curved surfaces was an obvious but technically difficult step. The problem was how to avoid the need for a separate reference surface for each radius of curvature tested and yet maintain high quality fringes. In 1967 the development of such an instrument was announced independently by SIRA in the UK and by the Perkin-Elmer Corporation of the USA.

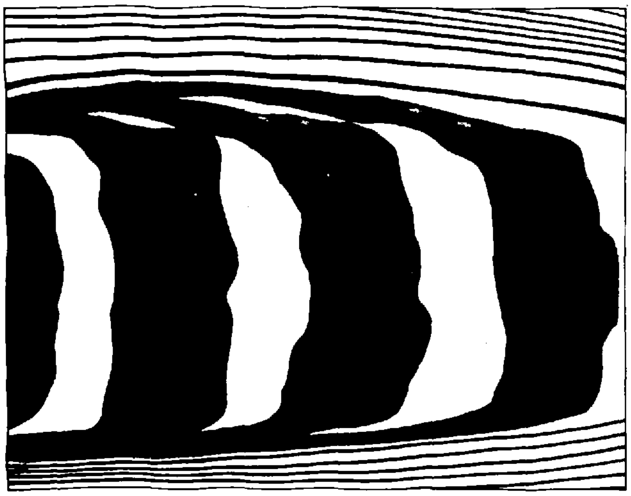


Fig. 2 A typical Lloyd moiré interferogram[12]

The SIRA version[14] uses a classical Fizeau arrangement with a conventional light source and with a gap between the curved reference surface and the test surface that can be adjusted between zero and 5 millimetres. This adjustable gap means that a comparatively small number of test plates (reference surfaces) may be used to test a large number of differently curved surfaces. The basic design of the interferometer is illustrated in Fig. 3, and further details can be found in a paper by Biddles[15]. A disadvantage of the SIRA interferometer is the need for a corrector lens when the surface under test is convex.

The Perkin-Elmer 'Spherical Wave Interferometer, (See Fig. 4) Multiple-beam' (SWIM), described by Heintze et al[16], is based on a concentric cavity formed by the reference and tested surfaces. The resulting large gap means that a laser must be used as the light source. If fully reflecting surfaces are to be tested, the small reference surface must be given a partially reflecting coating so that multiple-beam fringes are obtained. Walk-off of the beams is eliminated by incorporating a field lens at the common centre of curvature of the reference and test surfaces. The basic design of the interferometer was designed primarily for testing large spherical concave mirrors, but it can also be used with any autostigmatic system. Thus it can be used to test optical flats and converging lenses (with the aid of a standard concave mirror), infinite-conjugate lenses (with the aid of a good optical flat), and convex surfaces that have a radius of curvature smaller than that of the reference surface. Spherical Fizeau interferograms are interpreted in a way similar to ordinary Fizeau fringes (see above).

Transmission testing

The final testing of an optical system or component should always be under the conditions in which it is intended to be used. Thus, the manufacture of a plane parallel optical window can be assisted by the use of test plates or a Fizeau interferometer in order to check the flatness of the two surfaces, but, since the window will ultimately be used to transmit a wavefront rather than to reflect it, the final test should be on a wavefront that has passed through the window. This is because any lack of homogeneity in the glass will affect the quality of a transmitted wavefront but will not be shown up by the use of a test plate or a Fizeau interferometer, which merely test the quality of the surfaces. However, a Fizeau interferometer can be adapted to test windows in transmission simply by the use of two reference flats instead of one (see Fig. 5). If a slight wedge is introduced between the two high-quality flats, straight, parallel, equally-spaced fringes (tilt fringes) will be seen crossing the field of view. These fringes can be made into multiple-beam fringes of high finesse by having the first reference flat partially reflecting and the second fully reflecting. If the window to be tested is now introduced into the gap between the flats any deformation that it imparts to be transmitted plane wavefront will be observed as a corresponding distortion of the fringes, and the fringe pattern will be similar to that illustrated in Fig. 6. This distortion indicates the combined effects of surface flatness error and inhomogeneity, and Forman[17] has pointed out that both Fizeau and transmission tests are needed to separate the two effects. A modified Fizeau interferometer, working in the transmission mode, has been described by Saunders[18] and Post[19], and a reflection type by Ashton and Marchant[20]. A much larger version (750 mm aperture) has recently been reported by Roberts and Langenbeck[21]. The latter also describe a technique in which the same interferometer can be used for both the Fizeau and the transmission tests, thus avoiding the complications involved in changing the environment of the window under test.

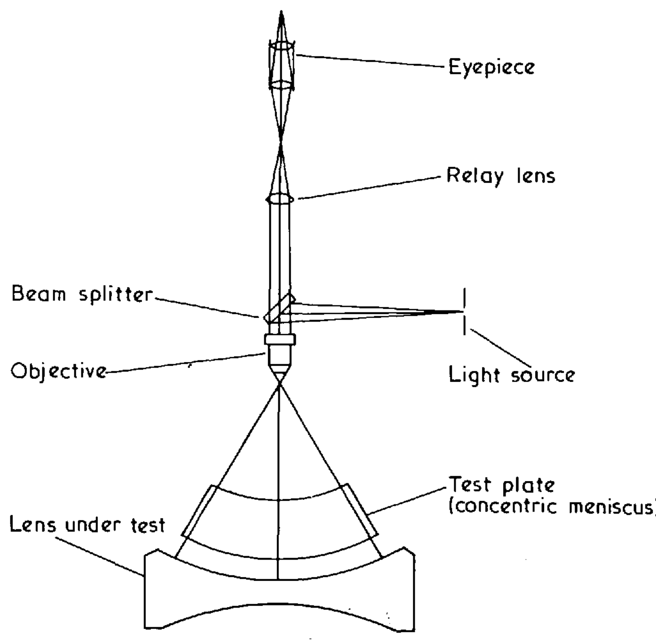


Fig. 3 The SIRA spherical wave interferometer[14]

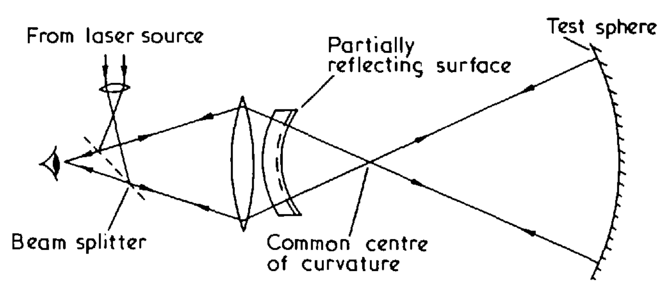


Fig. 4 The Perkin-Elmer spherical wave interferometer[31]

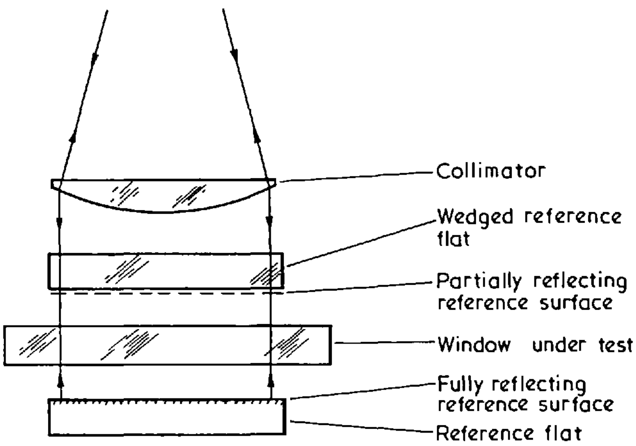


Fig. 5 Fizeau interferometer adapted to the transmission testing of optical windows

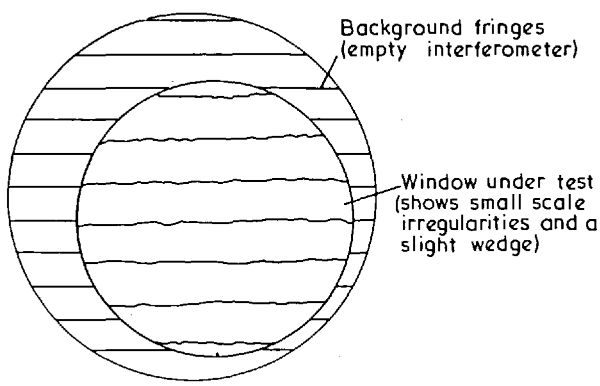


Fig. 6 Appearance of tilt fringes in a multiple-beam transmission interferometer

The transmission test can also be used to give a measure of the wedge angle in a window. If the two surfaces of the window are not perfectly parallel, the fringe spacing and/or orientation will be affected, and the wedge-angle can be calculated from this change in the fringe pattern. An alternative interferometric method of wedge-angle measurement has recently been described by Leppelmeier and Mullenhoff[22].

Very high precision measurement of the errors in glass windows can be obtained by a method suggested by Tynes and Bisbee[23]. They used a Twyman-Green interferometer (see below), the output of which constituted one arm of a two-beam optical measuring set, and claimed to be able to measure thickness variations of the order of $\lambda/1000$ and refractive index variations of the order of 10^{-7}. Various other methods of assessing the transmission properties of windows have been suggested, notably the shearing method of Hariharan and Sen[24] and the Kösters prism interferometer of Saunders[25].

Prisms can also, in principle, be tested by a modified Fizeau interferometer, but because of the difficulties introduced by large deviations of the light beam such tests are usually carried out on a Twyman-Green interferometer (see below). Windows could, in fact, also be tested on the Twyman-Green interferometer, but this instrument suffers two major disadvantages compared with the modified Fizeau—it is more expensive for a similar aperture, and it cannot be used with multiple-beam fringes.

The Twyman-Green interferometer

If the first piece of standard test equipment obtained by an optical workshop is usually a Fizeau interferometer, the second is undoubtedly a Twyman-Green interferometer. This instrument was adapted from the classical Michelson interferometer over fifty years ago[26,27], and was designed primarily as a device for testing lenses and prisms. The principle of the interferometer is illustrated in Fig. 7 and full details of its design and applications can be found in Twyman's book[4]. As with the Fizeau and transmission interferometers it can be used either in normal adjustment or with tilt.

The Twyman-Green interferometer can be used to test optical flats, windows, raw glass (using a liquid cell), convex and concave mirrors, lenses, prisms, microscope objectives, optical crystals, and many other components. Fig. 8 shows the modifications required for some of these applications. The greatest use the interferometer finds is in the field of lens testing. Basically any lens or lens system with at least one infinite conjugate can be tested very easily with the Twyman-Green interferometer—it is ideal for refracting telescope objectives. In practice, however, virtually any lens can be tested with the aid of an auxiliary lens, as shown in Fig. 8 for microscope objectives. When a lens is tested with the interferometer each type of aberration gives rise to a distinctive fringe pattern which can easily be identified by an operator. The introduction of tilt and of defocus each results in a distinctive modification of the pattern. An indication of the typical patterns which can be observed is given in Fig. 9, and further details, including photographs, can be found in a paper by Kingslake[28]. Hariharan and Sen[29] have pointed out that the even and odd aberrations can be separated by using the Twyman-Green interferometer in a double-pass arrangement, a facility that may be useful in some cases.

As with the Fizeau interferometer there are restrictions on the size and coherence of the light source used in order to obtain high-contrast fringes. The usual source in the past

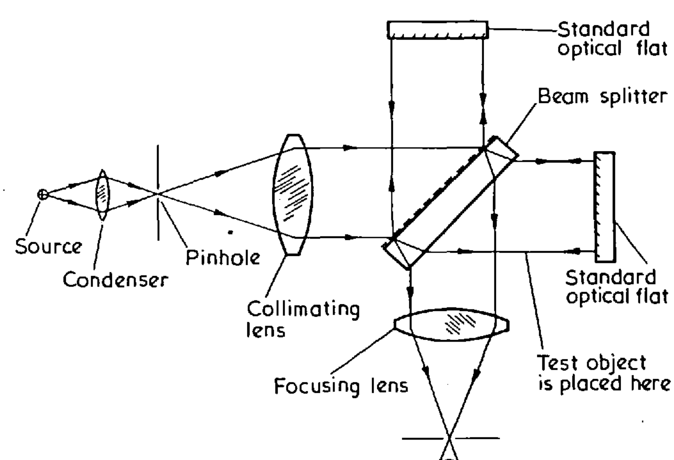


Fig. 7 Basic arrangement of the Twyman-Green interferometer

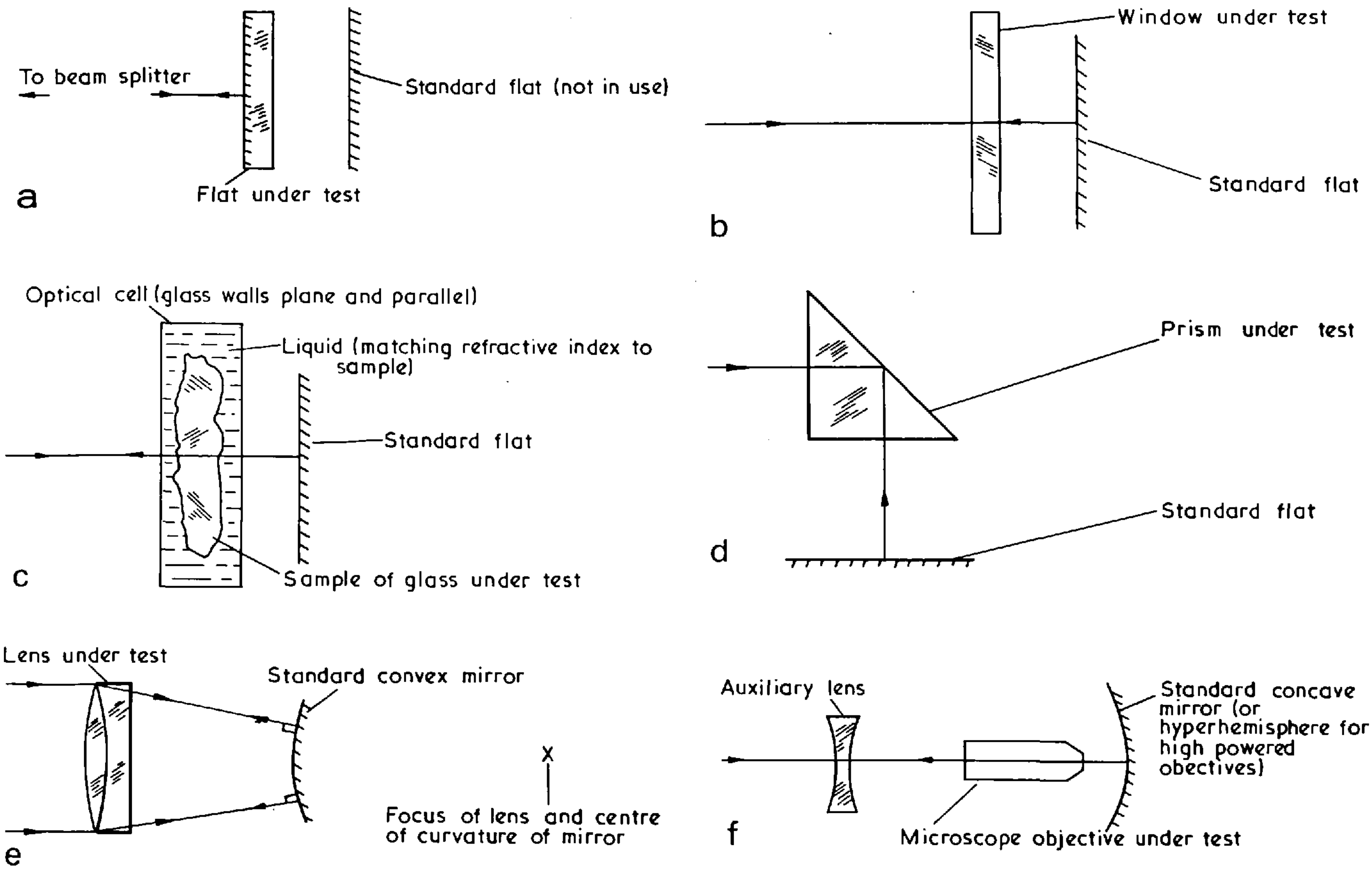


Fig. 8 Use of the Twyman-Green interferometer: (a) Testing an optical flat; (b) Testing an optical window; (c) Testing raw glass for homogeneity; (d) Testing a 90° prism; (e) Testing a telescope objective; (f) Testing a microscope objective

has been a low-pressure mercury lamp with a green filter to isolate the 546. 1 μm line. A pinhole is used to restrict the source size so that the first Haidinger fringe just fills the exit pupil of the interferometer. The advent of the laser in the early 1960's provided an intense, coherent source for the first time, and this was soon applied to interferometry. The increased coherence length of the laser enabled Houston et al[30] to design a version of the Twyman-Green interferometer with unequal path lengths in the two arms, and Munnerlyn et al[31] used the same approach when they adapted the instrument to the testing of spherical surfaces.

