

An Introduction to
OPTICS OF COHERENT AND NON-COHERENT ELECTROMAGNETIC RADIATIONS*

GEORGE W. STROKE
Professor of Electrical Engineering
Head, Electro-Optical Sciences Laboratory

The University of Michigan
Ann Arbor, Michigan

March 1965

- * Any use, reproduction or adaptation of any material in these lecture notes must bear the complete quotation of this cover page. These notes (like the May 1964 notes) contain a substantial amount of previously unpublished and therefore original scientific material and results.

The scientific results, theoretical and experimental, in particular those related to holography, were obtained in part under Grant GK-141 from the National Science Foundation.

I. INTRODUCTION

I.1. Emergence of Modern Optics as a Branch of Electrical Engineering.

Many of the most dramatic advances in the field of optics in the last decade or two were directly stimulated or originated by advances in electrical engineering, in its various branches of communication sciences, microwave electronics and radio-astronomy. The operational Fourier transform treatment of optical image forming processes and of spectroscopy, the introduction of resonant structures and of optical feedback control, the remarkable simplicity of optical computing, communication systems, and coherent-background (heterodyne) detection, the exploitation of the statistical and coherence properties of electromagnetic signals and radiations, as well as polarization in interferometry and astronomy, the dramatic development of light amplification and control in optical masers, and more recently, the newly dramatic achievements in "lensless" photography and "automatic" character recognition, and non-linear optics, are some of the more well known examples of the interdependence of theory and techniques throughout broad ranges of the electromagnetic domain, in astronomy, radio-astronomy, physics and electrical engineering. Skillful recognition and exploitation of basic similarities in pursuits throughout the entire electromagnetic domain is proving most fruitful in pinpointing new areas of research and of industrial applications in what may well be called the new field of "electro-optical science and engineering."

I.2. Mathematical Character of Electro-Optical Engineering.

Perhaps the single most important element in the rapid development of electro-optical sciences is the great experimental simplicity resulting from the deliberate use of sophisticated mathematical formulation.

To paraphrase C. H. Townes⁽¹⁾, one may say that the recent dramatic developments in electro-optical science, including the maser, "epitomize the great change that has recently come over the character of technological frontiers." The maser, non-linear optics, optical computers, interferometric gratings, lensless photography, optical filters and automatic "reading" systems, to mention only a few, were predicted and worked out "almost entirely on the basis of theoretical ideas of a rather complex and abstract nature." These are not inventions or developments "which could grow out of a basement workshop, or solely from the Edisonian approach of intuitive trial and error." They are rather creatures of our present scientific age which have come almost entirely from modern theory in physics, communication sciences and indeed in electro-optical engineering.

(1) C. H. TOWNES, "Masers" in The Age of Electronics, edited by Carl F. J. Overhage, McGraw Hill Book Co., Inc., New York (1962) p. 166.

I.3. Mathematical Methods of Modern Optics

Except for the solution of boundary-value problems, and in non-linear optics, where electromagnetic theory is basic, and in the study of the basic physics of radiations, where quantum theory and statistical theories are appropriate, a dominant part of electro-optical engineering is based on exploiting the simplicity of an operational, generally Fourier-transform treatment of the problems. Fourier-transform formulation, already used by Lord Rayleigh and A. A. Michelson near the turn of the 20th century, and more recently the theory of distribution, based on the work of Laurent Schwartz (1950-51) appear as uniquely powerful tools, not only in the analysis of more or less classical image-forming and communication systems, but indeed the prediction and synthesis of new devices and systems. The matrix formulation of image formation by lenses and mirrors has reduced lens-computation to very great mathematical simplicity, especially when used with electronic computers. Indeed, optical analogue correlators and computers, born out of the new mathematical formulations, have started to complement the sometimes much more complex digital computers. Optical computers are marked by simplicity and compactness, and have particular success because of their multi-channel, two-dimensional spatial capability (as compared to the single, time, dimension in purely electronic computers).

A list of the readily available mathematical references is given at the end of this chapter.

I.4. Some limitations of Operational Formulation of Optical Image Formation and the Need for Boundary-Value Solutions in the Study of "diffraction" of Electromagnetic Waves in Optics.

The emphasis placed on operational treatment of image-formation and communications systems in optics also calls for a clear understanding of its limitations, and indeed of domains where it is inapplicable.

For instance, the recent advances in attaining unprecedented high efficiencies in optical diffraction gratings⁽²⁾ were based on recognizing that polarization, and consequently, electromagnetic theory (rather than extensions of elementary scalar "diffraction theory") are basic in determining the energy distribution among the waves of various diffracted orders (or "modes"). Indeed, a complete, exact solution of the electro-magnetic boundary-value problem of optical gratings (added to the very few electromagnetic boundary-value solutions in existence) has been recently obtained in a series of papers by Stroke, Bousquet, Petit and Hadni⁽²⁾, based on the method of solution given by Stroke in 1960⁽³⁾.

The most important limitation in the use of "Huygens' principle", as given by its Fourier-transform expression, is generally not serious in most cases. The Fourier-transform relation between the complex amplitude distribution in the wave front and the complex amplitude at any given point in the image applies only to the near-vicinity of the center of the (aperture-limited) quasi-spherical image-forming

(2) See for example, G. W. STROKE, "Effects of Polarization on the Energy Distribution and Efficiency of Optical Diffraction Gratings", J. Opt. Soc. Am. 54, 846 (1964).

(3) G. W. STROKE, Rev. Optique, 39, 291-398 (1960).

wave front, for instance near the focus of a lens, although regardless of whether this be a principal or a secondary focus. The domain of applicability of "Huygens' principle" and of the Fourier-transform expression of image formation can be made quite obvious when the Fourier-transform expression is derived from Maxwell's equations⁽⁴⁾⁽⁵⁾.

However, it is not true, as sometimes assumed, that the scalar formulation of diffraction will always provide at least a qualitative description of optical diffraction. For instance, in the case of the grating illustrated in Figure 1.a, it is found on the basis of electromagnetic theory⁽³⁾ that the grating illustrated will behave like a perfect mirror, reflecting only a single wave back along the \bar{k} direction, showing none of the "classical" diffraction into any other side-orders, regardless of the narrowness of the horizontal facet (parallel to \bar{E}). Experiments with 3cm microwaves in the E_{\perp} polarization, shown in Figure 1.b (page 8) confirm this classically unexplainable, but electromagnetically easily foreseen prediction, made by Maréchal and Stroke in 1959⁽³⁾.

The basically electromagnetic nature of diffraction should be constantly born in mind when systematic inconsistencies between experiment and "theory" are encountered in the use of the otherwise so powerful Fourier-transform formulation of "diffraction" and image formation in optics.

-
- (4) A. MARECHAL and M. FRANÇON, "Diffraction", Revue d'Optique, Paris (1960).
- (5) G. W. STROKE, "Diffraction Gratings", in Handbuch der Physik, Vol 29, Edited by S. Flügge (Springer Verlag, Berlin and Heidelberg, in print).

In the most general sense, diffraction of electromagnetic waves results simply from the requirement that easily specified boundary conditions must be satisfied at the boundary of the "diffracting object" by the total field, incident plus diffracted.

Frequently, the great simplicity in the actual solution of a diffraction problem will result from writing the diffracted field as an integral sum of plane waves, with the direction cosines as parameters for the waves of different amplitudes. An important example is given in the following section, for the case of optical gratings.

The theory of gratings and interferometry is basic to much of modern optical image-formation theory⁽⁴⁾⁽⁵⁾.

I.5. Grating Equation, as Example of Boundary-Value Solution of a Diffraction Problem.⁽⁵⁾

Perhaps the most well-known equation describing the diffraction of light by gratings is the so called "grating equation"

$$\sin i + \sin i' = m \lambda / a \quad (\text{I.5.1})$$

where i and i' are the angles formed by the incident and the diffracted wavefronts with the mean grating surface, and a is the grating spacing constant.

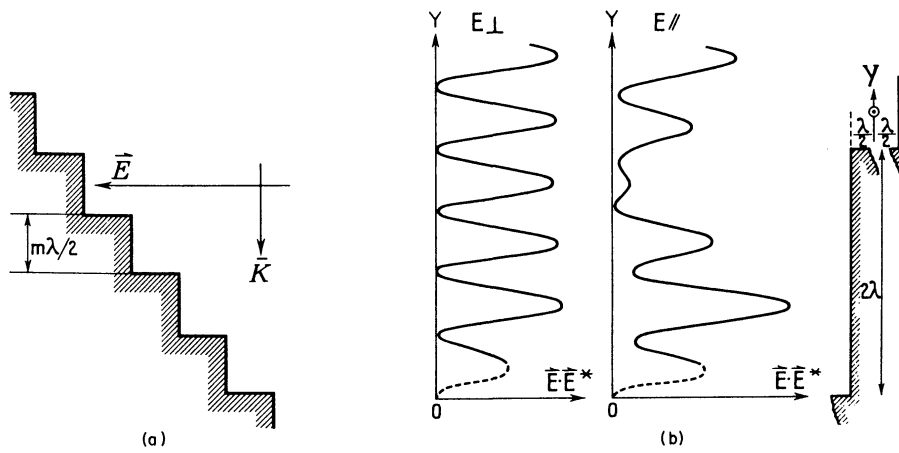
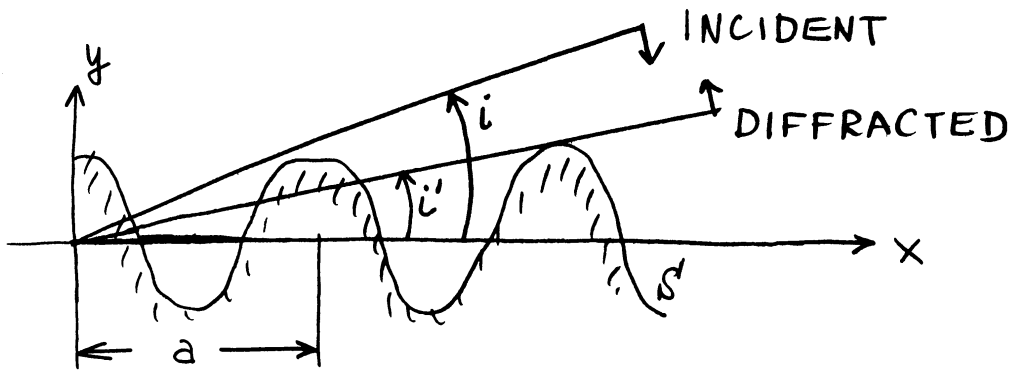


Fig.1.- Mirror-like behaviour of grating for indicated polarization, according Maréchal and Stroke (1959). Actual experimental recordings of Hertz-like standing waves in 3-cm microwave domain are shown in (b).



Parameters used in grating equation

$$\sin i + \sin i' = m \lambda / a$$

The derivation usually found in textbooks with the help of Huygens' principle is straightforward. However, the Huygens' solution makes it appear, incorrectly that the plane diffracted waves are "built up", as it were by "envelope" of little "wavelets".

More generally, it has been the practice to assume that, when a plane wave is incident on a plane grating, the diffracted waves are also plane and exist only in a discrete set. In fact, the existence of plane diffracted waves is merely the result of the periodic nature of the grating.

The proof of the existence of a discrete set of plane diffracted waves, satisfying the grating equation I.5.1 when produced by the incidence of a plane wave on a grating, follows.

Ruled gratings are essentially two-dimensional structures. As such, their surface can be described by a function $S = f(x, y)$, [Fig. p. 8], which is independent of one of the coordinates, say Z , but is periodic as a function of X ,

$$S_{x+pa} = S_x \quad (p=\text{integer}) \quad (\text{I.5.2})$$

The groove length is along Z . One next recalls the existence of the class of two-dimensional problems (here, independent of Z). Two-dimensional problems are essentially scalar in nature. However, this does not mean that they can be identified to non-electromagnetic problems (for example, to acoustical problems). Rather,

two-dimensional problems can be expressed in terms of only a single dependent electromagnetic field variable (for example E_z or H_z). It can be further shown that E_z or H_z , in these problems, satisfy the wave equation (written for E_z , as example):

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} + k^2 E_z = 0 \quad (\text{I.5.3})$$

The factor $\exp(-i\omega t)$ is implicit.

A fundamental elementary solution of (I.5.3) can be written as

$$\exp [ik(x\sin\theta + y\cos\theta)] \quad (\text{I.5.4})$$

which is the equation of a plane wave. $k = 2\pi/\lambda$. When θ is real, equation (I.5.4) represents an inhomogeneous or evanescent wave⁽⁵⁾⁽⁶⁾.

A complete solution of equation (I.5.3) is formed by an angular spectrum of plane waves, and can be represented by a Fourier integral

$$\int_{-\infty}^{+\infty} E(\theta) \exp [ik(x\sin\theta + y\cos\theta)] d\theta \quad (\text{I.5.5})$$

(6) For general background, see for example, P.C. CLEMMOW, "Rigorous diffraction theory" in M. Born and E. Wolf, Principles of Optics, Pergammon Press, London, New York (1959), p. 553-588.

where $E(\theta)$ are the amplitudes of the various waves. The coefficients $E(\theta)$ are generally complex, and are to be determined, in order to solve any special problem described by equation (I.5.5).

In the case of optical gratings, it is necessary to investigate the diffraction of polarized waves, and either \bar{E} or \bar{H} in the incident plane wave are chosen to be parallel to the groove length, along Z . Let E_z^i or H_z^i be the components of the incident field for the two cases. One has

$$E_z^i = \exp[-ik(x\sin i + y\cos i)] \quad (\text{I.5.6})$$

(The i in front of the k is clearly equal to $\sqrt{-1}$). Let E_z^d be the diffracted field. Since the total field, $E_z = E_z^i + E_z^d$ satisfies the wave equation, and since the incident field satisfies the same equation, the diffracted field E_z^d must also satisfy the wave equation. In its most general form, the diffracted field E_z^d can be represented by its angular spectrum of plane waves:

$$E_z^d = \int_{-\infty}^{+\infty} E_z^d(i') \exp[ik(x\sin i' + y\cos i')] di' \quad (\text{I.5.7})$$

Here, $E_z^d(i')$ are the amplitudes of the plane "diffracted" waves corresponding to the angles i' . According to equation (I.5.7) there exists an infinity of diffracted waves, distributed in continuous angular directions.

We shall next show that the periodic nature of the grating boundary, and the fact that the boundary condition on the surface of the grating must be satisfied by the total field \bar{E}_z (incident plus diffracted), restrict the continuous angular distribution of diffracted waves to only a discrete set of waves, satisfying the grating equation (I.5.1).

Indeed, one must have,

$$\left(E_z\right)_{x+pa} = \left(E_z\right)_x \quad (\text{I.5.8})$$

on the grating surface. Equation (I.5.6) and I.5.7), when introduced into (I.5.8) give

$$\begin{aligned} & \exp[-ik(x\sin i + y\cos i)] \exp[-ikpasini] \\ & + \int_{-\infty}^{+\infty} E_z^d(i') \exp[ik(x\sin i' + y\cos i')] \exp[ikpasini'] di' \\ & = \exp[-ik(x\sin i + y\cos i)] + \int_{-\infty}^{+\infty} E_z^d(i') \exp[ik(x\sin i' + y\cos i')] di' \end{aligned} \quad (\text{I.5.9})$$

After division by $\exp[-ik(x\sin i + y\cos i)]$ and factoring of $\exp[-ikpasini]$ equation (I.5.9) gives:

$$\exp[-ikpasini] \left\{ 1 + \int_{-\infty}^{+\infty} E_z^d(i') \exp[ik(x+pa)(sini + sini')] \exp[iky(\cosi + \cosi')] \right\} di' \quad (I.5.10)$$

$$= 1 + \int_{-\infty}^{+\infty} E_z^d(i') \exp[ikx(sini + sini')] \exp[iky(\cosi + \cosi')] di'$$

For any given k, p, a , and any given angle of incidence i , the factor $\exp[-ikpasini]$ is equal to a constant. Therefore, for equation (I.5.10) to be satisfied, one must have

$$kpa(sini + sini') = n2\pi \quad (I.5.11)$$

(n = integer)

that is

$$k(sini + sini') = m \frac{2\pi}{a} \quad (I.5.12)$$

(m = integer)

Recalling that $k = 2\pi/\lambda$, equation (I.5.12) gives finally

$$\sin i + \sin i' = m \lambda / a \quad (\text{I.5.1})$$

which is indeed the grating equation. We have just shown that a plane wave

$$E_z^i = \exp[-ik(x \sin i + y \cos i)] \quad (\text{I.5.13})$$

incident on a periodic surface, having a period a , gives rise to a diffracted field E_z^d formed of a discrete set of plane waves, such that

$$E_z^d = \sum_m (E_z^d)_m \exp[ik(x \sin i'_m + y \cos i'_m)] \quad (\text{I.5.14})$$

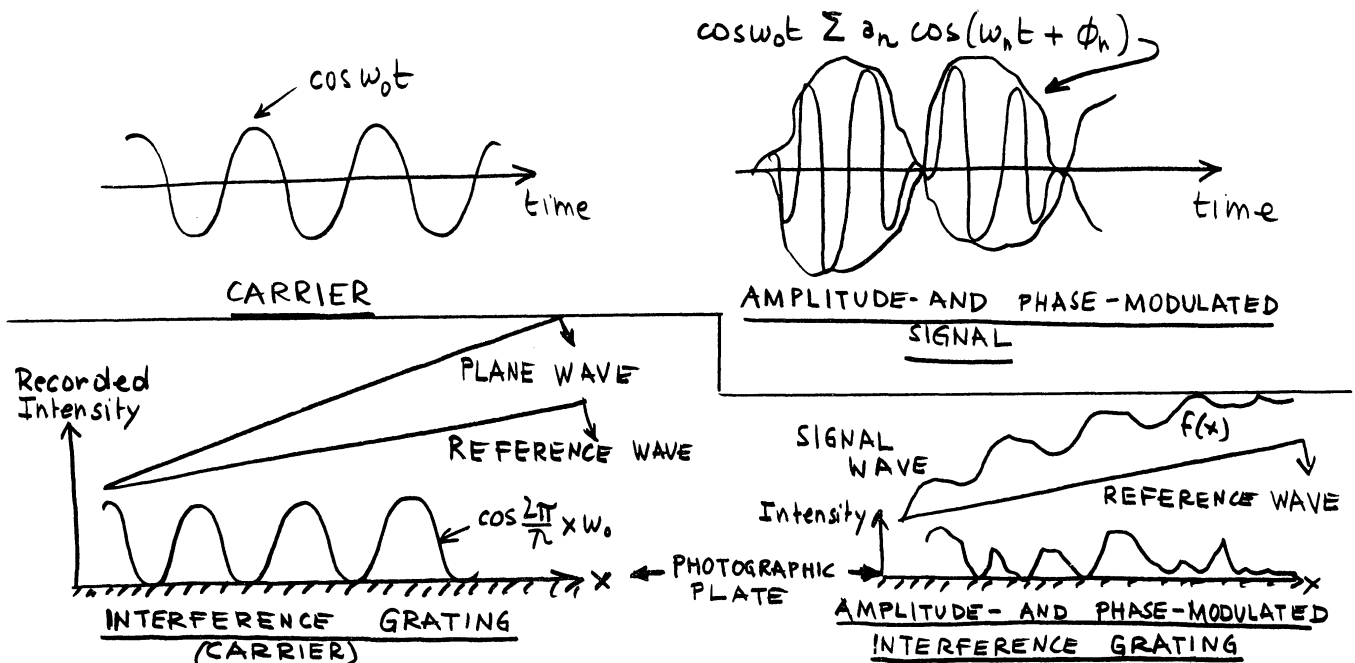
It is noted that it is truly the boundary condition and the periodicity of the grating which are at the origin of the diffraction of light by gratings. It is remarkable that the exact nature of the boundary condition does not enter the part of the solution required to demonstrate the existence of plane diffracted waves, satisfying the grating equation. Clearly, though, the amplitudes $E_z^d(i'_m)$ of the diffracted waves do depend on the exact nature of the boundary, and more particularly on its material nature

(dielectric, conductor, etc.), as well as on the groove shape. (See section I.4).

I.6. Gratings as Information-carriers in Optics.

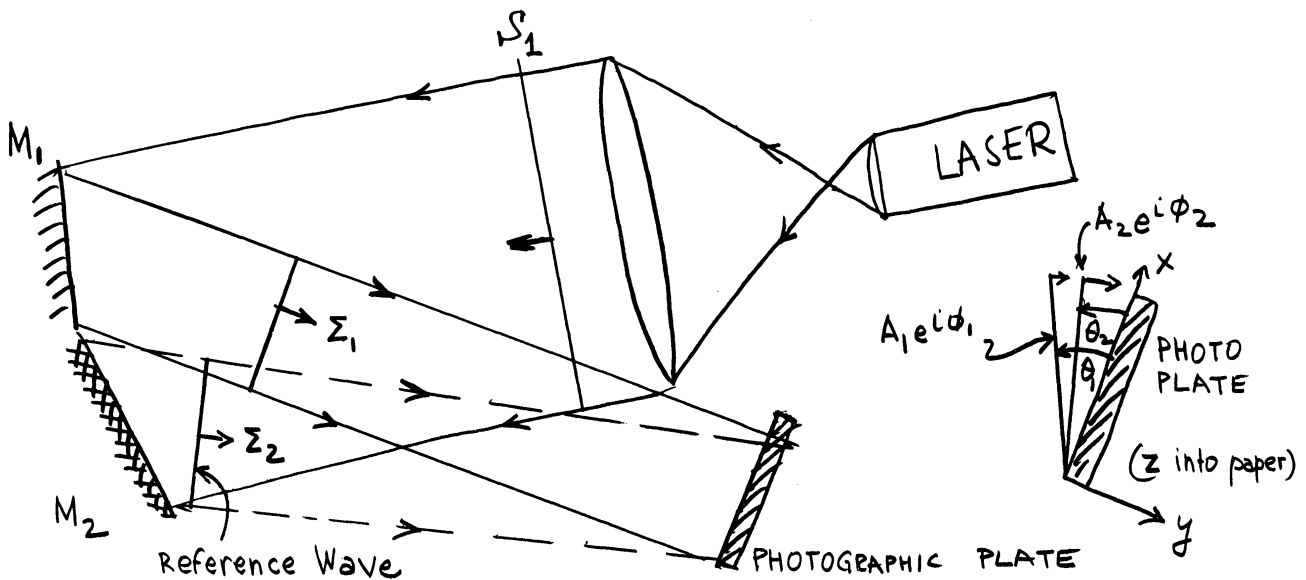
(Application to Wave-front "Reconstruction Imaging")*

Just like a sinusoidally time-varying wave is the basic carrier for time-varying signals in radio-communications, so a spatially varying grating (for instance the sinusoidal grating produced on a photographic plate by interference of two plane waves) is a basic carrier for spatially-varying images in optics. In both cases, when either space or time are the parameters, the signal (or image) involve variations of both amplitude and phase. Single or double side-band modulation can both be achieved, and the use of interference gratings as information carriers can be immediately understood with the help of elementary image formation theory.



* A more extensive treatment of the principles of optical holography (wavefront-reconstruction imaging) is given in VII

Consider first the case of two mirrors M_1 and M_2 illuminated by a plane wave S_1 . The two mirrors are inclined with respect to each other, so that the two waves $A_1 e^{i\phi_1}$ and $A_2 e^{i\phi_2}$ incident on a distant photographic plate form angles θ_1 , and θ_2 with the plate (that is, the two waves form an angle $\theta_1 - \theta_2$ with each other). Clearly the intensity recorded on the photographic plate is nothing but a sinusoidal interference grating, with straight-line fringes parallel to z . One has, with ($|A_1| = \text{constant}$, $|A_2| = \text{constant}$)



$$\begin{aligned} \bar{I}(x) &= (A_1 e^{i\phi_1} + A_2 e^{i\phi_2})(A_1^* e^{-i\phi_1} + A_2^* e^{-i\phi_2}) \\ &= |A_1|^2 + |A_2|^2 + |A_1||A_2| e^{-i(\phi_1-\phi_2)} + |A_1||A_2| e^{+i(\phi_1-\phi_2)} \end{aligned} \quad (\text{I.6.1})$$

where one recognizes

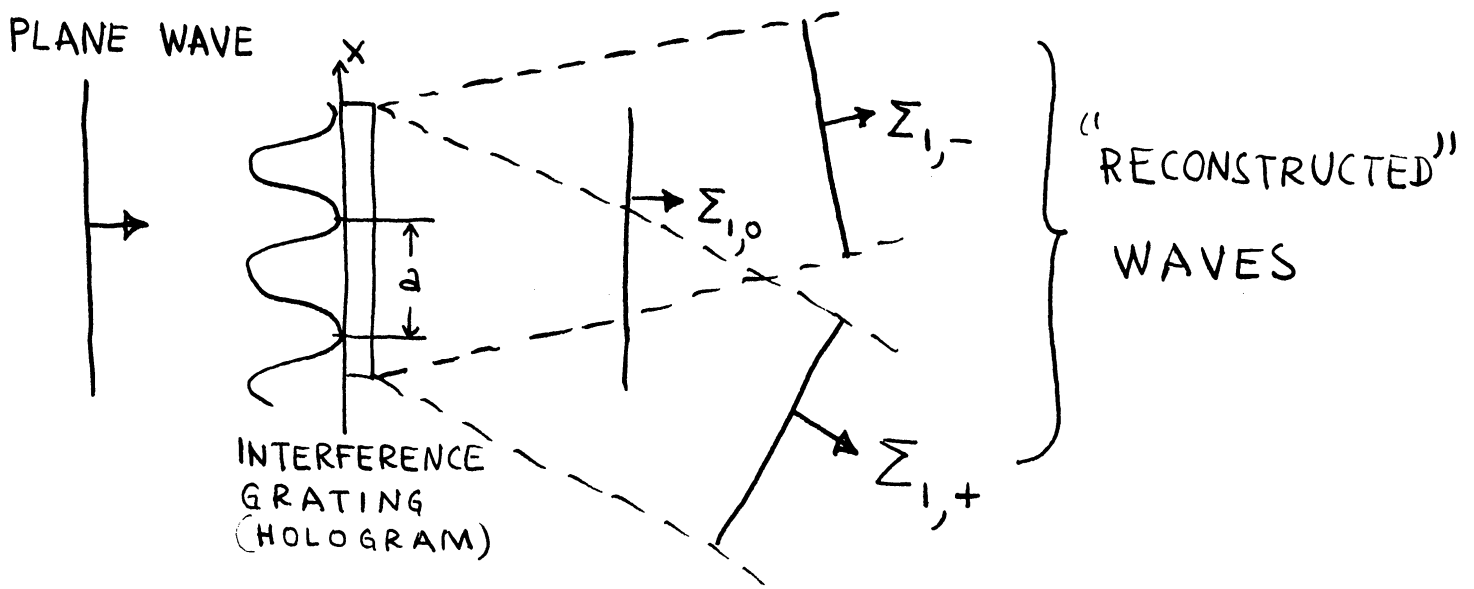
$$\bar{I}(x) = |A_1|^2 + |A_2|^2 + 2|A_1||A_2| \cos(\phi_1 - \phi_2) \quad (\text{I.6.2})$$

In this case one also notes

$$\phi_1 = \frac{2\pi}{\lambda} x \theta_1 \quad (\text{I.6.3a})$$

$$\phi_2 = \frac{2\pi}{\lambda} x \theta_2 \quad (\text{I.6.3b})$$

The period of the fringes is $a = \lambda / (\theta_1 - \theta_2)$. When the interference grating is now illuminated by a plane wave, equation (I.6.1) immediately shows that three waves will emerge from the plate: one wave on axis, and two waves forming angles $+(\theta_1 - \theta_2)$ and $-(\theta_1 - \theta_2)$ with the plate.



Indeed, one has from Fourier transform theory the equation:

$$\text{If } f(x) \rightarrow F(\omega)$$

$$\text{then } f(x)e^{2\pi i\theta_1 x} \rightarrow F(\omega - \theta_1) \quad (\text{I.6.4})$$

$$\text{and } f(x)e^{-2\pi i\theta_1 x} \rightarrow F(\omega + \theta_1)$$

where \rightarrow indicates "by Fourier transformation".

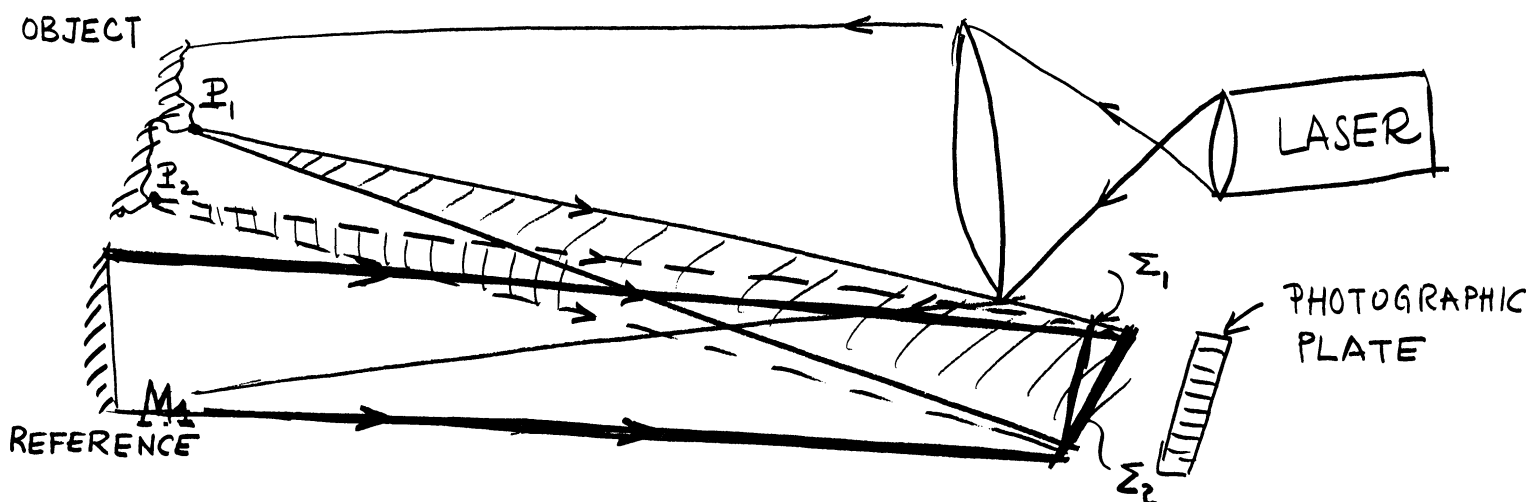
Also, equation (I.6.1) can be written as

$$I(x) = \left[|A_1|^2 + |A_2|^2 \right] + |A_1||A_2| e^{-i \frac{2\pi}{\lambda} x (\theta_1 - \theta_2)} + |A_1||A_2| e^{+i \frac{2\pi}{\lambda} x (\theta_1 - \theta_2)} \quad (\text{I.6.5})$$

where one recognizes that $\left[|A_1|^2 + |A_2|^2 \right]$ and $|A_1||A_2|$ are constants, within the aperture of the plate. It follows that the Fraunhofer "diffraction pattern" at infinity are three spectral lines, one on axis, and one each at $\pm(\theta_1 - \theta_2)$. (For a plate of finite width, the spectral lines have a $\sin\theta/\theta$ distribution.) To each of the spectral lines at infinity corresponds a plane wave at the emergence from the plate. Let one now consider one of the original waves reflected by the mirrors, and forming the interference grating, to be the "reference wave". Let Σ_2 be the reference wave. One can then say that the two waves $\Sigma_{,-}$ and $\Sigma_{,+}$ obtained by illuminating the interference grating with a plane wave are the reconstruction of the original "unknown" or "modulating" wave.

The importance of this analysis becomes clear when one examines how an arbitrary wave Σ_1 , reflected by an object of arbitrary amplitude and phase is recorded on the interference grating, as carrier.

Indeed, let Σ_1 be formed by the spherical waves originating by scattering at the various points P of the "object". In other words, Σ_1 can be taken as equal to the resultant complex amplitude produced in the plane forming an angle θ_1 , with the photographic plate, when the plate is illuminated by light scattered (or transmitted) by the object.



The intensity recorded is now given by

$$\begin{aligned}
 I(x) = & (A_1 A_m)(A_1 A_m)^* + A_2 A_2^* + [A_m^* e^{-i\phi_m}] [A_1^* A_2 e^{-i(\phi_1 - \phi_2)}] \\
 & + [A_m e^{i\phi_m}] [A_1 A_2^* e^{+i(\phi_1 - \phi_2)}]
 \end{aligned}
 \tag{I.6.6}$$

where Σ_1 has been written as a modulated plane wave

$$\Sigma_1 = A_m(x) e^{i\phi_m(x)} A_1 e^{i \frac{2\pi}{\lambda} x \theta_1}
 \tag{I.6.7}$$

As before, $|A_1| = \text{constant}$, $|A_2| = \text{constant}$ and

$$\Sigma_2 = A_2 e^{i \frac{2\pi}{\lambda} x \theta_2}
 \tag{I.6.8}$$

It is important to note that both $A_m(x)$ and $\phi_m(x)$ are now functions of x .

By comparing equation (I.6.6) to equation (I.6.1) and by noting equation (I.6.4) one immediately sees that

the modulating wave $A_m(x) e^{i\phi_m(x)}$ can now be considered as "carried" by the interference grating of the case described by equation (I.6.1). Therefore, by illuminating the plate with a plane wave, as before, one again obtains three waves a zero-angle (or zero-order) wave, carrying "no information," and two side-band waves, modulated by $A_m e^{i\phi_m}$ and by $A_m^* e^{-i\phi_m}$ respectively, and leaving the plate at angles $\pm(\theta_1 - \theta_2)$ respectively. Clearly, the side-band waves are complete "reconstructions" of the Σ_1 wave which, in turn, was formed by scattering of light by the original object. The two reconstructed waves will form a real and virtual image, identical to the original object.

The preceding scheme of wave-front reconstruction with linearly spaced interference gratings is of course, applicable to tri-dimensional objects. It has now been successfully verified in a number of laboratories* and was suggested by the author, and his associates, as a practical basis for truly image-forming x-ray microscopy.^{(10)**} The Fresnel zone interference gratings, used as carriers in Gabor's original hologram work⁽⁷⁾⁽⁸⁾⁽⁹⁾ tend to produce two reversed and overlapping images, and require other schemes for good separation of the image, such as those described in ref.(18)and(21).

-
- * See Chapter VII, as well as references (10) and (27)
- (7) D. GABOR, "A New Microscopic Principle", Nat. 161, 777 (1948)
- (8) D. GABOR, "Microscopy by Reconstructed Wavefronts", Proc. Roy. Soc. (London) A 197, 454-487 (1949)
- (9) D. GABOR, "Microscopy by Reconstructed Wavefronts II, Proc. Phy. Soc. B, 64, 449-469 (1951)
- (10) G. W. STROKE and D. G. FALCONER, "Theoretical and Experimental Foundations of Wavefront Reconstruction Imaging (Optical Holography)" in Optical Information Processing, (Symposium of Nov. 9-10, 1964), ed. by J. T. Tippett, L. C. Clapp, D. Berkowitz and C. J. Koester, (M.I.T. Press, in print).

** see also, G.W.STROKE and E.N.LEITH, private communication to the National Science Foundation(December 6,1963).

Some possibilities of optical interference gratings as carriers were explicitly considered at least as early as 1944 by P. M. Duffieux in his classical treatise⁽¹¹⁾ and again in 1958⁽¹²⁾. More recently, A. Lohmann⁽¹³⁾ had again explicitly suggested the use of the optical equivalents of single side-band modulation in the "lensless" hologram photography, originated by Gabor in 1948. (See also ref. 14, 15). A more extensive discussion of the theoretical and experimental foundations of holographic imaging can be found in ref.(10) as well as in Chapter VII.

There is of course, no need for placing the reference mirror next to the object, in actual practice, when recording the hologram, nor is it necessary to illuminate the object with a plane wave. For instance, illumination with a wave scattered by a ground glass is perfectly adequate*. The only obvious requirement in the recording is that there should exist the possibility of recording an interference grating, in case that the object and the reference beam both give rise to plane waves.