The main advantage of the Twyman-Green interferometer, especially as a lens-testing device, is its simplicity of use and interpretation. The adjustments are, in general, easy to make, and the residual aberrations of a lens can be identified at a glance. Detailed quantitative assessments of the aberrations can be made if desired. The major disadvantages are its limitations of aperture—anything more than 150 mm becomes prohibitively expensive, and 100 mm is a more usual size—and its inability to produce multiple-beam fringes. These factors limit its usefulness for testing optical flats and windows, which are better left to the Fizeau-type interferometers.

The Fizeau and Twyman-Green interferometers together will satisfy most of the routine testing requirements of an optical workshop. However, both are limited to the testing of components or systems with an overall aperture less than that of the interferometer (usually 150-200 mm for the Fizeau and 100 mm for the Twyman-Green). Hence they will not deal satisfactorily with the occasional larger items produced by an optical workshop, and other test methods must be used. One of these, the spherical Fizeau interferometer, which is applicable to spherical surfaces, has already been described. This instrument could be a useful addition to the test equipment of a workshop that is regularly called upon to produce large spherical mirrors or lenses (though in general only the individual surfaces of the latter could be tested and not the lens as a whole). Other methods of testing large-aperture optics are discussed below.

Wavefront shearing interferometry

Most interferometric tests require the use of a standard reference surface and, except in the cases of the spherical Fizeau interferometer[16] and the unequal path Twyman-Green interferometers[30,31] for testing spherical surfaces, the reference surface must be at least as large as the object under test. This places a restriction on the size of components that can be tested, since the cost of an interferometer rises prohibitively once a rather moderate aperture is exceeded. Because of this, much thought has gone into the devising of tests which do not require a reference surface. One result of this has been the development of the wavefront shearing interferometer (WSI), in which the wavefront under test is superimposed on a sheared image of itself. The shear may be lateral, rotary, reversal or radial.

In lateral shearing the wavefront is split into two by means of a beam-splitter, and one component is displaced laterally

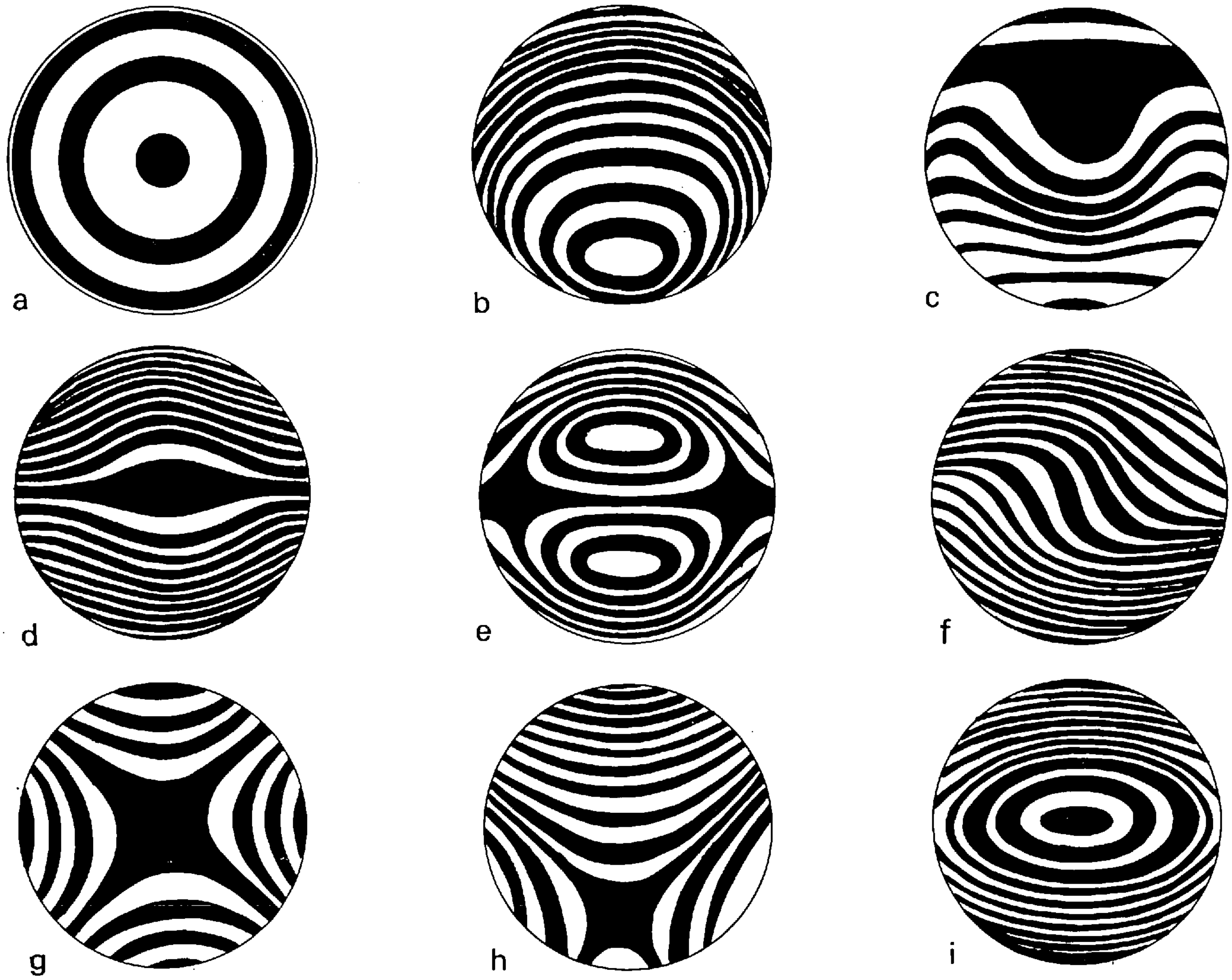


Fig. 9 Typical Twyman-Green interferograms[28]: **Spherical aberration:** (a) no tilt, paraxial focus; (b) with tilt, (c) with tilt and defocus; **Coma:** (d) no tilt, paraxial focus; (e) and (f) with tilt about orthogonal axes; **Astigmatism:** (g) no tilt, midway between tangential and sagittal foci; (h) with tilt; (i) no tilt, paraxial focus

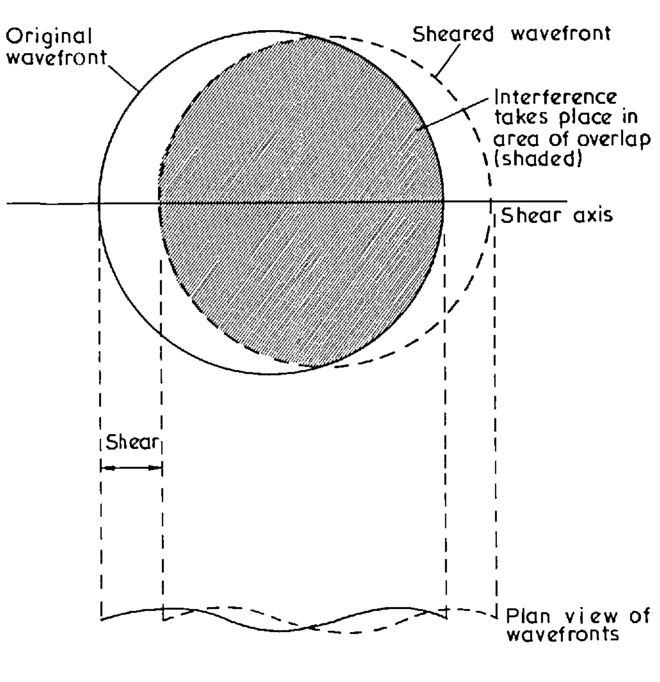


Fig. 10 Principle of lateral wavefront shearing

by an amount which is small compared with the diameter of the wavefront. When the wavefronts are recombined, interference occurs in the area of overlap (see Fig. 10). In practice, the shearing is usually accomplished by rotating one wavefront about its focus (centre of curvature), so that a perfect spherical wavefront is identical with its sheared self and no fringes are seen. Provision is usually made for one of the wavefronts to be tilted about a diameter, thus introducing tilt fringes (which would be straight, parallel and evenly-spaced for a perfect spherical wavefront). Further details of the theory of the WSI can be found in the literature[32-36]. The earliest example of wavefront shearing interferometry is the Ronchi test, developed in the 1920's, but this will be discussed separately (see below).

One of the earliest beam-splitters used for wavefront shearing was the Wollaston prism adopted by Lenouvel and Lenouvel[37]; this technique received further attention from Dyson[38,39] and Bartholomeyczyk[40]. The major disadvantage of these methods was that the tilt introduced, which was inherent in the type of beam-splitter used, was about an axis perpendicular to the shear axis (see Fig. 10). This makes the resulting interferogram very difficult to interpret quantitatively, and it was left to Bates[32] to design the first WSI to use the more convenient tilt about the shear axis.

The reader is referred to a paper by the author[36] for a more detailed discussion of this point. Bates' design was based on the Mach-Zehnder interferometer, and is illustrated in Fig. 11. The device was modified by Drew[41] in order to simplify its construction, and later by Brown[42], whose original version had fixed, rather than variable, shear and tilt. The latter device is now available commercially, the latest model having fixed shear (a choice of plug-in units) and variable tilt. The removal of adjustments for shear and tilt resulted in a greater robustness, and this advantage was later increased by designing the interferometer in the form of a solid glass prism consisting of two or more cemented components. Interferometers of this type have been described by Saunders[43-46] the author[36], and Murty[47]. The author's prism shearing interferometer (PSI) is illustrated in Fig. 12. A simplified version of the PSI has also been described by Saunders[48], and the possibilities of using simple plane-parallel glass plates as shearing interferometers have been investigated by Murty[49], Murty & Malacara[50], and Kelley & Hargreaves[51]. PSI's with variable shear have been suggested by Donath & Carlough[52] and by van Rooyen & van Houten[53]. Ashton & Marchant[54] have described a new type of lateral shearing interferometer in which the aberration is obtained directly by measuring the variation of phase difference at the centre of one of the component wavefronts as it is scanned across the other. A novel method of shear was suggested by Lohmann & Bryngdahl[55], who used two diffraction gratings as beam-splitters; variable shear was obtained by rotating the gratings in opposite directions. A 'cyclic' shearing interferometer designed by Hariharan & Sen[56] has the advantage that it avoids the necessity for matching optical paths inherent in the Bates design.

The interpretation of lateral shearing interferograms is not nearly as simple as that of, say, Twyman-Green interferograms, since the interference occurs between two imperfect wavefronts instead of between one imperfect wavefront (from the system under test) and one perfect wavefront (from the reference surface). The use of tilt about the shear axis, first introduced by Bates[32], greatly reduces the problem, but a significant amount of computation is still required. The analysis of lateral shearing interferograms has been discussed by Drew[41], Brown[57], Saunders[33,58], Malacara & Mendez[35], and others, and a method of automatically processing the data to give the wavefront

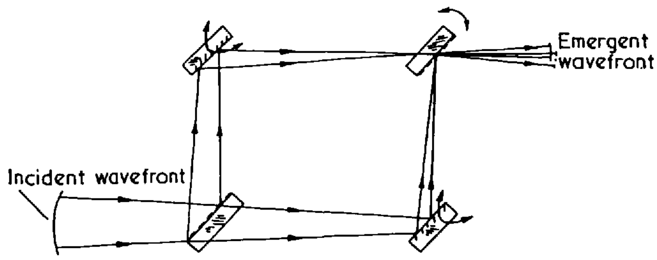


Fig. 11 Bates' shearing interferometer

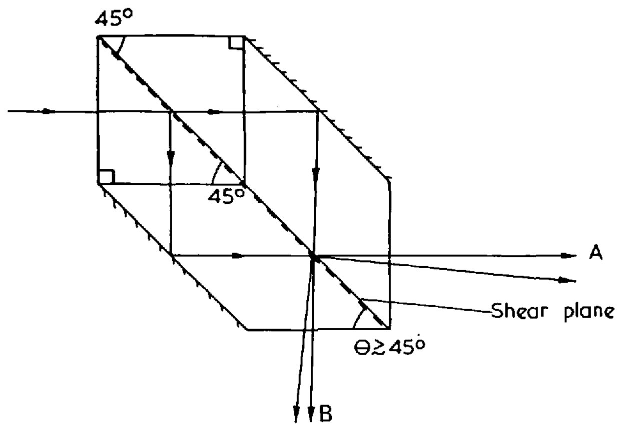


Fig. 12 Prism shearing interferometer

polynomial has been described by Dutton et al[59]. Typical examples of lateral shearing interferograms are given in Figs. 13 and 14.

Rotary, or angular, shearing is achieved by splitting the wavefront into two, rotating one component with respect to the other, and recombining the two components so that they interfere. Murty & Hagerott[60] have described a system based on the Jamin interferometer, while the Michelson and Sagnac interferometers were used as starting points by Armitage & Lohmann[61]. Both these papers point out the advantage of rotary shearing in separating the effects of different aberrations, and typical fringe patterns for coma and astigmatism are illustrated in Fig. 15. (Aberrations with rotational symmetry, such as spherical aberration, are not detected).

In the wavefront reversal interferometer (WRI) one half of the wavefront is folded back on to the other half such that

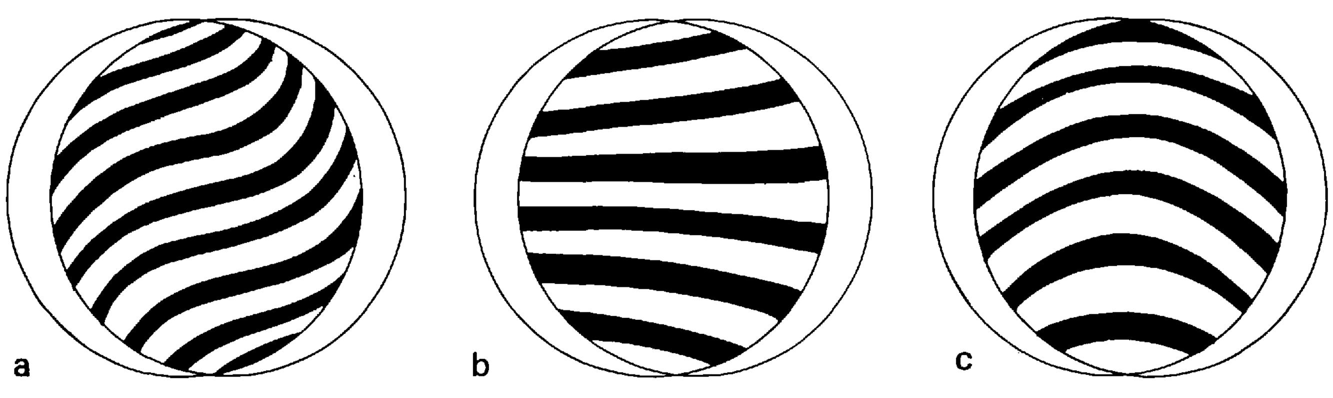


Fig. 13 Typical lateral shearing interferograms with tilt about the shear axis: (a) Spherical aberration; (b) Coma perpendicular to shear axis[41]. (c) Coma parallel to shear axis[41]. (Note: Astigmatism is detected by a change in fringe spacing and/or tilt when the interferometer is rotated through 90°)

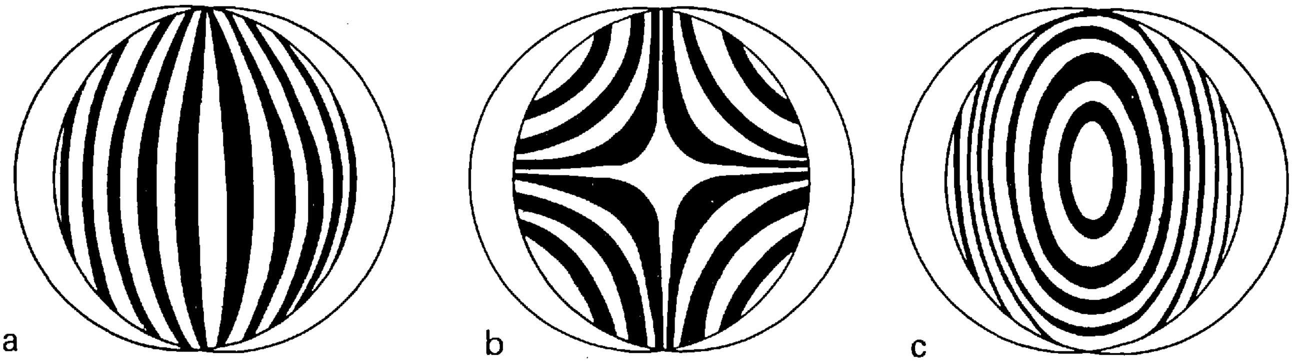


Fig. 14 Typical lateral shearing interferograms with tilt about an axis perpendicular to the shear axis: (a) Spherical aberration; (b) Coma perpendicular to shear axis[49]. (c) Coma parallel to shear axis[49]. (Note: Astigmatism is detected by a rotation of the fringes when the interferometer is moved along the optic axis)

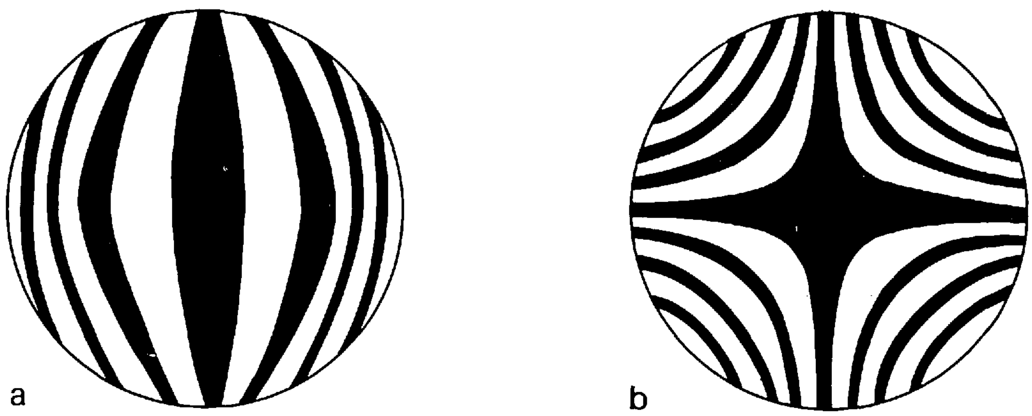


Fig. 15 Typical rotary shearing interferogram[60]: (a) Coma, rotational shear = 180°; (b) Astigmatism, rotational shear = 90°. (Note: Rotationally symmetrical aberrations such as spherical aberration cannot be detected by rotary shearing)

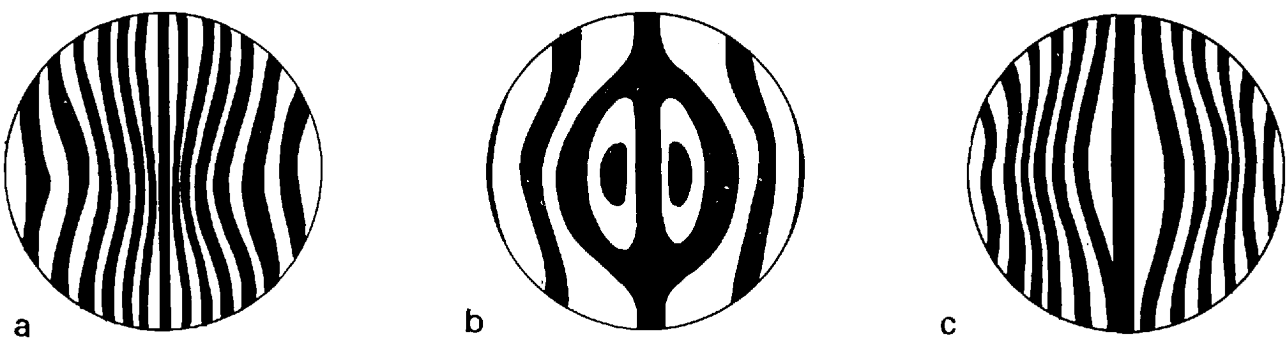


Fig. 16 Typical wavefront reversal interferograms for coma[63]. (Tilt increasing from left to right)

they interfere. The superposition may be exact, or there may be some residual lateral shear. Methods using the Kösters prism have been described by Gates[62,63] and Saunders[64], while Sen and Puntambekar[65] used a wavefront reversing system based on the Jamin interferometer and a wavefront inverting system based on the Fizeau interferometer. Rotationally symmetrical wavefronts give the same fringe patterns as with ordinary lateral shearing interferometers with the same amount of shear (see, for example, Fig. 13(a)). Off-axis aberrations, on the other hand, have a different effect, and even give a residual fringe pattern at zero shear. This can be very useful for isolating the off-axis effects. Examples of interferograms obtained with a comatic wavefront are given in Fig. 16.

A different type of wavefront reversal, in which the centre of the wavefront interferes with the periphery and vice versa, can be achieved by using an axicon lens or a circular grating to turn one of the components of the wavefront 'inside out'. This technique was used by Bryngdahl[66] to investigate the radial symmetry of rotationally symmetrical wavefronts.

Radial shearing occurs when one component of the split wavefront is expanded before superposition so that interference takes place between the whole wavefront and its central portion (see Fig. 17). A radial shear interferometer based on the Jamin interferometer was demonstrated by Brown[67] and details of this device, and of an 'exploded shear' system based on the Kösters prism, were given in a later paper by the same author[68]. Hariharan and Sen[69] designed a cyclic type of radial shear interferometer, while Murty's version[70] was based on a pentaprism with a hemispherical depression in one half. Steel[71] used a birefringent lens in his radial shear interferometer for testing microscope objectives. The laser has recently been applied to radial shear interferometry by Steel[72] and Som[73]. The main advantage of radial shearing is that the fringe patterns are approximately the same as Twyman-Green interferograms (see Fig. 9), the approximation improving with increasing shear.

The radial shear devices mentioned in the last paragraph all employ a magnification difference to produce the shear. Hence the radial shear is not constant, but varies from zero at the centre of the aperture to a maximum at the edge. Bryngdahl[74] has recently used axicon lenses and circular gratings to introduce a constant radial shear, thus obtaining (for small shears) the radial derivative of the wavefront under test.

To sum up, the main advantage of the WSI over the more 'classical' interferometers such as the Twyman-Green is the elimination of the need for a reference surface. This means that much larger components can be tested, and the methods are invaluable for inspecting such things as astronomical telescope mirrors. Another advantage is that, being a 'common path' device (both interfering beams traverse essentially the same path), the shearing interferometer is less susceptible to vibration than its classical counterpart. These advantages are shared with the other common path techniques, such as the Ronchi test and scatter fringe testing which will be described below. The main disadvantage of wavefront shearing interferometry is the notorious difficulty of interpreting the fringe patterns quantitatively. This problem is lessened somewhat if the tilt is arranged to be about the shear axis, and almost completely removed in radial shearing if large shears are used. A novel solution to the interpretation problem has recently been suggested by Langenbeck[75]. By introducing a very small pinhole stop into one of the two spatially separated point-source images produced by a shearing interferometer, he converts the device into an absolute interferometer (with a perfect reference beam), while preserving the advantages of wavefront shearing.

The Ronchi test

The Ronchi grating is nowadays regarded as a lateral wavefront shearing interferometer. However, it was originally developed (in the 1920's) as a geometrical test, and the geometrical interpretation is valid unless very fine gratings are used. For this reason, and because the test has become so widely used, the Ronchi test is covered separately. A comprehensive review of the history of the test has been given by the originator himself[76,77].

The basic form of the Ronchi test is illustrated in Fig. 18. Light from a slit source is converged by the component under test and a coarse grating (up to 20 lines per mm) is placed near the focus. The observer's eye can be placed behind the grating, as indicated in the diagram, or an image of the aperture of the component under test can be observed on a suitably placed screen. In either case the component is seen with a pattern of bars, or fringes, super-

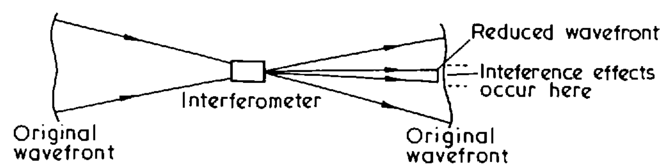


Fig. 17 Principle of radial shearing interferometry

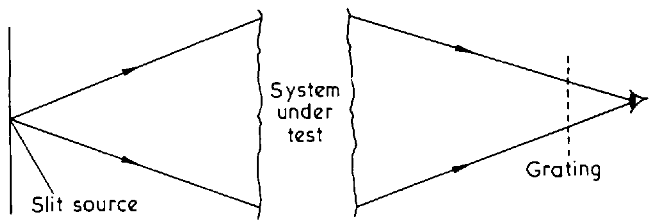


Fig. 18 Basic arrangement for the Ronchi test

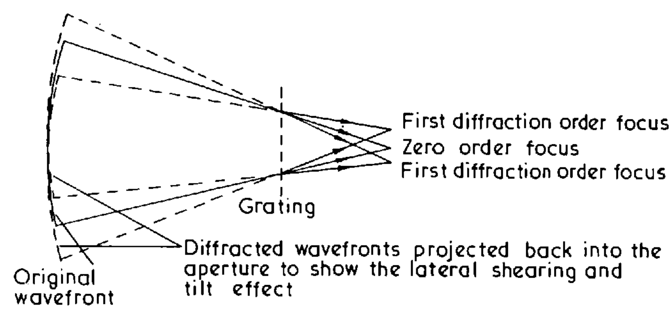


Fig. 19 The Ronchi grating as a lateral shearing interferometer

imposed on it. Geometrically, the lines of the grating act as a multiple Foucault knife-edge[78-81], and the lack of straightness is caused by the fact that in an aberrated wavefront different sets of rays have different foci, thus leading to distorted shadows. These distortions become greater the nearer the grating is to the main focus. This geometrical approach, which is valid for coarse gratings, was the one originally used by Ronchi[82], though he later adopted the interferometric approach[83]. This regards the Ronchi grating as a shearing interferometer, the diffracted orders overlapping and interfering with each other (see Fig. 19). The amount of shear is fixed by the line spacing of the grating, and tilt (and hence sensitivity) can be varied by moving the grating along the beam. An interesting point is that the fringe spacing is independent of wavelength, and hence white light can be used. The interference approach, of course, is valid for all grating spacings, but becomes involved at large spacings when several diffracted orders are all superimposed. The Ronchi grating falls into the class of lateral shearing interferometers in which the tilt axis is perpendicular to the shear axis, and hence the fringe patterns are difficult to interpret quantitatively (see above). Some workers[84-86] have, however, gone to the trouble of computing the fringe patterns that should be observed for certain aspheric mirrors, and De Vany[87] has pointed out the similarity between 'Ronchigrams' and Twyman-Green interferograms with tilt. The Ronchi test has also been applied to axial chromatic aberrations by Malacara and Cornejo[88]. Despite the difficulties of detailed interpretation, the simplicity of the test itself has been largely responsible for its popularity, and experienced operators can make very good use of the qualitative information it yields, especially if used in conjunction with the Foucault knife-edge test. Ronchigrams are identical with other lateral shearing interferograms of the same type (see Fig. 14).

Scatter fringe interferometry

Another type of common path interferometer, and one which theoretically offers significant advantages, was pioneered by Burch[89]. A 'scatter plate', or diffusing screen, is placed in a narrow, collimated beam as shown in Fig. 20. This plate allows a certain percentage of the light to pass through it unchanged, while the rest is scattered as shown in the diagram. The component under test is located so that the direct beam strikes it at its centre and the scattered beam over its full aperture. After passing through (or being reflected at) the test component, the beams are directed towards another scatter plate, identical to the first and similarly positioned and oriented. Again some of the light is transmitted directly and some scattered, and interference will occur between the beam that was transmitted by the first plate and scattered by the second and the beam that was scattered by the first and transmitted by the second (see Fig. 20). If the component under test is perfect, giving a wavefront with zero aberration, no interference fringes will be seen. If the wavefront is not perfectly spherical, however, the interference fringes seen will correspond to those of the Twyman-Green interferometer (see Fig. 9), since the wavefront from the whole component is being compared with a virtually perfect wavefront from the central zone of the component. Tilt fringes can be introduced by displacing one of the scatter plates in its own plane, and moving one of the plates along the axis introduces defocus fringes as in the Twyman-Green interferometer. The main disadvantage of the technique is the difficulty of making exactly identical scatter plates but this problem has now been largely overcome and the method applied successfully [90,91]. The similarity between scatter fringe testing and radial shearing interferometry is obvious, and the former can be regarded as a special case of the latter.

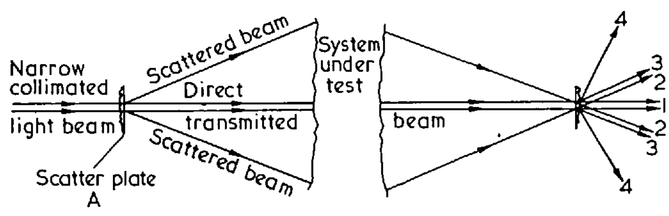


Fig. 20 Principle of scatter fringe interferometry. Beam 1 is directly transmitted by both scatter plates; beam 2 is transmitted by A and scattered by B; beam 3 is scattered by A then transmitted by B; beam 4 is scattered by both scatter plates. Beams 2 and 3 are exactly superimposed and of similar intensity, and interference occurs between them.

Interferometric testing of nonoptical surfaces

All the tests described so far, with the exception of the Lloyd moiré interferometer, require the surface(s) of the component or system under test to be optically polished, since the light must be either transmitted or specularly reflected from the test object. Compared with the grinding process, the removal of glass from a surface by polishing is a very slow business, and an indication of how close a ground glass surface is to the required figure would be very useful as a guide to when it is worthwhile to start polishing.

One obvious method of rendering a ground glass surface specularly reflecting is to coat it with a suitable substance. Some years ago, Waland[92] described the application of paraffin to a Schmidt plate at the grinding stage in order to test it with the Ronchi test, and more recently Moreau & Hopkins[93] used wax to test ground flats in a Fizeau interferometer. The main disadvantage of the technique is the care that must be taken to ensure an even coating of the substance. The choice of material is also important—Moreau & Hopkins found that the most successful one they tried was transparent shoe polish!

The use of longer wavelengths also enables rougher surfaces to be tested interferometrically, and Munnerlyn et al[94] have described a modified Twyman-Green interferometer using infra-red light from a carbon dioxide laser. Transducers of some kind must now be used to render visible the fringe patterns and the equipment required for this, as well as the laser, makes the method difficult and expensive.

Rowe & Welford[95] suggested a technique of projected fringes by illuminating the surface under test obliquely with two coherent light beams inclined to each other. The fringe spacing can be varied continuously down to 0.7λ. This could be a very powerful tool in surface topography studies since there is no restriction on the type of surface to which it can be applied. The fringe patterns are interpreted in the same way as Fizeau fringes except for a sensitivity factor which can be adjusted at will. Further details of the technique have been given by Welford[96].

Another method of performing interferometric tests on nonoptical surfaces, and particularly flat surfaces, is to use grazing incidence. At very high angles of incidence, even quite rough surfaces became specularly reflecting, and this enables techniques like the Lloyd moiré interferometer[12] to operate on such surfaces. The use of a prism to introduce light to a flat surface at grazing incidence has been described by Abramson[97,98] who used Fizeau-type fringes of reduced sensitivity, and by the author[99] who used Haidinger-type fringes. Both techniques can be used on matt surfaces such as ground glass or metal. Abramson's method is more easily interpreted quantitatively, but requires an optical system to produce the necessary collimated light and to redirect it to the eye of the observer; the technique described by the author uses only the prism, and is very quick and easy to use if only an estimate of the error is required. Both methods are limited to testing areas no larger than the base of prism used.

Holography

The process of holography was discovered by Gabor[100-102] and was developed rapidly in the 1960's by Leith and Upatnieks[103], and others, following the introduction of the laser. A full account of the state of the art of holography can be found in the standard textbooks on the subject[104-106] but the basic process is indicated in Fig. 21.

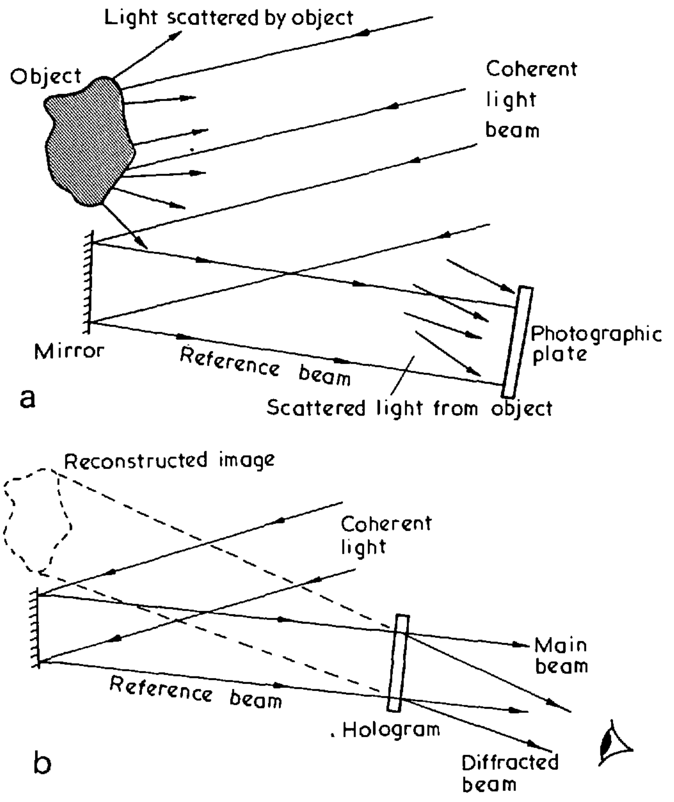


Fig. 21 Principle of holography: (a) Recording; (b) Reconstruction

The first application of holography to the testing of optical components was suggested by Hildebrand et al[107]. A hologram was made of the wavefront from the component to be tested, and this hologram was used in place of the component in Foucault knife-edge tests and in a common-path interferometer arrangement devised by Murty[108]. A more direct application of holographic techniques, however, became available with the introduction of holographic interferometry by Haines and Hildebrand[109]. In one method, the 'real-time' approach, the reconstructed image from a hologram of an object is superimposed on a second object (or the same object after changing its environment) with which it is to be compared. The technique is illustrated in Fig. 22. Any slight differences between the two objects are revealed by interference fringes appearing superimposed on the object. In the other main method, the double-exposure technique, holograms of the two objects are superimposed on the same photographic plate: again interference fringes betray the presence of any slight discrepancies between the two objects. Either technique can be used on opaque, three-dimensional objects, and hence can be used to compare an optical surface, even in the ground state, with a reference hologram. Further, since holograms can be generated by computer, the reference surface need exist only as a mathematical equation!