-
- (11) P. M. DUFFIEUX, "L'Intégrale de Fourier et ses Applications à l'Optique", Faculté des Sciences, Besançon (1946)
- (12) P. M. DUFFIEUX, "Les Franges d'interférences de deux systèmes d'ondes et la théorie de l'information", Rev. Opt. 37, 441-457 (1958)
- (13) A. LOHMANN, "Optische Einseitenbandübertragung angewandt auf das Gabor-Mikroskop", Optica Acta, 3, 97-99 (1956).
- (14) L. J. CUTRONA, E. N. LEITH, C. J. PALERMO, and L. PORCELLO "Optical Data Processing and Filtering Systems", IRE Trans. on Inform. Theory, IT-6, No. 3, page 386-400 (June 1960)
- (15) E. N. LEITH and J. UPATNIEKS, "Reconstructed Wavefronts and Communication Theory", J. Opt. Soc. Am. 52, 1123 (1962)

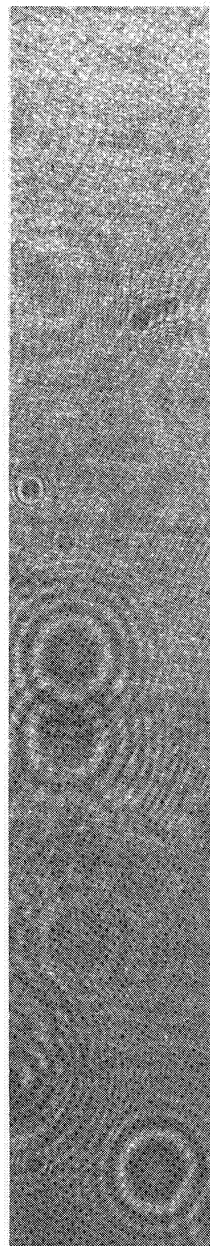
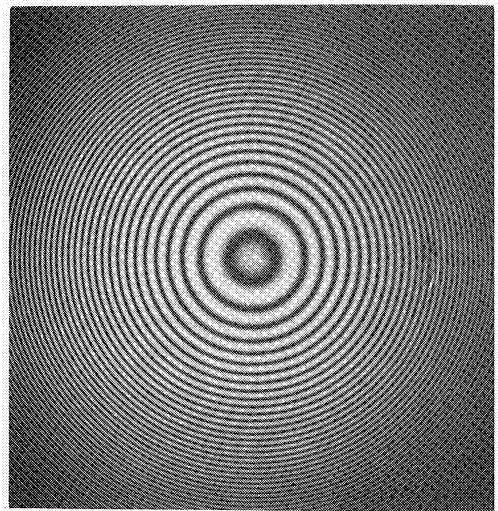
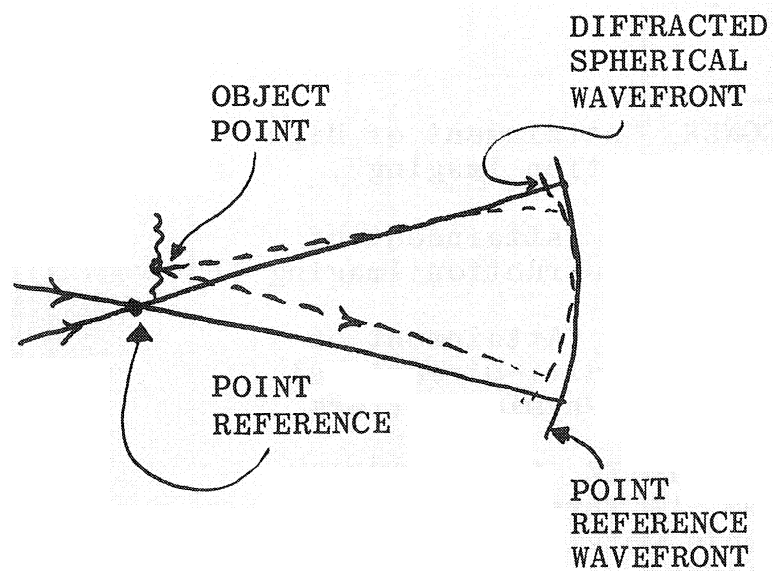
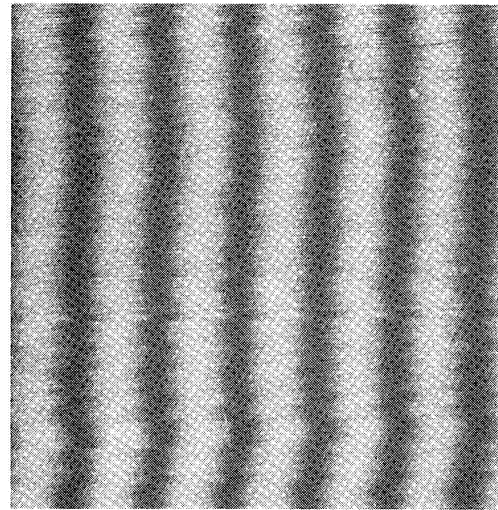
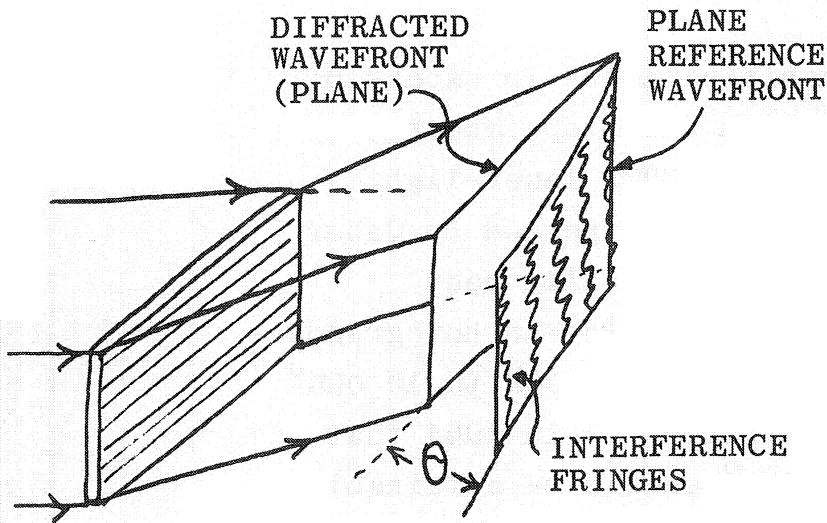
* "Diffused" or "scattered" light illumination in holography was first suggested by G.W.Stroke and explicitly described in the May 1964 edition of these LASER "Introduction to Optics of Coherent..." notes. This and other aspects of a collaborative effort between G.W.Stroke and E.N.Leith, extending from early 1962 to late 1964, were subsequently incorporated in ref.(22).

Extensive work on wavefront reconstruction, following the initial work by Gabor, was carried out by many authors, in particular by H. M. A. El-Sum^(18,19,20,21) and others (7 to 24). The "zone-plate" lens-like image forming properties of holograms were particularly noted by Rogers in 1950.⁽¹⁶⁾

-
- (16) G. ROGERS, "Gabor Diffraction Microscopy: the Hologram as a Generalized Zone-Plate", Nature 166, 273 (1950).
- (17) A. BAEZ, "A Study in Diffraction Microscopy with Special Reference to X-rays", J. Opt. Soc. Am. 42, 756 (1952).
- (18) See for instance, H. M. A. EL-SUM, "Reconstructed Wavefront Microscopy" Ph.D. Thesis, Stanford Univ. (Nov. 1952); available from University Microfilms, Inc. Ann Arbor (Dissertation Abstracts 4663, 1953).
- (19) A. V. BAEZ and H. M. A. EL-SUM, "Effect of Finite Source Size, Radiation Bandwidth and Object Transmission in Microscopy by Reconstructed Wavefronts", in X-ray Microscopy and Microradiography, Academic Press, Inc. New York (1957) p. 347-366.
- (20) See also: U. S. Patent No. 3, 083, 615, Granted April 2, 1963, (Filed May 16, 1960 by H. M. A. EL-SUM, Assignor to Lockheed Aircraft Corp.), Optical Apparatus for Making and Reconstructing Holograms.
- (21) P. KIRKPATRICK and H. M. A. EL-SUM, "Image Formation by Reconstructed Wavefronts. I. Physical Principles and Methods of Refinement", J. Opt. Soc. Am. 46, 825 (1956)
- (22) J. UPATNIEKS and E. LEITH, "Wavefront Reconstruction with Diffused Illumination and Three-Dimensional Objects", J. Opt. Soc. Am. 54, 1295 (1964)

The surprisingly large magnifications--in excess of 1 million attainable in holographic two-step imaging, in going from 1\AA x-ray construction to 6328\AA laser-light reconstructions were of course already stressed by Gabor in 1948. (7,8,9,16) However, it had generally been considered that these magnifications would--in holography--be "empty", (17) with resolutions only of 5000\AA to $10\ 000\text{\AA}$, until Stroke and Falconer demonstrated (23) in 1964 that high-resolutions on the order of 1\AA should be attainable in holographic imaging by new Fourier-transforming construction and reconstruction schemes, (24) with a generalization of the remarkable results obtained by the x-ray microscopy methods using digital computers (Kendrew) (25) or optical reconstruction (Buerger (26)).

-
- (23) G. W. STROKE and D. G. FALCONER, "Attainment of High Resolutions in Wavefront-Reconstruction Imaging", Physics Letters, 13, 306 (1964)
- (24) A) G. W. STROKE and D. G. FALCONER, "Attainment of High Resolutions in Wavefront-Reconstruction Imaging II" J. Opt. Soc. Am. 55, May 1965
B) G. W. STROKE and D. G. FALCONER, "Attainment of High Resolutions in Holography by Multi-Directional Illumination and Moving Scatters", Physics Letters, 15, 238(1965).
- (25) J. C. KENDREW, G. BODO, H. M. DINITZ, R. G. PARRISH, H. WYCKOFF and D. C. PHILLIPS, "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis", Nature 181, 662 (1958)
- (26) M. J. BUERGER, "Generalized Microscopy and the Two-Wavelength Microscope", J. Appl. Physics 21, 909-917, (1950)



HOLOGRAM

BUILDING BLOCKS FOR HOLOGRAMS *

* G.W. STROKE, "LENSLESS PHOTOGRAPHY", International Science and Technology, May 1965.

I.7. Optics and Communication Theory. Historical
Background.

Obvious similarities between the modulation and demodulation in the use of interference gratings as carriers and heterodyne methods used in communications are easily recognized. Quite basic similarities between the methods described, and the phase-contrast (coherent background) methods introduced into optics by Fritz Zernike in 1934^(27, 28, 29) should be recognized. The Nobel Prize in Physics was awarded to Zernike for this work in 1953. Indeed, many of the beginnings of this work may be traced back at least as far as Abbe (Arch. mikrosk. Anat. 9 (1873)) and Toppler (1867)⁽³⁰⁾.

It has been stressed by Duffieux (loc. cit. 1) in his now classic treatise reviving interest in Fourier transform treatment of optical processes, that there is little doubt that the modern foundations of operational methods in optical communications, image processing and spectroscopy can be found in the work of A. A. Michelson⁽³¹⁾ and of

-
- (27) F. ZERNIKE, Beugungs-theorie des Schneidenverfahrens und seiner verbesserten Form, der Phasen Kontrast Methode, Physica, Haag, 1, 43 (1934)
(28) F. ZERNIKE, "Das Phasenkontrastverfahren bei der mikroskopischen Beobachtung", Physik Z. 36, 848 (1935)
(29) F. ZERNIKE, Z. Tech. Phys. 16, 454, (1935)
(30) See H. Wolter, "Schlieren-, Phasenkontrast- und Lichtschnittverfahren," in Handbuch der Physik, Vol.24 (edited by S. Flügge), Springer Verlag, Berlin (1956) p. 555-645.
(31) A. A. MICHELSON, Phil. Mag. V, 34, 280 (1892)

Lord Rayleigh.⁽³²⁾ The attention to the use of Fourier methods in English speaking countries is largely due to a paper on "Optics and Communication Theory" published by Peter Elias⁽³³⁾ in 1953 and to the lecture notes on "Selected Topics in Optics and Communication Theory", first published by Edward L. O'Neill in 1956⁽³⁴⁾.

Extensive references and credit to the considerable work which has preceded their own papers is given by both Elias and O'Neill. A large share of the credit for stimulating modern developments in image formation theory and filtering goes to A. Maréchal^(4, 35), in optical communication theory to A. Blanc-Lapierre,⁽³⁶⁾ P. Duffieux (loc. cit. 11), and E. L. O'Neill (loc. cit.) and in spectroscopy to P. Jacquinot⁽³⁷⁾.

The foundations for the dramatic achievements in optical computing, filtering character recognition, and more generally for the new sophisticated optical materialization of communication theory principles at The University of Michigan have resulted to a large part from the work of Louis Cutrona and his associates.⁽¹⁴⁾ More recently the work by the author on interferometry and diffraction grating principles have played a major role in these advances.^(5,10,38)

(32) LORD RAYLEIGH, Phil. Mag. V., 34, 407 (1892)

(33) P. ELIAS, J. Opt. Soc. Am., 43, 229-232 (1953)

(34) E. L. O'NEILL, Opt. Res. Lab., Boston Univ. (1956)

(35) A. MARECHAL et P. CROCE, "Un Filtre de Fréquences Spatiales pour L'Amélioration du Contraste des Images Optiques", C.r.Ac.Sc. 237, 607 (1953)

(36) A. BLANC-LAPIERRE, "Considérations sur la Théorie de la Transmission de L'Information et Sur Son Application A Certains Domaines de la Physique", Ann. Inst. H. Poincaré, 13, 275 (1953).

(37) P. JACQUINOT, "New Development in Interference Spectroscopy", Reports on Progress in Physics, 23, 267-312 (1960)

(38) G. W. STROKE, "An Introduction to Optics of Coherent and Non-Coherent Electromagnetic Radiations", The Univ. of Mich., Engineering Summer Conf. on LASERS, Lecture Notes (77 pages), May 1964.

II. DIFFRACTION THEORY (Qualitative Introduction)

II.1. The Two Aspects of the Diffraction of Light

Two characters of the diffraction of light have been recognized ever since the early investigations of the properties of light in the 18th and 19th centuries (Fraunhofer, Fresnel, Young, Airy, etc):

1. The departures from straightness in the directions of propagation of waves when interacting with boundaries.

2. The formation of diffraction patterns in "image planes" when waves of finite aperture are brought to focus with the help of mirrors, lenses or other means.

More recently, in particular as a result of work of Maxwell, Lord Rayleigh, Sommerfeld, Maréchal and others, the two characters of diffraction of electromagnetic waves that need to be distinguished can be singled out as follows:

1. An electromagnetic character which has to do with the very origin of the diffraction of light by boundaries, and with the direction, polarization and amplitudes of the various waves.

2. A scalar aspect, having to do with the image-forming properties of the various diffracted waves, when they are limited in size by the aperture of the optics and limited in quality by inherent or manufacturing imperfections.

3. Moreover, the instrumental applications of optical elements (mirrors, lenses, prisms, diffraction gratings, interferometers and so on) for the formation and transformation of images, and in general for the analysis and processing of light, involve certain geometrical characteristics of optics (magnification, image-object distance equations, spectral dispersion in spectroscopic instruments, resolution, etc.)

II.2. Theoretical Calculation of Energy Distribution in Diffraction and of Spectral Diffraction Patterns

1. Electromagnetic boundary-value solutions.

Maxwell's electromagnetic equations and the boundary conditions that must be satisfied by the total field (incident plus diffracted) on the surfaces of boundaries in the electromagnetic field are sufficient to determine any electromagnetic diffraction problem in a homogeneous dielectric medium. In fact, if rigorous solutions of Maxwell's equations were easy to obtain in practice, all optical diffraction and image-formation problems would be attacked and solved with the help of these equations.

If the incident electric field vector is described by \vec{E}_i and the scattered field vector by \vec{E}_s , then one simply has for the scattered field

$$\bar{E}_s = \bar{E} - \bar{E}_i \quad (II.1)$$

where the total field vector

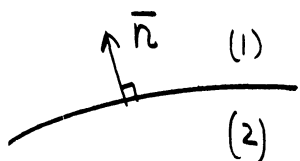
$$\bar{E} = \bar{E}_i + \bar{E}_s \quad (II.2)$$

has to satisfy appropriate boundary conditions on the diffracting surfaces or apertures and boundaries. Of course, one simple type of boundary is the infinitely extended perfectly conducting plane acting as a reflector. In this case, one of the boundary conditions is that the tangential component of the total field \bar{E} must vanish on the surface of the conductor. From this condition and Maxwell's equations, the laws of reflection of electromagnetic waves are immediately obtained.

Maxwell's equations written in differential form for free space are

$$\begin{aligned} \text{curl } \bar{E} &= -\mu_0 \frac{\partial \bar{H}}{\partial t} \\ \text{curl } \bar{H} &= \epsilon_0 \frac{\partial \bar{E}}{\partial t} \\ \text{div } \bar{E} &= 0 \\ \text{div } \bar{H} &= 0 \end{aligned} \quad (II.3)$$

To these equations one must always associate appropriate boundary conditions which are for the normal components \bar{E} and \bar{H}



Boundary Surface

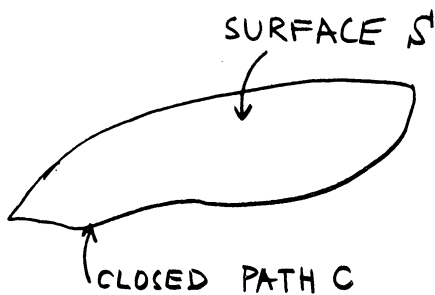
$$\begin{aligned} \bar{n} \cdot \mu_0 (\bar{H}_1 - \bar{H}_2) &= 0 \\ \bar{n} \cdot \epsilon_0 (\bar{E}_1 - \bar{E}_2) &= \sigma \end{aligned} \quad (II.4)$$

with σ = surface charge, and for the tangential components

$$\begin{aligned} \bar{n} \times (\bar{H}_1 - \bar{H}_2) &= \bar{K} \\ \bar{n} \times (\bar{E}_1 - \bar{E}_2) &= 0 \end{aligned} \quad (II.5)$$

where \bar{K} = surface current density.

When Maxwell's equations are written in integral form, the boundary conditions are implicit. One has



$$\oint_c \bar{E} \cdot d\bar{s} = - \frac{d}{dt} \int_S \mu_0 \bar{H} \cdot \bar{n} da$$

$$\oint_c \bar{H} \cdot d\bar{s} = \frac{d}{dt} \int_S \epsilon_0 \bar{E} \cdot \bar{n} da + \int_S \bar{J} \cdot \bar{n} da \quad (\text{II.6})$$

Maxwell's equations in the differential form and the associated boundary conditions are particularly suitable for the solution of the diffraction of electromagnetic waves (light) by boundaries of simple geometries, which result in separable solutions of the differential equations. This is the case of isolated cylinders, or developed cylindrical surfaces of infinite extent (for example diffraction gratings), spheres; ellipsoids, arrays of cylinders, spheres and so on. The solutions become particularly straightforward for "two-dimensional" surfaces which are a function of two coordinates only (cylinders, gratings).

However, even when they appear straightforward, rigorous solutions of electromagnetic boundary value problems have so far been obtained in only a very small number of cases (edge, slit, wedge, sphere, gratings of sinusoidal profile* etc). In reality, the mathematical complexity of electromagnetic boundary value problems is generally rather formidable, and comparable to that of boundary value problems in other domains of physics (quantum theory for example).

In general one is interested in the diffraction of waves having a simple harmonic frequency of the form

$$\bar{E}_i = (\bar{E}_i)_0 \exp(-i\omega t) \quad (\text{II.7})$$

or of a superposition of such waves. In this case, the \bar{E} and \bar{H} vectors are known to satisfy wave equations of the form

$$\nabla^2 \bar{E} + k^2 \bar{E} = 0$$

$$\nabla^2 \bar{H} + k^2 \bar{H} = 0 \quad (\text{II.8})$$

where $k = 2\pi/\lambda$ and $\omega = 2\pi f$. The frequency f and the wavelength λ are related by the equation

$$c = f\lambda \quad (\text{II.9})$$

*Recently, rigorous solutions for gratings of triangular profile have been obtained, based on the solution for the sinusoidal grating given by Stroke in 1960 (see G. W. Stroke, J. Opt. Soc. Am. 54, 846, (1964).)

where c is defined as the "velocity of light". For the case of monochromatic (single frequency) waves one has

$$\begin{aligned} \text{curl } \bar{E} &= i\omega\mu_0\bar{H} \\ \text{curl } \bar{H} &= -i\omega\epsilon_0\bar{E} \end{aligned} \quad (\text{II.10})$$

and at boundaries

$$\begin{aligned} \bar{E} &= \frac{1}{\sigma - i\omega\epsilon_0} \text{curl } \bar{H} \\ \bar{H} &= \frac{1}{i\omega\mu} \text{curl } \bar{E} \end{aligned} \quad (\text{II.11})$$

Clearly, at a perfect conductor, for which $\sigma \rightarrow \infty$ and $\bar{E}_2 \rightarrow 0, \bar{H} \rightarrow 0$ inside the conductor, one has from equation (II.5), $\bar{n} \times (\bar{E}_1 - \bar{E}_2) = 0$, the result

$$\bar{n} \times \bar{E}_2 = \bar{n} \times \bar{E}_1 = 0 \quad (\text{II.12a})$$

which is the previously stated boundary condition

$$\bar{E}_{\text{tangential}} = 0 \quad (\text{II.12b})$$

on the surface of the conductor.

The boundary condition for the \bar{H} field is particularly simple to obtain for two-dimensional surfaces which are independent of one coordinate. The method is used to illustrate some of the steps used in dealing with diffraction problems. Consider a plane polarized wave H_z^i incident along the $-y$ direction in vacuo (medium 1) on to a surface of equation $y = f(x)$ independent of z . Let H_z^i be parallel to z . By symmetry, the diffracted field will also only have a single magnetic component, H_z^d . Let the total magnetic field component in the medium 1 be $H_z = H_z^i + H_z^d$. One has for the total field in the medium 1

$$\text{curl } \bar{H} = -i\omega\epsilon_0\bar{E} \quad (\text{II.13})$$

and on the surface

$$\bar{n} \times \nabla \times \bar{H} = -i\omega \epsilon_0 \bar{n} \times \bar{E} = 0 \quad (\text{II.14})$$

from equation (II.12). By expanding $\bar{n} \times \nabla \times \bar{H}$ and noting that $H_x = H_y = 0$ and that $n_z = 0$, one obtains

$$n_x \frac{\partial H_z}{\partial x} + n_y \frac{\partial H_z}{\partial y} = 0 \quad (\text{II.15})$$

applicable to the total field. But

$$\frac{\partial H_z}{\partial n} = \bar{n} \cdot \nabla H_z = n_x \frac{\partial H_z}{\partial x} + n_y \frac{\partial H_z}{\partial y} + n_z \frac{\partial H_z}{\partial z} \quad (\text{II.16})$$

which gives the boundary condition

$$\frac{\partial H_z}{\partial n} = 0 \quad (\text{II.17})$$

We find that the boundary condition for the \bar{H} field is that the normal derivative of the component of the total magnetic field parallel to the surface coordinate z must be zero, in the case when the perfectly conducting surface is independent of the coordinate z to which the incident magnetic field vector is parallel. More generally, it can be shown that all normal derivatives of the covariant components of the magnetic field must vanish on the surface of a perfect conductor, for the class of surfaces for which the conductor surface coincided with the coordinate surface of a curvilinear orthogonal coordinate system.

In addition to Maxwell's equations and the boundary conditions, other conditions need to be considered and satisfied when attempting to solve electromagnetic boundary-value diffraction problems. One of these is the Sommerfeld "radiation condition at infinity", which has to do with the requirement that the amount of energy flowing from sources in a finite domain through an infinitely extended boundary at an infinite distance from the sources must tend to zero. (In fact the condition is somewhat stronger and states that the sources must be sources and not sinks of energy). Another condition results from the use of energy conservation laws and Poynting's theorem. Another still arises when the fields are expanded as Fourier-transform superpositions of plane waves, in which case all the waves not only with real but also with imaginary propagation constants must be included. Waves with imaginary propagation constants, or evanescent waves, and more generally inhomogeneous waves with complex propagation constants, have surfaces of constant phase and surfaces of constant amplitude which do not coincide: in a two-dimensional case, for example in an ordinary cylindrical lens, for which the glass-thickness variation results in a variation of absorption across the lens surface, the surfaces of constant phase and the surface of constant amplitude are orthogonal to each other. Inhomogeneous waves form the most general and most frequently encountered type of waves in optics.

2. Image Formation Solutions Using "Huygens' Principle"

Basically, it can be shown that the various "Huygens' principle" solutions used in optics (for example its "rigorous" vector form or its Fourier transform formulation) can be obtained from Maxwell's electromagnetic equations provided that certain important approximations are made. One of the fundamental approximations is that the Huygens' principle formulation applies only to the vicinity of the center of quasi-spherical image-forming wave fronts. It is essential to be aware of the approximate nature of the Huygens' principle expressions in their application to the solution of diffraction and image formation problems. The apparent simplicity and even reasonable form of Huygens' principle expressions, when evaluated heuristically on the basis of superposition principles and plane wave spectra, should not be mistakenly used as the basis for "more rigorous" solutions by the inclusion of higher order terms obtained from the initially approximate formulation. When correctly applied, however, Huygens' principle in its Fourier transform expression, forms a most powerful tool for dealing with image formation problems. In particular, it is being used for the calculation of the distribution of energy in diffraction patterns formed by wave fronts of finite size produced by reflection, transmission and diffraction from optical elements (mirrors, lenses, prisms and gratings).

A "rigorous" expression of Huygens' principle which we have obtained from Maxwell's equations is

$$\bar{E}_p(\alpha, \beta) = \frac{i}{\lambda} \frac{\exp[-ikR]}{R} \iint_{\text{APERTURE}} \bar{E}_0 \exp[ik\Delta(x,y)] \exp[-ik(\alpha x + \beta y)] dx dy \quad (\text{II.18})$$

where \bar{E}_p is the complex amplitude of the electric field vector at a point P in the image plane in the vicinity of the center of a quasi-spherical wave front of radius R. α and β are the direction cosines defining the position of P as seen from the aperture centered on the point 0 on the quasi-spherical wave front, x and y the coordinates defining the aperture, $\Delta(x,y)$ is the aberration of the wave front from sphericity, $k = 2\pi/\lambda$, $i = \sqrt{-1}$, \bar{E}_0 the amplitude $\bar{E}_0(x,y)$ in the wave front. The direction cosines α and β , and in fact the coordinates x,y refer to the quasi-plane wave front side of a perfect focussing system (mirror or lens). A perfect focussing system is defined as having the property of transforming a perfectly plane wave front into a perfectly spherical wave front. Consequently, it transforms a quasi-plane wave front into a quasi-spherical wave front. Moreover, the aberrations from the plane wave front are identical to the aberrations from the spherical image forming wave front within the approximations used.

Equation (II.18) has the form of a Fourier transformation and can be read as follows: the complex amplitude of the electric field vector at a point in the image plane is equal to the Fourier transform of the distribution of complex amplitude of the electric field within the image forming aperture. Clearly here, the field vectors in the aperture and the field vectors in the image plane project along the same direction in the image plane. One Fourier transformation needs to be carried out for each point in the diffraction pattern.

For example, for a perfectly uniform plane wave front within a rectangular aperture of width A along the x-axis, $|\bar{E}_0| = 1$, $\Delta = 0$,

$$\bar{E}_p(k\alpha) = \frac{i}{\lambda} \frac{\exp[-ikR]}{R} \int_{-A/2}^{+A/2} \exp[-ik\alpha x] dx \quad (\text{II.19})$$

which immediately gives upon integration

$$\bar{E}_p(k\alpha) = \left[\frac{i}{\lambda} \frac{\exp[-ikR]}{R} \frac{A}{2} \right] \left[\frac{\sin k\alpha \frac{A}{2}}{k\alpha \frac{A}{2}} \right] \quad (\text{II.20})$$

The complex amplitude $\bar{E}_p(k\alpha)$ thus varies according to the well-known $\sin x'/x'$ function along $x' \cong R\alpha = f\alpha$ in the image plane and has a first minimum for

$$\alpha_0 = \frac{\lambda}{A} \text{ radians} \quad (\text{II.21})$$

or

$$x'_0 = R \frac{\lambda}{A} = f \frac{\lambda}{A} \quad (\text{with } R=f=\text{focal length of focussing system}) \quad (\text{II.22})$$

In general only the intensity

$$I_p = \overline{\bar{E}_p \cdot \bar{E}_p^*} \quad (\text{II.23})$$

is detectable (with the help of photoelectric, photographic or other receivers). I_p varies as $(\sin x'/x')^2$ and has a first minimum at $(\lambda/A) \cdot f$ from the central maximum. I_p is generally called the diffraction pattern corresponding to the aperture A.

Equation (II.18) which expresses the diffraction at infinity by an aperture can be made heuristically plausible on the basis of superposition and "Huygens' principle" as it is generally understood. Consider a pupil in the x, y plane and a point M centered on an element of area $dx dy$ in that plane. The element of area centered on M emits an elementary wave $f(x, y) dx dy$ where $f(x, y)$ describes the scalar component of the \bar{E} vector in the pupil. In the direction defined by the direction cosines α, β, γ , the wave from M is dephased by $(2\pi/\lambda)(\alpha x + \beta y)$ with respect to an elementary wave emitted from the center of the pupil O. By superposition, we have in the direction α, β, γ the sum of these waves

$$\iint_{\text{PUPIL}} f(x, y) \exp\left[i \frac{2\pi}{\lambda} (\alpha x + \beta y)\right] dx dy \quad (\text{II.24})$$

All elementary waves emitted by the pupil (or transmitted by the pupil) in the direction α, β, γ come to focus at a single point $P'(x', y')$ in the focal plane of a perfect focussing system, such that

$$x' = \frac{f\alpha}{\gamma} \quad y' = \frac{f\beta}{\gamma} \quad (\text{II.25})$$

In as much as $\gamma \cong 1$ and $\alpha^2 + \beta^2 + \gamma^2 = 1$, we have

$$x' = f\alpha \quad y' = f\beta \quad (\text{II.26})$$

and if we define the coordinates in terms of "reduced" coordinates

$$u = \frac{x'}{f} \quad v = \frac{y'}{f} \quad (\text{II.27})$$

and take λ for the unit of length we have

$$E(u, v) = \iint_{\text{PUPIL}} f(x, y) \exp[2\pi i (ux + vy)] dx dy \quad (\text{II.28})$$

which is an expression for the diffraction pattern identical to that obtained in equation (II.18) from Maxwell's equations by using the there stated approximations.

In summary, a knowledge of the distribution of complex amplitude (amplitude and phase) of the electromagnetic field (or more exactly of a component of the electric field) within an aperture, no matter how this field is created in this aperture, permits to compute the diffraction patterns corresponding to the aperture and light distribution. A unique relation between the field distribution in the aperture and the light distribution in the diffraction patterns exists, within the stated approximations, and takes the form of a Fourier transformation. The powerful techniques of operational calculus have been extensively applied to optical image formation problems with very fruitful results. In particular, a basic similarity has been recognized between problems in electrical engineering and problems of optical image formation and spectroscopy, whenever superposition and operational methods are appropriate.

II.3 Image Formation Theory and Optical Signal Processing in Fourier-Transform Formulation

The well known operational theory of signal processing used when dealing with electrical and electronic systems can be immediately translated into optics, provided that the parameter "time" used in electrical engineering is translated as "space" when used in optics.

The diffraction patterns (intensity) can be considered as the impulse response of an optical system. The intensity distribution in the image as a function of spacial coordinates in the image space is simply given by the convolution integral of the intensity distribution

in the object (as geometrically imaged in the image plane) with the intensity distribution in the diffraction pattern (as it appears in the image plane). Also the Fourier transform of the image intensity distribution is equal to the product of the transform of the object intensity distribution by the transform of the diffraction pattern. The transform of the diffraction pattern is also called the "frequency response function" of the optical system, since it gives the distribution of light in the image of a spatially periodic intensity distribution in the object. Finally, the frequency response (function) of an optical system can be immediately shown to be equal to the spatial convolution of the light distribution (complex amplitude) in the aperture with itself. For example, for the uniformly illuminated rectangular aperture considered before, the frequency response function is immediately seen to have a triangular shape, as a function of spatial frequency, that is as a function of (distance)⁻¹.

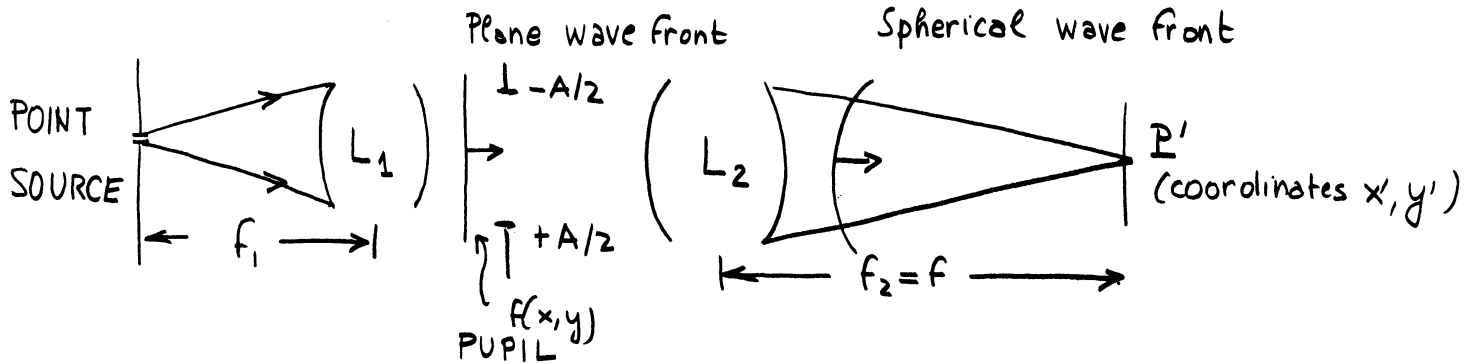
Clearly, optical systems, when used as elements of a signal processing system for electromagnetic waves, for example in television, radio-astronomy and similar systems, present a degree of two-dimensional flexibility unavailable in electrical systems. Moreover, the parameter "time" can be reintroduced and used to advantage.

Skillful use of the relations between object, aperture and image space permits to considerably simplify measurements of the desired distribution of light (complex amplitude and intensity) in any of these domains in practical applications.

III. IMAGE FORMATION IN NON-COHERENT LIGHT

(Elements and Definitions)

III.1 Image of Point Source



Under the conditions where the Fourier transform relation between the complex amplitude distribution in the pupil $f(x, y)$ and the complex amplitude distribution at a point in the focal plane, $F_p(u, v)$, holds, it was shown that

$$F_p(u, v) = \iint_{\text{PUPIL}} f(X, Y) \exp [2\pi i (uX + vY)] dX dY \quad (\text{III.1})$$

$F_p(u, v)$ is the electric field vector amplitude at the point P' .
 One has

$$u = \frac{X'}{f} = \frac{\alpha}{\lambda} \quad (\text{III.2})$$

$$v = \frac{Y'}{f} = \frac{\beta}{\lambda} \quad (\text{III.3})$$

with X', Y' in units of λ . One can write (III.1) as

$$F_p(u, v) = \mathcal{T} [f(X, Y)] \quad (\text{III.4})$$

or

$$f(x, y) \rightarrow F_p(u, v) \quad (\text{III.5})$$

to indicate the Fourier transform relationship between pupil and diffraction image. Note

$$x' = f\alpha = f\lambda u \quad (\text{III.6})$$

$$y' = f\beta = f\lambda v \quad (\text{III.7})$$

From Equation (III.1) it is seen that the image of a point source is not a point source.

In general, as a result of the electromagnetic nature of light and the finite dimensions of the pupil, the light from a point source is spread out into a "diffraction pattern".

A physical receptor (photo-electric cell, photographic plate, eye, etc.) can only detect (or produce a signal proportional to) the magnitude of $F_{P'} \equiv \bar{E}_P$, at best. The detectable quantity is then the intensity

$$I_{P'} = \bar{E} \cdot \bar{E}^* = |\bar{E}|^2 \quad (\text{III.8})$$

III.2 Summation of Light from Several Source Points Reaching One Image Point

It is very important to observe that light originating from more than one source point may reach any point in the image. Let \bar{E}_1 and \bar{E}_2 be the light-amplitude vectors corresponding to the light reaching a point P' from two different source points. Two extreme situations may arise. They are illustrated for two source points.

1. The two source points radiate coherently: \bar{E}_1 and \bar{E}_2 may interfere and the detected intensity is

$$[I_{P'}]_{\text{COHERENT}} = (\bar{E}_1 + \bar{E}_2) \cdot (\bar{E}_1 + \bar{E}_2)^* = |\bar{E}_1 + \bar{E}_2|^2 \quad (\text{III.9})$$

2. The two source points radiate completely non-coherently: \bar{E}_1 and \bar{E}_2 cannot interfere and the detected intensity is

$$[I_{P'}]_{\text{NON-COHERENT}} = (\bar{E}_1 \bar{E}_1^*) + (\bar{E}_2 \bar{E}_2^*) = |\bar{E}_1|^2 + |\bar{E}_2|^2 = I_1 + I_2 \quad (\text{III.10})$$

The distinction between the detection according to equation (III.9) and equation (III.10) is all important. It is basic for the further discussion of image formation.

Equation (III.8), (III.9), and (III.10) describe ideal situations. In general, only suitable time averages of the indicated intensities are detected. Clearly also, in case of the light from several source points reaching one image point, one has

$$[I_{P'}]_{\text{COHERENT}} = \left| \sum_i \bar{E}_i \right|^2 \quad (\text{III.11})$$

and

$$[I_p]_{\text{NON-COHERENT}} = \sum_i |\bar{E}_i|^2 = \sum_i I_i \quad (\text{III.12})$$

In coherent light: SUM AMPLITUDES AND TAKE SQUARE OF MAGNITUDE

In non coherent light: SUM INTENSITIES (i.e. SUM SQUARES OF MAGNITUDES)

It will be shown, when dealing with the mathematical formulation of coherence, partial coherence and incoherence, that the absence of interference is indeed sufficient to characterize non-coherence as described by the above.

NOTE: It is very important to note that the transition from complex amplitude \bar{E} to the detected intensity $I = \bar{E} \cdot \bar{E}^*$ occurs only at a detector.