Holographic interferometry has been successfully applied to the testing of glass surfaces by several workers[110-113]. The advantages of the technique are that it can be applied to unpolished surfaces, the test object need not be removed to a laboratory set-up (especially if the double-exposure method is used), and the reference hologram can be computer-generated if required, thus avoiding the need for a real reference surface. Disadvantages are the difficulty of interpreting the interference patterns, which are representative of three-dimensional differences, and the expensive and rather difficult techniques involved in holography. A major consideration is the requirement to maintain point-to-point correspondence of object and reconstruction. The technique holds its own mainly in the field of testing large aperture mirrors.

Choice of test

The choice of test for a particular situation is often a matter of personal taste. For example, a skilled operator who has had years of experience in using and interpreting the Foucault knife-edge test is likely to prefer that test to the introduction of a new design of shearing interferometer, even though the latter may be capable of giving an accurate quantitative assessment of the wavefront under test. However, there are certain broad principles that should, on the whole, be followed.

The first of these is the adoption, if possible, of a null test. The test should be designed so that a perfect system would give a zero result. This might reveal itself as a uniformly illuminated field of view, straight and parallel interference fringes, a symmetrical distribution of light, or some other easily recognised state. In practice, a null result usually occurs when the test is carried out on a perfect spherical (or plane) wavefront, so that the design of a null test implies arranging the system so that it gives, in theory, such a perfect wavefront. A spherical mirror may be null tested by using a point source of light at its centre of curvature and an image-forming lens system by testing it at the conjugates for which it was designed. A paraboloidal mirror requires either collimated incident light or the introduction of a high-quality optical flat to produce an autostigmatic test (see Fig. 23), and an ellipsoidal mirror can be null tested by using a point source at one focus and testing the wavefront converging to the other. Often, the design of a simple null test is not possible, but even then workers often prefer to go to the extent of designing and manufacturing a special correcting lens-system so that a null test can in fact be achieved[114-118]. This is usually preferable to a test in which the observer is trying to match some sophisticated pattern that has been calculated as being produced by a perfect system, though this latter method is often used[35,85,86]

The second consideration is that the system or component should be tested in the conditions (optical and environmental) for which it was designed. Thirdly, and especially if

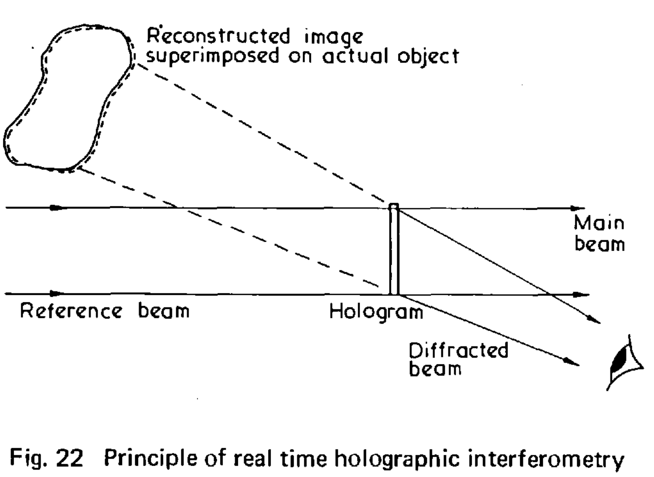


Fig. 22 Principle of real time holographic interferometry

a null test is not available, the test chosen should be easily interpreted, if possible quantitatively. And finally, the sensitivity of the chosen test should be known and understood—many optical tests are extremely sensitive and there is sometimes a danger of using a sledgehammer to crack a walnut. Other factors, such as cost, time and available equipment, must also be considered, as must the personal preferences of the operator mentioned above. Table 1 indicates which test methods, in general, are suitable for different types of optical components and systems. It does not claim to be exhaustive or exclusive, but is intended merely as a guide to the types of test that are usually adopted in various circumstances.

Conclusion

The testing of optical components and systems is of vital importance to the optical industry. It is the testing facilities that ultimately determine the degree of precision attainable by an operator. It is sad, therefore, to see so many optics companies all over the world ill-equipped with test

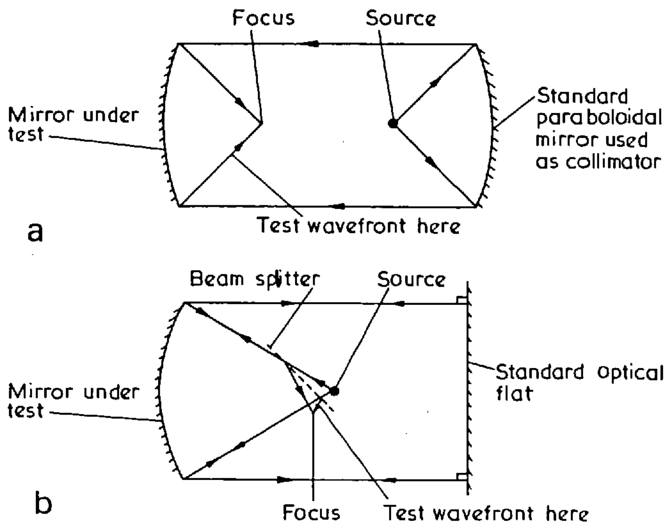


Fig. 23 Null testing a paraboloidal mirror: (a) Using a collimator; (b) Using an optical flat

Table 1 Choice of test

	Testplates	Fizeau	Lloyd moiré	Spherical Fizeau	Transmission	Twyman-Green	Wavefront shearing	Ronchi	Scatter fringe	Grazing incidence	Holography
Homogeneity of raw glass						x a					
Smoothed flats			x							x	
Optical flats (polished)	x	x				x					
Optical flats (large)			x				x	x			
Windows		x b			x	x	x	x			
Prisms						x					
Curved surfaces (unpolished)											x
Concave spherical mirrors				x		x c	x	x	x		
Convex spherical mirrors	x			x		x c					
Concave aspheric mirrors							x	x	x		
Convex aspheric mirrors (d)							x	x			
Lens surfaces	x			x		x					
Lens components						x	x	x	x		
Telescope objectives						x	x	x			
Microscope objectives						x	x				
Photographic lenses / Projection lenses						x	x	x	x		
Copying lenses							x	x	x		
Homogeneity of finished components					x	x					

(a) use liquid cell
(b) surface flatness test
(c) modified Twyman-Green
(d) use auxiliary optics

equipment and techniques, and relying on one or two, often unsuitable, test methods. On the other hand, a vast amount has been published on methods of optical testing that are limited to the one situation for which they were designed. Some sort of compromise is required, and it is hoped that this review has succeeded in its aim of describing the main interferometric test methods available, indicating to which types of optical component or system they are particularly suitable, and referring the reader to published papers for further details of individual techniques. No attempt has been made to compare interferometric test methods with ray-path techniques such as the Foucault knife-edge test and the Hartmann test, or with image evaluation techniques such as resolution tests and OTF. Such a task would be well outside the range of a single paper.

Acknowledgement

The author wishes to thank Dr M.C. Probine, Director of the Physics and Engineering Laboratory, for permission to publish this paper.

References

1 Candler, C. (1951). Modern Interferometers. Hilger and Watts
2 Francon, M. (1966). Optical Interferometry. Academic Press
3 Steel, W.H. (1967). Interferometry. CUP
4 Twyman, F. (1952). Prism and Lens Making (2nd Edition), Hilger and Watts, pp 375-380
5 Scott, R.M. (1965). Optical Manufacturing. Chapter 2 of: Applied Optics and Optical Engineering, Vol III (Ed R. Kingslake) Academic Press. 68-72
6 Fizeau, H. (1862). Recherches sur les modifications que subit la vitesse de la lumière dans le verre et plusieurs autres corps solides sous l'influence de la chaleur. Ann de Chim et de Phys **66**, 429-482
7 Laurent, L. (1883). Sur plusieurs appareils d'optique, destinés à controler les surfaces planes: parallèles, perpendiculaires et obliques. C r Acad Sci Paris **96**, 1035-1038
8 Dew, G.D. (1967). Optical flatness measurement – the construction and use of the Fizeau interferometer. NPL Rep No Op Met 1
9 Bünnagel, R. et al. (1968). Fizeau interferometer for measuring the flatness of optical surfaces. Appl Opt **7**, 331-335
10 Clapham, P.B. and Dew, G.D. (1967). Surface-coated reference flats for testing fully aluminized surfaces by means of the Fizeau interferometer. J Sci Instrum **44**, 899-902
11 Langenbeck, P. (1970). Fizeau interferometer – fringe sharpening. Appl Opt **9**, 2053-2058
12 Langenbeck, P.H. (1967). Lloyd interferometer applied to flatness testing. Appl Opt **6**, 1707-1714
13 Langenbeck, P. (1970). Higher-order Lloyd interferometer. Appl Opt 9, 1838-1841
14 SIRA (1967). New aid for the optical industry: an interferometer for testing deeply curved lens surfaces. SIRA News-sheet ('Impact') No 5
15 Biddles, B.J. (1969). A non-contacting interferometer for testing steeply curved surfaces. Opt Acta **16**, 137-157.
16 Heintze, L.R. et al. (1967). A multiple-beam interferometer for use with spherical wavefronts. Appl Opt **6**, 1924-1929
17 Forman, P.F. (1964). A note on possible errors due to thickness variations in testing nominally parallel plates. Appl Opt **3**, 646-647
18 Saunders, J.B. (1954). In-line interferometer. J Opt Soc Am **44**, 241-242
19 Post, D. (1954). Characteristics of the series interferometer. J Opt Soc Am 44, 243-249
20 Ashton, A. and Marchant, A.C. (1967). Note on the testing of large glass panels. Opt Acta **14**, 203-204
21 Roberts, F.E. and Langenbeck, P. (1969). Homogeneity evaluation of very large disks. Appl Opt **8**, 2311-2314
22 Leppelmeier, G.W. and Mullenhoff, D.J. (1970). A technique to measure the wedge angle of optical flats. Appl Opt 9, 509-510
23 Tynes, A.R. and Bisbee, D.L. (1967). Precise interferometry of glass plates. IEEE J Quantum Electron **QE3**, 459-463
24 Hariharan, P. and Sen, D. (1959). Triangular path macro-interferometer. J Opt Soc Am 49, 1105-1106
25 Saunders, J.B. (1969). An interferometer for measuring gradients in both refractive index and thickness of large or small optics. J Res Natn Bur Stands **73C**, 1-4
26 Twyman, F. and Green, A. (1917). Improvements in finishing prisms or lenses or combinations of the same and in apparatus therefor. Brit Pat No 103 832
27 Twyman, F. (1918). Interferometers for the experimental study of optical systems from the point of view of the wave theory. Phil Mag 35, 49-58
28 Kingslake, R. (1925-6). The interferometer patterns due to the primary aberrations. Trans Opt Soc London 27, 94-105
29 Hariharan, P. and Sen, D. (1961). The separation of symmetrical and asymmetrical wavefront aberrations in the Twyman interferometer. Proc Phys Soc 77, 328-334
30 Houston, J.B. et al. (1967). A laser unequal path interferometer for the optical shop. Appl Opt **6**, 1237-1242
31 Munnerlyn, C.R. et al. (1970). Interferometer to replace test glasses for optical production. Opt Soc Am Annual Meeting Paper No ThH13
32 Bates, W.J. (1947). A wavefront shearing interferometer. Proc Phys Soc **59** 940-950.
33 Saunders, J.B. (1961). Measurement of wavefronts without a reference standard: Part 1. The wavefront-shearing interferometer. J Res Natn Bur Stands **65B**, 239-244
34 Bryngdahl, O. (1965). Applications of shearing interferometry. Progress in Optics, Vol IV (Ed E. Wolf). North-Holland. 39-83
35 Malacara, D. and Mendez, M. (1968). Lateral shearing interferometry of wavefronts having rotational symmetry. Opt Acta **15**, 59-63
36 Briers, J.D. (1969). Prism shearing interferometer. Opt Technol 1, 196-202
37 Lenouvel, L. and Lenouvel, F. (1938). Étude des faisceaux convergents. Rev d'Opt 17, 350-361
38 Dyson, J. (1957). Common-path interferometer for testing purposes. J Opt Soc Am **47**, 386-390
39 Dyson, J. (1963). Very stable common-path interferometers and applications. J Opt Soc Am **53**, 690-694
40 Bartholomeyczyk, W. (1960). Zur Verwendung des Wollastonprismas als Strahlteiler in Differentialinterferometern. Z InstrumKde **68**, 208-213.
41 Drew, R.L. (1951). A simplified shearing interferometer. Proc Phys Soc **B64**, 1005-1010
42 Brown, D. (1954). A shearing interferometer with fixed shear and its application to some problems in the testing of astro-optics. Proc Phys Soc **B67**, 232-237.
43 Saunders, J.B. (1961). A simplified wavefront shearing interferometer and unique method of evaluating any converging wavefront. Opt Soc Am Annual Meeting Paper No TC 15
44 Saunders, J.B. (1964a). Wave front shearing prism interferometer. J Res Natn Bur Stands 68C, 155-171
45 Saunders, J.B. (1964b). Interferometer test of the 26-inch refractor at Leander McCormick Observatory. Astr J **69**, 449-451
46 Saunders, J.B. (1965). Some applications of the wave-front shearing interferometer. Jap J Appl Phys **4** (Supp 1), 99-105
47 Murty, M.V.R.K. (1970). A compact lateral shearing interferometer based on the Michelson interferometer. Appl Opt **9**, 1146-1148
48 Saunders, J.B. (1967). A simple, inexpensive wavefront shearing interferometer. Appl Opt **6**, 1581-1583
49 Murty, M.V.R.K. (1964). The use of a single plane parallel plate as a lateral shearing interferometer with a visible gas laser source. Appl Opt **3**, 531-534
50 Murty, M.V.R.K. and Malacara-Hernandez, D. (1965). Some applications of the gas laser as a source of light for the

testing of optical systems. Jap J Appl Phys 4, (Supp 1), 106-112
51 Kelley, J.G. and Hargreaves, R.A. (1970). A rugged inexpensive shearing interferometer. Appl Opt 9, 948-952
52 Donath, E. and Carlough, W. (1963) Radial shearing interferometer. J Opt Soc Am 53, 395
53 Van Rooyen, E. and van Houten, A.G. (1969). Design of a wavefront shearing interferometer useful for testing large aperture optical systems. Appl Opt 8, 91-94
54 Ashton, A. and Marchant, A.C. (1969). A scanning interferometer for wavefront aberration measurements. Appl Opt 8, 1953-1955
55 Lohmann, A. and Bryngdahl, O. (1967). A lateral wavefront shearing interferometer with variable shear. Appl Opt 6, 1934-1937
56 Hariharan, P. and Sen, D. (1960). Cyclic shearing interferometer. J Sci Instrum 37, 374-376
57 Brown, D.S. (1955). The application of shearing interferometry to routine optical testing. J Sci Instrum 32, 137-139.
58 Saunders, J.B. (1970). A simple interferometric method for workshop testing of optics. Appl Opt 9, 1623-1629
59 Dutton, D. et al. (1968). A semiautomatic method for interpreting shearing interferograms. Appl Opt 7, 125-131
60 Murty, M.V.R.K. and Hagerott, E.C. (1966). Rotational shearing interferometry. Appl Opt 5, 615-619
61 Armitage, J.D. and Lohmann, A. (1965). Rotary shearing interferometry. Opt Acta 12, 185-192
62 Gates, J.W. (1955). Reverse-shearing interferometry. Nature 176, 359-360
63 Gates, J.W. (1955). The measurement of aberrations by interferometry. Proc Phys Soc B68, 1065-1072
64 Saunders, J.B. (1962). Measurement of wavefronts without a reference standard: Part 2. The wave-front-reversing interferometer. J Res Natn Bur Stands 66B, 29-40
65 Sen, D. and Puntambekar, P.N. (1966). Shearing interferometers for testing corner cubes and right angle prisms. Appl Opt 5 1009-1114
66 Bryngdahl, O. (1970). Reversed-radial-shearing interferometry. J Opt Soc Am 60, 915-917
67 Brown, D.S. (1960). Radial shear interferograms. NPL Symposium No 11 (Interferometry), HMSO 253-256.
68 Brown, D.S. (1962). Radial shear interferometry. J Sci Instrum 39, 71-72.
69 Hariharan, P. and Sen, D. (1961). Radial shearing interferometer. J Sci Instrum 38, 428-432
70 Murty, M.V.R.K. (1964). A compact radial shearing interferometer based on the law of refraction. Appl Opt 3, 853-857
71 Steel, W.H. (1965). A radial shear interferometer for testing microscope objectives. J Sci Instrum 42, 102-104
72 Steel, W.H. (1970). A radial-shear interferometer for use with a laser source. Opt Acta 17, 721-724
73 Som, S.C. (1970). Theory of a compact radial shearing laser interferometer. Opt Acta 17, 107-113
74 Bryngdahl, O. (1971). Shearing interferometry with constant radial displacement. J Opt Soc Am 61, 169-172.
75 Langenbeck, P. (1971). Modifying a shear interferometer to obtain a neutral reference beam. J Opt Soc Am 61, 172-175
76 Ronchi, V. (1962). Forty years of 'gratings'. Atti Fond G Ronchi 17, 93-143
77 Ronchi, V. (1964). Forty years of history of a grating interferometer. Appl Opt 3, 437-451
78 Foucault, L. (1858). Description des procédés employés pour reconnaitre la configuration des surfaces optiques. C r Acad Sci Paris 47, 958-959
79 Gascoigne, S.C.B. (1944). The theory of the Foucault test. Mon Not Roy Astr Soc 104, 326-334.
80 Linfoot, E.H. (1946). A contribution to the theory of the Foucault test. Proc Roy Soc 186, 72-99
81 Welford, W.T. (1970). A note on the theory of the Foucault knife-edge test. Opt Commun 1, 443-445
82 Ronchi, V. (1923). Le frange di combinazione nello studio delle superficie e dei sistemi ottici. Riv d'ottica e Mecc 2, 9-35
83 Ronchi, V. (1926). Sur la nature interférentielle des franges d'ombre dans l'essai des systèmes optiques. Rev d'Opt 5, 441-449
84 Sherwood, A.A. (1959). Ronchi test charts for parabolic mirrors. J & Proc Roy Soc NSW 93, 19-23
85 Malacara, D. (1965). Geometrical Ronchi test of aspherical mirrors. Appl Opt 4, 1371-1374
86 Cornejo, A. and Malacara, D. (1970). Ronchi test of aspherical surfaces, analysis and accuracy. Appl Opt 9, 1897-1901
87 De Vany, A.S. (1970). Quasi-Ronchigrams as mirror transitive images of interferograms. Appl Opt 9, 1944-5
88 Malacara, D. & Cornejo, A. (1971). Modified Ronchi test to measure the axial chromatic aberrations in lenses. Appl Opt 10, 679-680
89 Burch, J.M. (1953). Scatter fringes of equal thickness. Nature 171, 889-890
90 Shoemaker, A.H. and Murty, M.V.R.K. (1966). Some further aspects of scatter-fringe interferometry. Appl Opt 5, 603-607
91 Polster, H.D. et al. (1969). New developments in interferometry. Appl Opt 8, 521-556
92 Waland, R.L. (1938). Note on figuring Schmidt correcting lenses. J Sci Instrum 15, 339-340
93 Moreau, B.G. and Hopkins, R.E. (1969). Application of wax to fine ground surfaces to simulate polish. Appl Opt 8, 2150, 2151
94 Munnerlyn, C.R. et al. (1969). Interferometric measurements of optically rough surfaces. IEEE J Quantum Electron QE5, 359
95 Rowe, S.H. and Welford, W.T. (1967). Surface topography of non-optical surfaces by projected interference fringes. Nature 216, 786, 787
96 Welford, W.T. (1969). Some applications of projected interference fringes. Opt Acta 16, 371-376
97 Abramson, N. (1969). The 'Interferoscope', a new type of interferometer with variable fringe separation. Optik 30, 56-71
98 Abramson, N. (1970). Das Interferoskop, ein neues Mittel zum Studium des Werkzeugverschleisses. Fertigung 2/3/70. 83-86
99 Briers, J.D. (1971). Interferometric flatness testing of nonoptical surfaces. Appl Opt 10, 519-524
100 Gabor, D. (1948). A new microscope principle. Nature 161, 777-778
101 Gabor, D. (1949). Microscopy by reconstructed wavefronts. Proc Roy Soc A197, 454-487
102 Gabor, D. (1951). Microscopy by reconstructed wavefronts: II Proc Phys Soc B64, 449-469
103 Leith, E.N. and Upatnieks, J. (1963). Wavefront reconstruction with continuous-tone objects. J Opt Soc Am 53, 1377-1381
104 De Velis, J.B. and Reynolds, G.O. (1967). Theory and Applications of Holography. Addison-Wesley
105 Smith, H.M. (1969). Principles of Holography. Wiley-Interscience
106 Stroke, G.W. (1969). An Introduction to Coherent Optics and Holography (2nd Edition). Academic Press
107 Hildebrand, B.P. et al. (1967). Holography as a tool in the testing of large aperture optics. Appl Opt 6, 1267-1269
108 Murty, M.V.R.K. (1963). Common path interferometer using Fresnel zone plates. J Opt Soc Am 53, 568-570
109 Haines, K.A. and Hildebrand, B.P. (1966). Surface deformation measurement using a wavefront reconstruction technique. Appl Opt 5, 595-602
110 Van Deelen, W. & Nisenson, P. (1969). Mirror blank testing by real time holographic interferometry. Appl Opt 8, 951-955
111 Fercher, A.F. and Torge, R. (1970). Holographic test glasses. Opt Laser Technol 2, 200-201
112 Snow, K. and Vandewarker, R. (1970). On using holograms for test glasses. Appl Opt 9, 822-827
113 MacGovern, A.J. and Wyant, J.C. (1971). Computer generated holograms for testing optical elements. Appl Opt 10, 619-624
114 Offner, A. (1963). A null corrector for paraboloidal mirrors. Appl Opt 2, 153-155
115 James, W.E. and Waterworth, M.D. (1965). A method for testing aspheric surfaces. Opt Acta 12, 223-227
116 Demarcq, J. et al. (1966). Controle des miroirs asphériques concaves. Rev d'Opt 45, 497-500
117 Shapochkin, B.A. et al. (1967). A simple interferometer for quality control of aspherical surfaces. Sov J Opt Technol 34, 579-582
118 Lytle, J.D. (1970). A suggested procedure for testing large Cassegrainian optical systems. Appl Opt 9, 2497-2500

Industrial and engineering applications of visible-light lasers

J. D. BRIERS
B.SC., M.SC., A.R.C.S., D.I.C., M.INST.P., (NON-MEMBER)

The laser is one of the most important and most dramatic inventions of recent years. Its arrival has revolutionised the science of optics, and its inflence is rapidly spreading to other disciplines. This review is an attempt to make engineers aware of the potential of this new tool, and to outline how the laser may be applied to their particular fields.

1. INTRODUCTION

THE invention of the laser in 1960[1] provided the world with an intense source of highly coherent light for the first time. Many techniques which had been impracticable with conventional light sources suddenly became possible. The laser is now an essential tool in many fields of scientific research, but this paper will concentrate on those applications of more direct interest to engineers and industrialists.

2. PROPERTIES OF LASER LIGHT

The two main differences between laser light and "ordinary" light are illustrated diagrammatically in Fig. 1. Light from a laser is emitted as a very narrow, almost parallel beam, while light from a conventional source is spread over a large area. Also, laser light is highly monochromatic and coherent. Monochromatic means single coloured, and this property of laser light results from the fact that the lasing action occurs only at very sharply defined wavelengths. Coherence and monochromaticity are linked, and coherence can be regarded as a measure of the purity of the light. Conventional light is emitted as a random series of short pulses, with no phase relationship between different pulses. This means that phenomena relying on phase relationships, such as interference effects, can only be observed over path lengths of the same order as the average pulse length. With laser light, on the other hand, the phase relationship is maintained over much longer distances: the light is said to be coherent.

3. TYPES OF LASER AVAILABLE

The laser in most common use is the helium-neon (He-Ne) laser, mainly because it is the cheapest and the most reliable. The light emitted is red (wavelength 633 nm), and many people seem to think that all laser light is red. This was certainly true a few years ago, but nowadays lasers cover the whole visible spectrum and extend into the ultraviolet and infrared regions. Infrared lasers, in fact, have several important industrial applications, but these lie outside the scope of this paper. (A colleague of the author hopes to publish a paper on the applications of infrared lasers in the near future[2].)

Table I gives a selection of typical lasers available today, with information on wavelengths, power outputs and approximate prices. The choice of laser will depend on the particular application for which it is required, but in most cases a low-powered helium-neon laser will be suitable.

J. D. BRIERS has been a scientist with the optics section, Physics and Engineering Laboratory, Department of Scientific and Industrial Research, at Lower Hutt since 1969. He was born in Lancashire, England, and in 1964 joined Pilkington Brothers Ltd., Lathom, Lancashire, as a research physicist. In 1967 he became project engineer with Pilkington Perkin-Elmer Ltd. at St. Asaph, Flintshire, and two years later joined P.E.L. in Lower Hutt. He has recently been awarded a D.S.I.R. study award to return to England to carry out post-graduate research in optics at Imperial College, London.

David Briers has been secretary-treasurer of the Wellington division of the N.Z. branch of the Institute of Physics from 1971 and has published a number of papers, most of which have been associated with optical interferometry.

This paper was received on 7 November 1972.

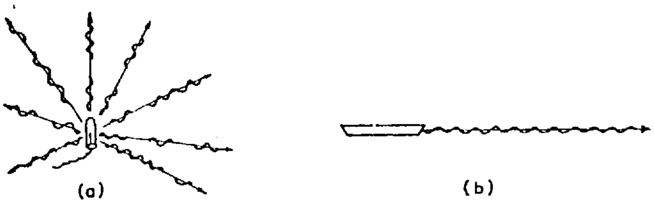


Fig. 1: Diagrammatic representation of the difference between (a) conventional light and (b) laser light.

A word of warning about laser safety might not be out of place here. The power outputs quoted in Table I will seem very low compared with the 100 W of an average household lamp. However, it must be realised that, with a laser, the power is concentrated into a very narrow beam instead of being spread over a large angle. In fact, the intensity of even a 2 mW laser beam is sufficient to cause irreparable eye damage, and the need for caution cannot be overemphasised.

4. GEOMETRICAL APPLICATIONS

To a first approximation, light travels in straight lines. The narrowness and intensity of the laser beam make it very useful for defining a straight line in space, and this property is the basis of several important laser applications.

4.1. Alignment

The laser beam makes an ideal "weightless string" for alignment purposes, and has been widely adopted for aligning machinery. The main advantage is that the spot of light formed where the beam hits a target is bright enough to be seen with the naked eye, without the need for sighting back along the beam. For high-precision alignment, optical techniques can be used to pinpoint the centre line of the laser beam[3].

Lasers can be used for alignment tasks over longer distances, providing an optical system is used to compensate for the slight beam divergence, and the effects of atmospheric refraction are taken into account[4, 5]. In these applications, the laser replaces the conventional surveying techniques using theodolites and levels, and greatly reduces the amount of labour and time involved in checking alignment and grade. Important fields of application include pipe-laying, bridge construction, dredger guidance, and tunnelling. Laser alignment systems have been used on several tunnelling projects in New Zealand, and have proved very simple and convenient to operate. It must be stressed, however, that laser light is subject to the same refraction effects as ordinary light, and, unlike the conventional two-way levelling techniques, the laser method does not compensate for such effects. For long path-lengths the recommended method is to employ two or more targets with holes, which are surveyed into position using conventional two-way techniques so that the line joining the holes defines the required line and grade; the laser beam is adjusted so that it passes through the holes in the targets, and the required line is then defined by the laser beam (see Fig. 2). Even then, care must be taken not to rely too much on the position of the beam at large distances from the last surveyed target, since refraction effects occurring after that target can cause the beam to deviate.

4.2. Navigational aids

Lasers can replace conventional light sources in various types of navigational aids, from harbour lights indicating the boundaries of safe channels, to full-scale lighthouses. Again the main advantages of the laser are its intensity and its ability to define a line very precisely, though opinion is divided as to whether the laser has sufficient advantage to replace conventional sources in this field.

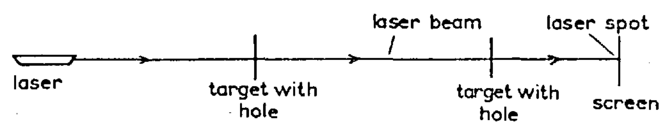


Fig. 2: Laser alignment technique—the two targets are surveyed into position using conventional methods.

TABLE I

Typical lasers available today

Type	*Wavelength* nm	*Colour*	*Power*	*Approx. price* $
Cadmium	325 442	UV Blue	5 mW 50 mW	7 500
Cadmium	442	Blue	20 mW	800
Argon	8 lines below ·515	Violet, blue, green	4 W total	10 000
Argon	4 lines below 515	Blue, green	25 mW total	4 300
Dye	360·650 (tunable)	Variable over whole spectrum	50μJ pulse	2 000
Frequency-doubled Nd doped YAG* or Nd in glass	530	Green	10 W	10 000
Frequency-doubled Nd doped YAG or Nd in glass	530	Green	500 mW	5 000
He-Ne	633	Red	0.5 mW	100
He-Ne	633	Red	2 mW	300
He-Ne	633	Red	15 mW	3 300
He-Ne	633	Red	50 mW	6 400
Ruby	694	Red	0.5 J pulse	4 000
Ruby	694	Red	70 J pulse	7 000

* Yttrium Aluminium Garnet.

4.3. Industrial inspection and control

Optical inspection and control systems have found widespread application in industry, their main advantage over mechanical methods being that they are non-contacting. The laser is often a suitable source for these techniques, especially when a small area needs to be defined or the speed of the industrial process makes an intense source desirable.

Practical applications of laser monitoring and control have included the automatic gauging of products such as glass tubing, checking the motion of a whirling spindle[6], and the measurement of angles and rotation[7].

4.4. Laser communications

The fact that laser light can propagate over quite large distances as a narrow, parallel beam with little divergence has stimulated interest in its use for communications. As with any electromagnetic radiation, a light beam can act as a carrier wave and can be modulated to carry information. The main problems are atmospheric refraction and attenuation by rain and fog, and it seems unlikely that any useful system will be developed involving point-to-point laser communications in the atmosphere. However, there are great prospects for lasers in the field of space communication, where they offer the advantages of high power density and low divergence. Also, the bandwidth limitations of conventional communication systems have encouraged research into overcoming the problems of terrestrial laser communications, and the most promising line at this moment seems to be the use of fibre optics as waveguides. Perhaps it will not be too long before telephone wires are being manufactured from glass!

5. INTERFEROMETRIC APPLICATIONS

Everyone is familiar with optical interference phenomena in the form of coloured patterns in soap bubbles or in oil films on water. Light reflected from the two surfaces of a thin film interferes, resulting in the destruction of some wavelengths (colours) and the enhancement of others. With conventional light sources such effects can only be observed over very small path-lengths—the thickness of the film of oil or soap solution in the above examples. Even with this restriction, much use has been made of the interference properties of light in the fields of metrology and optical testing. By using a spectral source such as a mercury or sodium lamp, the path-length can be increased to a few centimetres.

The basis of the technique is that, if the path-difference between the two interfering light beams is a whole number of wavelengths of the light used, the beams are superimposed in phase and their amplitudes added; on the other hand, if the path-difference is an odd number of half-wavelengths, the beams are superimposed in anti-phase (crest on trough) and cancel out. Between these two extremes there is a sinusoidal variation of intensity of the superimposed beams. This effect gives a measurement scale whose unit is the wavelength of light (about 0.5 μm), and which can be used to measure small path-differences to tolerances of 0.01 μm or better.

The advent of the laser has allowed the extension of interferometric techniques to situations involving large path-differences, while still retaining the fine tolerances quoted above.

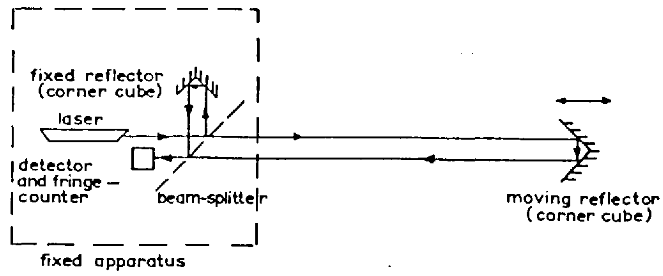


Fig. 3: Laser interferometer for distance measurement.