III.3 Spread Function

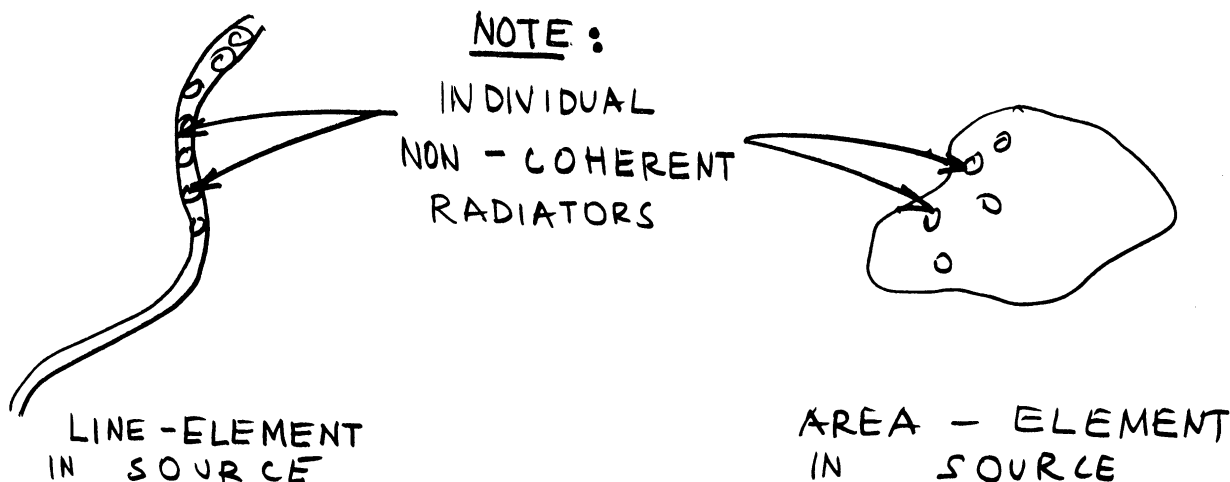
NOTE: In non-coherent light, a point source at infinity will produce not a point image, but rather an intensity distribution

$$I_p(u, v) = \overline{F_p(u, v)} \cdot \overline{F_p(u, v)}^* = |F_p(u, v)|^2 \quad (\text{III.13})$$

The intensity distribution in the image of a point source at infinity is defined as the Spread Function, such that

$$s(u, v) = |F_p(u, v)|^2, \text{ or better } s(u, v) = \overline{F(u, v)} \cdot \overline{F(u, v)}^* \quad (\text{III.14})$$

III.4 Image of Extended Source in Non-coherent Light



An extended source can be considered as formed by individual radiators (dipole, atomic, molecular etc.). This is true for self-radiating or illuminated surfaces, gases, lasers etc.

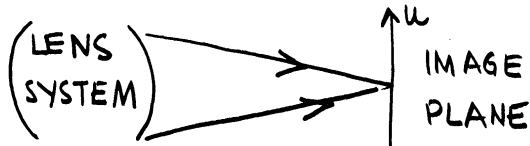
For the purposes of image-formation theory, we shall concern ourselves at first with sources in areas at right angle to the optical axis of a lens system. Two conclusions reached previously are basic to what follows:

1. The light from a point source is spread out into a "spread function". As a corollary, it is clear that light from two neighbouring points in the source will spill over from one of the "geometrical optics" images to the other, and vice-versa.

2. In non-coherent light, the contribution to the light intensity at any image point is obtained by a suitable summation of the intensity contributions to the particular image point of the light "spilled" over or centered there, as a result of its origin in the discrete object points.

METHOD

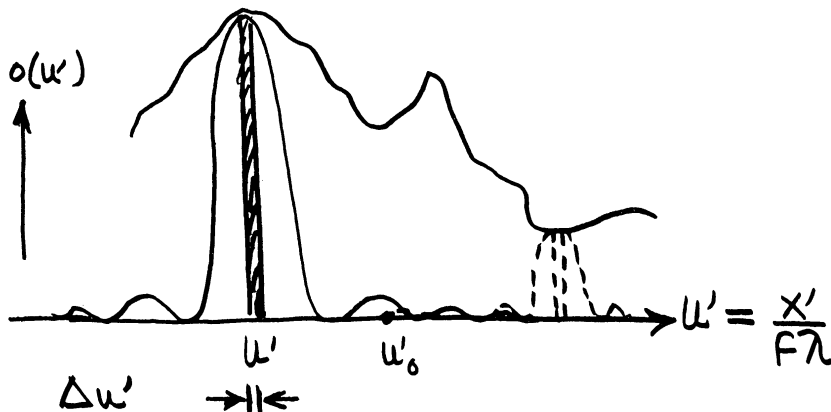
The light intensity at an image point in the presence of "spreading" can be computed by considering the image formed according to geometrical optics (called "geometrical optics image" henceforth) and by computing the weighted contribution by starting from the geometrical optics image rather than from the object.



Let $o(u, v)$ be the image according to geometrical optics.

Let $s(u, v)$ be the spread function according to equation (III.14). $s(u, v)$ is the intensity distribution in the image of a point source at infinity. (Later the concept of spread function and convolution can be easily extended to objects at finite distance).

Let $u' \equiv u$ for more convenient description of the image plane.



We need to compute the intensity

$i(u'_0, v'_0)$
 at a point (u'_0, v'_0) in the image. Clearly, the contribution of the image of a source element of width $\Delta u'$ and having an amplitude $\frac{\Delta u'}{o(u')}$ when centered at the point u' is given by

$$\Delta i(u'_0) = s(u'_0 - u') o(u') \Delta u' \quad (\text{III.15})$$

Equation (III.15) is nothing but the diffraction pattern (or spread function) weighted by the intensity $o(u')$ at the point u' where it is centered. Since even an elementary radiator (atom, molecule.....) has some finite width $\Delta u'$, the spread function maximum intensity must be weighted by $o(u') \Delta u'$. If several source (image) points contribute to the intensity at u'_0 , one has

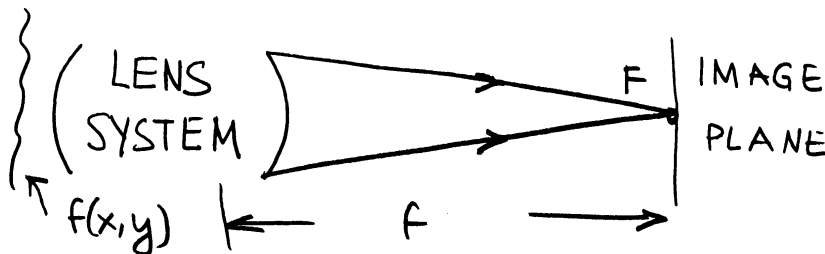
$$i(u'_0) = \sum_i o(u')_i s(u'_0 - u') du' \quad (\text{III.16})$$

and in the limit of a great number of source points

$$i(u'_0) = \int_{-\infty}^{+\infty} o(u') s(u'_0 - u') du' \quad (\text{III.17})$$

Equation (III.17) is immediately recognized as having the form of a convolution integral, i.e.

$$i(u'_0) = o(u') \otimes s(u') \quad (\text{III.17})$$



The two important relations established so far can be written as

SPREAD FUNCTION (III.18)

$$s(u', v') = T[f(x, y)] \cdot T^*[f(x, y)]$$

IMAGE INTENSITY DISTRIBUTION (III.19)

$$i(u'_0, v'_0) = o(u', v') \otimes s(u', v')$$

where $T =$ Fourier transform, i.e.

$$F(u, v) = \iint_{-\infty}^{+\infty} f(x, y) \exp [2\pi i (uX + vY)] dXdY \quad (\text{III.20})$$

and \otimes = convolution integral, i.e.

$$i(u'_0, v'_0) = o(u', v') \otimes s(u', v') \quad (III.21)$$

$$= \iint_{-\infty}^{+\infty} o(u', v') s(u'_0 - u', v'_0 - v') du' dv'$$

Equation (III.21) is the two-dimensional equivalent of equation (III.17).

III.5 Analysis of Image Formation in Terms of Spatial Harmonics

Certain important advantages are obtained by analyzing the image formation problem, as expressed by equations (III.18) and (III.19) in terms of its expression in the "spatial frequency" domain. A spatial frequency has been defined as the number of light intensity cycles (of a sinusoidal intensity distribution) per unit length. (Recall that the dimensions of u, v , respectively here u', v' are $(\text{LENGTH})^{-1}$) Consider

$$i(u'_0, v'_0) = o(u', v') \otimes s(u', v') \quad (III.19)$$

Take the inverse transform,

$$I.T. = \iint (\quad) \exp[-2\pi i(\omega_u u' + \omega_v v')] du' dv' \quad \text{on both sides of equation (III.19)}$$

In order to do this, recall one of the expressions of the Convolution Theorem, which gives

$$I.T. [i(u'_0, v'_0)] = I.T. [o(u', v')] \cdot I.T. [s(u', v')] \quad (III.22)$$

Also note that

$$s(u', v') = T[F(x, y)] \cdot T^* [F(x, y)] \quad (III.18)$$

where * indicates the complex conjugate. By applying the second expression of the Convolution Theorem to equation (III.18) one has

$$I.T. [s(u', v')] = I.T. \{ T[F(x, y)] \cdot T^* [F(x, y)] \} = F(x, y) \otimes F^*(-x, -y) \quad (III.23)$$

and finally equation (III.23) gives with equation (III.22) the equation

$$\text{GENERAL EXPRESSION } I.T. [i(u'_0, v'_0)] = I.T. [o(u', v')] \cdot [F(x, y) \otimes F^*(-x, -y)] \quad (III.24)$$

Note: For real symmetrical $f(x, y)$, equation (III.24) becomes

$$I.T. [i(u'_0, v'_0)] = I.T. [o(u', v')] \cdot [F(x, y) \otimes f(x, y)] \quad (III.25)$$

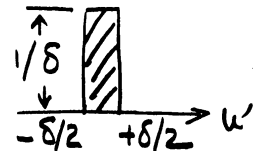
III.6 Physical Significance of "Spread" Function and of Spatial Harmonic Analysis of Image Formation

1. Point Source (Dirac Delta Function) → Spread Function

Consider a point source with a geometrical optics point source image $\delta(u')$

$$\delta(u') = \lim_{\delta \rightarrow 0} \int_{-\delta/2}^{+\delta/2} (1/\delta) du'$$

as shown.



According to equation (III.17)

$$i(u'_0) = \delta(u') \otimes s(u') = s(u') \quad (III.26)$$

and one concludes that

$$s(u') = \text{image of point source}$$

as in fact assumed in establishing equation (III.17). Note that $s(u')$ does not in any way have to be the "ordinary" diffraction pattern of the simplest type (for example the $(\sin u'/u')^2$ distribution or its equivalent). In its most general form $s(u')$ and $s(u', v')$ are the images of a slit or point source respectively. In fact $s(u', v')$ or $s(u')$ do not have to be in the focal plane, as far as equation (III.17) goes.

2. Sinusoidal Intensity Object → Spatial Frequency Response Function

Let

$$o(u', v') = \text{Re} \left\{ \exp [2\pi i (x_u u' + y_v v')] \right\} \quad (III.27)$$

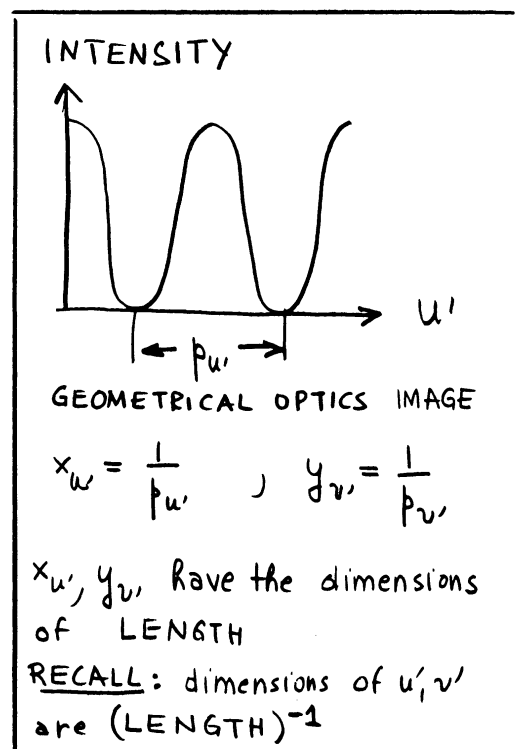
Let us deal with

$$o(u', v') = \exp [2\pi i (x_u u' + y_v v')] \quad (III.28)$$

and take the real part at the end. $o(u', v')$ describes the sinusoidal object (geometrical optics image) represented schematically in FIGURE. Use Equation (III.19)

$$i(u'_0, v'_0) = o(u', v') \otimes s(u', v') \quad (III.19)$$

It has been shown that the convolution integral is commutative i.e.



$$o(u',v') \otimes s(u',v') = s(u',v') \otimes o(u',v') \quad (\text{III.29})$$

and therefore

$$i(u'_0, v'_0) = s(u',v') \otimes o(u',v') \quad (\text{III.30})$$

i.e.

$$i(u'_0, v'_0) = \iint_{-\infty}^{+\infty} s(u',v') \exp\{2\pi i [x_{u'}(u'_0 - u') + y_{v'}(v'_0 - v')]\} du' dv' \quad (\text{III.31})$$

By noting that the integration is with respect to u', v' equation (III.31) gives

$$i(u'_0, v'_0) = \underbrace{\exp[2\pi i (x_{u'} u'_0 + y_{v'} v'_0)]}_{\substack{\text{COMPLEX} \\ \text{IMAGE INTENSITY} \\ \text{DISTRIBUTION}}} \underbrace{\iint_{-\infty}^{+\infty} s(u',v') \exp[-2\pi i (x_{u'} u' + y_{v'} v')]}_{\substack{\text{COMPLEX} \\ \text{OBJECT INTENSITY} \\ \text{DISTRIBUTION}}} du' dv' \quad (\text{III.32})$$

Let us define $\mathcal{I.T.} [s(u',v')] \equiv \tau(x,y)$

$$\mathcal{I.T.} [s(u',v')] = \iint_{-\infty}^{+\infty} s(u',v') \exp[-2\pi i (x_{u'} u' + y_{v'} v')] du' dv' \equiv \tau(x_{u'}, y_{v'}) = \tau(x,y) \quad (\text{III.33})$$

as the complex spatial frequency response function of the optical system. Note that the function $\tau(x,y)$ is a function of the pupil coordinates x, y which have inverse dimensions from the image coordinates u and v (respectively u' and v'). A sinusoid in the u, v domain is represented by a frequency in the x, y domain. The meaning of $\tau(x,y)$ indicated in equation (III.32) can be further clarified by taking the magnitude and phase of $\tau(x,y)$. One has with equations (III.28), (III.32), and (III.33)

$$\tau(x,y) = \frac{i(u',v')}{o(u',v')} = \frac{\text{Frequency response function for a complex (exponential) object}}{\quad} \quad (\text{III.34})$$

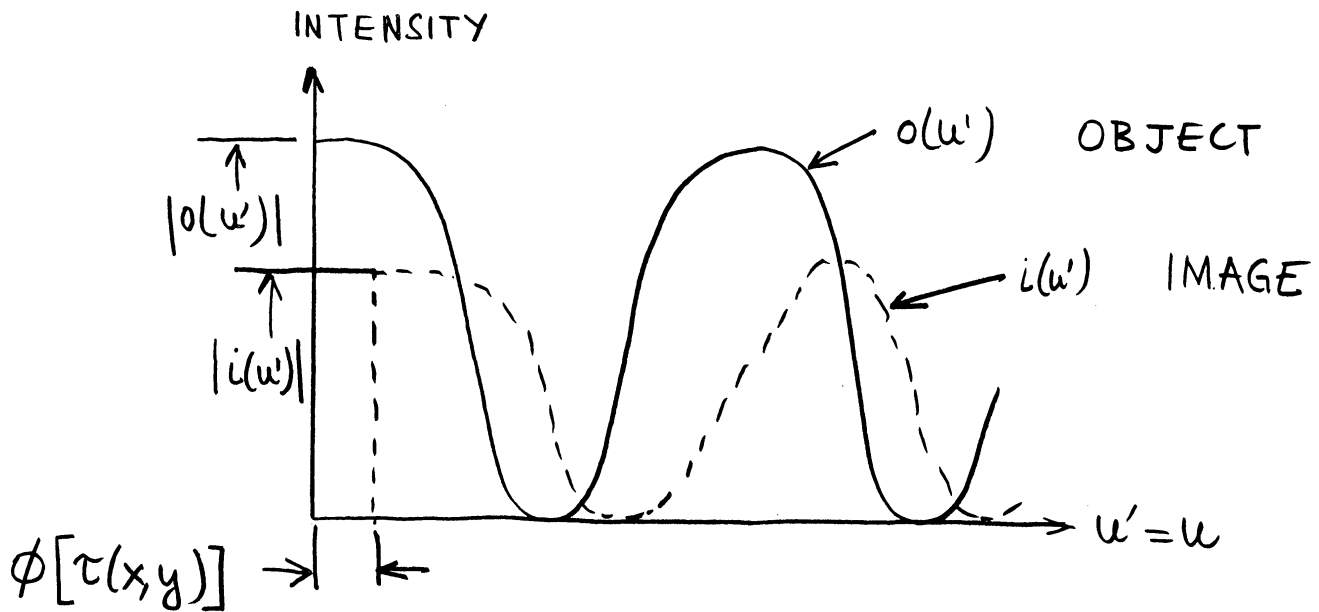
Equation (III.34) is important and shows that $\tau(x,y)$ is nothing but the gain of the optical system. By taking the real part and phase in (III.34) one has

$$|\text{Re} [\tau(x,y)]| = \frac{|\text{Re} [i(u',v')]|}{|\text{Re} [o(u',v')]|} \quad \left(\text{for a sinusoidal object } o(u',v') \right) \quad (\text{III.35})$$

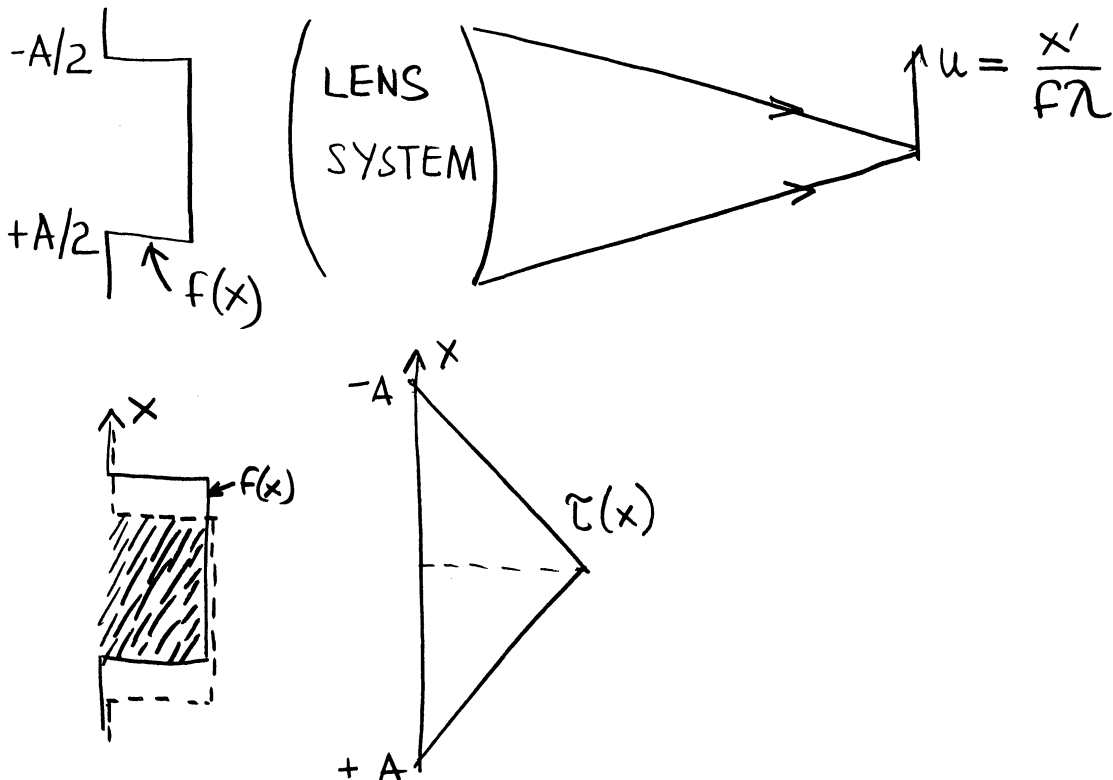
$$\phi[\tau(x,y)] = \phi[i(u',v')] - \phi[o(u',v')] \quad (\text{III.36})$$

Recall

$$\tau(x,y) = f(x,y) \otimes f^*(-x,-y) \quad (\text{III.24})$$



EXAMPLE



IV. COHERENCE CHARACTERISTICS OF LIGHT
 (Experimental Characterization)

IV.1 Introduction

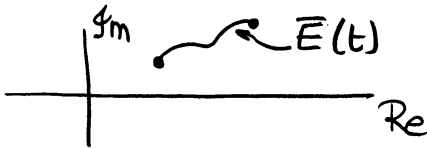
It appears that it is experimentally possible to distinguish at least the following types of coherence:

1. Spectral (temporal or phase) coherence
2. Spatial coherence
3. Frequency stability in time.

In general, light must be described by probability functions. In general also light from ordinary sources, and even from lasers, is either unpolarized or partially polarized, and has a finite frequency spread and an amplitude which varies in time.

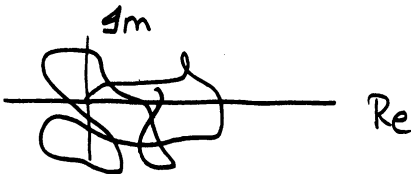
The narrowest definition for light will describe it as a single-frequency, polarized electromagnetic wave of constant amplitude. Let $\vec{E}(t)$ describe the instantaneous value of the light vector.

Over short periods of time one has



$$\vec{E}(t) = E_0(t)e^{i\omega t}$$

Over longer periods of time, the mean $\vec{E}(t) = 0$



$$\vec{E}(t) = \frac{1}{\Delta T} \int_0^{\Delta T} E(t) dt$$

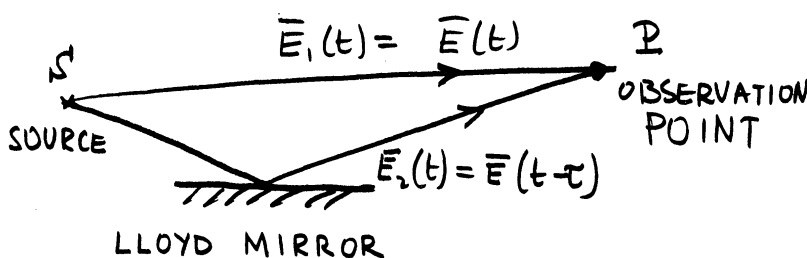
$$\lim_{\Delta T \rightarrow \infty} \vec{E}(t) = 0$$

However the mean square $\overline{E(t)E(t)^*}$ is equal to the energy and not zero:

$$\overline{E(t)E(t)^*} = \lim_{\Delta T \rightarrow \infty} \int_0^{\Delta T} E(t)E(t)^* dt \neq 0$$

IV.2. Characterization of time-coherence

1. Correlation method



Consider a light source S and an observation point P .

It is assumed that the light arriving at P via the two paths shown has colinear polarizations.

The resultant field at \underline{P} is

$$\bar{E}_R(t) = \bar{E}_1(t) + \bar{E}_2(t) = \bar{E}(t) + \bar{E}(t-\tau) \quad (\text{IV.1})$$

Assuming quadratic detection with long time integration, the observable quantity is

$$\begin{aligned} I(t) &= \overline{E_R(t) E_R(t)^*} \\ &= \frac{1}{\Delta T} \int_0^{\Delta T} E_R(t) E_R(t)^* dt \neq 0 \end{aligned} \quad (\text{IV.2})$$

If one further assumes that the source is stationary, one has

$$I(t) = I(\tau) \quad \text{only} \quad (\text{IV.3})$$

where τ is the time difference between the two beams. One has

$$\begin{aligned} I(\tau) &= \overline{|E(t) + E(t-\tau)|^2} \\ &= \overline{|E(t)|^2} + \overline{|E(t-\tau)|^2} + 2 \overline{|E(t) E(t-\tau)|} \end{aligned} \quad (\text{IV.4})$$

The three terms in equation (IV.4) are recognized as auto-correlation functions $\varphi(\tau)$. One also has

$$\overline{|E(t)|^2} = \overline{|E(t-\tau)|^2} = \varphi(0) \quad (\text{IV.5})$$

because the source is assumed stationary. One finally has for the observed intensity

$$I(\tau) = 2\varphi(0) + 2\varphi(\tau) \quad (\text{IV.6})$$

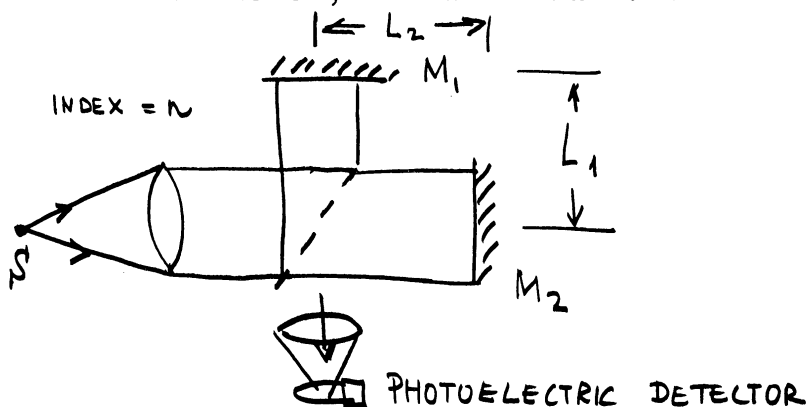
where the varying quantity of interest is $\varphi(\tau)$, the auto-correlation function of the electric field in the light beam. Equation (IV.6) will appear particularly interesting when it is compared to the expression for $I(\tau)$ obtained from interference theory.

2. Alternate way of looking at time coherence (Interference theory).

a) Monochromatic (single-frequency) light

Consider any ordinary two-beam interferometer: for example,

the Lloyd mirror illustrated under 1, or a Michelson Twyman-Green interferometer, illustrated here



Consider the $I(\tau)$ obtained with a finite delay τ at a given frequency ω .

One has

$$I(\tau) = 1 + 1 + 2 \cos 2\pi \frac{\tau}{T_\omega} = 2 (1 + \cos 2\pi \omega \tau) \quad (\text{IV.7})$$

with $T_\omega = 1/\omega$.

In an actual interferometer, one has

Path difference $\Delta = n(L_2 - L_1)$

Phase difference $\phi = 2\pi \Delta / \lambda$

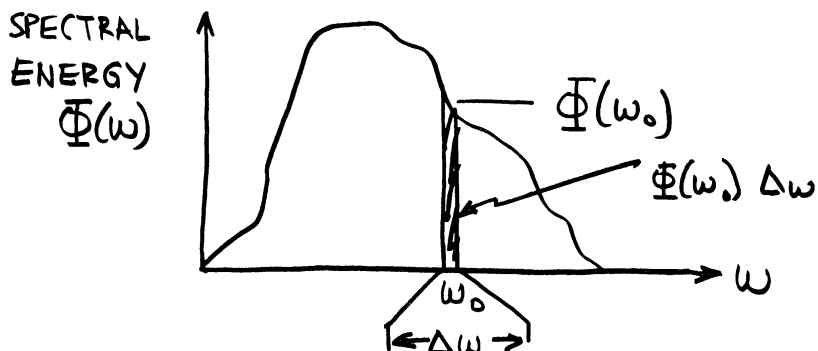
Time difference $\tau = \Delta / c$

c = velocity of light in vacu**u**m

n = refractive index.



b) Polychromatic light (Non-coherent Source).



Let $\Phi(\omega)$ be the spectral energy distribution in the light source. It is assumed that a "non-coherent source" is such that the different spectral frequencies in the source are statistically uncorrelated, so that no "interference" in the

ordinary sense is possible between waves of different frequencies. (The case of beat-frequencies between waves of different frequencies is analyzed below). Under this assumption, one has for

$$\begin{aligned} \omega_1 & : I_1(\tau) = \Phi(\omega_1) \Delta\omega 2(1 + \cos 2\pi\omega_1\tau) \\ \omega_2 & : I_2(\tau) = \Phi(\omega_2) \Delta\omega 2(1 + \cos 2\pi\omega_2\tau) \\ \text{etc.} & : \end{aligned}$$

and since the different frequencies are non-coherent with each other, the intensity observed is

$$I(\tau) = 2 \int_{\omega} (1 + \cos 2\pi\omega\tau) \Phi(\omega) d\omega \quad (\text{IV.8})$$

where $\int_{\omega} = \int_0^{\infty}$.



3. Compare correlation method with interferometric method.

Equation (IV.7) is

$$I(\tau) = 2\varphi(0) + 2\varphi(\tau) \quad (\text{IV.7})$$

Equation (IV.8) is

$$I(\tau) = 2 \int_{\omega} \Phi(\omega) d\omega + 2 \int_{\omega} \Phi(\omega) \cos(2\pi\omega\tau) d\omega \quad (\text{IV.8})$$

One immediately recognizes the physical meaning of the auto-correlation functions. One has

$$\varphi(0) = \int_{\omega} \Phi(\omega) d\omega \quad (\text{IV.9})$$

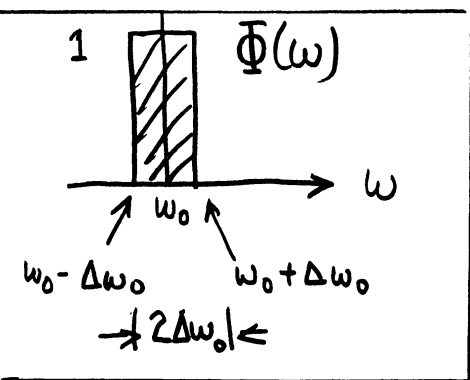
and

$$\varphi(\tau) = \int_{\omega} \Phi(\omega) \cos(2\pi\omega\tau) d\omega \quad (IV.10)$$

Equation (IV.10) shows that the auto-correlation function of the electric field in the light beam is equal to the Fourier cosine transform of the energy distribution in the spectrum.

Equation (IV.10) has found important applications in Fourier-transform spectroscopic analysis of light, and also in radio-astronomy.

4. Narrow spectrum. (Flat-tope line) (Special case. Important for laser analysis)



Consider the idealized line shown. One has for the intensity

$$I(\tau) = 2 \int_{\omega_0 - \Delta\omega_0}^{\omega_0 + \Delta\omega_0} d\omega + 2 \int_{\omega_0 - \Delta\omega_0}^{\omega_0 + \Delta\omega_0} \cos(2\pi\omega\tau) d\omega \quad (IV.11)$$

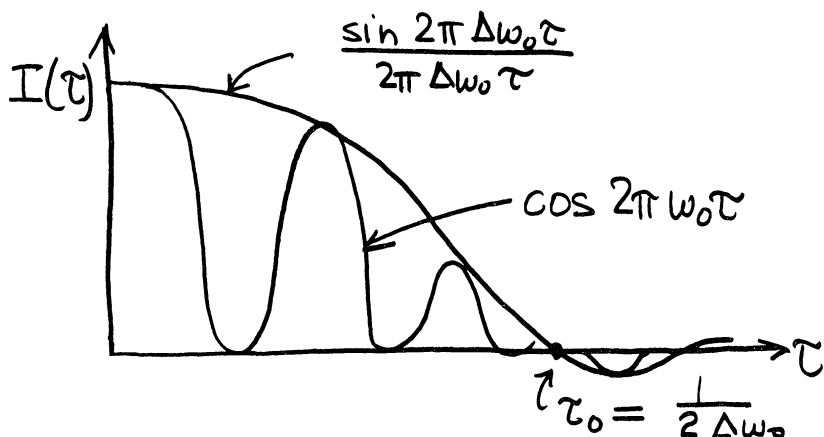
i.e.

$$I(\tau) = 4 \Delta\omega_0 + \frac{2}{2\pi\tau} \left[\sin 2\pi\omega\tau \right]_{\omega_0 - \Delta\omega_0}^{\omega_0 + \Delta\omega_0} \quad (IV.12)$$

and finally

$$I(\tau) = 4 \Delta\omega_0 \left[1 + \frac{\sin 2\pi \Delta\omega_0 \tau}{2\pi \Delta\omega_0 \tau} \cos 2\pi \omega_0 \tau \right] \quad (IV.13)$$

The $I(\tau)$ described in equation (IV.13) is shown in the figure



The first minimum of the $\frac{\sin x}{x}$ envelope occurs for $2\pi \Delta\omega_0 \tau_0 = \pi$.

The $\sin x/x$ envelope is sometimes described as the fringe-visibility curve. It is seen that the decrease of fringe visibility with path difference is inversely proportional to the width of the spectral line.

5. Photoelectric interferometry with Gaussian line-shapes.

For Gaussian line-shapes, the envelope is also Gaussian.

Indeed, let

$$g(\omega) = e^{-a^2 \omega^2} \quad (\text{IV.14})$$

where $a = \text{constant}$. By Fourier transformation

$$g(\omega) = e^{-a^2 \omega^2} \rightarrow G(x) = \int_{-\infty}^{+\infty} e^{-a^2 \omega^2} e^{2\pi i \omega x} d\omega \quad (\text{IV.15})$$

where

$$G(x) = \int_{-\infty}^{+\infty} e^{-a^2 \omega^2} \cos(2\pi \omega x) d\omega + i \int_{-\infty}^{+\infty} e^{-a^2 \omega^2} \sin(2\pi \omega x) d\omega \quad (\text{IV.16})$$

The second term is zero because $g(\omega)$ is even, and therefore

$$G(x) = 2 \int_0^{\infty} e^{-a^2 \omega^2} \cos(2\pi \omega x) d\omega = \frac{\sqrt{\pi}}{a} e^{-\frac{\pi^2 x^2}{a^2}} \quad (\text{IV.17})$$

In other words

$$g(\omega) = e^{-(a\omega)^2} \rightarrow G(x) = \frac{\sqrt{\pi}}{a} e^{-\left(\frac{\pi x}{a}\right)^2} \quad (\text{IV.18})$$

Equation (IV.18) shows the remarkable feature that the Fourier transform of a Gaussian function is also a Gaussian function, with a width that is inversely proportional to the width of the original function.

More generally, as shown by Stroke⁽¹⁾, the complete expression for the photoelectric flux F_G , for the case of a line of Gaussian line-shape

(1) G. W. STROKE, "Photoelectric Fringe Signal Visibility and Range in Interferometers with Moving Mirrors", J. Opt. Soc. Am. 47, 1097-1103 (1957).

$$\Phi(\sigma) = \Phi_0 e^{-\Delta\sigma^2/\sigma^2} \quad (IV.15)$$

where

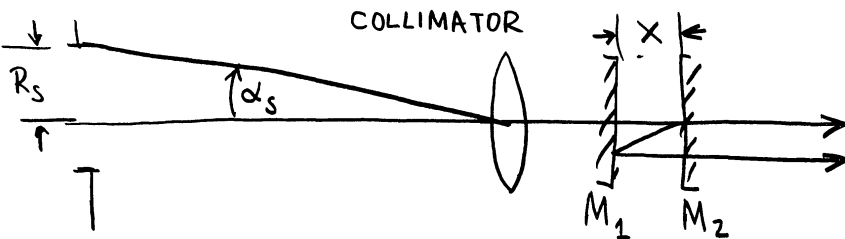
$$\Phi_\sigma = \Phi_0 \exp \left\{ - (mc^2/2kT) (\sigma - \sigma_0)^2 / \sigma^2 \right\}$$

is

$$P_G(x) = 4\pi^{3/2} R_s^2 a^2 \sigma \left\{ 1 + e^{-4\pi^2 x^2 \sigma^2 n^2} \frac{\sin(\pi x \alpha_s^2 n / \lambda)}{(\pi x \alpha_s^2 n / \lambda)} \cos \left[\frac{2\pi x (1 - \alpha_s^2 / 4)}{(\lambda/n)/2} \right] \right\} \quad (IV.16)$$

where account has been taken not only of the line shape, but also of the effect of finite source aperture α_s . The line shape effect is contained in the $\exp[-4\pi^2 x^2 \sigma^2 n^2]$ factor, while the effect of the finite

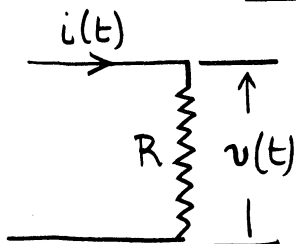
source aperture is contained in the $\sin x/x$ factor. (Note that this factor results here from another cause than the factor in equation (IV.13). With a finite source aperture, beams go through the



interferometer not only normally, but also at angles within a finite cone. In practice, the geometrical $\sin x/x$ factor can be made negligibly small compared to the line-shape factor for distances up to the order of meters. For instance, with $\alpha_s \approx 6 \times 10^{-4}$ radians, the first zero of the $\sin x/x$ factor appears at $x \approx 1.5$ meters. However, as shown by Stroke (Loc. cit. 1) a non-negligible length correction needs to be made in the use of interferometers with finite apertures. For instance, with the same $\alpha_s \approx 6 \times 10^{-4}$, the $\alpha_s^2/4 \approx 1 \times 10^{-7}$ in the cosine term, which is far from negligible in practice (an error of $.05 \lambda/2$ would result in $x = 136$ mm, if no account were taken of this correction).