5.1. Distance measurement

Laser interferometers have been applied to the precise measurement of distances. A typical instrument uses a fixed laser head and reference beam and a moving retro-reflector (see Fig. 3). As the reflector is moved away from the laser head, intensity maxima and minima are counted[8] and converted electronically to a distance measurement which is displayed digitally. Resolutions of 0.1 μm over distances up to 10 metres are possible. For distances greater than 10 metres, beam modulation techniques are usually used.

5.2. Velocity measurement

By combining Doppler techniques with laser interferometry it is possible to measure velocities. This approach is very useful when a non-contacting method is required, and practical applications have included systems to measure the velocity of hot aluminium extrusions and hot steel bars[9].

5.3. Metrology

For some time metrologists have been using optical interferometry as a measuring tool. Now that lasers are available, such techniques can be extended to cover applications involving longer path differences. The United States National Bureau of Standards has recently announced a laser interferometer which can measure the diameters of spheres and cylinders to an accuracy better than 1 μm[10]. Other applications reported include the accurate sizing of machined parts[11] and the control and calibration of machine tools[12].

5.4. Optical testing

The optical industry has used interferometry as a testing technique for many years. The tolerances on optical components are such that the wavelength of light is a convenient unit to use, and numerous types of interferometer have been developed for testing flatness, sphericity, homogeneity, parallelism and optical aberrations[13]. However, the coherence and intensity limitations of conventional light sources have placed some restrictions on the use of the technique. The laser has overcome these restrictions, and the optical industry now makes widespread use of laser interferometers[14].

6. DIFFRACTION APPLICATIONS

When a light beam is partially cut off by an object, the object casts a shadow. However, some of the light that just misses the edge of the object is bent slightly by the proximity of the object and spreads into the shadow region. Thus the edge of a shadow is never the sharp cut-off which would be predicted by geometrical optics, but is somewhat diffuse. This bending of light rays by objects in their path is called diffraction, and an everyday example is provided by the ghost images observed when scenes are viewed through close-weave net curtains. Diffraction effects also cause the coloured haloes around street lamps when viewed through fog or a misted-up window.

When the object intercepting the light beam is narrow enough, the light diffracted around both edges of the object will overlap and will interfere. The fringe spacing of the resulting "diffraction pattern" depends on the width of the object, narrower objects forming more widely spaced fringes. Also, the direction in which the light is diffracted is perpendicular to the edge of the object at the point considered. The result of all this is that a complex object will form a complex but distinctive diffraction pattern. Moreover, this diffraction pattern is centred on the optic axis of the optical system used, and is symmetrical about any diameter.

Diffraction, like interference, is wavelength-dependent, and white light sources give confused multi-coloured diffraction patterns which are difficult to analyse quantitatively. Filtering out a specific wavelength drastically reduces the intensity of light available, and once again the intense monochromatic light of the laser provides a very useful source for diffraction phenomena.

6.1. Remote gauging

Since the fringe spacing in a diffraction pattern increases as the scale of the object decreases, it follows that the smaller the object the easier it is to measure its size by monitoring its diffraction pattern. This leads to a valuable non-contacting technique for the automatic gauging of thin wires and similar objects[15, 16].

6.2. Edge inspection

The diffraction pattern from a nominally straight edge can be used to detect deviations from straightness of the edge. Large-scale irregularities will cause a rotation of the diffraction pattern, since light is diffracted into a line perpendicular to the edge. This rotation can easily be detected by either the human eye or an array of photocells. Smaller irregularities will cause more subtle changes in the diffraction pattern, but optical techniques are available to detect and measure these changes. The method can be used to test the edges of razor blades and similar objects[17, 18].

6.3. Pattern recognition

A given object produces a distinctive diffraction pattern, and identification of the diffraction pattern results in identification of the object. It is easier to design a system to recognise the symmetrical diffraction pattern than to design one to recognise the object itself, which might be completely asymmetrical, and randomly positioned and orientated. Automatic pattern recognition systems for simple shapes such as printed alphanumeric characters have been developed[19], and the technique can be extended to complex patterns such as biological samples[20], signatures and fingerprints. Typical laser diffraction patterns are illustrated in Fig. 4.

7. SPECKLE APPLICATIONS

Anyone who has seen a laser in operation will have noticed the granular effect laser light gives to surfaces on which it falls. As the observer moves his head the surface appears to scintillate. The effect, known as "speckle", is an interference effect resulting from the high coherence of laser light. Speckle has annoyed research workers ever since lasers were in-

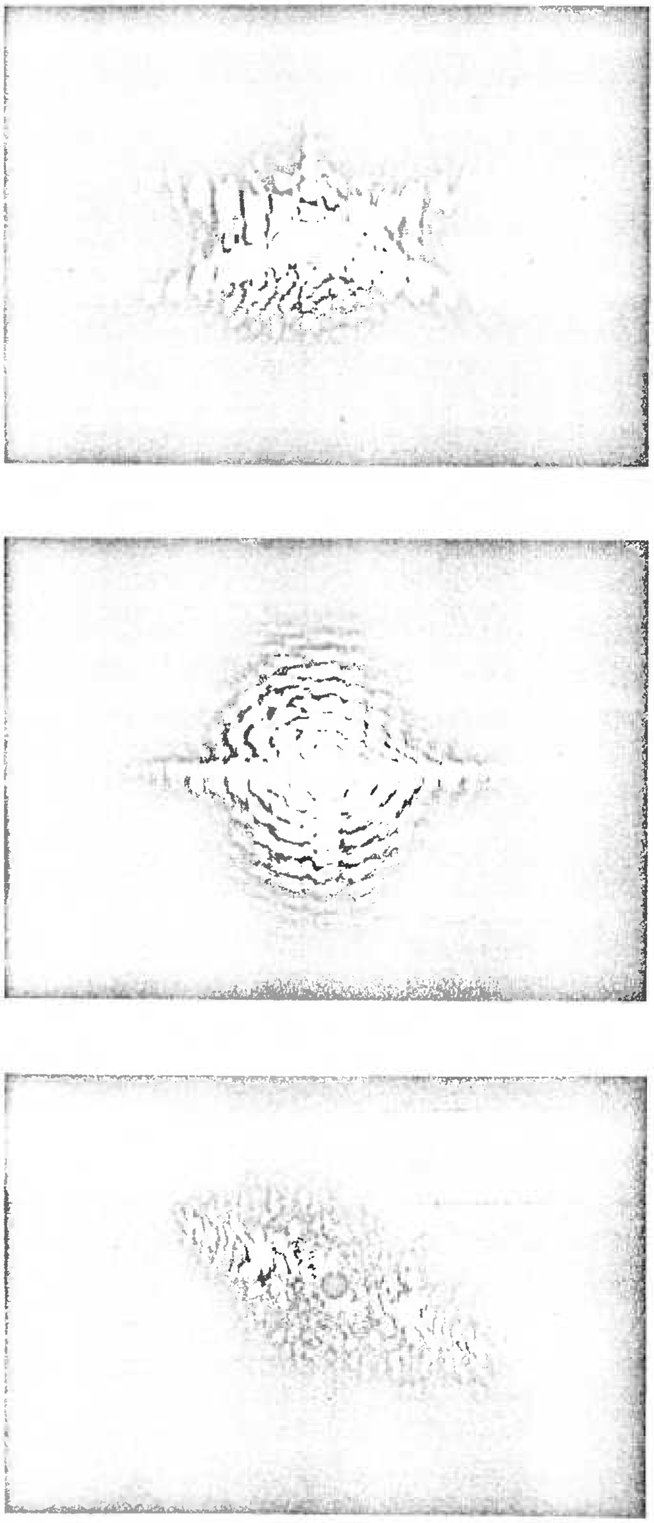


Fig. 4: Laser diffraction patterns of (a) the printed letter "w", (b) the printed letter "d", (c) the author's signature. (Note: *in (c) the bright central spot has been masked by an opaque disc to prevent excessive light-scattering in the emulsion.)*
(P.E.L. photographs)

vented, as it tends to interfere with observations and measurements. Recently, however, the speckle effect has been turned to good use with the development of "speckle interferometry"[21], after it was found that the speckle pattern varied with surface finish and with motion of the surface.

7.1. Vibration analysis

Motion of the illuminated surface tends to blur out the speckle pattern, and this fact can be used in the investigation of vibrational modes[21, 22, 23, 24]. Vibrating objects illuminated with diffuse laser light display patterns similar to the familiar Chladni sand patterns. In this way the speckle effect can be used to determine the mode of vibration, nodal lines, and the amplitude and the direction of the vibration.

7.2. Displacement and strain measurement

Double-exposure photographs of the speckle pattern from a surface taken before and after the surface is displaced yield interference fringes from which the in-plane component of the displacement can be calculated[25, 26]. The technique has also been extended to the determination of the normal component of the displacement[27]. Since applied stresses will produce local displacements or strains in an object, speckle interferometry can also be used as a stress analysis tool[28].

7.3. Velocity measurement

If the surface illuminated by a laser beam has a component of velocity in the plane perpendicular to the beam the speckle pattern will scintillate. A detector monitoring the speckle pattern will give a signal which includes a frequency proportional to the velocity. This approach has been suggested as the basis of a non-contact technique for measuring velocities, and has been used on strip passing through a rolling mill[29].

7.4. Surface roughness measurement

The appearance of the speckle pattern depends on the microstructure of the surface being illuminated. The speckle pattern can thus be used as a quality control tool for surface finish[30, 31], though care must be taken to eliminate other factors affecting the appearance of the speckle pattern, such as movement or vibration of the surface or the aperture of the viewing system.

8. HOLOGRAPHY

Holography, or "wavefront reconstruction", was invented in 1948[32, 33] in an effort to increase the resolution of the electron microscope. The technique, which requires coherent light to work satisfactorily, did not succeed in its original aim, but when lasers became available in the 1960s holography became a field of rapid growth[34, 35].

In simple terms, holography is a method of three-dimensional photography which requires coherent light but does not need photographic lenses. The difference between conventional photography and holography can be expressed in the following way. In ordinary photography a lens forms an image of an object on a photographic film so that information from any one point on the object is recorded at one point on the film—there is a one-to-one relationship between image and object, and the result is a two-dimensional projection of the object as seen from one direction of view defined by the centre of the photographic lens. In holography, on the other hand, light from all points on the object is allowed to reach all parts of the film —each point on the film receives information from all points of the object, and the result is a three-dimensional, full-parallax image, the perspective of which changes according to which part of the hologram is viewed. A conventional photograph is viewed by looking *at* the photograph, but a hologram is viewed by looking *through* the film, and the image appears as though it is being viewed through a window the same size as the hologram. If a conventional photograph is cut into pieces, each piece will contain only part of the image. Since all points on a hologram contain information about the whole of the object, when a hologram is cut into pieces each piece still contains an image of the whole object—the only things lost are angle of view (since the hologram-window is smaller) and some resolution.

The secret of the hologram's three-dimensional achievements lies in the fact that whereas a conventional photograph records only the amplitude of the light from the object, the hologram records both the amplitude and the phase. This is achieved by illuminating the photographic plate with a "reference beam" of laser light at the same time as it is receiving reflected laser light from the object (see Fig. 5). A simplified physical interpretation of what happens is that the light from the object interferes with the reference beam to form a complex interference pattern on the

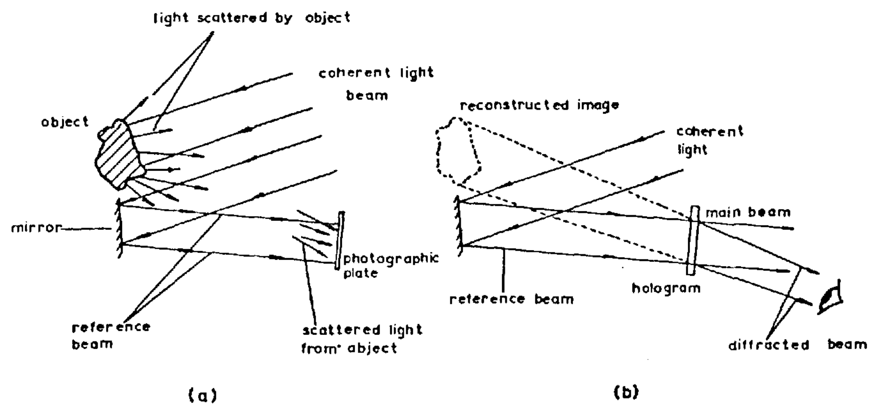


Fig. 5: Principle of holography: (a) recording, (b) reconstruction.

photographic film. When the film is developed this interference pattern provides a record of the wavefront coming from the object, coded with the reference beam. If the hologram is now illuminated with the reference beam only, it acts as a complex diffraction grating in such a way that the diffracted beam is indistinguishable from the original object beam that produced the hologram (see Fig. 5). Thus the wavefront from the object is reconstructed in both amplitude and phase, and an observer looking into this reconstructed wavefront sees a reconstruction of the original object just as though he were looking at the object itself.

8.1. Holographic display

The most obvious application of holography is three-dimensional display. The three-dimensional effect is very realistic, and most people are greatly impressed the first time they see a hologram. By using three lasers of different colours, and thick photographic emulsions to produce so-called "volume holograms", it is possible to make holograms that will reconstruct in ordinary white light to give full-colour three-dimensional images—a very dramatic effect[36, 37].

Holograms can be of value in the fields of advertising and education. Training simulators provide another possible application[38], and there have even been suggestions of holographic road signs[39].

8.2. Data processing

Compared with conventional magnetic tapes and discs, holograms can store a vast amount of easily accessible data. If volume holograms using thick emulsions are used, packing densities as high as 10^{15} bits in a 10 cm by 10 cm plate are feasible[40], a feature which would revolutionise data processing. The main stumbling-block at the moment is that holograms are "read-only" memories with no simple method of erasing and re-recording. However, several American companies are committed to the development of holographic data processing, and it may not be very long before the first "optical computer" is developed, using light instead of electric currents and holograms in place of magnetic discs.

8.3. Particle analysis

Unlike conventional photography, holography has no depth-of-field limitations, and the whole image is in focus at once. This tremendous advantage has led to the use of holography in the field of particle analysis[41]. Holography has been used to measure the sizes and distributions of particles in fogs and aerosols[42, 43] and has proved of value in bubble-chamber photography as an aid to nuclear research[44].

8.4. Image enhancement

Any amateur photographer will know the frustration of obtaining a "perfect" shot of an unrepeatable event and then having the result ruined by bad focusing, camera-shake, subject movement, or simply the limitations of the lens used. The unfortunate photographer would probably not believe us when we told him that all was not lost, and that by using holography the image on the original negative could be sharpened and improved[45]. Yet this seemingly impossible task can be achieved providing we know, or can make a reasonable guess at, the nature and magnitude of the effect causing the unsharpness. The blurred photograph can be regarded as containing information about the subject combined with information about the blurring function. If this blurring function is known, holographic techniques can be used to subtract it from the combined information map, leaving only the information about the subject[46]. In this way, a blurred photograph can often be sharpened up considerably, and previously hidden details resolved. The method was used with some success on the Gemini XII mission, when some of the astronaut's photographs were badly affected by camera movement[47]. Research is still proceeding to improve the technique.

8.5. Holographic interferometry

In the mid-1960s the techniques of holography and interferometry were combined[48] to provide a very powerful tool for non-destructive testing. For the first time interferometry could be applied to opaque, non-reflecting, arbitrarily shaped, three-dimensional objects[49].

If a hologram is made of an object, and, after processing, the plate is replaced on its original position, the holographic image formed by the hologram will be superimposed exactly on the original object itself. If the object is now moved or deformed very slightly, the wavefront reflected from it will interfere with the reconstructed wavefront from the hologram, and an interference pattern will appear on the object. This pattern can be analysed to give the magnitude and direction of movement of different parts of the object[50].

The technique just described is known as "real-time" or "live-fringe" holographic interferometry, but an alternative and easier method is to take a double-exposure hologram with the object moved or deformed between the exposures[51]. This "double-exposure" or "frozen-fringe" technique has the significant advantage of not requiring very precise relocation of the hologram after processing, but the benefits of real-time operation are lost. Double-exposure holographic interferometry is used when an object is to be compared with itself, before and after it is subjected to some kind of stress or deformation (see Fig. 6). Industrial and engineering applications reported to date include stress analysis on opaque objects[50], the inspection of car tyres[52], the investigation of surface corrosion[53], wind-tunnel studies[48], and the detection of faults in rubber-to-metal and metal-to-metal joints[52].

If it is required to compare a test object with a master object it is necessary to use the real-time approach, and work is going on at present into solving the problem of very accurate relocation of the hologram. One approach is to use *in situ* processing of the photographic plates[54], but a more promising technique is probably the use of alternative recording media which require no processing[55]. Real-time holographic interferometry has been used for dynamic stress analysis[56], flow analysis[57], and the testing of large optical components[58]. Another interesting point about the technique is that since the hologram of a given object can be computed and produced artificially, the "master object" need only exist as a mathematical equation! This can be of great value when no reference master is available, as in the case of large one-off optical components[59].

A third technique is known as "time-averaged" holographic interferometry, and is used for the study of vibrating systems[60]. In this case, the fringes indicate the distribution of the nodes and antinodes in the same manner as speckle interferometry (see section 7.1). Engineering applications have included the vibration analysis of turbine blades and aero engine components[61].

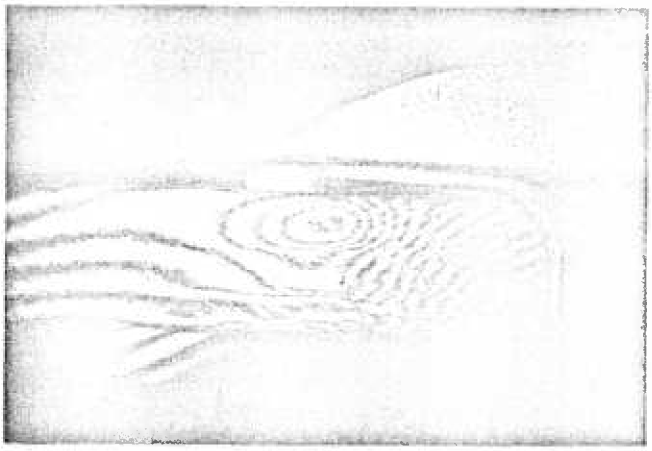


Fig. 6: Holographic interferometry. The G-clamp was tightened by a fraction of a turn between the two exposures, and the resulting fringe pattern indicates how the piece of timber has been strained. Note the stress concentration immediately below the point of contact of the clamp.
(P.E.L. photograph)

Finally, if laser light of two wavelengths is used to record a single-exposure hologram, the result is a reconstructed image with superimposed contour fringes[62]. The effect depends on the fact that the scale of a holographic image depends on the wavelength of the light used for the reconstruction. The technique is useful for reducing the sensitivity of holographic interferometry.

8.6. Miscellaneous applications of holography

Apart from the uses discussed above, holographic techniques are being applied to an ever widening range of activities. More sophisticated methods of pattern recognition (see section 6.3) have been developed using holography, and with the increasing availability of lasers in new wavelength ranges, interest is spreading to infrared and X-ray holography. Work is also proceeding on the use of holography to increase the resolution of the electron microscope, the original aim of the technique. Acoustic holography, too, is becoming a very valuable tool in non-destructive testing; in this technique sound waves are used to produce the hologram, which is reconstructed optically. Finally, mention should be made of the work going on in the entertainment field. Holographic cinema and holographic television are both under investigation, and although present signs are not very encouraging, it may well happen that before long three-dimensional colour television will be with us, and we shall be treated to miniature but realistic Coronation Streets being reconstructed in our living rooms!

9. CONCLUSIONS

It is hoped that this article has given some indication of the impact that the laser has had on technology in the first 12 years of its life. A review of this nature must of necessity be rather cursory, but further details of any particular application can be found in the cited literature. The author will have achieved his purpose if he has succeeded in making engineers and industrialists aware of the potential of the laser, and he hopes that they will be prompted to investigate the possibilities further. Lasers are now much cheaper and much more reliable than they were a few years ago, and will in general give trouble-free service with lives of up to 10 000 hours operation or more. They are no longer something to be afraid of.

10. ACKNOWLEDGMENT

The author wishes to thank Dr M. C. Probine, director of the Physics and Engineering Laboratory, D.S.I.R., for permission to publish this paper.

11. REFERENCES

[1] MAIMAN, T. H. (1960): "Stimulated optical radiation in ruby", *Nature 187*, p. 493.

[2] DEVINE, S. D. (——): Industrial and engineering applications of infrared lasers (in preparation).

[3] HERMANNSFELDT, W. B. (1965): *Linear alignment techniques, Stanford Linear Acceleration Centre Publication PUB-82.*

[4] DYSON, J. (1967): "Correction for atmospheric refraction in surveying and alignment", *Nature 216*, p. 782.

[5] HARRISON, P. W., *et al* (1972): "The laser for long distance alignment—a practical assessment", *Proc. Instn. Civ. Engrs. 52* (May), p. 1.

[6] BUTTERS, J. N., and DENBY, D. (1970): "Some practical uses of laser beam photography in engineering," *J. Phot. Sci. 18*, p. 60.

[7] SINGH, M., *et al* (1971): "Measurement of angles using a gas laser", *Optik 33*, p. 457.

[8] ROWLEY, W. R. C. (1966): "Some aspects of fringe-counting in laser interferometers", *Trans I.E.E.E., IM-15*, p. 146.

[9] BOTCHERBY, S. C. L., and BARTLEY-DENNIS, G. A. (1969): "Length and velocity measurement by laser", *Opt. Technol. 1*, p. 85.

[10] NATIONAL BUREAU OF STANDARDS (1972): "Interferometer developed to measure sphere and cylinder diameters", *NBS Tech. News Bull.* (Jan.), p. 4.

[11] MINKOWITZ, S. (1968): "Industrial laser interferometry", *Opt. Spect. 2* 3, p. 64.

[12] RUDE, A. (1970): "The two-frequency laser interferometer and its application to the control and calibration of machine tools", *Electron. Ind. 138*, p. 685.

[13] BRIERS, J. D. (1972): "Interferometric testing of optical systems and components: a review", *Opt. Laser Technol. 4*, p. 28.

[14] ROBERTS, F. E., and LANGENBECK, P. (1969): "Homogeneity evaluation of very large disks", *Appl. Opt. 8*, p. 2311.

[15] WOLINSKI, W., *et al* (1969): "Continuous measurement of the diameter of thin wires and the width of small slits by laser beam diffraction", *Electron. Technol. 2* 2/3, p. 223.

[16] WATKINS, L. S., and ZALECKAS, V. J. (1970): "Dynamic wire measurements by automatically monitoring the laser diffraction pattern", *I.E.E.E. J. Quantum Elec. QE6* 1, p. 2.

[17] DEW, G. D. (1970): "The application of spatial filtering techniques to profile inspection, and an associated interference phenomenon", *Opt. Acta 17*, p. 237.

[18] ROSS, A. R. L., and WEST, P. (1972): "Fourier transform techniques in optical metrology", *Metron 4*, p. 87.

[19] REDMAN, J. D. (1967): "Optical aids to pattern recognition", *Procs. S.I.R.A. Conf. on New Developments in Optics and their Applications in Industry*, Paper 4.

[20] SCOTT, R. M. (1967): "Aspects of optical information processing", *Procs. S.I.R.A. Conf. on New Developments in Optics and their Applications in Industry*, Paper 3.

[21] ARCHBOLD, E., *et al* (1969): "Visual observation of surface vibration nodal patterns", *Nature 222*, p. 263.

[22] FERNELIUS, N., and TOME, C. (1971): "Vibration-analysis studies using changes of laser speckle", *J. Opt. Soc. Am. 61, p.* 556.

[23] ELIASSON, B., and MOTTIER, F. M. (1971): "Determination of the granular radiance distribution of a diffuser and its use for vibration analysis", *J. Opt. Soc. Am. 61*, p. 559.

[24]TIZIANI, H. J. (1971): "Application of speckling to in-plane vibration analysis", *Opt. Acta 18*, p. 891.
[25]ARCHBOLD, E., *et al* (1970): "Recording of in-plane surface displacement by double-exposure speckle photography", *Opt. Acta 17*, p. 883.
[26]BUTTERS, J. N., and LEENDERTZ, J. A. (1971): "A double exposure technique for speckle pattern interferometry", *J. Phys. E. 4*, p. 277.
[27]LEENDERTZ, J. A. (1970): "Interferometric displacement measurement on scattering surfaces utilising speckle effect", *J. Phys. E. 3*, p. 214.
[28]KOPF, U. (1972): "Ein koharent-optisches Verfahren zur beruhrungslosen Messung mechanische Verformungen und Schwingungen", *Optik 35*, p. 144.
[29]ROBERTSON, C. W. (1967): "Lasers in industrial measurement", *Procs. S.I.R.A. Conf. on New Developments in Optics and their Applications in Industry*, Paper 13.
[30]CRANE, R. B. (1970): "Use of a laser-produced speckle pattern to determine surface roughness", *J. Opt. Soc. Am. 60*, p. 1658.
[31]ARSENAULT, H. H. (1971): "Roughness determination with laser speckle", *J. Opt. Soc. Am. 61*, p. 1425.
[32]GABOR, D. (1948): "A new microscope principle", *Nature 161*, p. 777.
[33]GABOR, D. (1949): "Microscopy by reconstructed wavefronts", *Proc. Roy. Soc. A197*, p. 454.
[34]LEITH, E. N., and UPATNIEKS, J. (1963): "Wavefront reconstruction with continuous tone objects", *J. Opt. Soc. Am. 53*, p. 1377.
[35]LEITH, E. N., and UPATNIEKS, J. (1964): "Wavefront reconstruction with diffused illumination and three-dimensional objects", *J. Opt. Soc. Am. 54*, p. 1295.
[36]PENNINGTON, K. S., and LIN, L. H. (1965): "Multicolour wavefront reconstruction", *Appl. Phys. Letters 7*, p. 56.
[37]FRIESEM, A. A., and FEDOROWICZ, R. J. (1966): "Recent advances in multicolor wavefront reconstruction", *Appl. Opt. 5*, p. 1085.
[38]LEITH, E. N., and UPATNIEKS, J. (1965): "True 3-D image from laser photography", *Elec. World* (Oct.), pp. 34 and 58.
[39]FORSTER, H. (1968): "Holosigns", *Traffic Engng.* (Apr.), p. 20.
[40]VAN HEERDEN, P. J. (1963): "A new optical method of storing and retrieving information", *Appl. Opt. 2*, p. 387.
[41]THOMPSON, B. J. *et al* (1967): "Applications of hologram techniques for particle size analysis", *Appl. Opt. 6*, p. 519.
[42]SILVERMAN, B. A., *et al* (1964): "A laser fog disdrometer", *J. Appl. Met. 3*, p. 792.
[43]ROYER, H., *et al* (1971): "Visualisation des brouillards par holographie", *Opt. Commun. 4*, p. 75.
[44]WITHRINGTON, R. J. (1970): "Application of holography to bubble-chambers—a review", *Opt. Technol. 2*, p. 132.
[45]UPATNIEKS, J. (1969): "Practical holography", paper presented at the ann. mtg. of the Opt. Soc. of America.
[46]STROKE, G. W. (1969): "Image deblurring and aperture synthesis using 'a posteriori' processing by Fourier-transform holography", *Opt. Acta 16*, p. 401.
[47]STROKE, G. W. (1969): *An Introduction to Coherent Optics and Holography*, 2nd edn., p. 254.
[48]HORMAN, M. H. (1965): "An application of wavefront reconstruction to interferometry", *Appl. Opt. 4*, p. 333.
[49]COLLIER, R. J., *et al* (1965): "Application of moire techniques to holography", *Appl. Phys. Letters 7*, p. 223.
[50]HAINES, K. A., and HILDEBRAND, B. P. (1966): "Surface-deformation measurement using the wavefront reconstruction technique", *Appl. Opt. 5*, p. 595.
[51]HEFLINGER, L. O., *et al* (1966): "Holographic interferometry", *J. Appl. Phys. 37*, p. 642.
[52]GRANT, R. M., and BROWN, G. M. (1970): "Holographic nondestructive testing (HNDT) in the automotive industry", *S.A.E. Trans. 78*, p. 305.
[53]ASHTON, R. A., *et al* (1971): "Interferometric holography applied to elastic stress and surface corrosion", *Appl. Opt. 10*, p. 440.
[54]BOLSTAD, J. D. (1967): "Holograms and spatial filters processed and copied in position", *Appl. Opt. 6*, p. 170.
[55]KIRK, J. P. (1966): "Holograms on photochromic glass", *Appl. Opt. 5*, p. 1684.
[56]FUNKHOUSER, A. T., and MIELENZ, K. D. (1970): "High-speed holographic interferometry", *Appl. Opt. 9*, p. 1215.
[57]SMIGIELSKI, P. (1971): "Holographie des objets de phase. Etude de quelques methodes d'exploitation de l'onde restituee", *Opt. Acta 18*, p. 483.
[58]VAN DEELEN, W., and SISENSON, P. (1969): "Mirror blank testing by real-time holographic interferometry", *Appl. Opt. 8*, p. 951.
[59]MACGOVERN, A. J., and WYANT, J. C. (1971): "Computer generated holograms for testing optical elements", *Appl. Opt 10*, p. 619.
[60]POWELL, R. L., and STETSON, K. A. (1965): "Interferometric vibration analysis by wavefront reconstruction", *J. Opt. Soc. Am. 55*, p. 1593.
[61]HOCKLEY, B. S., and BUTTERS, J. N. (1970): "Coherent photography (holography) as an aid to engineering design", *J. Phot. Soc. 18*, p. 16.
[62]HAINES, K. A., and HILDEBRAND, B. P. (1965): "Contour generation by wavefront reconstruction", *Phys. Letters 19*, p. 10.

▽

WAVELENGTH DEPENDENCE OF INTENSITY FLUCTUATIONS IN LASER SPECKLE PATTERNS FROM BIOLOGICAL SPECIMENS

J.D. BRIERS*

Applied Optics Section, Department of Physics, Imperial College of Science and Technology, London SW7 2BZ, UK

Received 7 October 1974

Speckle patterns obtained when botanical specimens are illuminated with laser light are observed to fluctuate at a rate which depends on the wavelength of the light used. It is suggested that this wavelength dependence may be of value as an additional degree of freedom in some applications of intensity fluctuation spectroscopy.

When laser light is incident on a diffusely reflecting surface, a granular "speckle pattern" is produced [1]. This pattern may be observed either on the surface concerned, or in the far field by intercepting the light scattered from the surface. The two cases may be called "image speckle" and "far-field speckle" respectively. (In Gabor's terminology [2] they are referred to as "subjective" and "objective" speckle.)

Speckle patterns are caused by interference between rays reflected from different parts of the object. In the case of image speckle, the rays contributing to each speckle come from a very small part of the object, whereas in far-field speckle the whole illuminated surface contributes to each speckle. However, in both cases the processes involved are random and can only be analysed by statistical techniques [3,4].

When the illuminated object is a living entity such as a fruit, the speckle pattern is seen to fluctuate. This effect can be observed with both image speckle and far-field speckle. (In the former case, observation through a pinhole enhances the effect by "magnifying" the speckles [4].) The rate of fluctuation depends on the wavelength of the light used, the colour of the fruit, and the viewing direction. For example, a red tomato produces more rapid fluctuations than a green tomato when observed in the red light from a helium–neon laser (wavelength 633 nm), but the effect is reversed when the green line of an argon-ion laser (wavelength 514 nm) is used; and in both cases the fluctuations are much less rapid in the neighbourhood of the "highlight" caused by specular reflection. Fig. 1 shows typical microdensitometer traces taken across photographs of a red and a green tomato illuminated by (a) argon-ion laser light and (b) helium–neon laser light. The modulation of the traces reflects the contrast of the speckle pattern recorded on the film, and this contrast is in turn a measure of the fluctuation that has occurred in the speckle pattern during the exposure time of 15 s. The smoother the trace in fig. 1, the more rapid is the fluctuation of the speckle.

A tentative explanation of these phenomena has been formulated, based on the fact that the colour of a tomato is due to the presence of discrete, pigmented bodies (plastids) that selectively scatter light of that colour. When a red tomato is illuminated with red light, most of the light is scattered from red chromoplasts inside the cells of the tomato. When green light is used this is absorbed by the chromoplasts, and most of the light reaching an observer or a detector arises from specular reflection at the skin of the tomato. For a green tomato the situation is reversed: the green chloroplasts within the cells preferentially absorb red light and scatter green. Thus it is suggested that light from an illuminated tomato consists of two components, a component scattered from the plastids, and a quasi-specular component reflected from the skin.

* On leave from the Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Lower Hutt, New Zealand.

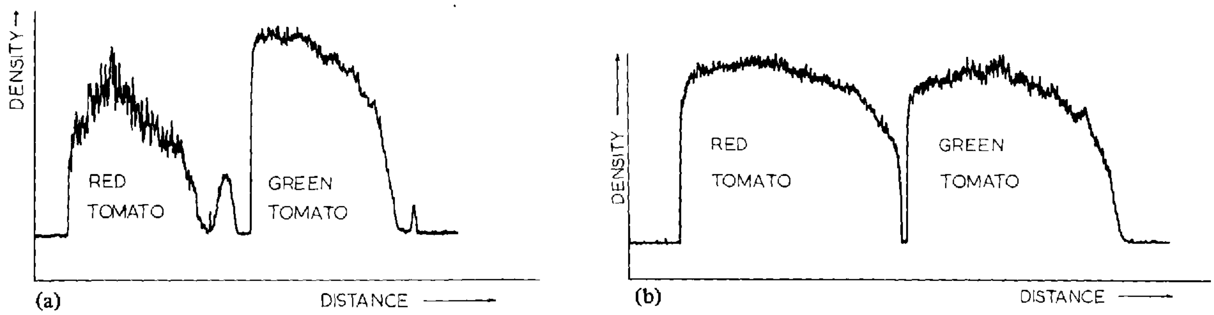


Fig. 1. Microdensitometer scans of photographs of a red and a green tomato illuminated with (a) green light (wavelength 514 nm) from an argon-ion laser, and (b) red light (wavelength 633 nm) from a helium–neon laser.