6. Physical significance of power spectra

a) Case of a single-frequency signal



Consider the circuit shown. Let

$$i(t) = I_0 \cos \omega_0 t \quad (IV.17)$$

and therefore

$$v(t) = R I_0 \cos \omega t \quad (IV.18)$$

It follows that the power is

$$\begin{aligned} w(t) &= R I_0^2 \cos^2 \omega_0 t \\ &= \frac{1}{2} R I_0^2 + \frac{1}{2} R I_0^2 \cos 2\omega_0 t \end{aligned} \quad (\text{IV.19})$$

In the frequency domain, one has

$$I(\omega) = \mathcal{T}[i(t)] = \frac{1}{2} I_0 \delta(\omega - \omega_0) + \frac{1}{2} I_0 \delta(\omega + \omega_0) \quad (\text{IV.20})$$

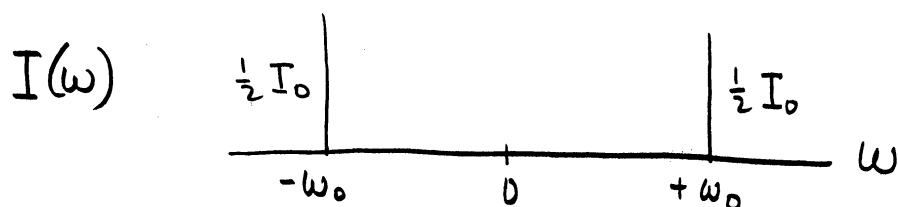
where the delta function is defined by the scalar products

$$\int_{-\infty}^{+\infty} f(x) \delta(x) dx = f(0) \quad (\text{IV.21})$$

and

$$\int_{-\infty}^{+\infty} f(x) \delta(x-a) dx = f(a) \quad (\text{IV.22})$$

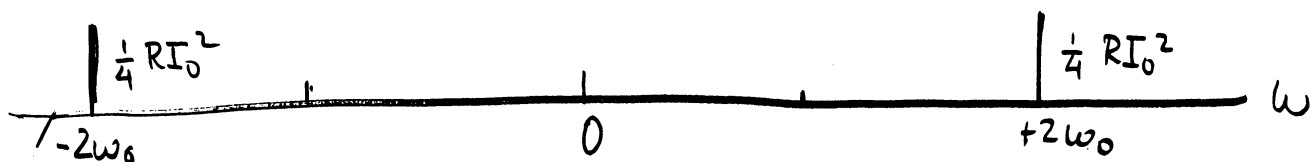
The spectral representation of the current, according to equation (IV.20) is



The frequency-domain representation of the power is

$$\begin{aligned} W(\omega) = \mathcal{T}[w(t)] &= \frac{1}{2} R I_0^2 \delta(0) \\ &+ \frac{1}{4} R I_0^2 \delta(\omega - 2\omega_0) + \frac{1}{4} R I_0^2 \delta(\omega + 2\omega_0) \end{aligned} \quad (\text{IV.23})$$

$W(\omega)$



One notes that the spectral representation $W(\omega)$ of the power $w(t)$ as described in equation (IV.19) shows energy at frequencies where there is no current! One concludes that it is necessary to seek a more suitable representation of the "power spectrum".

Consider the autocorrelation function of the current. Examine

$$\frac{1}{2T} \int_{-T}^{+T} i^2(t) dt = \frac{1}{2} I_0^2 \frac{\sin 2\omega_0 T}{2\omega_0 T} \quad (\text{IV.24})$$

In the limit, one has

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} i^2(t) dt = \frac{1}{2} I_0^2 \quad (\text{IV.25})$$

and therefore, an autocorrelation function $\varphi(\tau)$ exists.

$$\varphi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} [I_0 \cos \omega_0 t] [I_0 \cos \omega_0 (t - \tau)] dt \quad (\text{IV.26})$$

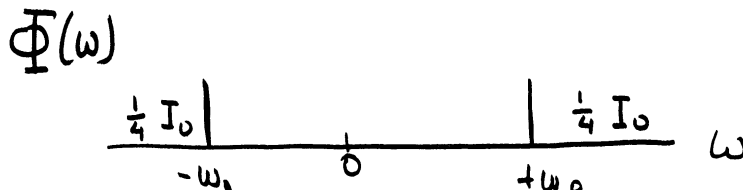
i.e.

$$\varphi(\tau) = \frac{1}{2} I_0 \cos \omega_0 \tau \quad (\text{IV.27})$$

The Fourier transform $\Phi(\omega)$ of the autocorrelation function is

$$\Phi(\omega) = \mathcal{T}[\varphi(\tau)] = \frac{1}{4} I_0 [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] \quad (\text{IV.28})$$

$\Phi(\omega)$ is defined as the power spectrum or the spectral energy distribution of the current. It is readily seen that $\Phi(\omega)$ has "power" at the same frequencies at which the current exists.



One is led to conclude that the power spectrum (equal to the Fourier transform of the autocorrelation function of the signal) is a meaningful representation of the spectral energy distribution in the signal (here, the current).

b) Case of a multiple-frequency signal.

Let

$$f(t) = \sum a_R \exp(i\omega_R t) \quad (\text{IV.29})$$

be the signal. One has

$$F(\omega) = \sum a_k \delta(\omega - \omega_k) \quad (IV.30)$$

and the power spectrum is

$$\Phi(\omega) = \mathcal{T}[y(\tau)] = \sum |a_k|^2 \delta(\omega - \omega_k) \quad (IV.31)$$

according to Fourier-transform theory, and the theory of distributions.

7. Heterodyne analysis of signals, beat frequencies, etc.

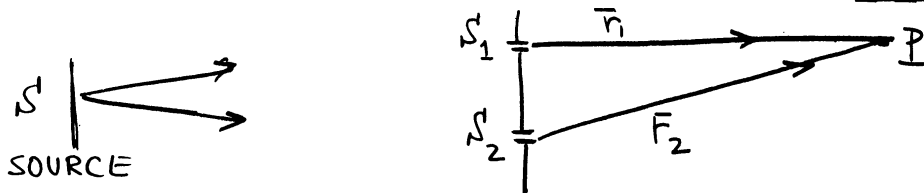
Will be examined separately.

8. Spatial Coherence

The spatial coherence in a light beam generally has to do with the "coherence" between two points in the field illuminated by one or more light sources. In its elementary sense, the degree of coherence between the two points simply describes the contrast of the interference fringes that are obtained when the two points are taken as secondary sources.

The meaning of spatial coherence can best be understood with the help of Young's two-slit experiment, or some equivalent. The important, and generally unexpected conclusion to be demonstrated is that even an extended source, composed of millions of statistically incoherent oscillators, can produce a coherent field, provided only that the angular diameter of the source, as seen from the plane of the two slits, is small compared to the slit separation.

We next derive the equation defining the partial coherence factor.



Let S_1 and S_2 be the two slits illuminated by an extended, non-coherent source. Let $E_1(t)$ and $E_2(t)$ be the electromagnetic field amplitudes at S_1 and S_2 . The amplitude at P is then

$$E_R(t) = E_1(t) \exp(i \frac{\phi}{2}) + E_2(t) \exp(-i \frac{\phi}{2}) \quad (IV.32)$$

where

$$\phi = \frac{2\pi}{\lambda} (|r_1| - |r_2|) \quad (IV.33)$$

The observable quantity at P is

$$\overline{E_R(t)E_R(t)^*} = \frac{1}{2T} \int_{-T}^{+T} E_R(t)E_R(t)^* dt \quad (\text{IV.34})$$

where $T \gg T_0$ (T_0 period of oscillation).

One has

$$\begin{aligned} \overline{E_R(t)E_R(t)^*} &= \overline{\left[E_1 e^{i\phi/2} + E_2 e^{-i\phi/2} \right] \left[E_1^* e^{-i\phi/2} + E_2^* e^{i\phi/2} \right]} \\ &= \overline{E_1 E_1^*} + \overline{E_2 E_2^*} + \overline{E_1 E_2^* e^{i\phi}} + \overline{E_1^* E_2 e^{-i\phi}} \end{aligned} \quad (\text{IV.35})$$

Let

$$\overline{E_1 E_2^*} = \left| \overline{E_1 E_2^*} \right| e^{i\theta} \quad (\text{IV.36})$$

then

$$\overline{E_1^* E_2} = \left| \overline{E_1^* E_2} \right| e^{-i\theta} \quad (\text{IV.36a})$$

and

$$\overline{E_R E_R^*} = \overline{E_1 E_1^*} + \overline{E_2 E_2^*} + \left| \overline{E_1 E_2^*} \right| e^{i\phi} e^{i\theta} + \left| \overline{E_1^* E_2} \right| e^{-i\phi} e^{-i\theta} \quad (\text{IV.37})$$

If

$$\left| \overline{E_1 E_2^*} \right| = \left| \overline{E_1^* E_2} \right| \quad (\text{DEFINITION OF "NON-COHERENCE" OF SOURCE S}) \quad (\text{IV.38})$$

it follows that

$$\overline{E_R(t)E_R(t)^*} = \overline{E_1 E_1^*} + \overline{E_2 E_2^*} + 2 \left| \overline{E_1 E_2^*} \right| \cos(\phi + \theta) \quad (\text{IV.39})$$

It is now appropriate to compare equation (IV.39) to the interference fringe equation, obtained in Young's experiment with two perfectly coherent sources, namely

$$I = |E_1|^2 + |E_2|^2 + 2 \sqrt{E_1 E_1^*} \sqrt{E_2 E_2^*} \cos \phi \quad (\text{IV.40})$$

or

$$I = I_1 + I_2 + 2\sqrt{I_1}\sqrt{I_2}\cos\phi \quad (\text{IV.40a})$$

It is seen from equation (IV.40) that one can write equation (IV.39) as

$$\overline{E_R(t)E_R(t)^*} = \overline{E_1E_1^*} + \overline{E_2E_2^*} + 2|\gamma_{12}|\sqrt{\overline{E_1E_1^*}}\sqrt{\overline{E_2E_2^*}}\cos(\phi+\theta) \quad (\text{IV.41})$$

where we have defined a

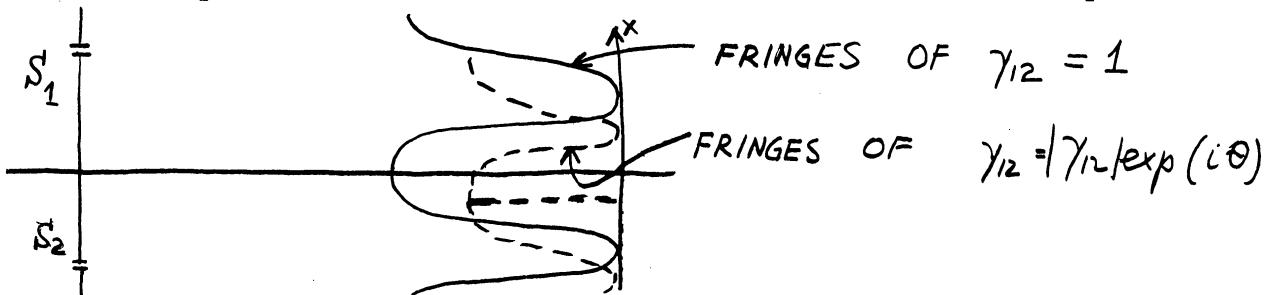
PARTIAL COHERENCE FACTOR

$$\gamma_{12} = \frac{\overline{E_1E_2^*}}{\sqrt{\overline{E_1E_1^*}}\sqrt{\overline{E_2E_2^*}}} = \frac{|\overline{E_1E_2^*}| \exp(i\theta)}{\sqrt{\overline{E_1E_1^*}}\sqrt{\overline{E_2E_2^*}}} \quad (\text{IV.42})$$

One can also write equation (IV.41) as

$$\overline{E_R(t)E_R(t)^*} = I_1 + I_2 + 2|\gamma_{12}|\sqrt{I_1}\sqrt{I_2}\cos(\phi+\theta) \quad (\text{IV.41a})$$

The meaning of the partial coherence factor γ_{12} becomes immediately apparent by comparing equation (IV.41) to equation (IV.40). It is seen that the magnitude $|\gamma_{12}|$ of the partial coherence factor is a measure of the contrast in the two slit interference fringe system formed by S_1 and S_2 when they are illuminated by the source S . The phase θ of the partial coherence factor is a measure of the phase shift



of the fringe maxima, compared to the phase obtained by perfectly coherent slits. For instance, θ measures the position-shift along x in the observation plane.

It is clear that

$$\begin{aligned} \gamma_{12} = 1 &\longrightarrow \text{perfect coherence} \\ \gamma_{12} = 0 &\longrightarrow \text{perfect incoherence} \\ 0 < |\gamma_{12}| < 1 &\longrightarrow \text{partial coherence} \end{aligned} \quad (\text{IV.43})$$

The meaning of θ can be illustrated for a few special cases. For example:

$$\phi + \theta = 2m\pi \quad \rightarrow \quad \cos(\phi + \theta) = \text{MAX} = +1$$

$$\text{and } \overline{E_R E_R^*} = (\overline{E_R E_R^*})_{\text{MAX}} = I_1 + I_2 + 2|\gamma_{12}| \sqrt{I_1} \sqrt{I_2}$$

$$\phi + \theta = (2m+1)\pi \quad \rightarrow \quad \cos(\phi + \theta) = \text{MIN} = -1$$

$$\text{and } \overline{E_R E_R^*} = (\overline{E_R E_R^*})_{\text{MIN}} = I_1 + I_2 - 2|\gamma_{12}| \sqrt{I_1} \sqrt{I_2}$$

The meaning of $|\gamma_{12}|$ can be illustrated by comparison to the case of perfect coherence, described by equation (IV.40) and (IV.40a). Assuming $I_1 = I_2$ one has

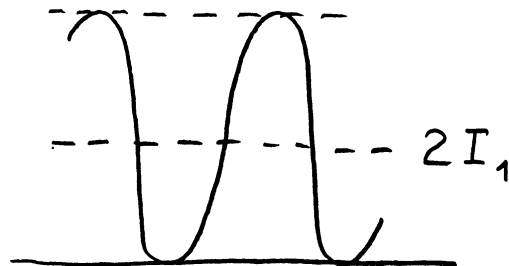
$$I_C = 2I_1 (1 + \cos \phi) \quad (\text{IV.44})$$

and therefore

$$I_{C, \text{MAX}} = 4I_1$$

and

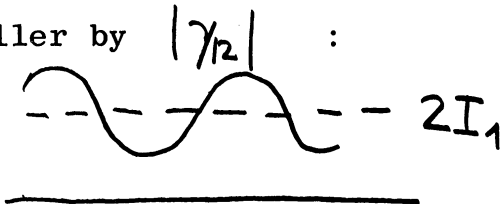
$$I_{C, \text{MIN}} = 0$$



For the case of $|\gamma_{12}| < 1$ the partially coherent fringes can be described (for the case $|\gamma_{12}| = \sqrt{I_1} = \sqrt{I_2}$) by

$$I_{\text{PC}} = 2I_1 |\gamma_{12}| [1 + \cos(\phi + \theta)] \quad (\text{IV.45})$$

The fringe contrast is then smaller by $|\gamma_{12}|$:

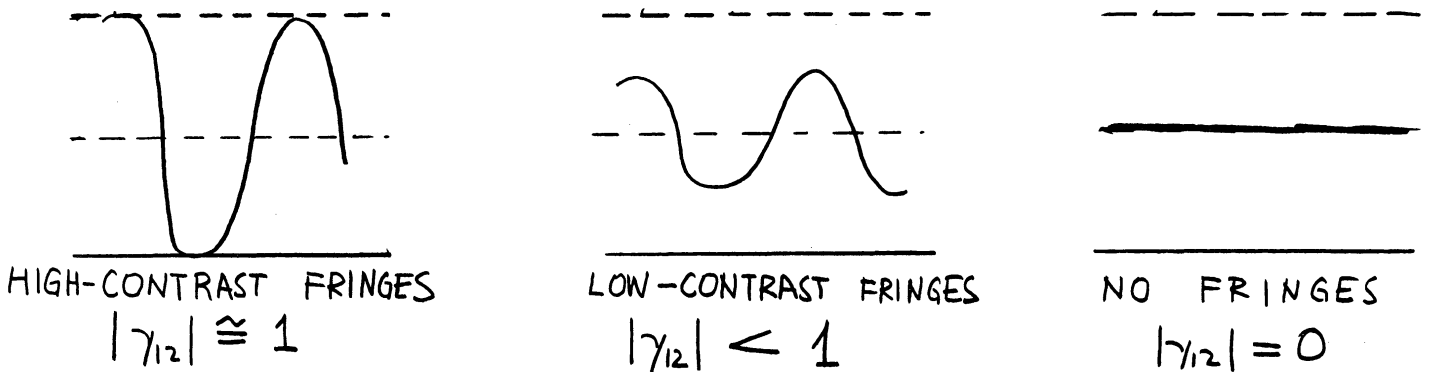


For completely incoherent sources, $|\gamma_{12}| = 0$ and

$$(\overline{E_1 E_2^*})_{\text{MAX}} = (\overline{E_1 E_2^*})_{\text{MIN}} \quad \text{---} 2I_1$$

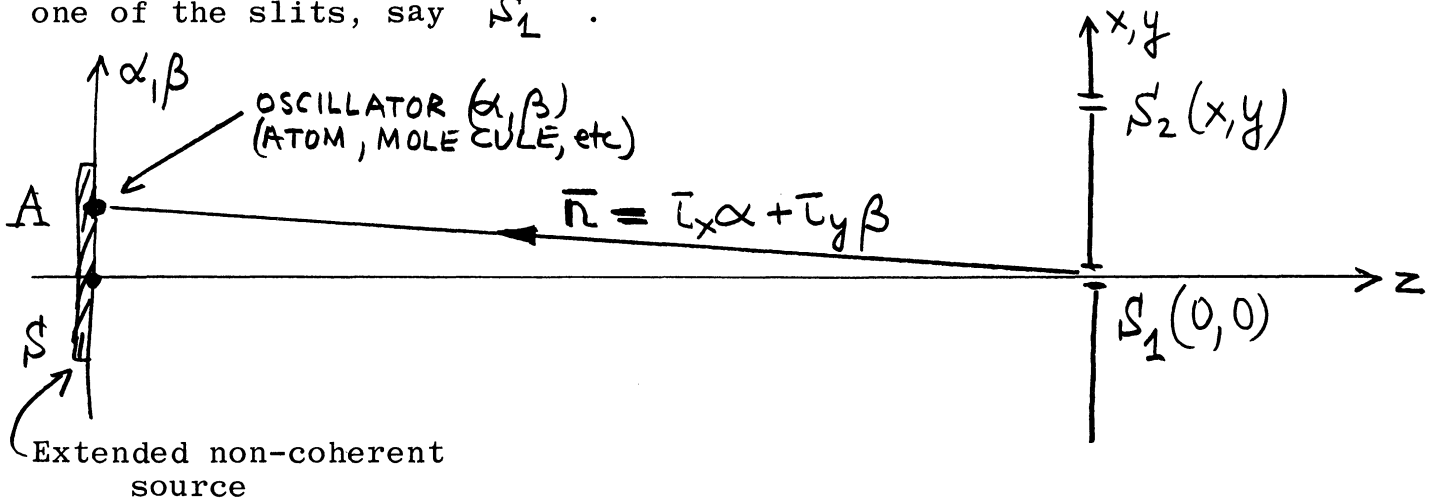
Consequently, the fringe contrast is zero.

As a general remark, it should be noted that the mere observation of fringes, by placing two slits into the field (for example at the exit of a laser only shows that there is partial coherence in the field. In order to truly describe the degree of coherence in the field, it is necessary to measure the fringe contrast.

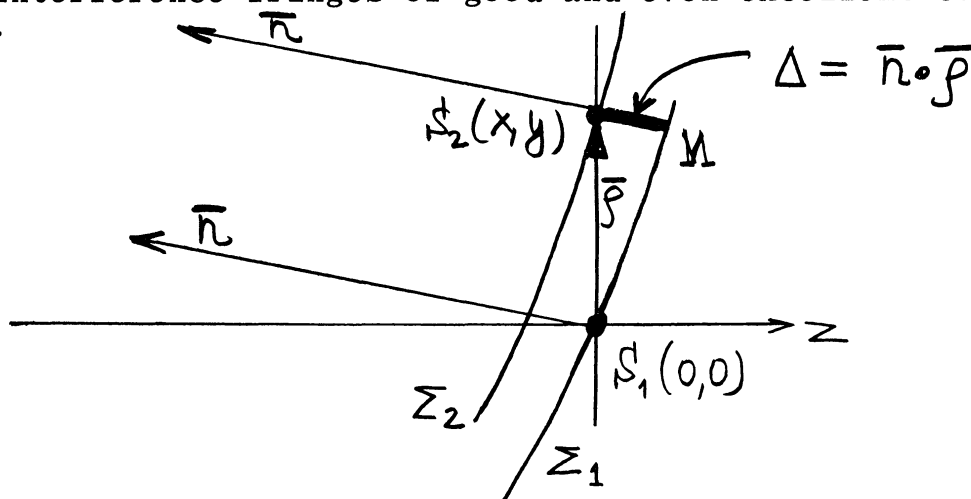


9. Partial coherence with extended non-coherent source.

Let a source S illuminate the two slits S_1, S_2 , as shown. The source is centered on the z -axis, normal to the xy plane at one of the slits, say S_1 .



The source is perfectly non-coherent, according to the preceding section. That is to say that no interference fringes can be obtained by placing two slits in the plane of the source. It will be shown, however, that if the two slits are placed far enough away from the non-coherent source, interference fringes of good and even excellent contrast can be obtained.



Consider the two wave-fronts Σ_1 , and Σ_2 originating from and oscillator in the direction \bar{n} as seen from S_1 . One has

$$\bar{n} = \bar{l}_x \alpha + \bar{l}_y \beta \quad (\text{IV.46})$$

where α and β are the direction cosines, and \bar{l}_x, \bar{l}_y the unit vectors. Let

$$\bar{\rho} = \bar{l}_x x + \bar{l}_y y \quad (\text{IV.47})$$

describe the coordinates of $S_2(x, y)$ with respect to $S_1(0, 0)$. The path difference between the two wavefronts Σ_1 and Σ_2 from A is

$$\Delta = \bar{n} \cdot \bar{\rho} = \alpha x + \beta y \quad (\text{IV.48})$$

The electric fields at S_1 and S_2 , resulting from the oscillator A are therefore

$$\text{At } S_1 : \quad E_{s_1}(t) = E(t)$$

$$\text{At } S_2 : \quad E_{s_2}(t) = E(t) \exp[-ik(\alpha x + \beta y)] \quad (\text{IV.49})$$

In case of many atoms A_i , one has

$$\text{At } S_1 : \quad E_{s_1}(t) = \sum E_i(t)$$

$$\text{At } S_2 : \quad E_{s_2}(t) = \sum E_i(t) \exp[-ik(\alpha_i x + \beta_i y)] \quad (\text{IV.50})$$

Monochromatic light is assumed. In practice, only reasonable monochromaticity is necessary for the observation of fringes in two-slit measurements, when the fringes are observed with reasonably small total σ from the source to the observation plane. (Filters are used in the case of polychromatic sources).

In order to evaluate the fringe contrast, it is necessary to obtain an expression for

$$\overline{E_{s_1}(t) E_{s_2}(t)^*}$$

One has

$$E_{s1}(t) = E_1(t) + E_2(t) + \dots$$

$$E_{s2}(t) = E_1(t) \exp(-i\phi_1) + E_2(t) \exp(-i\phi_2) + \dots \quad (\text{IV.51})$$

and therefore

$$\overline{E_{s1}(t) E_{s2}(t)^*} = \overline{[E_1 + E_2 + \dots][E_1^* \exp(+i\phi_1) + E_2^* \exp(+i\phi_2) + \dots]}$$

i.e.

$$\overline{E_{s1}(t) + E_{s2}(t)^*} = \overline{E_1 E_1^* \exp(i\phi_1)} + \overline{E_2 E_2^* \exp(i\phi_2)} \quad (\text{IV.52})$$

$$+ \left[\overline{E_2 E_1^* \exp(i\phi_1)} + \overline{E_3 E_1^* \exp(i\phi_1)} + \text{other cross terms} \right]$$

However, all of the cross terms are equal to zero, as a result of the non-coherence of the source.

The only terms remaining in (IV.52) are

$$\overline{E_{s1}(t) E_{s2}(t)^*} = \overline{E_1 E_1^* \exp(i\phi_1)} + \overline{E_2 E_2^* \exp(i\phi_2)} + \dots$$

i.e. (IV.53)

$$\overline{E_{s1}(t) E_{s2}(t)^*} = \sum E_i(t) E_i(t)^* \exp[+ik(\alpha_i x + \beta_i y)]$$

In the limit of an infinitely great number of oscillators, equation (IV.53) becomes

$$\overline{E_{s1}(t) E_{s2}(t)^*} = \iint_{\text{SOURCE}} I_s(\alpha, \beta) \exp[ik(\alpha x + \beta y)] d\alpha d\beta \quad (\text{IV.54})$$

where

$$I_s(\alpha, \beta) = \frac{\overline{E_i E_i^*}}{d\alpha d\beta} = \text{energy per unit angular area of the source at given } \alpha, \beta \quad (\text{IV.55})$$

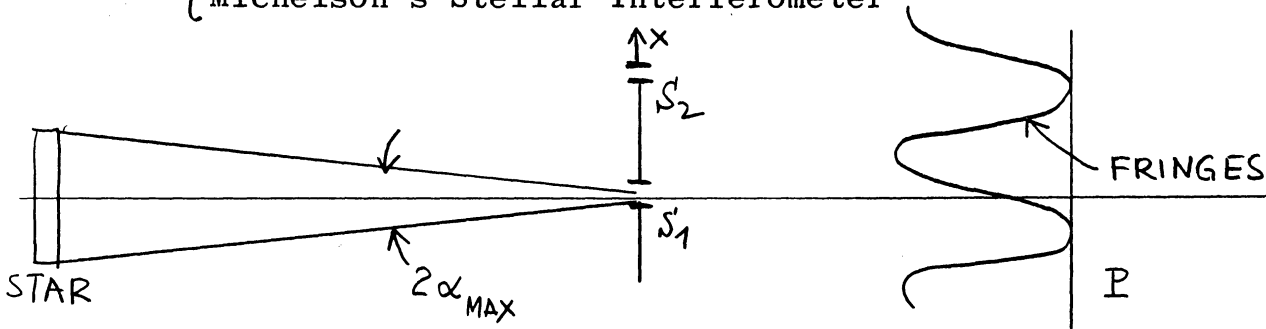
In other words, $I_s(\alpha, \beta)$ is the apparent intensity distribution in the source as seen from S_1 .

It follows immediately from (IV.54) and (IV.55) that the partial coherence factor γ , describing the coherence between S_1 and S_2 when illuminated by the extended non-coherent source is given by

$$\gamma = \frac{\iint_{SOURCE} I_s(\alpha, \beta) \exp[ik(\alpha x + \beta y)] d\alpha d\beta}{\iint_{SOURCE} I_s(\alpha, \beta) d\alpha d\beta} \quad (IV.56)$$

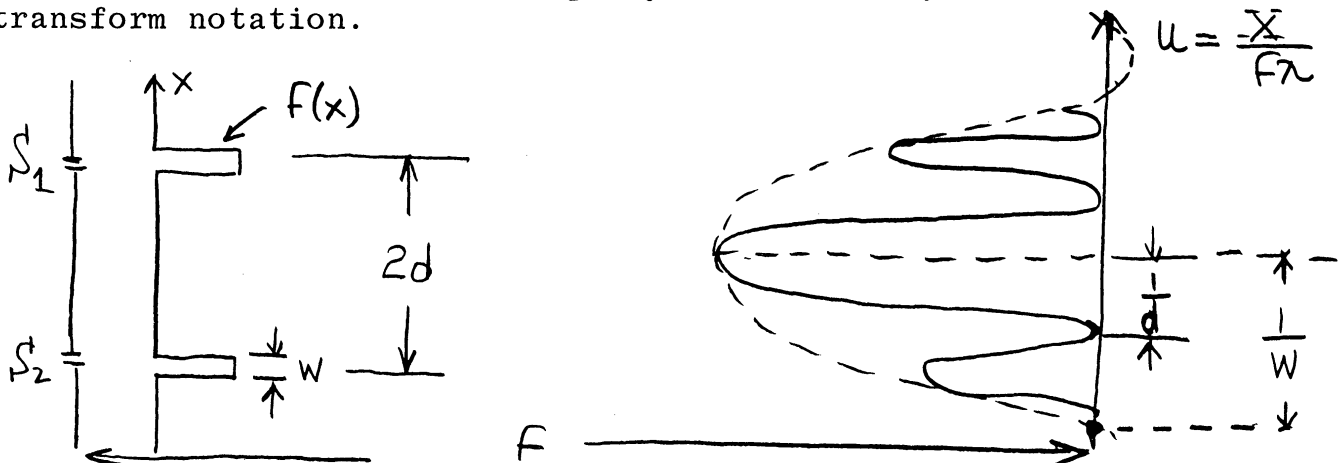
Equation (IV.56) is very remarkable indeed. It shows that: the degree of partial coherence between two points S_1 and S_2 illuminated by an extended, non-coherent source is given by the Fourier transform of the intensity distribution $I_s(\alpha, \beta)$ of the source as seen from the $S_1 S_2$ plane. It is quite essential to note that the source is centered on S_1 . This will be clarified in the following example.

EXAMPLE: { Two Slit Interference.
 Michelson's Stellar Interferometer



Let $I_s(\alpha, \beta) = I_s$, a constant. Experiment shows that an interference fringe system is observed, under suitable circumstances, in the plane P.

The nature of the fringe system is easily described in Fourier transform notation.



Indeed, if

$$f(x) \longrightarrow F(u)$$

then

$$f(x-d) \longrightarrow F(u) \exp[2\pi i u d] \quad (\text{IV.57})$$

and

$$f(x+d) \longrightarrow F(u) \exp[-2\pi i u d]$$

Therefore

$$f(x+d) + f(x-d) \longrightarrow 2 F(u) \cos 2\pi u d \quad (\text{IV.58})$$

and the observed intensity is

$$I(u) = 4 |F(u)|^2 \cos^2 2\pi u d \quad (\text{IV.59})$$

that is

$$I(u) = 2 |F(u)|^2 (1 + \cos 4\pi u d) \quad (\text{IV.60})$$

where the envelope

$$\begin{aligned} F(u) &= \mathcal{T}[f(x)] \\ &= \int_{-w/2}^{+w/2} \exp(2\pi i u x) dx \\ &= w \frac{\sin(\pi i u w)}{\pi i u w} \end{aligned} \quad (\text{IV.61})$$

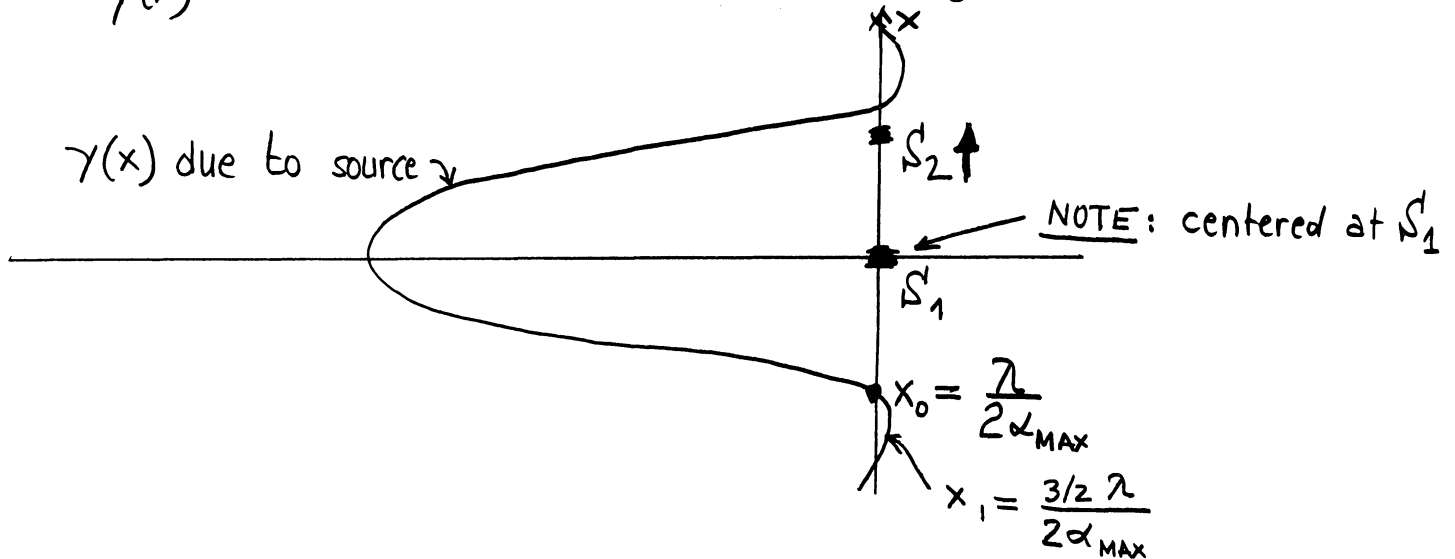
and

$$|F(u)|^2 = w^2 \left[\frac{\sin(\pi i u w)}{\pi i u w} \right]^2 \quad (\text{IV.62})$$

The contrast in this fringe system is obtained as

$$\begin{aligned} \gamma(x) &= \frac{\int_{-\alpha_{\text{MAX}}}^{\alpha_{\text{MAX}}} I_s \exp[-i k \alpha x] d\alpha}{\int_{-\alpha_{\text{MAX}}}^{\alpha_{\text{MAX}}} I_s d\alpha} \\ &= \frac{\sin(k \alpha_{\text{MAX}} x)}{k \alpha_{\text{MAX}} x} = \frac{\sin\left(\frac{2\pi}{\lambda} \alpha_{\text{MAX}} x\right)}{\left(\frac{2\pi}{\lambda} \alpha_{\text{MAX}} x\right)} \end{aligned} \quad (\text{IV.63})$$

The $\gamma(x)$ due to the star is shown in the figure



It is clear that the apparent angular diameter of the star can be obtained by simply determining the distance $2d$ of S_2 from S_1 , such that the fringes disappear.

The remarkable feature of this experiment is that there is always a separation S_1, S_2 at which good fringe contrast is obtained with an extended non-coherent source.

It is sometimes said that the coherence in light beams increases with distance "by mere propagation". What is happening is simply, with all dimensions (source, S_1, S_2) constant, that the separation S_1, S_2 becomes increasingly smaller compared to $\gamma(x)$ as the distance of the slits from the source increases.

10. Intensity correlations in partially coherent fields.

Much more can be said about the coherence properties of light than is sufficient for the experimental characterization given above. The reader should consult, for example.

1. L. Mandel, "Fluctuations of Light Beams", in Progress in Optics, Vol. II., edited by E. Wolf, North-Holland Publishing Co. (1963) pp 183-248.
2. "Quantum and Statistical Aspects of Light." Selected Reprints. Published by the American Institute of Physics (1963), \$2.00.
3. "Quantum Electronics III", edited by P. Grivet and N. Bloembergen, Columbia University Press (1964).

Much of the basic theory of coherence was developed by

P. H. Van Cittert, Physica 1, 201 (1934)

F. Zernike, Physica 5, 785 (1938)

H. H. Hopkins, Proc. Roy. Soc. A208, 263 (1951);
ibid. A217, 408 (1953).

A. Blanc-Lapierre and P. Dumontet, Rev. Optique,
34, 1 (1955)

E. Wolf, Nuovo Cimento 12, 884 (1954)

A. Maréchal and M. Françon, "Diffraction",
 ed. Revue d'Optique (1960)

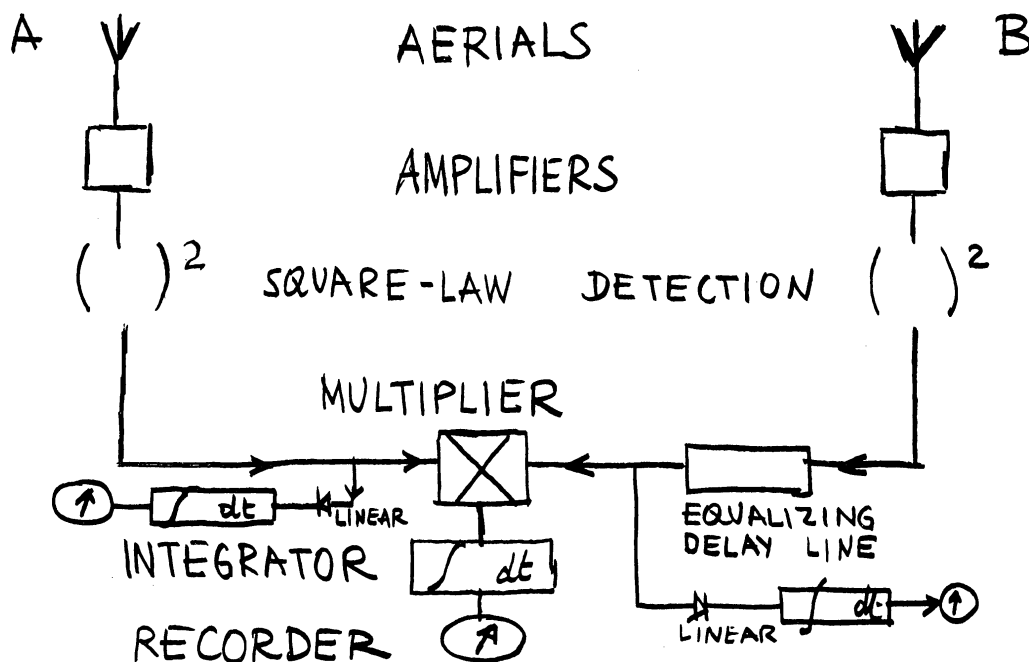
as well as by A. Einstein, and others.