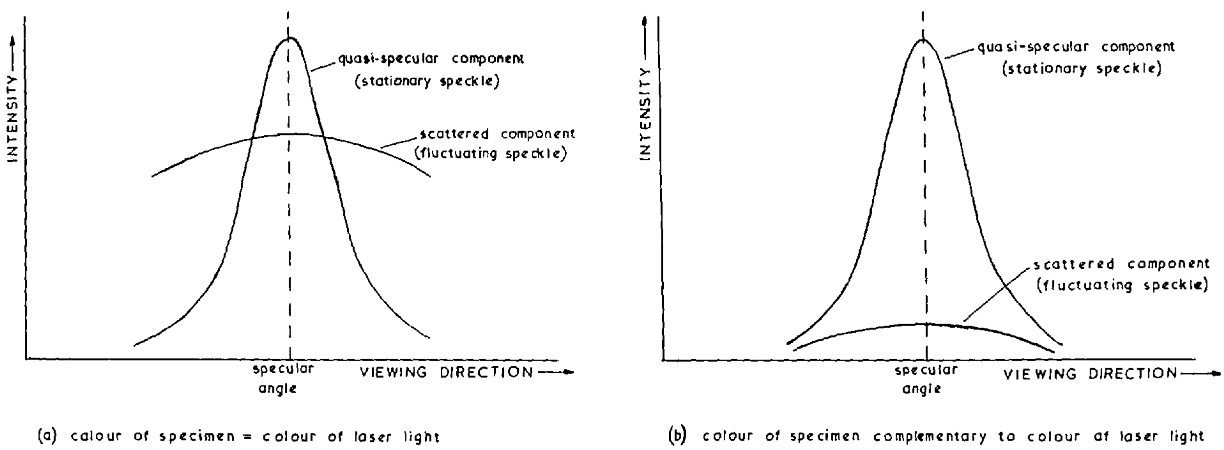


Fig. 2. Suggested explanation of the wavelength dependence.

The specular component will show much more angular dependence than the scattered component, as illustrated in fig. 2.

This model can be used to explain all the phenomena described above, if it is further assumed that the plastids are in motion and that the skin of the tomato is stationary. This would result in fluctuating speckle in the light scattered by the plastids (due to the time-variations in the optical path lengths from plastid to detector), and stationary speckle in the specular component. When the colour of the tomato is the same as the colour of the light, most of the observed speckle pattern will be due to the scattered light from the plastids, and will hence show fluctuations; in one particular direction, however, the specular component will be predominant (fig. 2a) and the fluctuation less pronounced. When the colour of the fruit is complementary to that of the laser light, the specular component, and hence stationary speckle, will always predominate (fig. 2b).

No movement of plastids has been observed during microscopic examination of tomato sections, but it can be argued that the taking of a section interferes with the life processes and might arrest any motion. In fact, motion of chloroplasts (cyclosis) has long been established for thin biological specimens which can be examined in vivo under a microscope, the classic example being canadian pondweed (*Elodea canadensis*). When a leaf of this plant, mounted in water on a microscope slide, was illuminated with the argon-ion laser beam, the far-field speckle pattern showed fluctuations whose rate varied as the beam was scanned across the leaf, the fluctuations being most rapid near the central vein of the leaf. Microscopic examination of the same leaf revealed that the chloroplasts in the cells near the central vascular tissue were much more active than those nearer the edge of the leaf. The same leaf was examined some time later, when it was found that all cyclosis had ceased; the far-field speckle pattern was again observed, and this time showed no signs of any fluctuations.

Order-of-magnitude measurements also tend to support the validity of the model. The chloroplasts of *Elodea canadensis* were observed under the microscope to be moving with velocities varying between zero and 4 $\mu m\, s^{-1}$, with a heavy bias towards the lower end of this range. A rough estimate put the mean velocity between 0.1 and 0.4 $\mu m\, s^{-1}$. The intensity fluctuations of a single speckle were observed, and the time constant of the fluctuations was of the order of 5 s: This can be associated, to an order-of-magnitude approximation, with an average movement of the scattering particles of one wavelength, and hence with an average velocity of about 0.1 $\mu m\, s^{-1}$. Similarly, the intensity envelope of the speckle pattern is determined by the size of the scattering particles, and diffraction theory gave the size in the present case as approximately 8 μm. Microscopic measurements gave the diameters of the chloroplasts as 4 to 6 μm.

It should be noted that there is a close connection between the phenomenon of fluctuating speckle and the techniques of intensity fluctuation spectroscopy which have come into prominence in recent years [5]. Although the approach of workers in that field is somewhat different, being based usually on the statistics of photon-counting rather than those of speckle patterns, the same phenomenon is essentially being observed in both cases. Intensity fluctuation spectroscopy has, in fact, been used for several biological applications [5]. Hence the ideal method of extending the present studies would be to use an autocorrelator of the type described by Pike and Jakeman [6] to analyse the statistics of the time-varying speckle patterns. If the interpretation outlined in this paper is correct, a method would then be available of monitoring cell activity in vivo.

The experiments described and proposed in this paper provide another potential application for the techniques of intensity fluctuation spectroscopy. However, attention is particularly drawn to the possibility of using different wavelengths of laser light to separate out motion of coloured particles such as plastids; it is suggested that this extra degree of freedom might also be useful in other applications of intensity fluctuation spectroscopy.

The author would like to record his gratitude to Dr. N. Abramson (Royal Institute of Technology, Stockholm), whose chance remark, that "when an apple is illuminated with laser light the speckles move!", initiated this investigation. Valuable discussions with Professor W.T. Welford and other members of the Applied Optics Section, and with Dr. J.W. Hannay of the Department of Botany and Plant Technology at Imperial College, are also gratefully acknowledged. The work is financed by a Study Award from the Department of Scientific and Industrial Research, New Zealand, and this paper is published with the permission of Dr. M.C. Probine, Director of the Physics and Engineering Laboratory, Lower Hutt, New Zealand.

References

[1] J.D. Rigden and E.I. Gordon, Proc. IRE 50 (1962) 2367.
[2] D. Gabor, IBM J. Res. Dev. 14 (1970) 509.
[3] J.W. Goodman, Stanford Electron. Lab. Tech. Rep. No. 2303–1 (1963).
[4] J.C. Dainty, Opt. Acta 17 (1970) 761.
[5] H.Z. Cummins and E.R. Pike (Eds.), NATO Adv. Study Inst. Series B: Physics, Vol. 3 (1974).
[6] E.R. Pike and E. Jakeman, Adv. in Qu. Electron. 2, Ed. D. Goodwin (Academic Press, 1973).

Short Communication

A note on the statistics of laser speckle patterns added to coherent and incoherent uniform background fields, and a possible application for the case of incoherent addition

1. Coherent addition of a speckle pattern and a uniform background

The statistics of the light field resulting from the coherent addition of a speckle pattern and a uniform background have been investigated by Dainty [1]. Using a result previously obtained by Goodman [2], Dainty gave the following expression for the standard deviation, σ, of the resultant pattern:

$$\frac{\sigma}{\langle I\rangle} = \frac{(2r+1)^{\frac{1}{2}}}{r+1} \tag{1}$$

where $\langle I\rangle$ = the average total intensity of the combined beams, r = the ratio of the intensity of the background beam to the mean intensity of the speckle pattern and $=I_D/\langle I_N\rangle$ in Dainty's notation.

When discussing the statistics of speckle patterns, it is more usual to talk in terms of the variance, σ^2, rather than the standard deviation, and to use the ratio $\sigma^2/\langle I\rangle^2$ as a measure of the contrast in the pattern. It is well known [3] that a normal speckle pattern has a negative exponential intensity probability distribution, and hence that $\sigma^2/\langle I\rangle^2 = 1$.

Squaring Equation 1 gives the following expression for the variance of the pattern resulting from the coherent addition of a speckle pattern and a uniform background:

$$\frac{\sigma^2}{\langle I\rangle^2} = \frac{2r+1}{(r+1)^2}\ . \tag{2}$$

We now propose a simpler and more convenient form of this expression. This is obtained merely by using the ratio of the intensity of the background beam to the average *total* intensity of the combined beams as the parameter, instead of the ratio used by Dainty. Using Dainty's notation, we define

$$\rho = I_D/\langle I\rangle = I_D/(I_D + \langle I_N\rangle) \tag{3}$$

and we shall use this parameter instead of $r = I_D/\langle I_N\rangle$ as used by Dainty.

Making this change results in the following expression for the variance:

$$\frac{\sigma^2}{\langle I\rangle^2} = 1 - \rho^{2*}\ . \tag{4}$$

This should be compared with the rather more complex Equation 2. Equation 4 has the additional advantage that it is more convenient to invert to give ρ (and hence the ratio of intensities of the speckle pattern and the background beam) in terms of the variance:

$$\rho = \left(1 - \frac{\sigma^2}{\langle I\rangle^2}\right)^{\frac{1}{2}}\ . \tag{5}$$

2. Incoherent addition of a speckle pattern and a uniform background

The case of a uniform background intensity being added *incoherently* to a speckle pattern can best be regarded from a purely physical viewpoint. It is intuitively obvious that the effect is merely to add a constant intensity I_D to the whole area of the speckle pattern, and hence that the variance σ^2 will remain unchanged while the mean intensity of the combined pattern will be increased by an amount I_D. Since for the original speckle pattern alone $\sigma^2 = \langle I_N\rangle^2$, it follows that the variance for the combined field is given by:

$$\frac{\sigma^2}{\langle I\rangle^2} = \frac{\sigma^2}{(\langle I_N\rangle + I_D)^2} = \frac{\langle I_N\rangle^2}{(\langle I_N\rangle + I_D)^2}\ . \tag{6}$$

Substituting $\rho = I_D/(\langle I_N\rangle + I_D)$ as before, this becomes:

$$\frac{\sigma^2}{\langle I\rangle^2} = (1-\rho)^2\ .\dagger \tag{7}$$

*An equivalent form of equation 4 has been used recently by Pedersen [8].

†Equation 7 can also be derived analytically from the expression given by Burch [5] for the probability distribution of intensity in the combined field. In Dainty's notation this expression takes the form:

$$P(I) = \frac{1}{\langle I_N\rangle}\exp\left[-\left(\frac{I - I_D}{\langle I_N\rangle}\right)\right]\ .$$

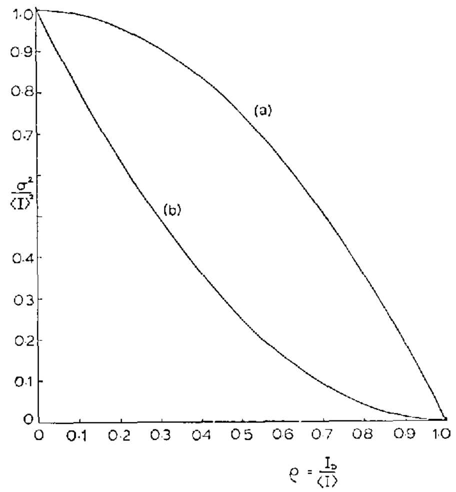


Figure 1 Variation of $\sigma^2/\langle I \rangle^2$ with ρ for addition of speckle pattern and uniform background. (a) Coherent addition; (b) Incoherent addition.

This should be compared with the coherent case given by Equation 4:

$$\frac{\sigma^2}{\langle I \rangle^2} = 1 - \rho^2 \ .$$

These two expressions are plotted for comparison in Fig. 1.

3. An application of the incoherent case: moving and stationary scatterers

Consider a scattering medium consisting of a mixture of stationary and moving scatterers, and illuminated with laser light. Assume that the motion of the moving scatterers is random. The resulting far-field speckle pattern produced by the light scattered by the medium can be considered as consisting of two independent components, one from the stationary scatterers and one from the moving scatterers. At any instant in time these two speckle patterns will add *coherently* to produce a third speckle pattern with normal speckle statistics (negative exponential distribution of intensities, $\sigma = \langle I \rangle$, etc.) [3]. Let the mean intensities of the individual speckle patterns from the stationary and moving scatterers be $\langle I_S \rangle$ and $\langle I_M \rangle$ respectively, and let their complex amplitudes at a given time and at a given point in the far field be A_S and A_M respectively. Since the two beams combine *coherently*, we must add complex amplitudes rather than intensities in order to determine the resultant intensity. Thus, at the given point in the far field, the instantaneous intensity is given by:

$$I(x, y, t) = |A_S + A_M|^2$$

where x, y designates a point in the far field and t designates a particular time.

The time-averaged intensity at the given point, such as would be obtained by exposing a photographic plate to the far-field speckle pattern for a time which is long compared with the time taken for the slowest moving scatterer to move through a distance equal to one wavelength of the laser light used, is given by:

$$\langle I(x, y) \rangle_t = \langle |A_S + A_M|^2 \rangle_t$$

where $\langle \rangle_t$, denotes time-average.

Expanding this expression, and noting that $\langle A_M^* \rangle_t = \langle A_M \rangle_t = 0$, we obtain:

$$\langle I(x, y) \rangle_t = \langle I_S(x, y) + \langle I_M(x, y) \rangle_t \ .$$

Since the addition of *intensities* rather than amplitudes is involved, it follows that the time-averaged speckle pattern is produced by the *incoherent* addition of the speckle patterns from the stationary and the moving scatterers. Further, since $I_M(x, y)$ is varying randomly over all permissible values (providing the exposure or integration time of the detector is long enough), it is apparent that

$$\begin{aligned} \langle I_M(x, y) \rangle_t &= \langle I_M(t) \rangle_{x,y} \quad (\langle \rangle_{x,y} = \text{spatial average}) \\ &= \langle I_M \rangle \quad \text{(the ensemble average)} \\ &= \text{constant at all points in the far field.} \end{aligned}$$

Hence the resultant time-averaged speckle pattern is equivalent to the incoherent addition of the speckle pattern due to the stationary scatterers (mean intensity $\langle I_S(x, y) \rangle_{x,y}$) and a uniform background intensity $\langle I_M \rangle$. (This result is also used in speckle interferometry [4]).

From Equation 7, therefore, the variance of the time-averaged speckle pattern is given by:

$$\frac{\sigma^2}{\langle I \rangle^2} = (1 - \rho)^2 \ .$$

Thus the ratio, ρ, of the mean intensity of the light from the moving scatterers to the total intensity of the scattered light, is given in terms of the contrast of the time-averaged resultant speckle pattern by the following expression:

$$\rho = 1 - \frac{\sigma}{\langle I \rangle} \cdot \qquad (8)$$

If it can further be assumed that the moving and stationary scatterers are identical in scattering properties and that all the light scattered into the far field is scattered by these particles, then ρ also represents the ratio of the *number* of moving scatterers to the total number of scatterers involved. Even if these last two assumptions are not valid, comparative measurements of ρ made on the same or similar systems will give an indication of how the relative number of moving scatterers varies from system to system, or from time to time in the same system. Information about the velocity distribution of the moving scatterers could also be obtained, by taking measurements at different exposure or integration times. This latter investigation, of course, would be much more easily achieved by using a digital autocorrelator of the type described by Pike and Jakeman [6], but if such an instrument is not available the proposed method might provide some useful information about the number of moving scatterers in a sample, even if only on a comparative or semi-quantitative level. The method might be of use, for example, with the fluctuating speckle patterns observed when laser light is scattered from biological specimens; the fluctuations are believed to be due, at least in part, to the motion of discrete bodies [7], and it should be possible to detect changes in the level of activity of these bodies.

Acknowledgements
The author would like to thank Professor W. T. Welford, Dr T. S. McKechnie and G. Parry for useful discussions. The work is financed by a Study Award from the Department of Scientific and Industrial Research, New Zealand, and the paper is published with the permission of Dr M. C. Probine, Director of the Physics and Engineering Laboratory, Lower Hutt, New Zealand.

References
1. J. C. DAINTY, *J. Opt. Soc. Am.* **62** (1972) 595-596.
2. J. W. GOODMAN, *ibid* **57** (1967) 493-502.
3. J. C. DAINTY, *Opt. Acta* **17** (1970) 761-772.
4. E. ARCHBOLD, A. E. ENNOS and P. A. TAYLOR, in Optical Instruments and Techniques (ed. J. HOME DICKSON), Oriel Press (1970) 265-275.
5. J. M. BURCH, *ibid*, 213-229.
6. E. R. PIKE and E. JAKEMAN, 'Advances in Quantum Electronics', Vol. 2, (ed. D. GOODWIN), Academic Press (1973).
7. J. D. BRIERS, *Opt. Comm.* **13** (1975) 324-326.
8. H. M. PEDERSEN, *ibid*, **12** (1974) 156–159.

Received 11 April 1975

J. D. BRIERS*
Optics Section
Department of Physics
Imperial College
London SW7

*On leave from the Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Lower Hutt, New Zealand.