Only one more of the many aspects of coherence will be examined, in order to further illustrate the simplicity which results from the use of modern image-formation theory and Fourier-transform methods in dealing with coherence.

11. Intensity Interferometers

A great deal of interest in the understanding of the coherence characteristics of light was created by some classically unexpected and not easily understandable results of an experiment reported by Hanbury Brown and Twiss in Nature, 177, 28 (1956). Basically, the experimental arrangement is quite similar to the two-slit experiments described above. However, instead of observing the interference fringe-system formed at some distance behind the two slits S_1 and S_2 , photoelectric receivers are placed directly behind the slits, and their outputs are correlated. A perfectly equivalent arrangement consists of correlating the outputs of two radio-astronomical antennae, following square-law detection. The interpretation of the experiment in terms of image formation theory, is immediate and perfectly straightforward*.

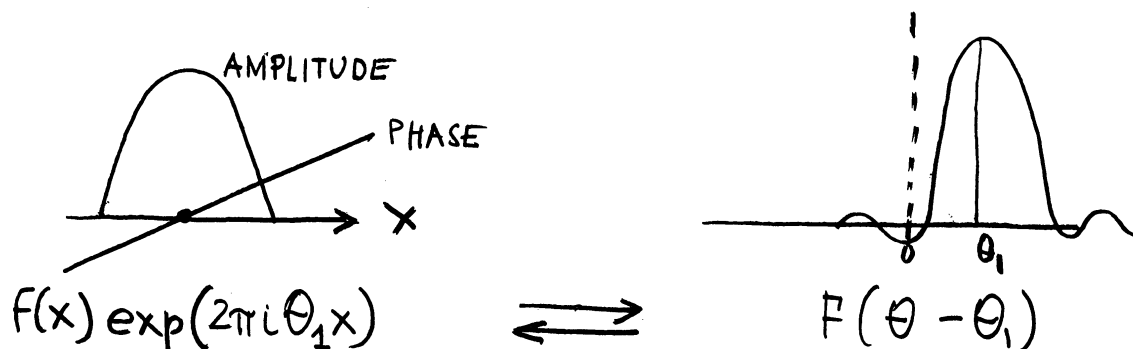
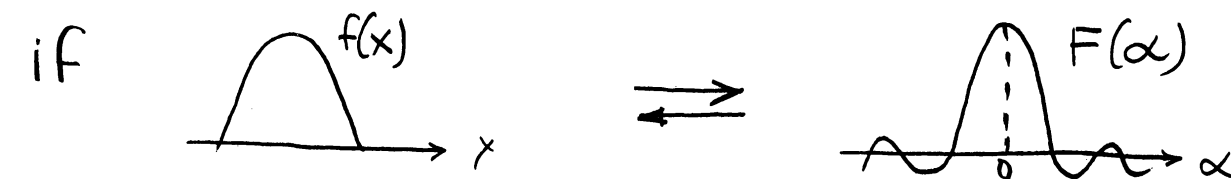
Consider the radio-astronomical experimental arrangement illustrated:



*See, for example, R. C. Jennison, "Fourier transforms and convolutions for the experimentalist", Pergamon-Press, The Macmillan Co., New York (1961).

It will be shown that the intensity (or post-detector) interferometer permits to measure the $\gamma(x)$ resulting from the star, just like Michelson's stellar interferometer. The significant difference between the two interferometers is that the intensity interferometer is much less subject to atmospheric fluctuations and considerably less subject to the mechanical and optical stability required in Michelson's interferometer.

First note that



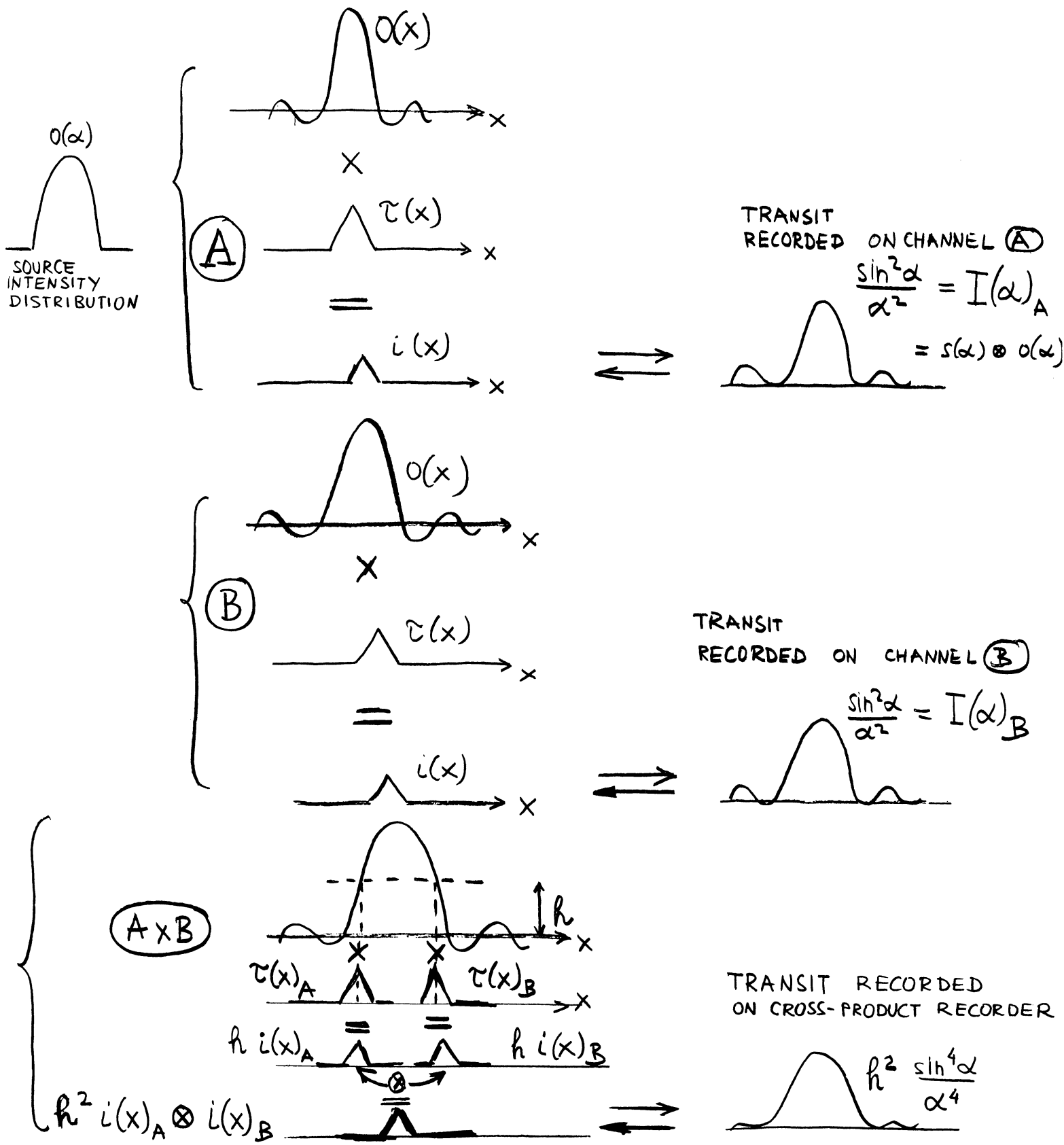
This is to say that if $f(x)$ is the image of the source as seen from one of the antennae, its image as seen from the other antenna is $f(x) \exp(2\pi i \theta_1 x)$. The corresponding functions in the frequency domain are shown on the right. The way in which the stellar diameter can be measured will now be explained, with the help of non-coherent image-formation theory.

In fact, a graphical explanation is most suitable.

The analysis is based on taking the transfer functions for the geometrical image of the star, $o(x)$, separately for each antenna, then convolving the two "images" $h_i(x)_A$ and $h_i(x)_B$, according to the operation of the multiplier and integrator. The intensity recorded on the cross-product recorder is shown to be $R^2(\sin^2 \alpha) / \alpha^4$ from which the intensity distribution across the star is immediately obtained by varying the antenna-distance (which in turn, results in the desired variation of R).

The graphical analysis is given on the following page.

INTENSITY INTERFEROMETER
 ANALYSIS



V. IMAGE FORMATION IN COHERENT LIGHT

V.1. Introduction

The basic characteristic of coherent-light image forming systems is that complex amplitudes, rather than intensities in the field add before recording. Of course, just like in the case of non-coherent image-formation, only the intensity $\overline{EE^*}$ of the resultant field can be recorded.

It is essential to note, however, that what is recorded is the resultant intensity. There is no law that prevents a coherent background field, heterodyning and interferometric methods from being used to suitably modulate or indeed demodulate the "coherent" signal field.

A basic example of an interferometric heterodyning method, as used in wave-front reconstruction, was described in the introduction (Section I.6.).

In the final analysis, coherent-light optical systems still appear to permit more flexible image processing, for instance in communication systems, than non-coherent systems.

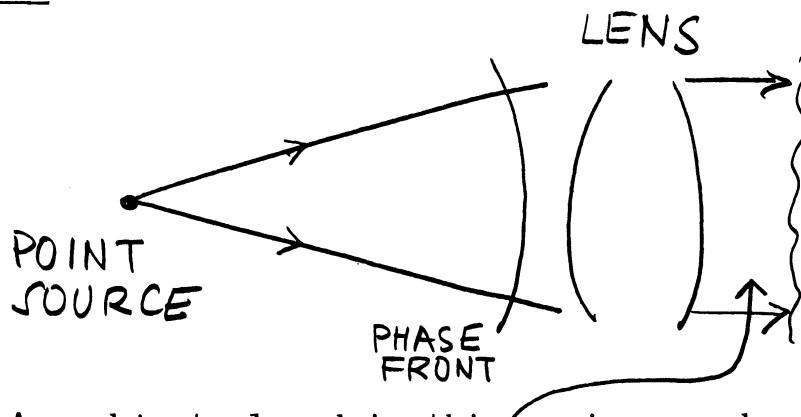
Only some of the most basic concepts are introduced here.

V.2. "Coherent" Illumination

Coherence, in this chapter, is referred to principally in terms of spatial coherence.

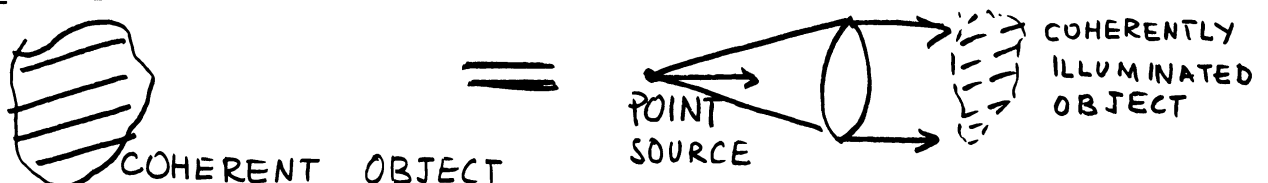
Three examples illustrate coherent illumination:

Example 1



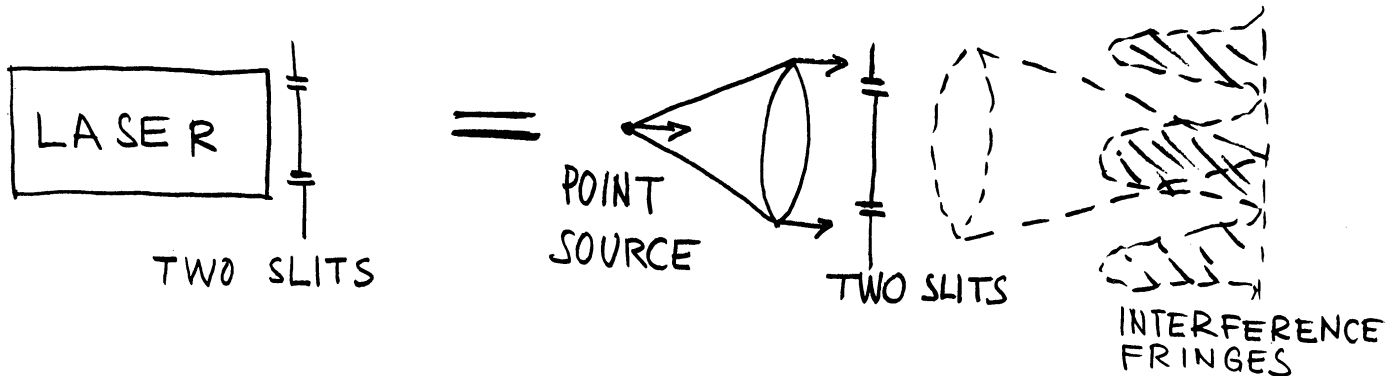
Any object placed in this region can be considered as coherently illuminated.

Example 2: Equivalence



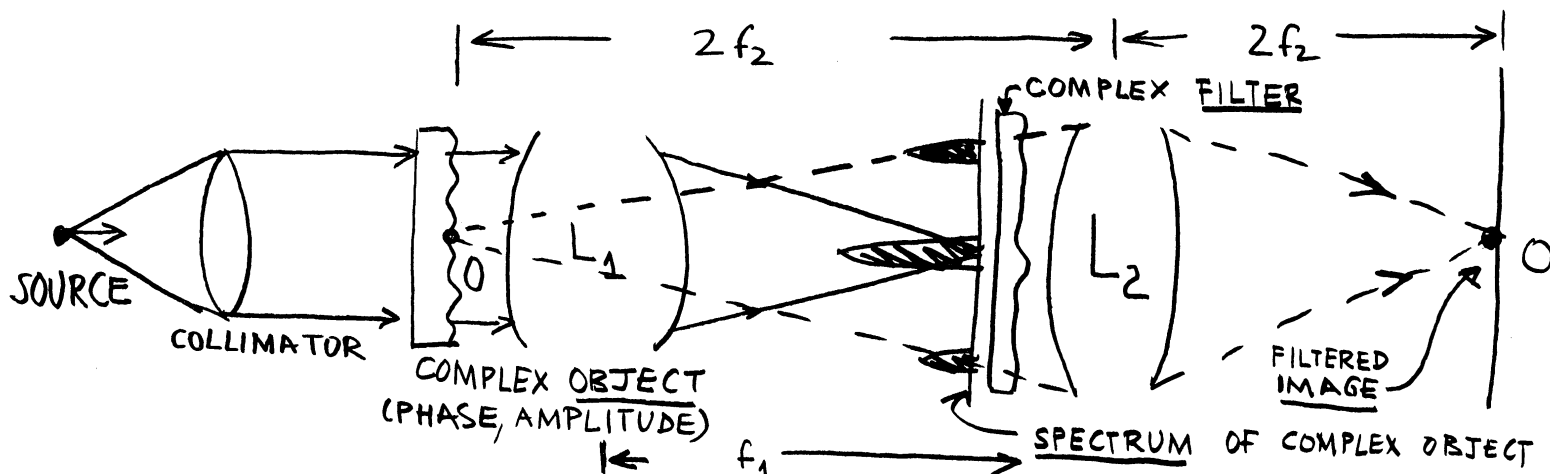
This coherently illuminated object must conserve uniform phase transmission, if it is primarily an amplitude object.

Example 3 Equivalence



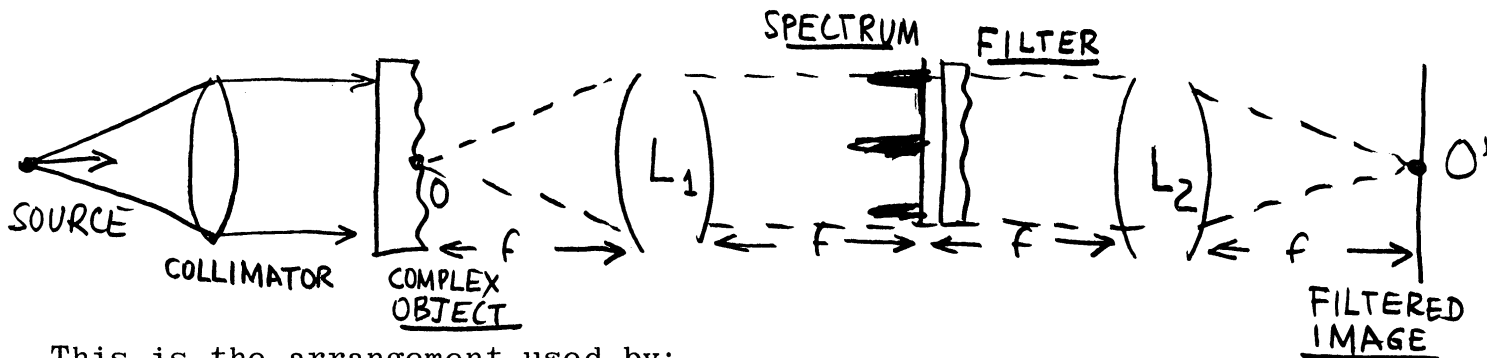
V.3. Image Formation in Coherent Light, Considered as Double Diffraction.

In many cases, it is convenient to consider image formation in coherent light as a "double-diffraction" process.



This is the arrangement used by:

- Zernike (1935)
- Maréchal and Croce (1953)
- E. L. O'Neill (1956)

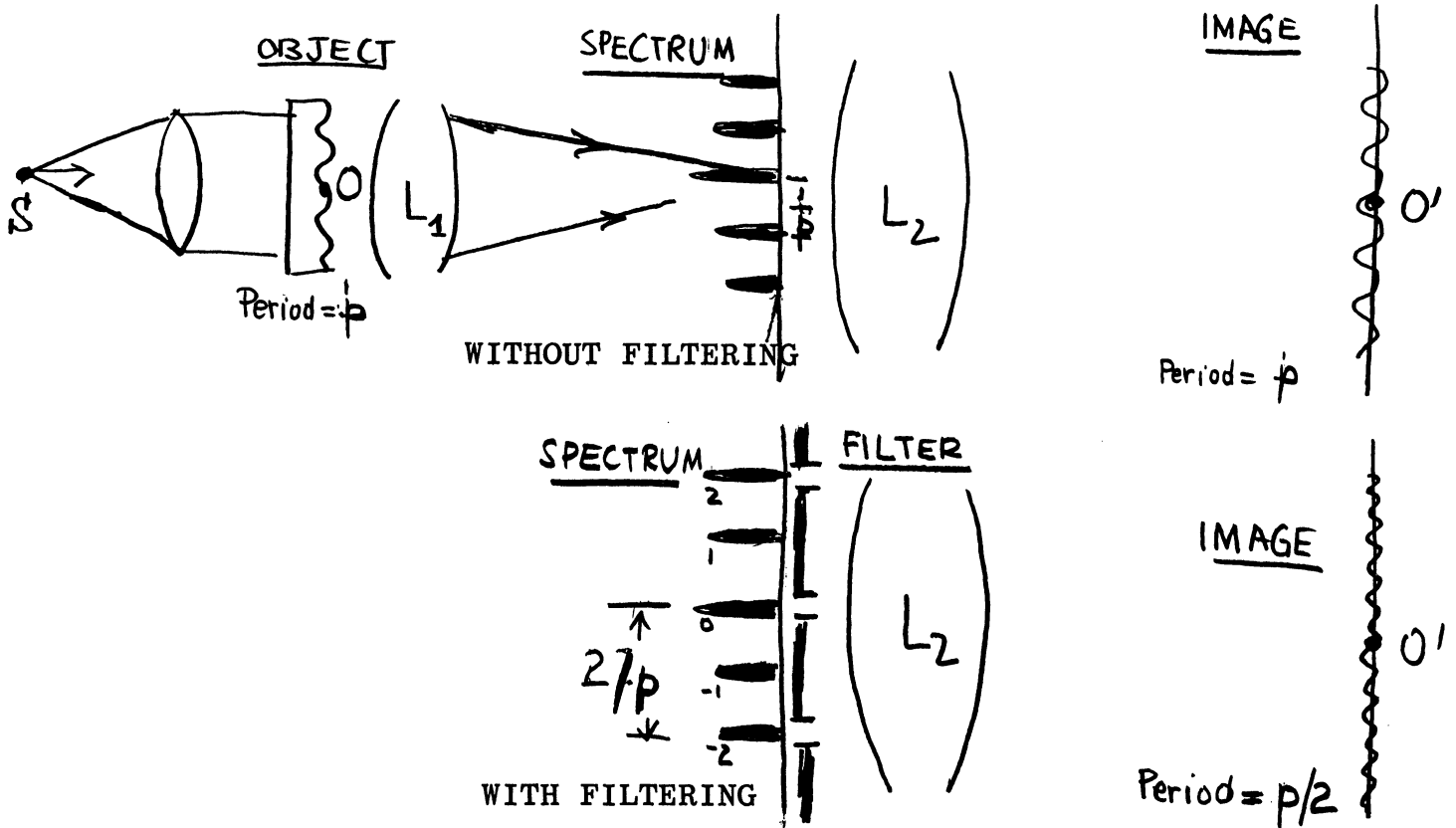


This is the arrangement used by:

- Cutrona, Leith, Palermo and Porcello (1960)

An example will illustrate the double diffraction concept. (Abbe, Wood, Maréchal).

Consider the phase grating shown.



The filter (in this case a simple mask) has singled out only the 0, 2, 4, etc. harmonics in the spectrum.

V.4. Abbe Resolution Criterion
 (Another example of double diffraction concept)

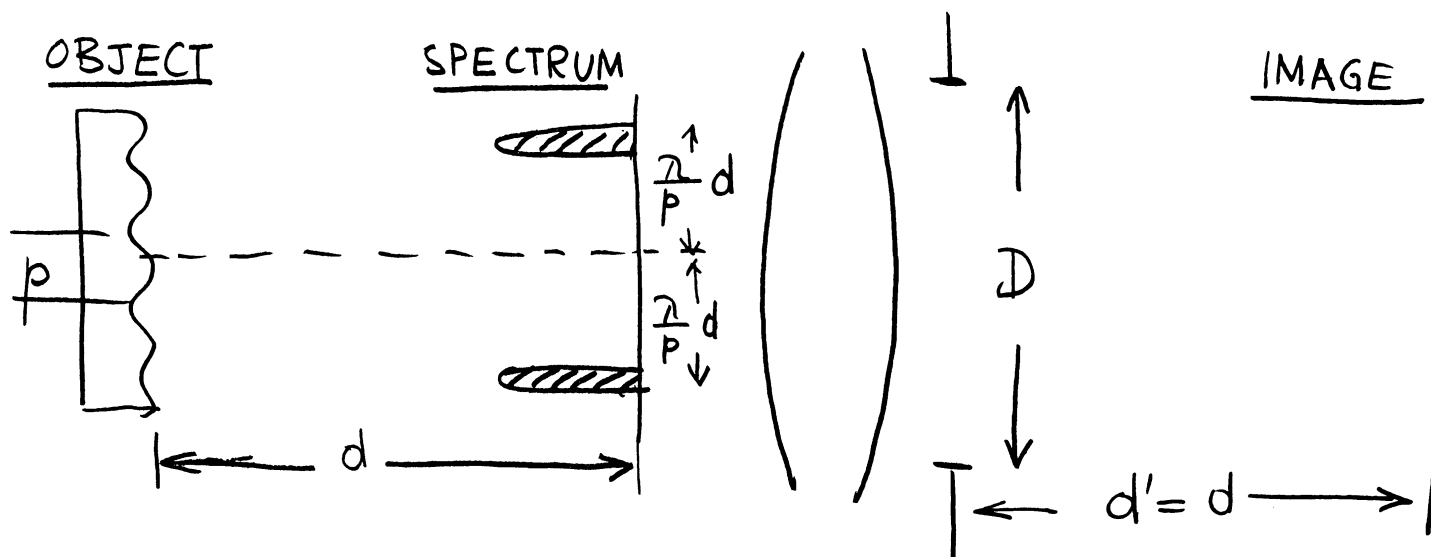
Consider the coherently illuminated periodic amplitude object of period p

$$2 F(u) \cos\left(2\pi \frac{u}{p}\right)$$

producing a diffraction image (spectrum),

$$f\left(x + \frac{1}{p}\right) + f\left(x - \frac{1}{p}\right)$$

according to Fourier transform theory.



Let D be the aperture of the objective lens looking at the object. If the two spectral "lines" at $\pm (\lambda/p) d$ get transmitted through the aperture, the amplitude in the image, according to Fourier-transform theory is

$$A(u) = T \left[f\left(x + \frac{1}{p}\right) + f\left(x - \frac{1}{p}\right) \right]$$

$$= 2 F(u) \cos\left(2\pi \frac{u}{p}\right)$$

and the intensity in the image is

$$I(u) = A(u) A(u)^*$$

$$= 4 |F(u)|^2 \cos^2\left(2\pi \frac{u}{p}\right)$$

$$= 2 |F(u)|^2 \left(1 + \cos 2\pi \frac{u}{p}\right)$$

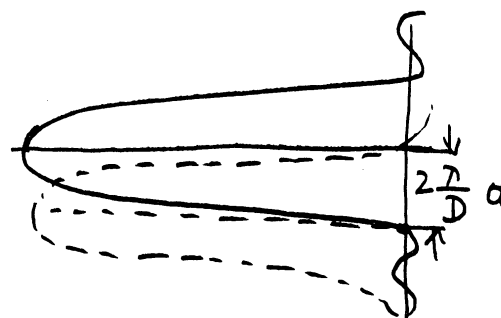
It is seen that the "image" has the same periodicity as the amplitude object. In fact, the object is fully "resolved".

As long as the two spectral lines pass through the aperture, complete resolution will be obtained. It is seen that resolution, without loss of contrast, is obtained up to the value of

$$\frac{\lambda}{p} d \leq \frac{D}{2}$$

that is up to

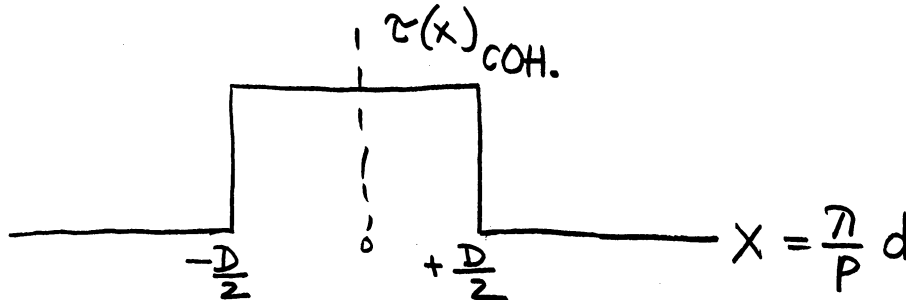
$$p \geq \frac{2\lambda}{D} d$$



in agreement with the Rayleigh criterion (see figure, recall operation in coherent light here).

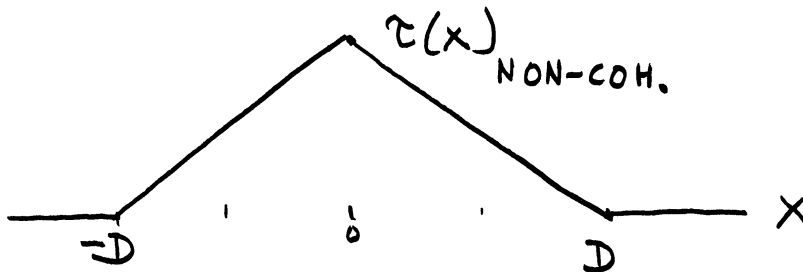
V.5. Transfer Functions in Coherent and in Non-Coherent Light.

From the preceding section, we can represent the frequency transfer function $\tau(x)$ by the graph

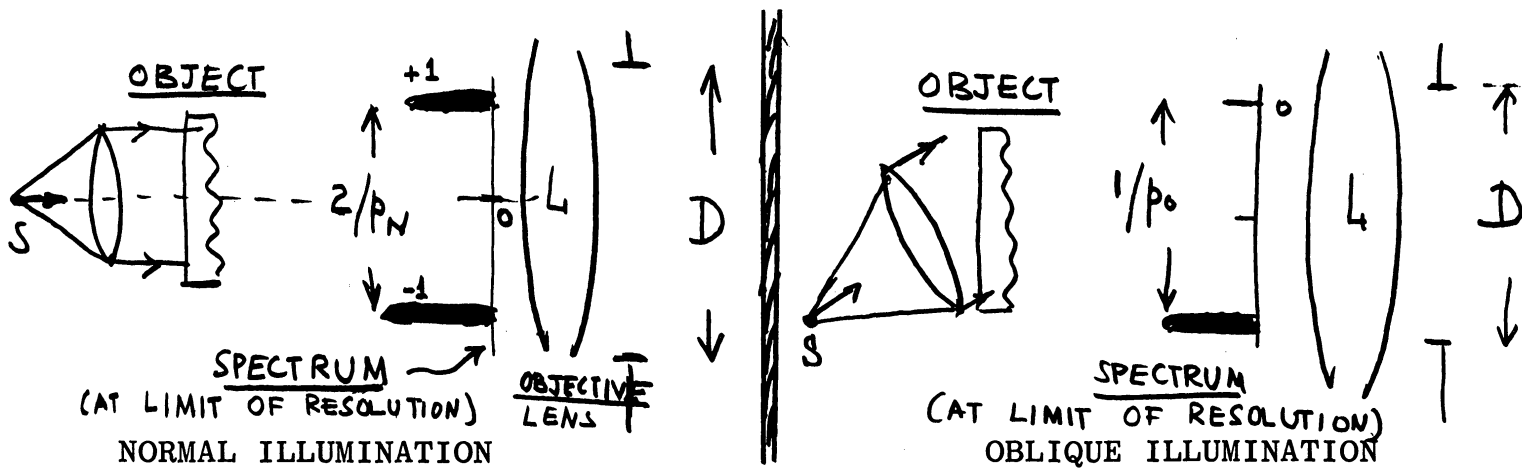


showing that transmission without loss of contrast is obtained from $x=0$ to $x = \frac{D}{2} = \frac{\pi}{p} d$, and then the contrast is zero.

We recall, for comparison, the frequency response for non-coherent light:



Finally, in terms of the double-diffraction concept, it is easy to understand why the resolution appears to be doubled, in coherent light with the use of oblique illumination



Analysis shows that the single side-band imaging obtained with oblique illumination does not represent faithful imaging. However, there is no doubt that increased "resolution" is obtained.

V.6. Phase-contrast Filtering

The concepts of double diffraction permit a very simple interpretation of the phase-contrast filtering used in microscopy, in order to transform unobservable small phase variations into easily observable intensity variations. We recall that F. Zernike received the Nobel Prize for this work in 1953.

Consider an object $f(x,y)$ which is a pure phase object, i.e.

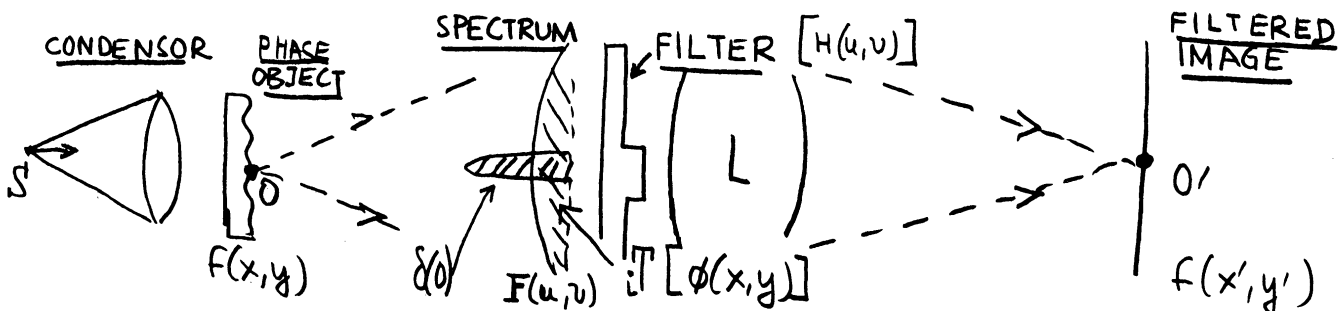
$$f(x,y) = \exp [i \phi(x,y)] \tag{V.1}$$

Let the phase $\phi(x,y)$ have sufficiently small values, so that one can write

$$f(x,y) \cong 1 + i \phi(x,y) + \dots \tag{V.2}$$

It follows that

$$f(x,y) \rightarrow F(u,v) = \delta(0) + iT [\phi(x,y)] \tag{V.3}$$



Let the filter $H(u,v)$ be chosen so that

$$\left. \begin{aligned} H(u,v) &= it && \text{over the region } \delta(0) \\ &= 1 && \text{over the remaining spectrum} \end{aligned} \right\} \quad (V.4)$$

where t = amplitude transmission factor ($0 \leq t \leq 1$).

The "filtered spectrum" is then

$$\begin{aligned} G(u,v) &= P(u,v) H(u,v) \\ &= it \delta(0) + i T [\phi(x,y)] \end{aligned} \quad (V.5)$$

By the second "diffraction", the image $f(x',y')$ is

$$\begin{aligned} f(x',y') &= T [G(u,v)] \\ &= it + i \phi(x,y) \end{aligned}$$

that is

$$f(x',y') = i [t + \phi(x,y)] \quad (V.6)$$

The intensity in the image is

$$i(x',y') = f(x',y') \cdot f(x',y')^* = t^2 + \phi^2 + 2\phi t \cong t^2 + 2\phi t \quad (V.7)$$

The intensity of the background is t^2 and the contrast of the phase portions of the image is

$$C \cong \frac{2\phi t}{t^2 + 2\phi t} = \frac{2\phi}{t + 2\phi} \quad (V.8)$$

With $t \leq 1$, it is seen that a very high contrast is obtained even for extremely small values of ϕ . (For instance, with $t = 10^{-3}$, $\phi = 10^{-3}$ radians, $\rightarrow C = 2/3$).

Equation (V.8) is to be compared with the contrast which would be obtained without filtering. One would have

$$i(x',y')_{\text{UNFILTERED}} \cong 1 + \phi^2(x,y) \quad (V.9)$$

and the contrast of the phase portion would be

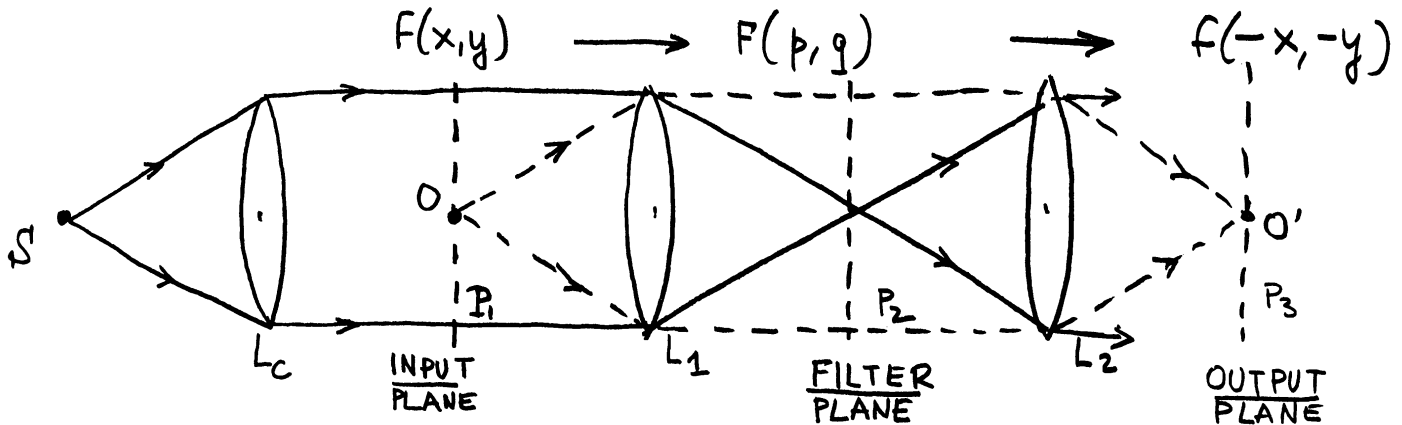
$$C_{UNFILTERED} \approx \frac{\phi^2}{1} = \phi^2 \quad (V.10)$$

giving, for $\phi = 10^{-3} \rightarrow C = 10^{-6}$.

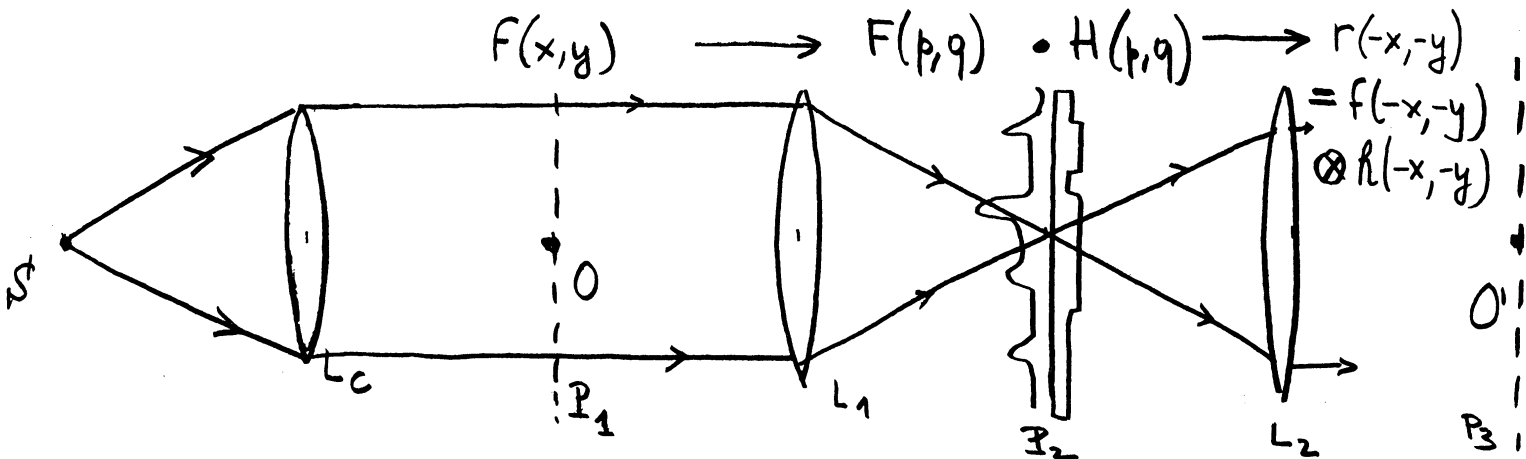
The phase contrast method of selective filtering is a basic example of the "matched" optical filtering.

V.7. Optical Filtering with Interferometrically Matched Spatial Filters.

Consider the system shown in the figure:



The system shown is recognized as one of the coherent-light imaging systems. The object O , in the P_1 plane is imaged one to one in the P_3 plane. If a suitable spatial frequency filter is used in the P_2 plane, it is possible to selectively filter out of the image O' any desired frequency, in a way quite comparable in principle to the filtering methods used in electrical communication systems.



OPTICAL IMAGE FILTERING SYSTEM

It should be clear that the entire process between S and O' must be carried out in coherent light, and, therefore, that the spatial filter $H(p,q)$ must be a ^acomplex filter.

As an example of a filtering operation, one may wish to determine the location of the word CHRIST in a page of the "Letter of Paul to the Philippians". The location of the word CHRIST will be marked by a bright dot of light, corresponding to the convolution

$$f(-x,-y) \otimes h(-x,-y)$$

indicated on page 76, where

$$\text{and } H(p,q) = \mathcal{R}(x,y) = \mathcal{T}[H(p,q)]$$

complex filter for the word CHRIST.

The photographs on page 81, illustrate the three principal steps in a coherent-light filtering operation. The top figure shows the section of the transparency in which the location of the word CHRIST is to be determined. The center photograph (center-left) shows the "filter" obtained by recording the diffraction pattern of the word CHRIST in the Fourier-transforming arrangement illustrated below, together with a coherent background. The similarity between the "complex filter" and a "Fourier-transform hologram" is readily recognized. A "Fourier-transform reconstruction" of the word CHRIST, obtained in the focal plane of a lens by illuminating the filter in collimated light is shown in the (Center-right) figure.

Finally, the filtered image is shown at the (bottom-right) together with a zero-order image, both in the "output plane".

The entire work was carried out on P/N Polaroid Film.

This work shown in these photographs was carried out in the Electro-Optical Sciences Laboratory, at The University of Michigan, under the direction of Professor George W. Stroke.

Realization of the Complex Filter and of Filtering Operation

Let the desired filter be

$$H(p, q) = |H| e^{j\theta} \quad (V.11)$$

Let $h(x, y)$ be a transparency of the word CHRIST which is to be filtered,

$$h(x, y) = \text{CHRIST} = \text{"filter signal"} \quad (V.12)$$

Let

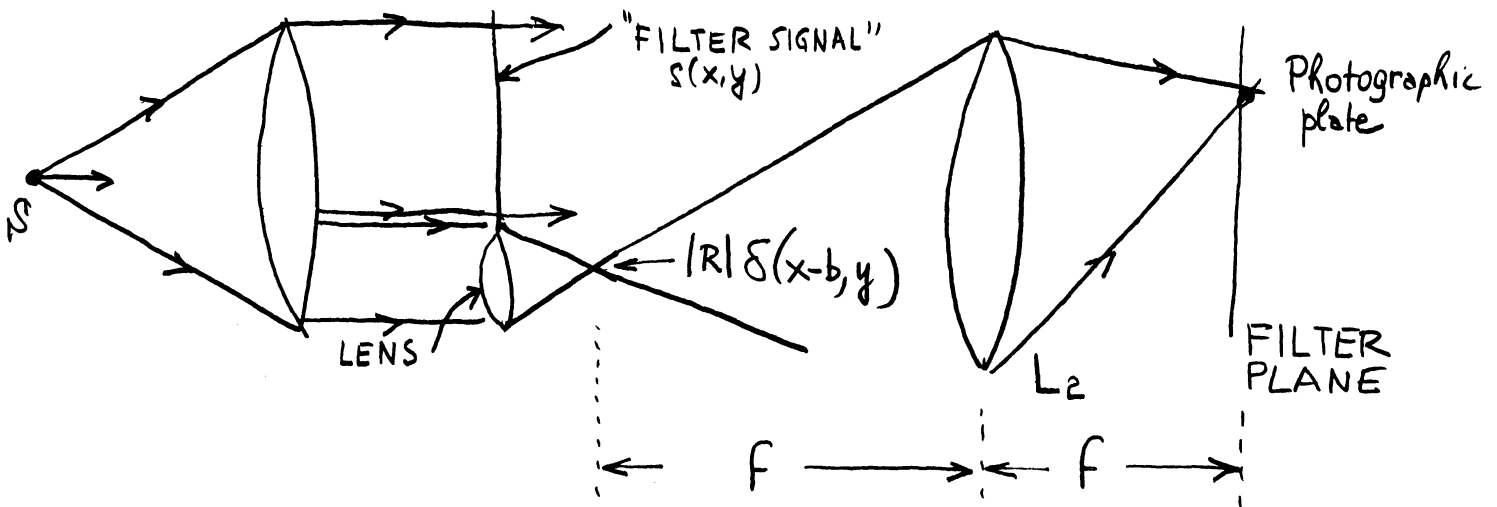
$$h(x, y) \rightarrow H(p, q) = |H| e^{j\theta} \quad (V.13)$$

by Fourier transformation, in the focal plane of L_2

Let the reference signal be

$$|R| e^{j\phi} \quad (V.14)$$

produced as shown in the figure.



One has, in the focal plane of L_2 :

$$h(x, y) + |R|\delta(x-b, y) \rightarrow |H| e^{j\theta} + |R| e^{j\phi} \quad (V.15)$$

It is clear that

$$|R|e^{j\phi} = |R| \exp [2\pi j (pb + q)] \quad (V.16)$$

where

$$\phi = pb + q = \text{linear phase function} \quad (V.17)$$

The intensity $G(p, q)$ recorded on the filter is

$$G(p, q) = [|R|e^{j\phi} + |H|e^{j\theta}] [|R|e^{-j\phi} + |H|e^{-j\theta}] \quad (V.18)$$

that is

$$G(p, q) = |R|^2 + |H|^2 + |R||H|e^{j(\phi-\theta)} + |R||H|e^{-j(\phi-\theta)} \quad (V.19)$$

In other words, the recorded filter is

$$G(p, q) = [|R|^2 + |H|^2] + |H|e^{j\theta} |R|e^{-j\phi} + |H|e^{-j\theta} |R|e^{j\phi} \quad (V.20)$$

However, the desired filter is $|H|e^{j\theta}$.

Question: How can the desired filter $|H|e^{j\theta}$ be separated from the recorded filter $G(p, q)$?

Answer: The separation is carried out "automatically" by the lens L_2 in the optical image-filtering system shown above (p.76).

Indeed, let $|R| = k = \text{some constant, say } 1$. Note that the phase is a linear phase function

$$\phi = 2\pi (bp + cq)$$

As a result, it is seen that the desired filter function $|H|e^{j\theta}$ has been placed on an interferometric grating carrier, such that the filtered output will appear off axis by amounts $x=b, y=c$. Indeed,

$$|H|e^{j\theta} e^{j\phi} \longrightarrow h(x+b, y+c) \quad (V.21)$$

and therefore,

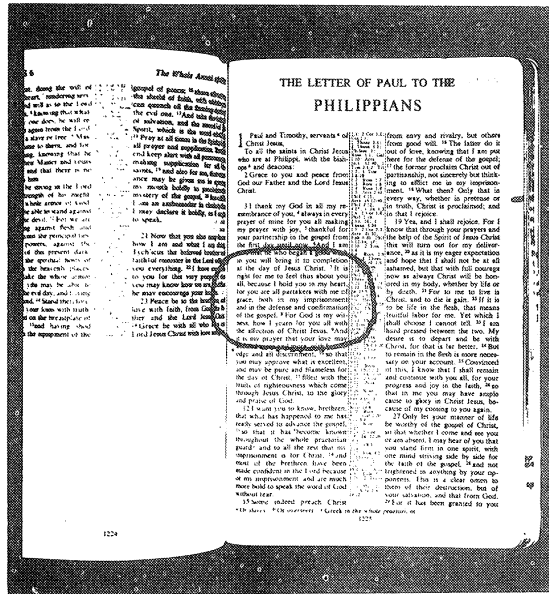
$$F(p, q) \cdot |H| e^{j\theta} e^{j\phi} \longrightarrow f(-x, -y) \otimes h(x+b, y+c) \quad (V.22)$$

Comparison of equation (V.22) with the Optical Image Filtering System" figure (p.76) shows that the image has indeed been filtered as desired.

The other terms "filtered" by $G(p, q)$, some undesired and some desired will appear respectively on axis (the $|R|^2$ and $|S|^2$ terms), and on the other side of the axis (the $|H| e^{-j\theta} |R| e^{j\phi}$ term).

The close relation of spatial filtering work to aspects of optical holography will be further stressed in Chapter VII.

TEXT
BEING FILTERED

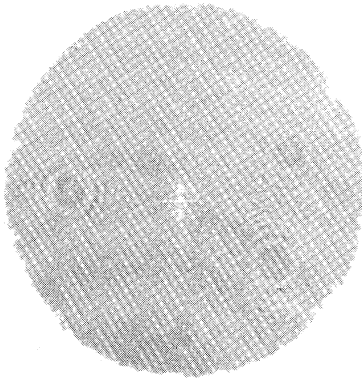


INPUT PLANE

$f(x,y)$

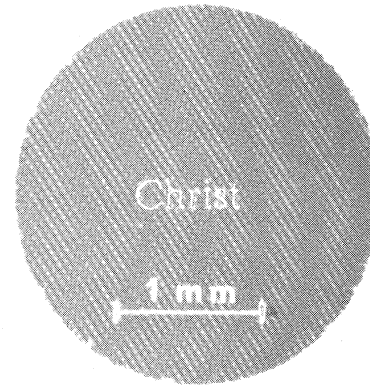
TEXT BEING FILTERED

FILTER
PLANE



$H(p,q)$

OUTPUT
PLANE

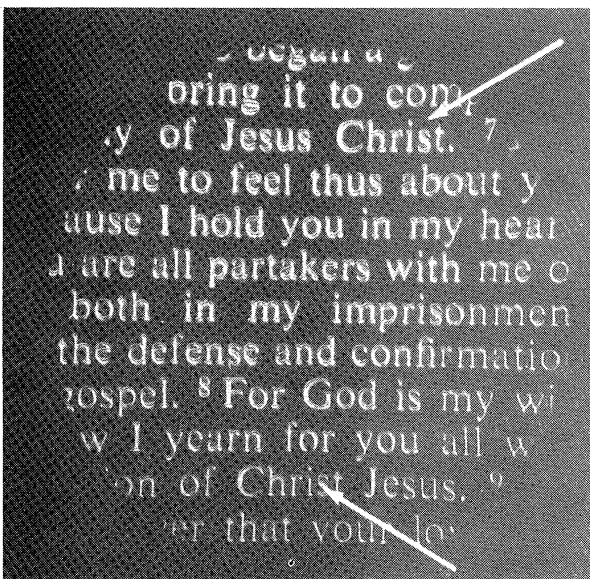


$h(-x,-y)$

FILTER
FOR WORD "CHRIST"

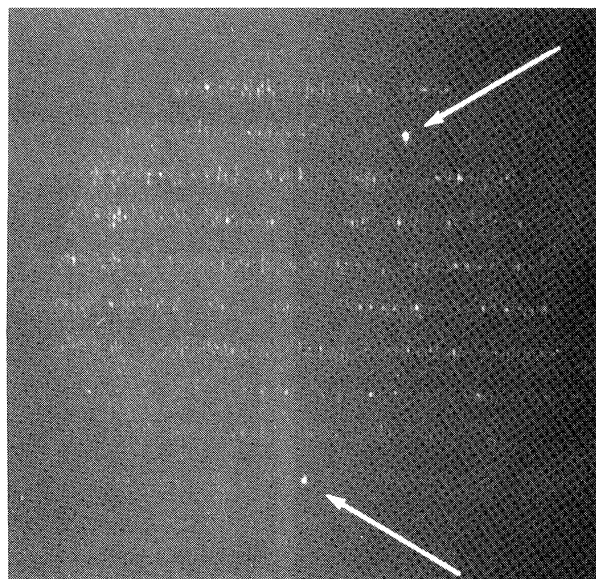
"RECONSTRUCTION"
FROM FILTER OF WORD "CHRIST"

INPUT
PLANE



ZERO-ORDER IMAGE

OUTPUT
PLANE



FILTERED IMAGE
(RIGHT-HAND SIDE BAND)

$f(-x,-y)$

$\otimes h(x+b,y+c)$

V.8. Optical Computing, Correlating and Signal Processing

Two principal types of optical signal processing systems, using coherent imaging, have been developed at the University of Michigan and at the Conduccion Corporation in Ann Arbor, as a result of the pioneering work of Professor Louis Cutrona and his associates.

These are:

1. Spectrum analysers.
2. Optical cross correlators.

A brief description of these systems now follows.

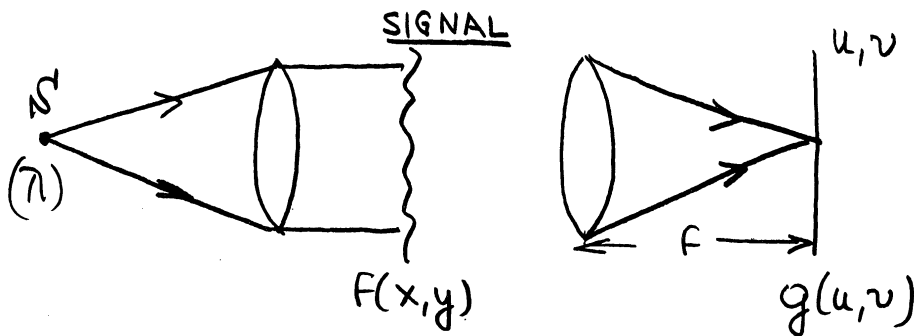
1. Spectrum analysers

Basically, an optical spectrum analyzer is comparable to an optical grating spectrometer using monochromatic radiation and an imperfectly spaced diffraction grating.

In an optical spectrum analyzer, the grating "imperfections" contain the significant, desired information, recorded on photographic film, for instance with the help of a cathode ray oscilloscope. Seismic vibration recordings are an example of signals recorded and analyzed in this manner. The analysis aims at determining the power "spectrum" of the recorded signal, forming the imperfect grating.

A complete theory of optical grating imperfections was given by G. W. Stroke in 1960*. Its results are immediately applicable to the spectral analysis as carried out in optical spectrum analyzers.

A schematic diagram of a spectrum analyzer is shown here:

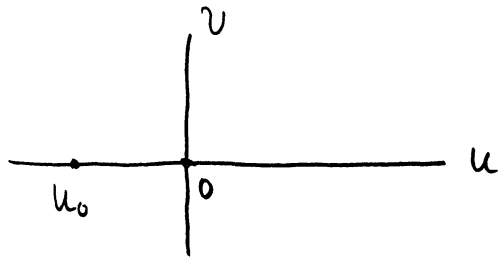


One has

$$g(u,v) = \iint_{-\infty}^{+\infty} F(x,y) e^{-\frac{2\pi i}{\lambda f} (ux+vy)} dx dy \quad (V.24)$$

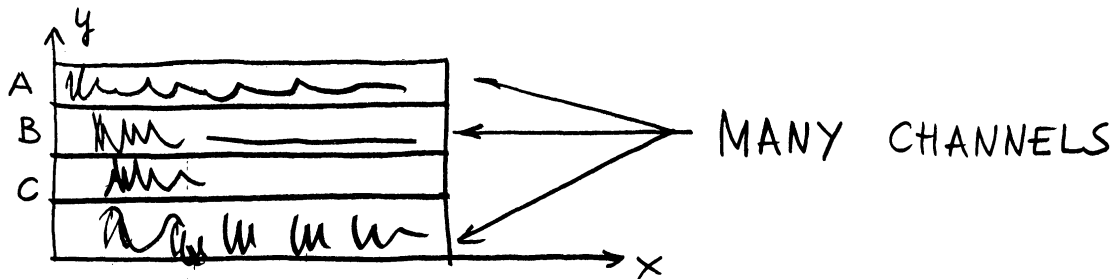
*G. W. Stroke, "Etude théorique et expérimentale de deux aspects de la diffraction par les réseaux optiques. L'évolution des défauts dans les figures de diffraction, et l'origine électromagnétique de la répartition entre les ordres", Revue d'Optique 39, 291-398, (1960).

In the (u, v) plane one has k_0 lines/mm

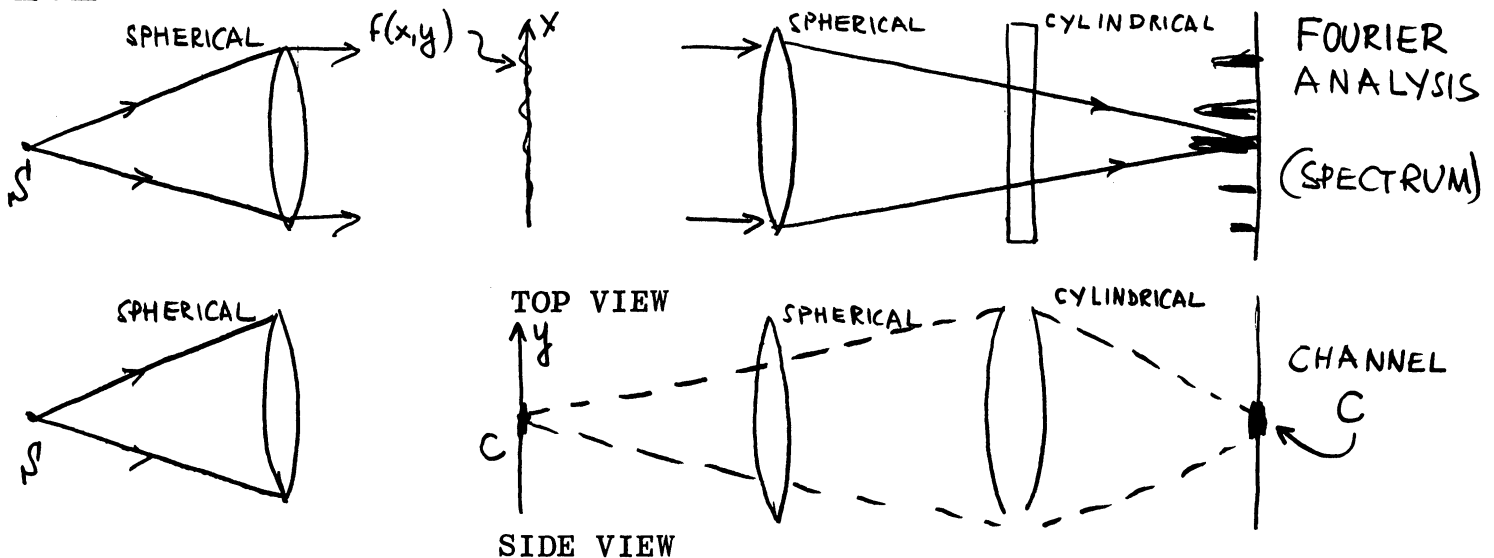


such that $u_0 = \lambda f k_0$. The difference with spectrometers used in the analysis of optical radiations is that here one investigates k_0 rather than λ

The important feature of optical spectrum analyzers is that they can be designed to perform the spectral analysis simultaneously for dozens, hundreds, and in principle thousands of channels.



The simultaneous analysis is based on the use of cylindrical optics. A two-plane view of a multi-channel analyzer is shown.

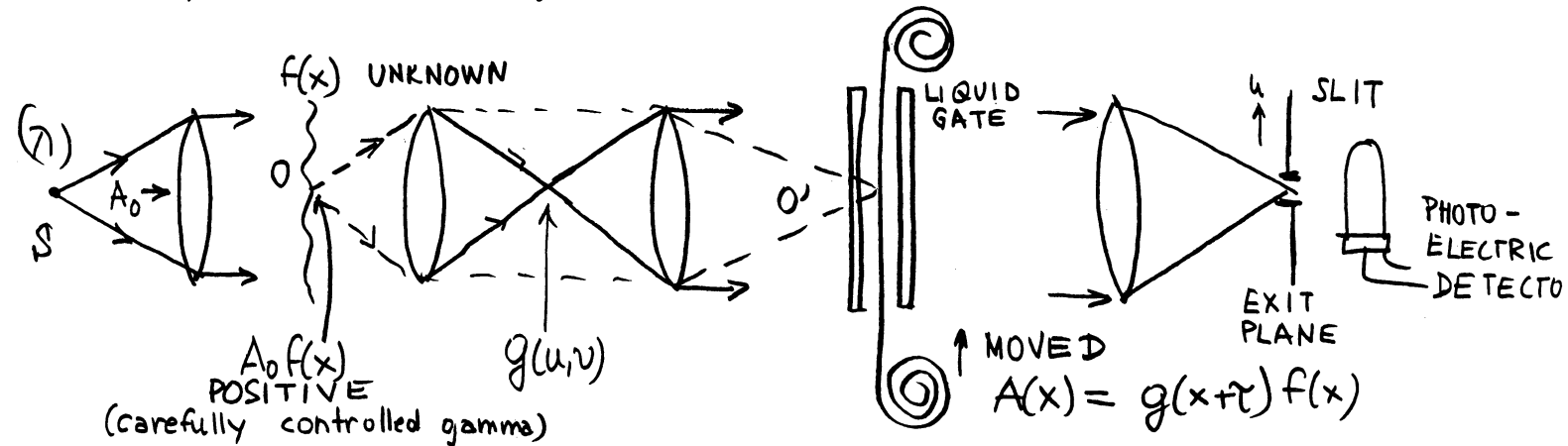


For example, 120 channel analysis has been carried out in some oil-industry application.

2. Optical cross-correlators

Cross correlation problems arise in such diverse fields as statistics, pharmaceutical product development and indeed target detection and identification.

In an optical cross-correlator, using coherent light, both the "known" and the "unknown" signals are recorded on film. Both films are generally immersed in liquid gates, and indeed one of the films, say the reference film $g(x)$ is moved through the liquid gate, at very uniform rate, in order to carry out the cross-correlation.



One obtains in the exit plane the function

$$A(u) = \mathcal{T}[A(x)] = \int_{-\infty}^{+\infty} f(x) g(x+\tau) e^{-\frac{2\pi i}{\lambda f} ux} dx \quad (V.25)$$

called the ambiguity function.

Note that if the exponential term is made equal to unity, then the output is equal to

$$A(u)_0 = \int_{-\infty}^{+\infty} f(x) g(x+\tau) dx \quad (V.26)$$

which is the desired cross-correlation. The desired reduction of the exponential term to unity is very simply accomplished by placing a slit (pinhole) on axis. Indeed, in this case $u=0$ and $e^0 = 1$. For multichannel cross-correlations, cylindrical optics and a slit in the exit plan are used, together with multiple photo-detectors, to obtain simultaneous analysis.

V.9. Note

Some additional background and details on material described in this chapter are given in University of Michigan technical reports by A. B. VanderLugt (Signal detection by complex spatial filtering) and by A. B. VanderLugt and F. B. Rotz (Data reduction by coherent optical systems), as well as in G. W. Stroke (University of Michigan, EE475 lecture notes, "Optics of Coherent and Non-Coherent Electromagnetic Radiations), and in technical literature available from the Conduccion Corporation in Ann Arbor (L. Cutrona and C. E. Thomas).

VI. CONVOLUTIONS, SPECTRAL ANALYSIS
AND THE THEORY OF DISTRIBUTIONS

VI.1. Introduction

The theory of distributions, as first described by Laurent Schwartz in 1950-51, and developed by many authors (in particular J. Arzac, A. Erdélyi, M. J. Lighthill, G. Temple and others) is proving to be an increasingly powerful tool in modern optics and electro-optical science. The reader is referred to one of the new treatises for even the basic definitions. However, a few examples will be used to illustrate the power of the method.

VI.2. Distributions

A scalar product of two functions $f(x)$ and $g(x)$ is defined by

$$\int_{-\infty}^{+\infty} f(x) g(x) dx = \langle f(x), g(x) \rangle \quad (\text{VI.1})$$

Even if a function $f(x)$ does not have a Fourier transform, one can define

$$\langle f(x), g(x) \rangle = \text{some number} \quad (\text{VI.2})$$

provided that $\int_{-\infty}^{+\infty} g(x) dx$ exists. In this case, $f(x)$ is called a distribution. It is recalled that a function can have a Fourier transform in the strictest sense only if it is absolutely integrable, that is if

$$\int_{-\infty}^{+\infty} |f(x)| dx \quad (\text{VI.3})$$

For instance, $f(x) = 1$ does not have a F.T. in this sense. However, let $f(x) = \delta(x)$ such that

$$\langle \delta(x) g(x) \rangle = g(0) \quad (\text{VI.4})$$

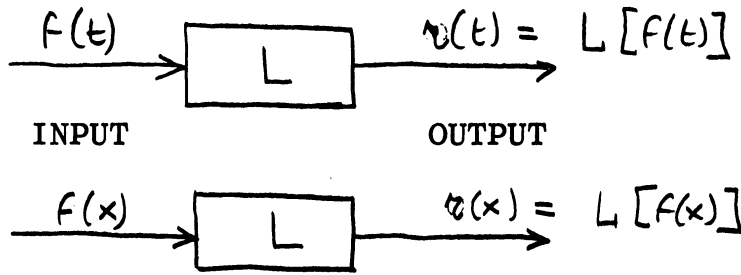
and

$$\langle \delta(x-a), g(x) \rangle = g(a) \quad (\text{VI.5})$$

Equations (VI.4) and (VI.5) define the Dirac delta function as a distribution.

VI.3. Impulse Response of a Linear System

Let L be a linear operator, operating on $f(t)$ or $f(x)$, such that



We require the linear system to have the properties of:

Linearity:

$$L[a_1 f_1(t) + a_2 f_2(t)] = a_1 L[f_1(t)] + a_2 L[f_2(t)] \quad (\text{VI.6})$$

with $a_1, a_2 = \text{constants}$ and

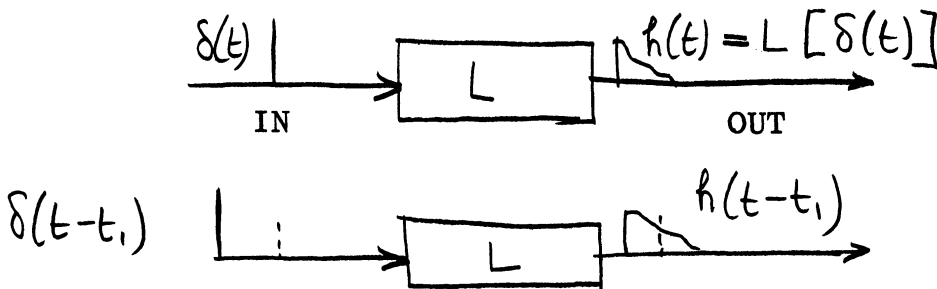
Time (or space) invariance

$$L[f(t-t_1)] = r(t-t_1)$$

$$L[f(x-x_1)] = r(x-x_1)$$

(VI.7)

The impulse response of the linear system (optical system, electrical network, etc.), is represented by $h(t)$ or $h(x)$ such that



where $h(t) = \text{impulse response of the system.}$

For an optical filter (f.ex. interference filter, optical network*) $h(t)$, the "time" impulse response is in order. For an optical image forming system, $h(x)$, the "space" impulse response is in order. (The "space" impulse response $h(x)$ is nothing but what is ordinarily called the "diffraction pattern").

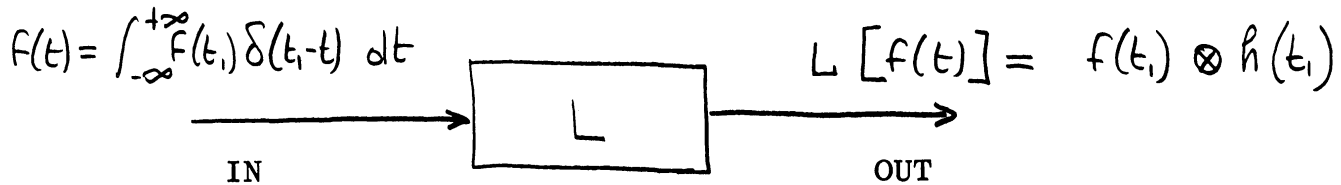
*See, for example, G. W. Stroke, "Optical Network and Filter Synthesis Using the Theory of Distributions", Proc. IEEE, 52, No.7(July 1964).

VI.4. Response of Linear System to Any Input Function

Now let the input function $f(t)$ be formed of a sum of unit impulses:

$$f(t) = \int_{-\infty}^{+\infty} f(t_1) \delta(t_1 - t) dt \quad (\text{VI.8})$$

Recalling that $\delta(t_1 - t) f(t_1) = f(t)$ according to (VI.5)



One immediately finds for the output function

$$\begin{aligned} L[f(t)] &= L \left[\int_{-\infty}^{+\infty} f(t_1) \delta(t_1 - t) dt_1 \right] \\ &= \int_{-\infty}^{+\infty} f(t_1) L[\delta(t_1 - t)] dt_1 \end{aligned} \quad (\text{VI.9})$$

because of the fact that L is linear. It follows that

$$L[f(t)] = r(t) = \int_{-\infty}^{+\infty} f(t_1) h(t_1 - t) dt_1 \quad (\text{VI.10})$$

showing the remarkable fact that the output $r(t)$ is equal to the convolution of the impulse response function with the input function.

In case of non-coherent image formation, for which the "diffraction pattern" or spread function is

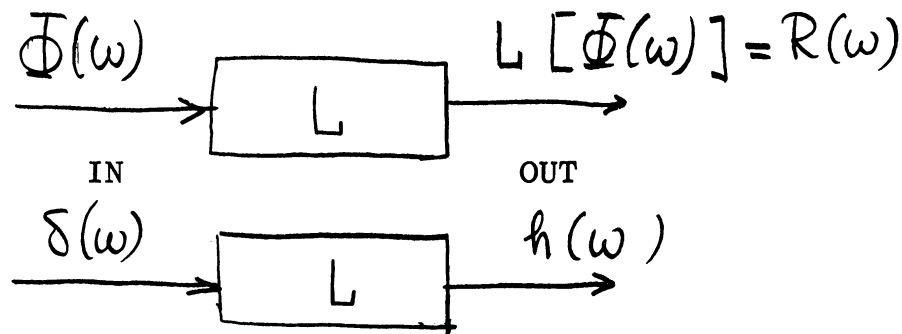
$$h(x) = s(x)$$

one has, for an (input geometrical image) $f(x) = o(x)$ the output $r(x) = l(x)$

$$l(x) = o(x) \otimes s(x) \quad (\text{VI.11})$$

VI.5. Spectral Analysis

Let $\Phi(\omega)$ be the spectral energy distribution in the input. The output $R(\omega) = L[\Phi(\omega)]$ is as shown



$R(\omega)$ may be the spectrum as recorded in a grating spectrometer, as a function of ω or λ , or the response in an optical spectrum analyzer using heterodyning detection.

Let the impulse response of the spectrometer be $h(\omega)$. Also, let the input $\Phi(\omega)$ be

$$\Phi(\omega) = \int_{-\infty}^{+\infty} \Phi(\omega_0) \delta(\omega_0 - \omega) d\omega_0 \quad (\text{VI.12})$$

One immediately has for the output

$$R(\omega) = \int_{-\infty}^{+\infty} \Phi(\omega_0) h(\omega_0 - \omega) d\omega_0 \quad (\text{VI.13})$$

that is

$$R(\omega) = \Phi(\omega) \otimes h(\omega) \quad (\text{VI.14})$$

which is recognized as the basic equation governing the behavior of spectrometers. $R(\omega)$ is the quantity actually recorded in the spectrometer. One has by Fourier transformation

$$\mathcal{T}[R(\omega)] = \mathcal{T}[\Phi(\omega)] \cdot \mathcal{T}[h(\omega)] \quad (\text{VI.15})$$

It is recalled, for optical as well as for electrical network systems, that

$$\mathcal{T}[R(\omega)] = \text{frequency response of system} \quad (\text{VI.16})$$

VII. THEORETICAL AND EXPERIMENTAL FOUNDATIONS OF OPTICAL
HOLOGRAPHY (WAVEFRONT-RECONSTRUCTION IMAGING) *

Introduction

Recent developments have generated a new interest in extensions of the wavefront reconstruction method first described by D. Gabor in 1948.⁽¹⁾⁽²⁾

It now appears possible to describe conditions under which and "x-ray microscope" might indeed be made to work so as to produce from x-ray holograms images which would be "true" well-resolved pictures of crystals, proteins, and molecules.

A great deal of work has been devoted to wavefront reconstruction since 1948.¹⁻²² Much progress has been made in developing methods for superposing the coherent background on the field scattered by the object, and for obtaining well-separated reconstructed "wavefronts" and images". By and large, the optimism and foresight expressed by Gabor in his 33-page paper in the 1949 Proceedings of the Royal Society have appeared to be justified in the 15-odd years which have followed his work.¹⁻⁵

It now appears that some of the principles required for new extensions of wavefront reconstruction methods may require additional clarification.

It is the purpose of this paper to discuss the theoretical and experimental principles of wavefront reconstruction imaging, as they now appear, in the light of new theoretical and experimental evidence. Four advances, two experimental and two theoretical, may be particularly singled out.

*This chapter is substantially the material presented on Nov 9, 1964 in Boston at the Symposium on Optical Information Processing.... and forming the chapter "THEORETICAL AND EXPERIMENTAL FOUNDATIONS OF OPTICAL HOLOGRAPHY (WAVEFRONT RECONSTRUCTION IMAGING)" by G. W. Stroke and D. G. Falconer in "Optical Information Processing", edited by J. T. Tippett, L. C. Clapp, D. Berkowitz and C. J. Koester (M.I.T. Press, 1964).

1. The successful reconstruction of wavefronts scattered from three-dimensional macroscopic scenes illuminated with 6328Å laser light.

2. The attainment by "lensless" wavefront reconstruction of greatly magnified ($\approx 150 \times$) microphotographs of biological samples, illuminated in 6328Å laser light without the aid of any auxiliary lenses.

3. The new theoretical evidence, which we indicate, that considerably greater resolving can be obtained with x-ray holograms than had in the past been considered possible. Real rather than "empty" magnifications on the order of 1 million and more appear attainable, and should permit one to obtain the highly resolved x-ray pictures which have been sought.

4. The simple interpretation of the spatial and temporal coherence requirements, which led to the three-dimensional laser holograms, and which is necessary for the extensions of the method to such problems as x-ray microscopy.

It might be in order to recognize Gabor's unique role in introducing a new method of image formation in optics.¹⁻⁵ In analogy with photography where lenses are used to form images, we suggest "holography" as the description for a process where holograms are used as aids of image formation

VII.1. Background and Experimental Foundations

The term "wavefront reconstruction" refers to a process in which the amplitude and phase of a scattered electromagnetic wavefront is recorded (usually photographically) together with a suitable coherent background in such a way that it is possible to produce at a later time a reproduction of the electromagnetic field distribution of the original wavefront. The coherent background is necessary for the separation of the "reconstructed" wavefront from the rest of the field scattered by the hologram. The various wavefront recording methods differ by the manner in which the coherent background is supplied, although the general idea of introducing a coherent background may be shown to be directly related to the methods introduced by Fritz Zernike in 1934 with specific application to phase-contrast microscopy.²⁴⁻²⁷

Much similarity can be found between the manner in which the "phase" in the scattered field is recorded in a hologram, on one hand, and the manner in which the phase is being recorded in an ordinary two-beam interferogram, on the other (Fig. 1). This analogy is almost complete in the method which we use for illustration.

In fact, it can be readily shown²²(Fig. 2 and Fig.4) that an "interference grating" is formed on the photographic plate, both in the case of a two-beam interferogram (Fig. 2), and in the case of a hologram (Fig. 4) where the "background" or "reference" wave is made to fall at a suitable angle on the plate, with respect to the scattered wave. In the case of a plane scattered wave (such as that reflected by a mirror) and a plane background wave (Fig.2), the "hologram" or "interferogram" is simply formed of a grating with sinusoidally varying, spatial straight-line interference fringes.

When this "grating", forming the hologram, is illuminated by a "plane" wave (Fig. 3), it will produce by diffraction a set of plane diffracted waves, which are readily seen to be the "reconstructions" of the originally plane "scattered" wave.

In the case of a scattered wave containing both amplitude and phase variations, the fringes in the "hologram" will still form a "grating" in the general sense (Fig. 4). The fringes in the grating will be suitably modulated in position and in intensity, according to the distribution of the electromagnetic field in the scattered wave near the photographic plate. When the "modulated interference-fringe grating" is now illuminated by a plane wave, it will reproduce, in two distinct sets of diffracted waves, precisely the phase and amplitude modulations which were present in the original scattered wave at the plate. An observer "looking" at one of the diffracted waves would "see" the same object which he would have seen by looking at the waves scattered by the original object. The other wave has the property of actually forming a real image of the object without the aid of any auxiliary lenses.

Interference and diffraction principles of gratings are not only basic, but sufficient to explain the physical aspects of wavefront reconstruction.^{22,23} It may frequently appear convenient to visualize the generalized hologram as "modulated interference gratings", or as "diffractograms".

The Fresnel-zone plate interpretation of some aspects of holography was already recognized by Gabor,^{1,2} and later clarified by several authors, among whom are Rogers⁶ and El-Sum.¹⁴

For the purposes of clarity we shall briefly review those elements in the theory which are required for the further development of the theory which we present.

VII.2. Theoretical Foundations

Strictly speaking, rigorous electromagnetic theory of scattering, diffraction and polarization is required for an exact treatment of holography. Under the conditions discussed elsewhere,²³ the "physical optics" approximations used in this paper are generally found to be sufficient.

Let Σ_2 be a wavefront, such as the scattered by an illuminated object (Fig.4). Then the complex amplitude of an electromagnetic wavefront may be decomposed into two parts: its magnitude $A(x)$ and its phase $\phi(x)$, each of which is essential to the structure of the wavefront. In order to be able to reconstruct this wavefront at a later time, care must be taken to lose neither the magnitude nor the phase of the scattered amplitude during the recording process. The magnitude, or rather some power of it, can be captured by simply photographing the wavefront; however, the phase is invariably lost in such a process, since photographic emulsions are sensitive only to the absolute value of the scattered amplitude. Fortunately, there exist several methods for recording the phase as well as the magnitude of a scattered wavefront.

Perhaps the best-known methods for recording the phase in a wavefront are the methods of two-beam interferometry. For instance, it is possible to photograph the phase distribution in a wavefront diffracted by a ruled optical grating (Fig. 5) by means of the interferometer shown in (Fig.6). The spatial displacements of the interference fringes in the interferogram are linearly related to the phase distribution in the diffracted wavefront one fringespace corresponding to 2π . Using this interferogram, it is possible to reconstruct the diffraction pattern, either empirically, or by Fourier-transform computation (Fig. 7). Many other two-beam interferometer systems are known to permit similar recordings of the phase in a wavefront. It may suffice to recall the Lloyd mirror, the Fizeau interferometer, the Michelson - Twyman-Green interferometer, and so on. Heterodyning methods, using lasers, also have many points in common with two-beam interferometry.²²

An important method of recording both the amplitude and the phase distribution in scattered wavefronts was introduced by Frits Zernike in 1934 in connection with microscopy.²⁴⁻²⁷ Zernike's method of phase-contrast microscopy is based on bringing a suitably phased and attenuated background to interfere with the wavefront scattered by the object, the interference taking place before the recording. The coherent background originates in the object itself, and is in fact, nothing but the "undiffracted" portion of the field scattered by the object (Fig.8). The coherent background is superposed on the widely scattered field, produced by small-size object regions in the object. The superposition as it occurs in the Fourier space before the second imaging by the lens L_2 , is shown in Fig. 8. The entire process then amounts to a one-step imaging process. The role of the complex filter in the Fourier space is to suitably shift the phase and to attenuate the generally strong coherent background with respect to the field scattered by the small object regions under

study. Principles similar to those illustrated in Fig. 8 are basic to the methods of spatial filtering, originated by Maréchal in 1953,²⁸ and developed by many authors.²⁹⁻³⁰

In a general sense, the two-step hologram imaging process introduced by Gabor was already noted by him to have some significant basic similarities with Zernike's phase contrast microscopy.² As in Zernike's method, the coherent background is introduced by means of scattering from the object itself. The important difference between the two methods results from Gabor's successful prediction and demonstration that the "diffractogram" or "interferogram" obtained by interference between the background and the scattered light can in fact be first photographed and subsequently used to form an image by a second diffraction, rather than proceeding directly from the "diffractogram" to the image, as in the method by Zernike.

Unfortunately, the in-line or forward introduction of the coherent background in Gabor's method has proven to be a basic limitation in some applications of optical holography, in that radiation from the "twin image" interferes with the reconstructed wavefront, and hence reduces its fidelity with respect to the original wavefront. This is a very serious problem when actual image formation is desired, for instance in photography and microphotography. A method for eliminating the difficulties associated with the twin images was hinted at by Gabor in the conclusion of his original (1949) paper,² and was subsequently developed by Lohmann¹⁸ in 1956 in direct reference to wavefront reconstruction work as well as by Cutrona et al³¹ in 1960. Additional theoretical and experimental work²² and various spectacular verifications were reported by Leith and Upatnieks^{20,21} in 1963 and 1964. The method can best be described as the "interferometric wavefront reconstruction method."²² Its physical principles as given above can be described without elaborate mathematical formulation. The various aspects of this method will be further developed in our discussion of the theoretical foundations of optical holography.

A. The Recording Process

The magnitude and the phase of a scattered wavefront can be recorded photographically by superposing a coherent "reference beam" or background-wave on the field striking the photographic plate. Perhaps the simplest technique for carrying out this superposition is the one illustrated in Figs. 4 and 9, wherein a plane wave illuminates a region containing scattering object and a plane mirror²² or simple triangular prism²⁰ respectively. The object, of course, diffracts the incident radiation to generate a field with some magnitude $A(x)$ and some phase $\phi(x)$ at the photographic plate, while the prism simply turns the incident plane wave through an angle θ to contribute a field with a uniform magnitude A_0 and a linear phase variation αx , where α is a constant relating the angle θ and the wavelength λ according to

$$\alpha \lambda = 2\pi\theta \quad (1)$$

Thus the total amplitude striking the plate is

$$A_0 e^{-i\alpha x} + A(x) e^{i\phi(x)} \quad (2)$$

Hence, the intensity, i.e., the quantity to which the emulsion is sensitive, is

$$I(x) = A_0^2 + A(x)^2 - 2A_0 A(x) \cos[\alpha x + \phi(x)] \quad (3)$$

It will be noted that the phase $\phi(x)$ of the scattered wavefront has not been lost in computing the intensity, as it would be if the reference beam were not present.

The emulsion of course, records some power of the intensity; that is, the amplitude transmittance $T(x)$ of the resulting photographic plate, providing one works in the linear range of the H-D curve, is proportional to

$$\begin{aligned}
& \propto [I(x)]^{-\gamma/2} \\
& = [A_0^2 + A(x)^2 - 2A_0 A(x) \cos[\alpha x + \phi(x)]]^{-\gamma/2} \\
& \approx A_0^{-\gamma-2} [A_0^2 - \frac{1}{2}\gamma A(x)^2 + \gamma A_0 A(x) \cos(\alpha x + \phi(x))] \\
& \propto 2A_0^2 - \gamma A(x)^2 + 2\gamma A_0 A(x) \cos[\alpha x + \phi(x)] \tag{4} \\
& = 2A_0^2 - \gamma A(x)^2 + \gamma A_0 A(x) e^{i\phi(x) + i\alpha x} + \gamma A_0 A(x) e^{-i\phi(x) - i\alpha x}
\end{aligned}$$

where γ is the slope of the H-D curve. It has been assumed that the intensity of the reference beam greatly exceeds that of the radiation scattered by the object, so that the approximation made in dropping the higher order terms of the binomial expansion is justified. The photograph described by Eq. (4) is called a hologram after Gabor.^{1,2}

There are two aspects of Eq. 4 that should be pointed out. The first involves the role of γ : Contrary to the requirements of many similar processes, neither the sign nor the exact magnitude of γ is of any consequence in the recording process; that is, making a contact print of the hologram, which is equivalent to changing the sign of γ , serves only to shift the phase of the nonconstant portion of the transmittance of an inconsequential 180° , while changing slightly the magnitude of γ serves only to enhance or to suppress the magnitude of this same portion of the transmittance. The second facet involves the relationship between $A(x)$ and $\phi(x)$: it will be noted that the magnitude $A(x)$ and the phase $\phi(x)$ of the scattered wave appear in the natural way, i.e., as $A(x) \exp[i\phi(x)]$, in the third term of Eq. 4, and with the sign of the phase reversed in the fourth term. The

fact that each of these terms has been multiplied by $\exp[\pm i\alpha x]$ is a consequence of the presence of the coherent reference beam, which solves the otherwise vexing problem of the "twin" or "virtual" image.

B. The Reconstruction Process

With the hologram of Eq. 4 the reconstruction process is simple, involving no lens systems, schlieren disks, half-plane filters and the like. In fact, to reconstruct the original wavefront it is only necessary to illuminate the hologram with a plane wave of radiation, as illustrated in Fig. 10. As the plane wave passes through the photographic plate it is multiplied by the transmittance $T(x)$, producing thusly, four distinct components of radiation corresponding to four terms of Eq. 4. The first term, being a constant, attenuates the parallel beam uniformly, but otherwise does not alter it. The second term also attenuates the beam, but not uniformly, so that the plane wave suffers some diffraction as it passes through the photograph.

The patterns produced by the third and fourth terms are more complicated. To understand how they affect the incident plane wave, it is necessary to recall that a common triangular prism shifts the phase of an incident ray by an amount proportional to its thickness at the point of incidence (Fig. 11), a positive phase shift deflecting the ray upward and a negative one deflecting the ray downward. Thus, the third term of Eq. 4 may be interpreted as the product of the amplitude of the scattered wavefront and a positive prismatic phase shift; similarly, the fourth term of Eq. 4 may be viewed as a composite of the complex conjugate of the amplitude of the original wavefront and a negative prismatic phase shift.

By virtue of these prismatic phase shifts the third and fourth terms of Eq. 4 deflect the incident beam respectively upward and downward through an angle θ , as defined through Eq. 1.

Moreover, in the case of the third term, the upward deflected beam is also multiplied by the scattered amplitude $A(x) \exp[i\phi(x)]$, and hence reconstructs a copy of this wavefront. On the other hand, the fourth term multiplies the downward beam by the complex conjugate of the scattered amplitude, and hence constructs a copy of the scattered wavefront except that it travels backward in time and consequently constructs a three-dimensional image of the scattering object. (The physical principles underlying this process are explained in the next section.)

C. Physics of the Method

The physical principles of the process described above can be illustrated by tracing the history of a vanishing small object through the recording and reconstruction process. This method was originally introduced by Gabor² and later clarified by Rogers⁶⁻⁹ and El-Sum.¹⁴

Our approach is similar to theirs in that we suppose that the object used in the recording process is an opaque plate with a very small hole in it. When this aperture is illuminated with a plane wave it will act as a simple spherical radiator according to "Huygen's principle." Thus, the amplitude striking the photographic plate will be of the form

$$A_0 e^{-i\alpha x} + A \exp\left(i \frac{\pi}{\lambda f} x^2\right) \quad (5)$$

where A is now some constant, λ is the wavelength of the radiation, and where f is defined in (Fig.12). Hence according to formula 4 the transmittance of the hologram corresponding to this elementary source will be of the form

$$\begin{aligned} T(x) \propto & 2 A_0^2 - \gamma A^2 \\ & + \gamma A_0 A \exp\left(i \frac{\pi}{\lambda f} x^2 + i\alpha x\right) \\ & + \gamma A_0 A \exp\left(-i \frac{\pi}{\lambda f} x^2 - i\alpha x\right) \end{aligned} \quad (6)$$

The relative simplicity of this formula together with that of the diffracting object permits one to discover the mechanism of the reconstruction process in the following way. When the hologram described by Eq. 6 is placed in a parallel beam three distinct components of radiation are generated, as shown in Fig. 13. The first component arises from the first two terms of Eq. 6, which, being constants, uniformly attenuate the incident waves producing another parallel beam to the right of the hologram. The third and fourth terms produce two additional components by deflecting the incident waves upward and downward, respectively, by virtue of the linear phase shift in their exponents.

To understand how the quadratic phase shifts in the terms $[+\gamma A_0 A \exp(i \frac{\pi}{\lambda f} x + i \alpha x)]$ and $[+\gamma A_0 A \exp(-i \frac{\pi}{\lambda f} x^2 - i \alpha x)]$ act on the incident radiation one simply recalls that a thin spherical lens shifts the phase of an impinging ray by an amount proportional to the square of the distance between the optic axis and the point of incidence, a positive phase shift describing the action of a positive lens Fig. 14. Thus, the third term of Eq. 6 acts not only as an upward deflecting prism, but also as a negative lens on incident radiation; i.e., an incident plane wave will be deflected upward and given a convex spherical wavefront, and this spherical wavefront will in fact be identical to the one which exposed the hologram. Similarly, the fourth term of Eq. 6 acts not only as a downward deflecting prism, but also as a positive lens on incident radiation; i.e., an incident beam will be deflected downward and given a concave spherical wavefront, and this spherical wavefront will in the normal way come to focus at a distance f from the hologram. Thus, in illuminating the hologram with a plane wave one not only reconstructs the scattered wavefront, but also obtains as a "bonus" an image of the object, which in this case was a point source. The image formation

with point-source illumination of holograms, and the associated very considerable magnification characteristics inherent in holography are discussed in Sec. IID.

D. Magnification

A surprisingly large magnification is attainable with wavefront reconstruction systems, especially if one uses a longer wavelength radiation in the reconstruction process than in the recording process. To obtain a formula for the degree of magnification we suppose that the object is again an opaque plate, but now with two identical and vanishingly small holes in it, which are separated by a distance 2δ . Then, since each of the holes will act as a simple spherical radiator according to Huygen's principle, the amplitude of the wavefront at the photographic plate will be

$$A_0 \exp(-i\alpha x) + A \exp\left[i \frac{\pi}{\lambda f} (x-\delta)^2\right] + A \exp\left[i \frac{\pi}{\lambda f} (x+\delta)^2\right] \quad (7)$$

Hence, according to Eq. 4 the corresponding hologram will have a transmittance of the form:

$$\begin{aligned} T(x) \propto & 2A_0^2 - 2\gamma A^2 \left[1 - \cos\left(\frac{4\pi}{\lambda f} \delta x\right) \right] \\ & + \gamma A_0 A \left\{ \exp\left[i \frac{\pi}{\lambda f} (x-\delta)^2\right] + \exp\left[i \frac{\pi}{\lambda f} (x+\delta)^2\right] \right\} \exp(i\alpha x) \\ & + \gamma A_0 A \left\{ \exp\left[-i \frac{\pi}{\lambda f} (x-\delta)^2\right] + \exp\left[-i \frac{\pi}{\lambda f} (x+\delta)^2\right] \right\} \exp(-i\alpha x) \end{aligned} \quad (8)$$

At this point we depart from the usual method of reconstruction and use instead the system shown in Fig. 15. Here a point source of wavelength λ' is used to illuminate the hologram, rather than a plane wave of wavelength λ . The fourth term of Eq. 8 still focuses the radiation from the point source according to the lens plus prism interpretation introduced in Sec. IIC. However, since

we have changed the wavelength used in the reconstruction the focal length of the "lens" will no longer be f but f' where

$$\lambda' f' = \lambda f \quad (9)$$

Of course (see Fig. 16), the object distance p will be related to the image distance q according to the classical formula, namely,

$$\frac{1}{f'} = \frac{1}{p} + \frac{1}{q} = \frac{\lambda'}{\lambda f} \quad (10)$$

Moreover, the linear magnification M attained with this method of reconstruction will obviously be Fig. 16.

$$M = \frac{2\Delta}{2\delta} \quad (11)$$

In order to reduce this formula to one involving known parameters we observe from (Fig. 16) that

$$\frac{2\Delta}{2\delta} = \frac{p+q}{p} \quad (12)$$

because of the similar triangles involved. (The prism deflects the two rays through an equal angle and thus does not affect the similar triangles argument.) Hence, we immediately obtain the following formula for the linear magnification:

$$M = \frac{\lambda'}{\lambda} \frac{q}{f} \quad (13)$$

In this equation, f is the distance of the original object from the hologram (Fig. 12), when photographed in the wavelength λ , and q is the distance of the hologram from the final image plane (Fig. 16), when the wavelength used in the reconstruction is λ' . This number could exceed 10^6 in applications to x-ray microscopy.

E. Resolution

It is well known that magnification alone is "empty" unless it is accompanied by a corresponding degree of resolution. As a number of writers, notably Baez^{12,13} and El-Sum,¹⁴ have pointed out a resolution capability of conventional projection holography¹⁻²¹ is limited ultimately by (1) to that of the photographic plate used to make the hologram and by (2) the diameter of the source used in the recording process. Since holographic recording systems are fundamentally interferometers, the source diameter and emulsion limitations enter into holography as they do into interferometry in general (see reference in Fig. 9.6). The intensity which strikes the photographic plate in recording the hologram for a single point is with a plane-wave reference beam

$$I(x) = A_0^2 + A^2 - 2A_0A \cos\left(\alpha x + \frac{\pi}{\lambda f} x^2\right) \quad (14)$$

Now, the frequency γ of the term of interest, i.e., the third term, is a function of x since by definition

$$\begin{aligned} \gamma(x) &= \frac{1}{2\pi} \frac{d}{dx} \left(\alpha x + \frac{\pi x^2}{\lambda f} \right) \\ &= \frac{\alpha}{2\pi} + \frac{x}{\lambda f} \end{aligned} \quad (15)$$

Thus, if the emulsion used in the recording process has a resolution of N lines per unit length, the only frequencies which will register on the plate will be those which satisfy

$$|\gamma(x)| = \left| \frac{\alpha}{2\pi} + \frac{x}{\lambda f} \right| \leq N \quad (16)$$

In other words, an oscillating pattern will be recorded only if x falls in the range defined by

$$-N_0 - \frac{\alpha}{2\pi} \leq \frac{x}{\lambda f} \leq N_0 - \frac{\alpha}{2\pi} \quad (17)$$

where the length d of this range is obviously

$$d = 2 N \lambda F \quad (18)$$

Physically, the finiteness of this range means that the positive "lens" which brings the incident plane wave to focus in Fig. 13 has a finite diameter d . Moreover, as is well known from classical diffraction theory, the diameter \mathcal{E} of the spot produced by a lens of focal length F and diameter d when illuminated by a plane wave is

$$\frac{\lambda}{d} = \frac{\mathcal{E}}{f} \quad (19)$$

Thus, in light of Equation 18, we see that

$$\mathcal{E} = \frac{1}{2N} \quad (20)$$

Eq. 20 applied also to the more general case where an off-axis spherical wavefront is used as a reference-beam, and where the source-diameter d exceeds the resolution-limit ($1/N$) of the emulsion. Indeed, with presently available pinhole-sources and photographic emulsions d is of the order of ($1/N$).

Consequently, the ultimate resolution of the conventional projection wavefront reconstruction technique is seen to be approximately one-half that of the recording media. Since the best emulsions, e.g., Kodak spectroscopic plate 649 F, have a resolution of about one-half micron, the conventional projection wavefront reconstruction technique is limited to resolutions to the order of one micron independently of the wavelength used in the recording process. However, our recent theoretical investigations³² have revealed that considerably higher real resolutions may be attainable in certain cases where resolution-preserving methods of recording are employed.³² Attainment of high resolutions in holography is discussed in Sec. IV. below.

F. Temporal Coherence Requirements

Simple interferometric considerations²² readily show that a primary coherence requirement in holography is that the source be a steady-state source over the range of extreme path differences between the object and background. In fact the term "coherent" background implies this condition.

The degree of temporal coherence required to carry out the recording of the hologram may also be determined by a simple argument. In order that the interference pattern produced by the superposition of the reference beam and the scattered waves not be destroyed, it is necessary to arrange the prism, object, and photographic plate in such a way that the maximum path difference between any two interfering waves will be less than, say, a quarter of the coherence length of the radiation. Now, according to the (Fig. 17) the maximum path difference in this arrangement is approximately

$$F - \sqrt{F^2 + (2\ell)^2} \approx \frac{2\ell^2}{F} \quad (21)$$

where ℓ is a measure of the dimensions of the object. Thus, since the coherence length is equal $\lambda^2/4\pi\Delta\lambda$, where $\Delta\lambda$ is the rms bandwidth about the central wavelength λ , we obtain the condition

or
$$\frac{2\ell^2}{F} < \frac{1}{4} \cdot \frac{\lambda^2}{4\pi\Delta\lambda}$$

$$32\pi\ell^2\Delta\lambda < F\lambda^2 \quad (22)$$

G. Steady-State Coherence Requirements

The "Coherence" requirements discussed above must exist between the coherent background on one hand, and each object point (i.e., re-radiator point) on the other. Unlike some arguments which have been made to this effect the relative phase

difference between the various object points calls for no special requirements. Accordingly, as first suggested by Stroke²² in 1963, one can, for instance, illuminate the "object" through a stationary or moving diffusing glass or other diffusor, if desirable, provided only that the reference background is indeed a continuous wave (for example, plane or spherical) as shown above⁴⁴. The three-dimensional holograms (Fig. 19 and 20) and the photomicrographs (Figure 21 to 27) made following these and other relevant considerations²² have indeed permitted one to successfully verify the various holography principles which we discuss in this paper.

VII.3. Summary and Results

The theoretical and experimental foundations of holography are implicit in the discussions of the previous sections. We recall here for clarity the principles involved, notably:

1. Interferometry.
2. Diffraction gratings.
3. Phase-contrast methods using coherent background.
4. Coherence requirements.

In view of extensions, such as the x-ray domain, it is important to note that the interferometric criteria permit an a priori evaluation of the likelihood of success of a holography method, provided that suitable experimental evidence is available.

In Fig. 18 we show, according to Kellstrom,³⁴ interference fringes produced with $\lambda = 8.33\text{\AA}$ x rays, by means of a Lloyd's mirror system. Such evidence can be taken as an indication of the orders of magnitude of attainable coherence, and therefore of the attainable holograms, according to the principles discussed in the preceding section, and notable in Figs. 1 to 4.

Reconstructions of a three-dimensional scene illuminated according to the principles which we describe^{21,23,32,33} are also shown in Fig. 19 and 20. Holograms corresponding to three-dimensional scenes are also shown in Fig 20 (holograms and reconstruction in 6328 Å laser light).

A photomicrograph of crystal-like grating (magnification $\simeq 6x$) obtained by holography in our laboratory, entirely without lenses, according to the principles discussed in Section II.D. is shown in Figure 21A, and the corresponding hologram, also obtained without any lenses is shown in Figure 21B. A schematic diagram of the arrangement used in obtaining the hologram is shown in Figure 22, and a photograph of the apparatus used in Figure 23. The lensless reconstruction of the real image of Figure 21A was obtained in the arrangement sketched in Figure 24A and a photograph of the actual apparatus used is shown in Figure 24B. In extending holography to x-ray microscopy applications, it may be necessary to illuminate the object by means of a moving mirror or scatterer. A photographic reproduction of a lensless reconstruction obtained from a hologram photographed by means of illumination with a moving scatterer is shown in Figure 25, together with a reproduction of the actual hologram and object, all to scale. A sketch of the apparatus used for obtaining the hologram of Figure 25 is shown in Figure 26, and a photograph of the apparatus in Figure 27. Among the principal experimental requirements for the attainment of good holograms are:

(1) interferometric (mechanical and thermal) stability between the object, reference mirror(or lens) and the photographic plate;

(2) an intensity in the reference beam about 5 to 10 times greater than the intensity in the scattered field, at the photographic plate, so as to permit to maintain only the first side-band terms in eq. (4) above.

Waxing down of the various elements on solid support, and waiting for the object, mirror and photographic plate to come to a thermal equilibrium has been found to be as important in holography as it is in interferometry in general.

VII.4. Attainment of High Resolutions

The classical resolution limitations^{12,13,15} in "projection holography" can be shown to result from the non-recordability of the interference fringes with available photographic emulsions and source sizes. For instance, an emulsion with a 1/2 micron resolution will only resolve about 10,000 Å at 1 Å x-ray wavelengths (if a minimum of one fringe is recorded in the 1/2 micron distance, assuming plane-wave illumination).

A gain of 10^4 in resolution and the possibility of resolving 1 Å at x-ray wavelengths should result if the hologram is recorded in a "Fourier-transforming" rather than in the conventional "projection" holography arrangement.

A Fourier-transforming holographic recording arrangement is obtained, for instance, by placing the "point-reference" into the plane of the "object-grating" in Fig 22.

More generally, the transmission of a high-resolution hologram should correspond to the recordable equation

$$H(x) \propto 2A_0^2 + \gamma \sum_n A_n^2 \text{rect}(x/l - fn\beta/k\ell) - 2A_0\gamma \sum_n \text{rect}(x/l - fn\beta/k\ell) \cos(\alpha x + \phi_n - fn^2\beta^2/2k)$$

(25)

rather than to Eq. 4 characterizing conventional projection holograms.

Any one of a number of Fourier-transforming recording arrangements should permit one to realize Eq. 25, in addition to the one indicated above ^(Fig. 28). Regardless of the manner in which the high-resolution hologram of Eq 25. is recorded, the reconstruction is then clearly obtained in the focal plane of a Fourier-transforming lens. The high magnification of Eq 13. is of course maintained in the use of high resolution holograms.

VII.5. Electron-Microscopy and X-Ray Hologram Microscopy

Some remarks may be in order concerning the possible applications of hologram microscopy, in the light of the remarkable results obtained with modern electron microscopes.

Modern electron microscopes appear to readily provide resolving powers on the order of the Angström (10^{-10} meter). Even the most optimistic present estimate makes it unlikely that these values would be exceeded by an x-ray hologram microscope, or for that matter by any x-ray microscope, using the wavelengths of the x rays contemplated. On the other hand, it is not at all unlikely that even these remarkable resolutions will be exceeded by future electron microscopes, and that sample-heating problems will be satisfactorily eliminated.

There are, however, several aspects in which an x-ray hologram microscope, if developed, would fill a role which cannot at the moment be filled by electron microscopes. Considerably greater penetrations without heating of the samples can be obtained with x-rays than even with very-high-energy electron beams. This would be of a particular interest in areas such as metallurgy, and especially biophysics, in particular perhaps with "live" tissues and so on. It is also of interest to note that an x-ray microscope would not necessarily require a vacuum, which is a necessity in the electron microscopes. Considerably greater resolutions could be attained if the hologram technique could be extended to gamma rays.

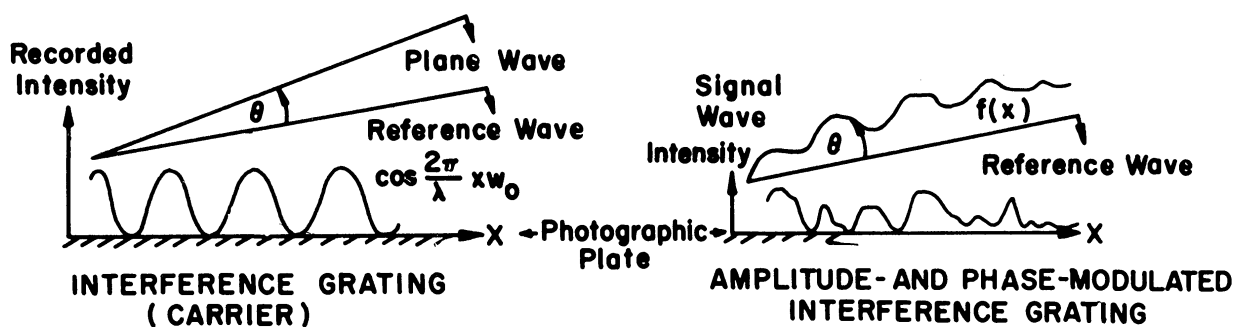


FIG.1. Amplitude- and phase-modulated interference grating.

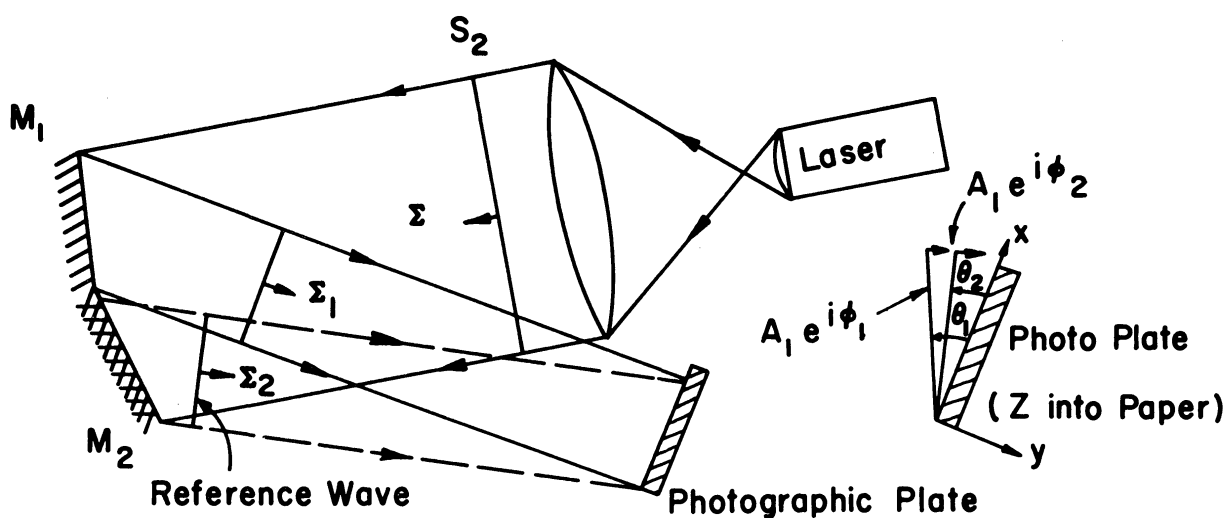


FIG.2. Recording of hologram(modulated interference grating) for the case of a plane-wave generating object(M_1)

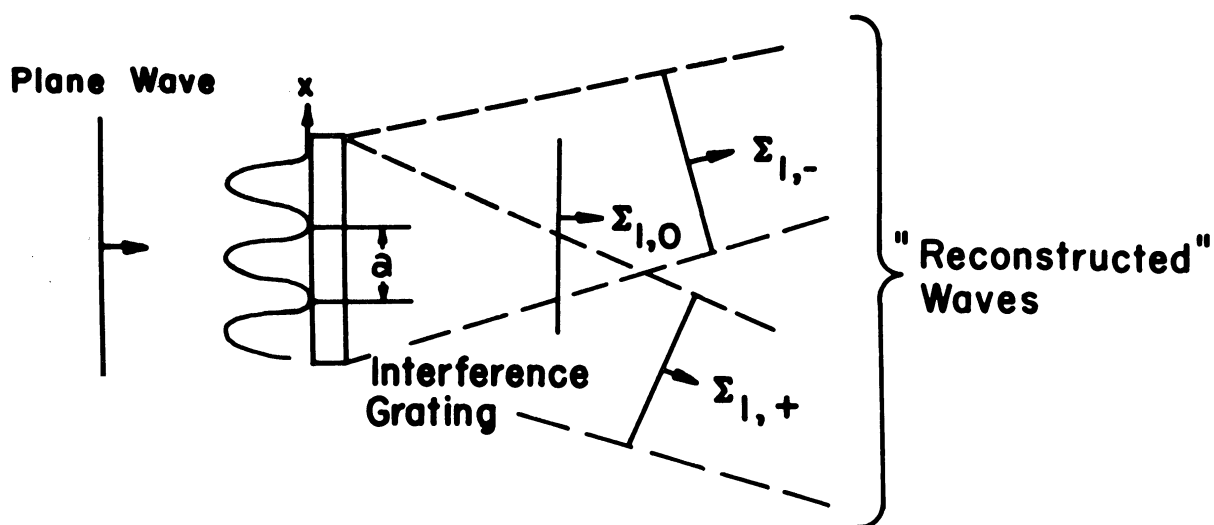
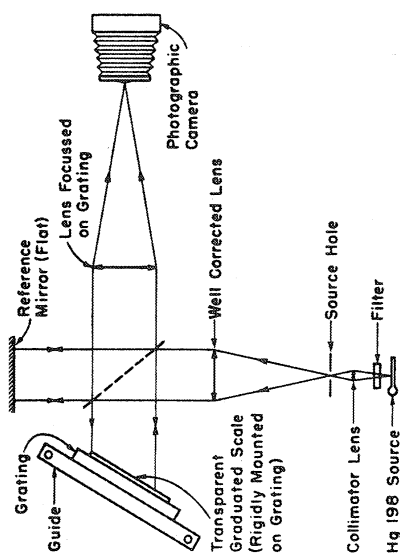
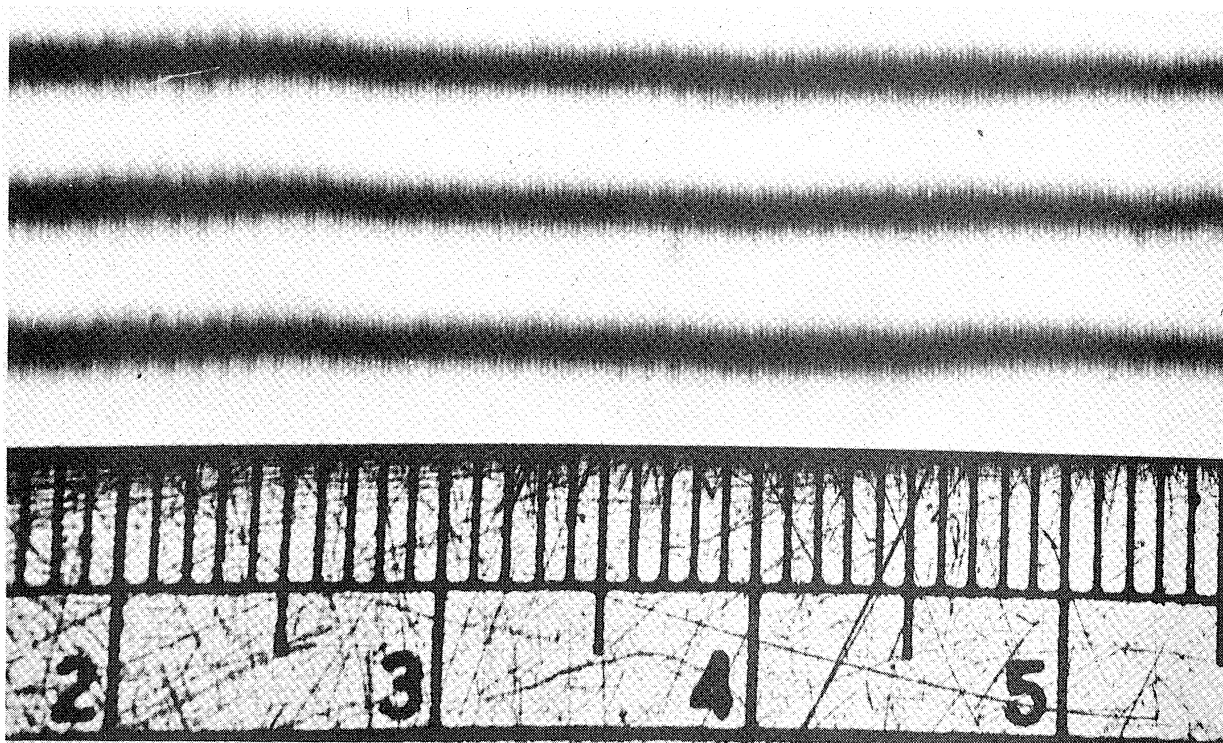
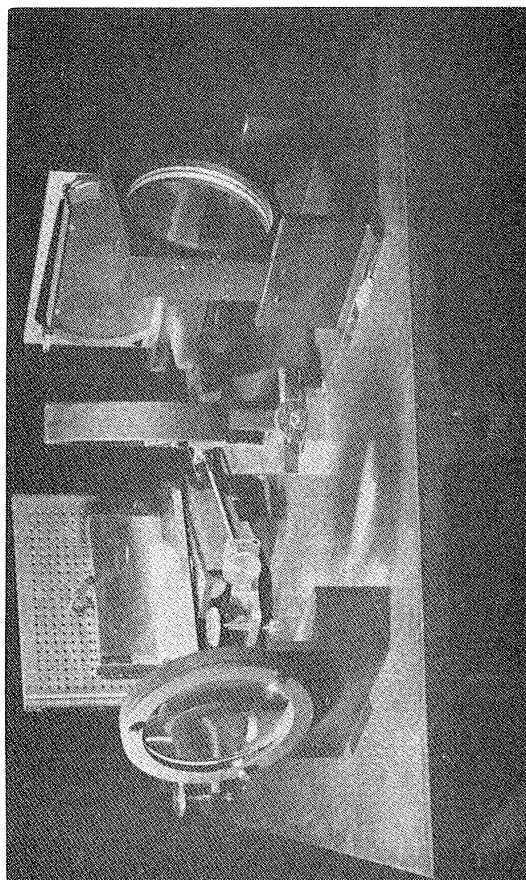


FIG.3. Reconstruction of plane wave from hologram of FIG.2.



Grating testing interferometer. The gratings should be examined at the highest angle permitted by the blaze. For good fringe contrast, the reference mirror needs to be carefully autocollimated on the source hole. The grating is so adjusted, with the help of leveling screws, that the fringes of the mean wave front appear at right angles to the groove length (STROKE [1955, 1960, 1961]).



Grating testing interferometer. From left to right: camera lens, 8-inch grating (shown here at low angle for clarity), beam splitter, reference mirror, collimator. As shown, this interferometer permits testing of gratings up to approximately 12 inches in width at the highest angles (76°) in autocollimation. Mechanical vibrations in such interferometers are easily avoided if the mechanical mounts are fairly heavy, if possible made out of 'meehanite', and have well-machined, flat supporting surfaces.

FIG.5. and FIG.6. Interferometric display of wavefront diffracted by optical gratings according to G.W.Stroke(J.Opt.Soc.Am.45,30,(1955)). Above: interferogram of a 300 groove/mm grating. Below: interferometer.Note similarity with hologram recording.

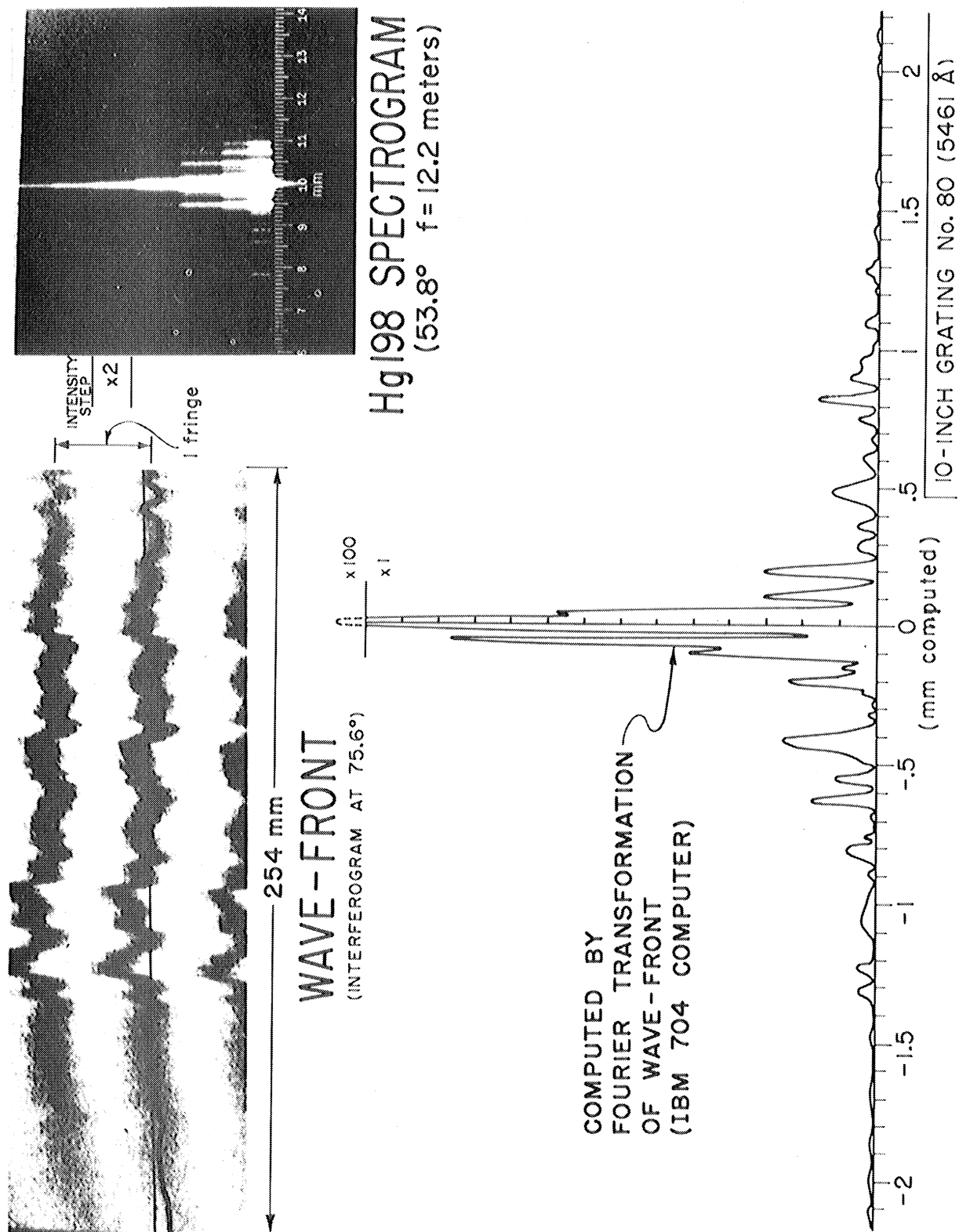


FIG. 7. Fourier-transform computation of point-source image (diffraction pattern) corresponding to diffracted-wavefront interferogram of a 300 groove/mm grating shown (according to G.W. Stroke, *Revue d'Optique*, 39, 291 -1960- and *J. Opt. Soc. Am.* 51, 1321 -1961-). The spectrogram of green line (5461 Å) of Hg198 is shown. Note similarity of relation between diffraction pattern ("spectrum") and wavefront with optical data processing of data recorded on film, for instance in "diffractive processing of geophysical data" (P.L. Jackson, *Applied Optics* 4, 419, -1965-).

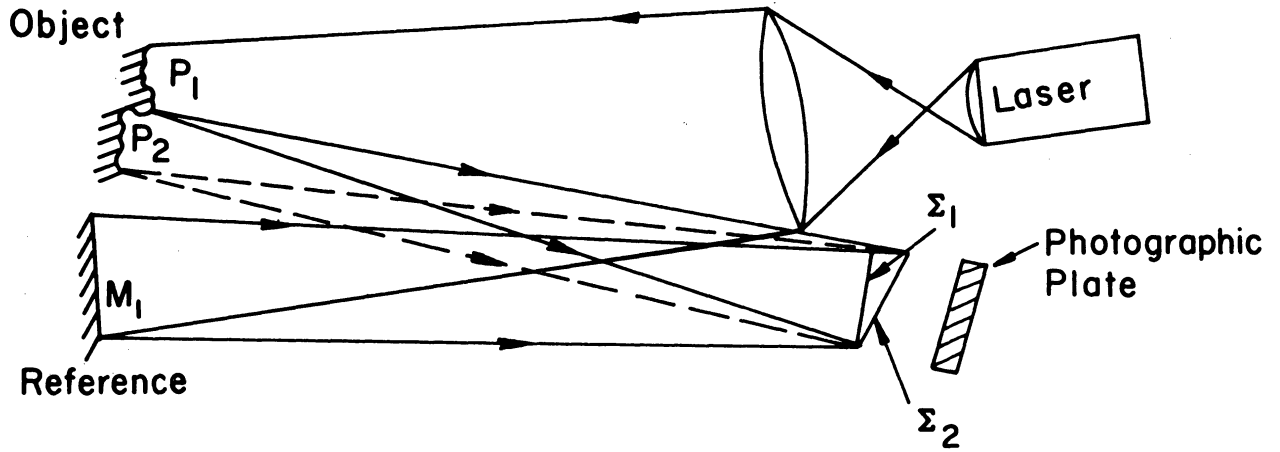


FIG.4. Recording of hologram in case of three-dimensional object according to Stroke²².

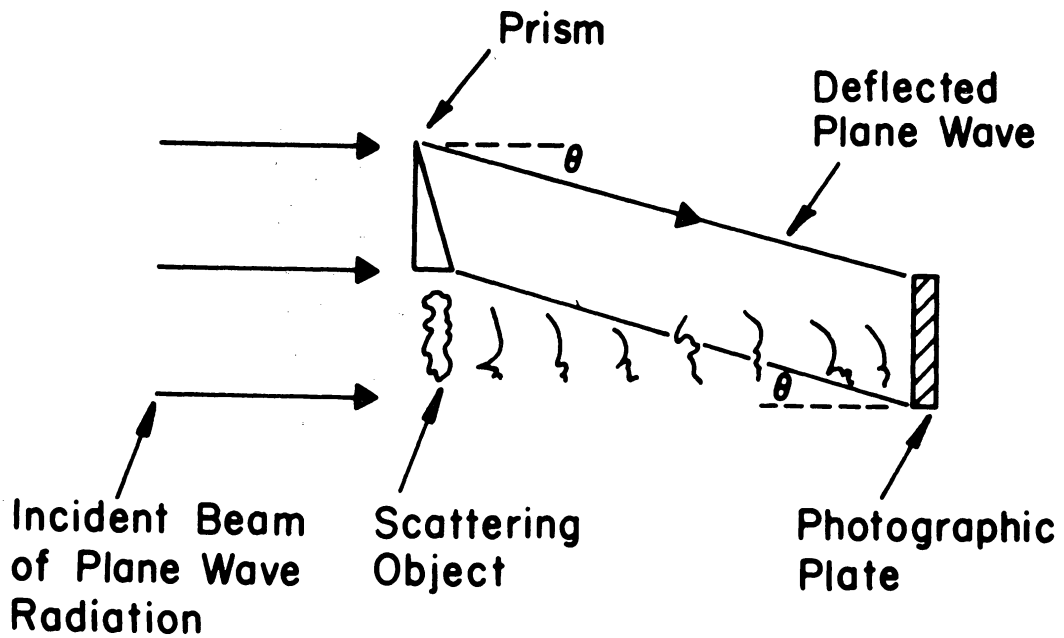


FIG.9. Schematic arrangement to illustrate recording of hologram (See also FIG.2. and FIG.4)

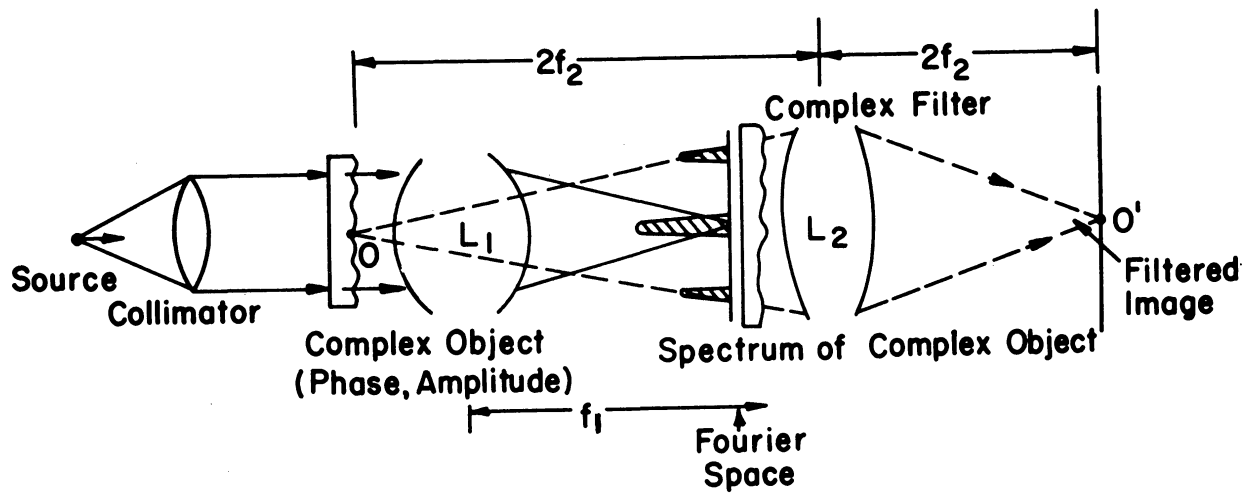


FIG.8. Arrangement used in phase-contrast imaging according to Zernike²⁴⁻²⁷
and in "spatial filtering" according to Marechal and Croce²⁸
(C.r.Ac.Sc.237,607 - 1953-)

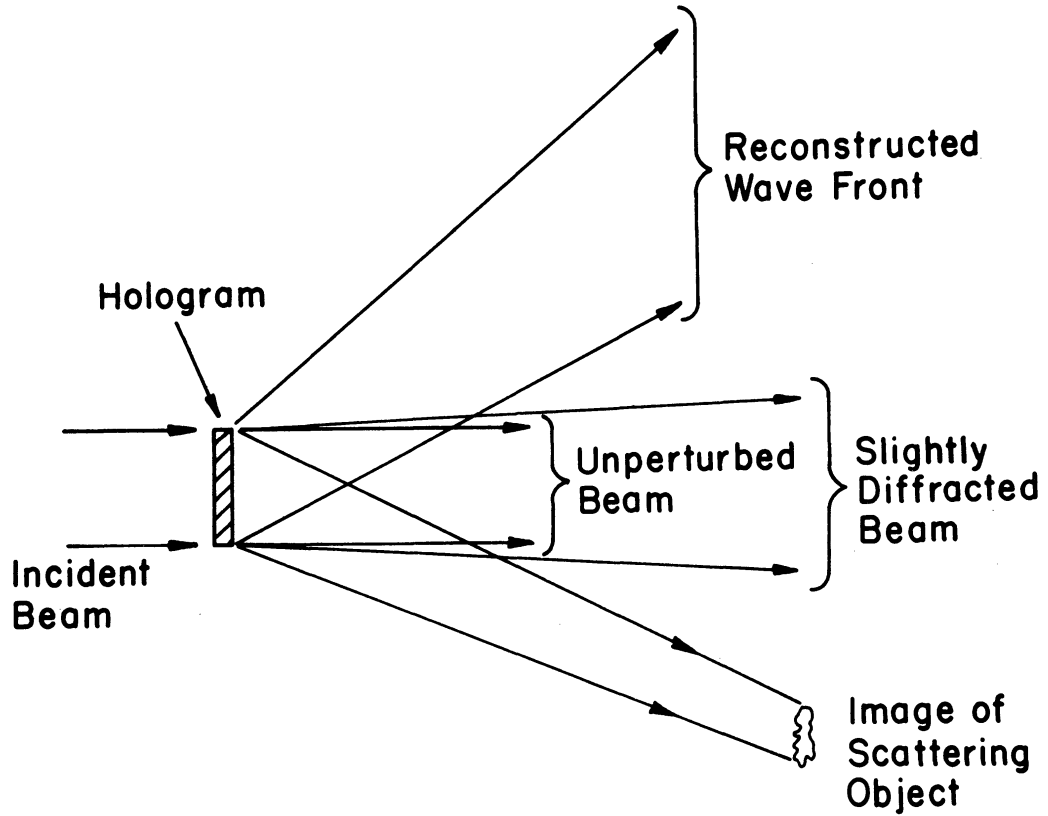


FIG.10. Wavefront reconstruction and image formation from a hologram in case of plane-wave illumination.

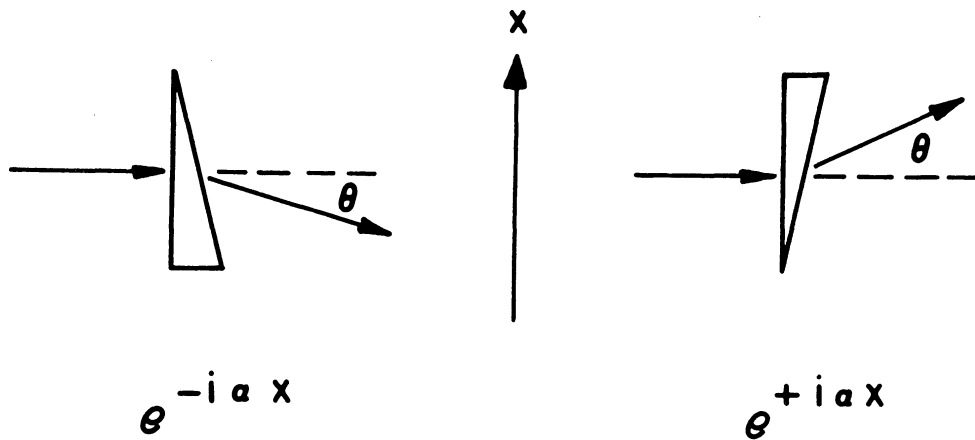


FIG.11. Phase shifts and deflection angles corresponding to the "prism" terms in Eq.4.

Note: $\alpha \lambda = 2\pi\theta$

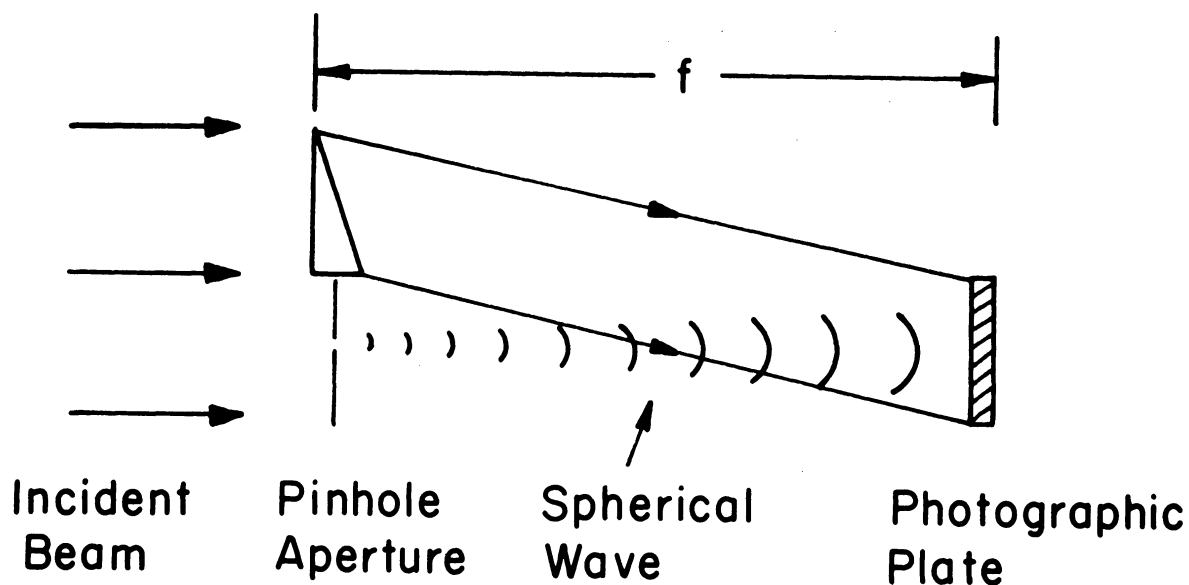


FIG.12. Hologram of point object(pinhole aperture)
 Arrangement used in discussion of magnification in Sec.II.D.,in the discussion of Eq.6.,and in the discussion regarding FIG.13.and FIG.14.

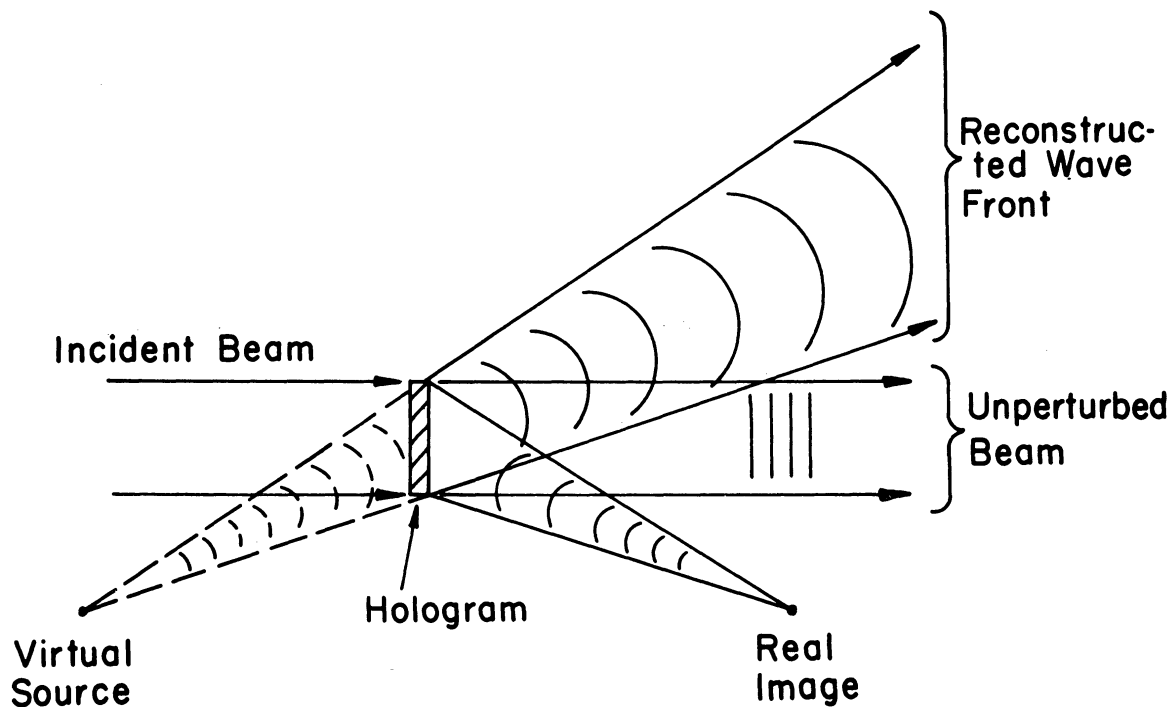


FIG.13. Wavefront reconstruction and image formation from hologram of a point scatterer(see FIG.12.)

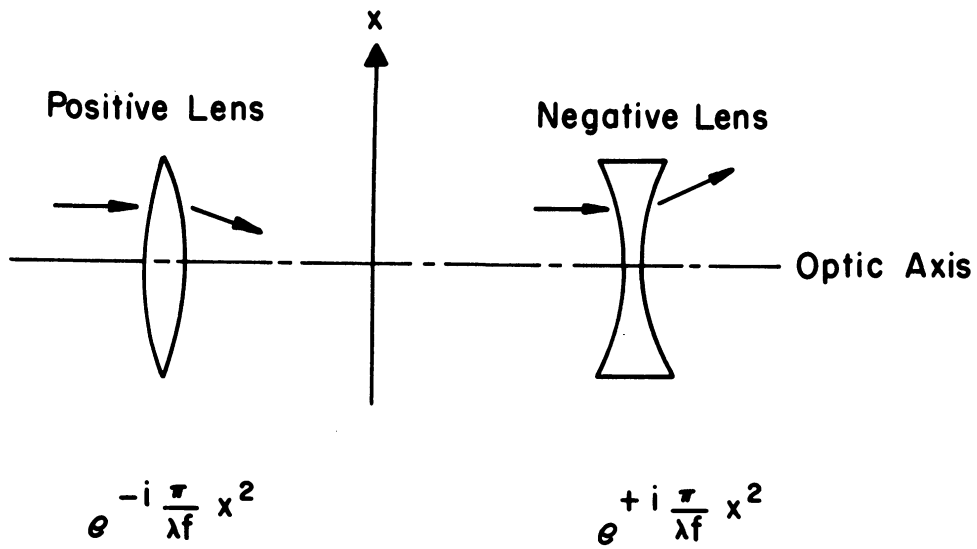


FIG.14. Lens-like effects and phase shifts associated with terms in Eq.6.

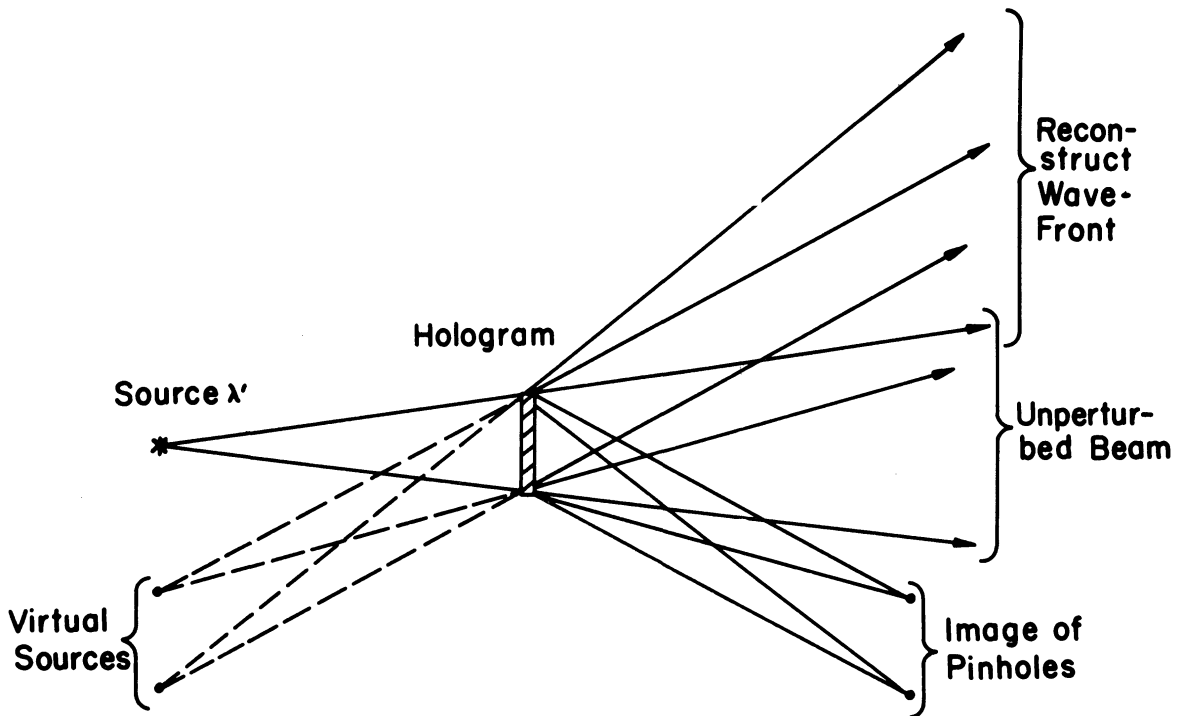


FIG.15. Wavefront reconstruction and image formation with point-source illumination of hologram, in the case of an object formed by two point scatterers (See also FIG.16)

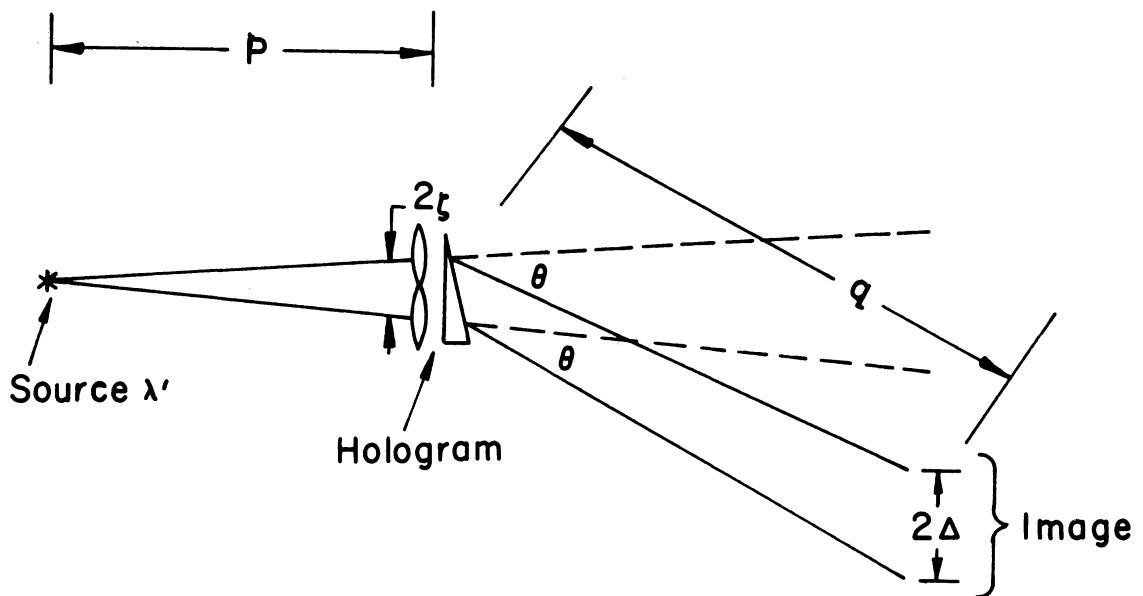


FIG.16. Magnification property inherent in holograms. In comparing the geometry shown with FIG.15-, it should be noted that that the hologram of FIG.15. has been replaced by its equivalent lens-prism arrangement. (The central and upper beam components have been omitted for clarity.)

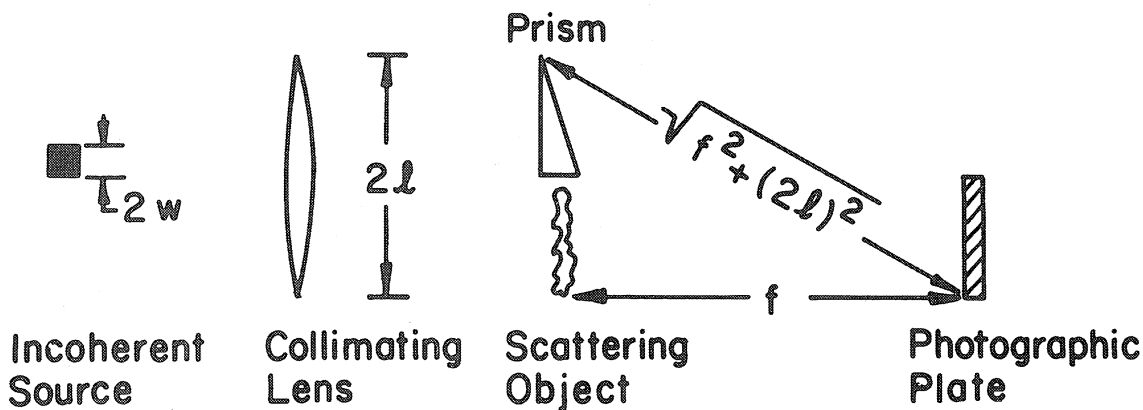


FIG. 17. Geometry used in the discussion of spatial and temporal coherence requirements in holography.

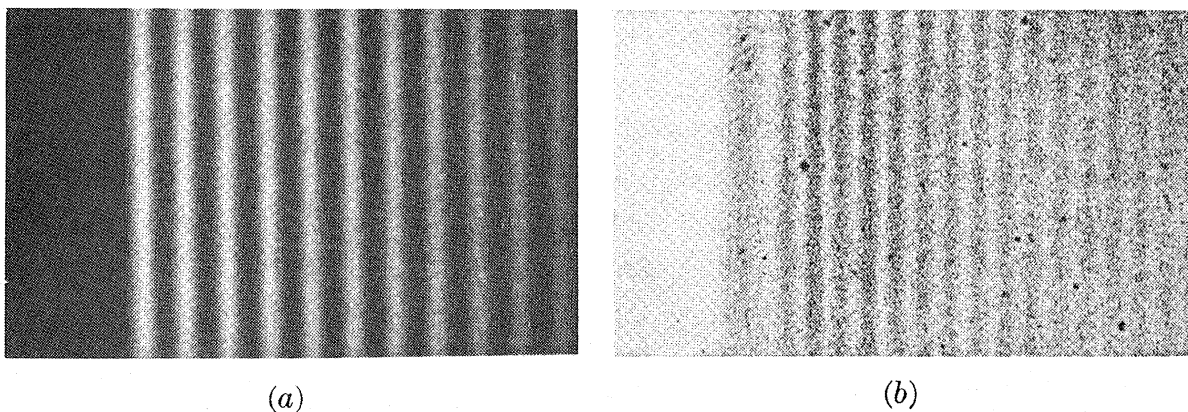
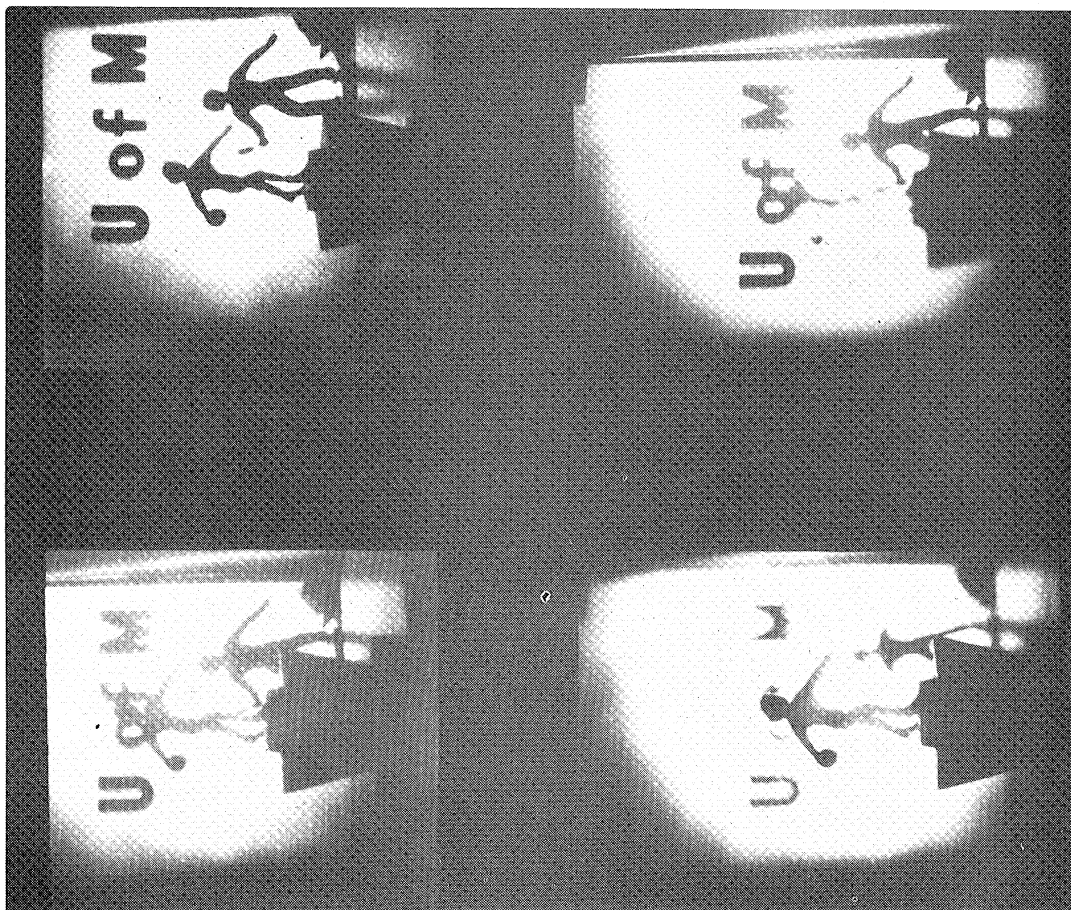
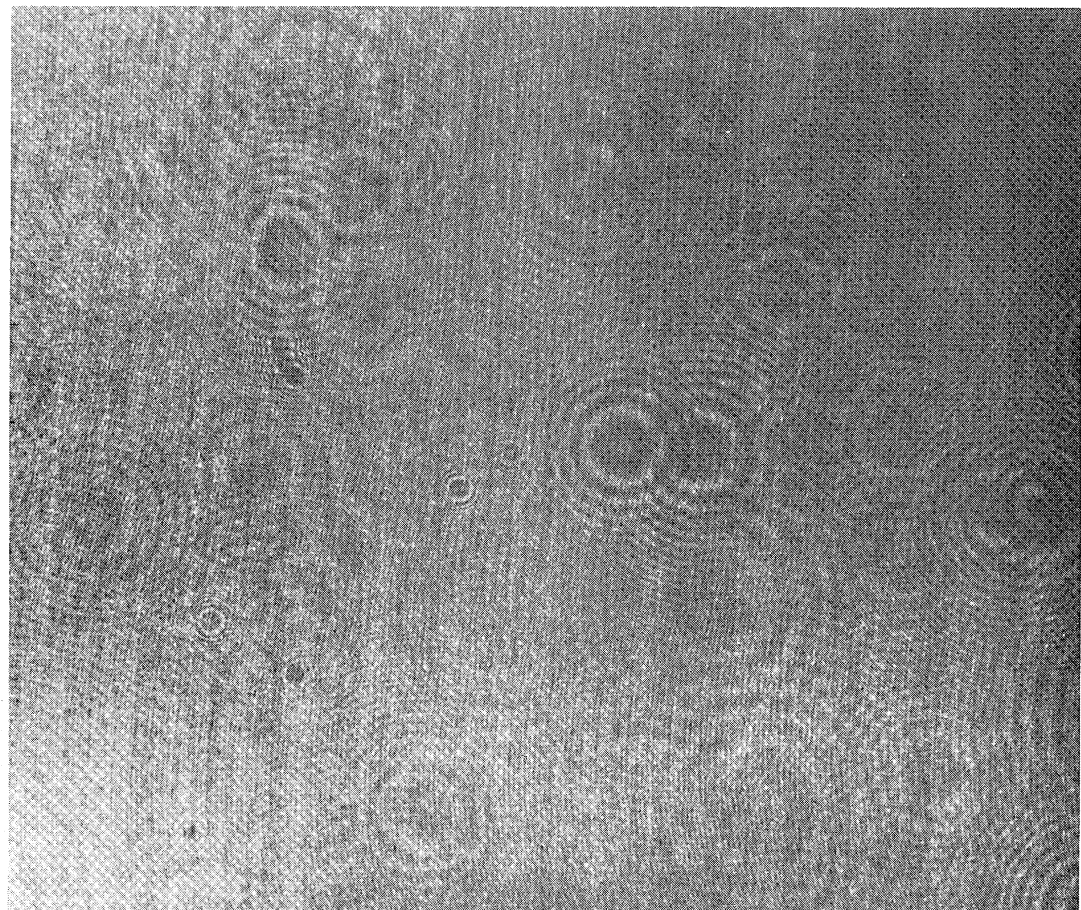


FIG. 13M. Interference fringes produced with Lloyd's mirror. (a) Taken with visible light, $\lambda = 4358 \text{ \AA}$. (After White.) (b) Taken with X rays, $\lambda = 8.33 \text{ \AA}$. (After Kellstrom.)

FIG.18. Interference fringes at x-ray wavelengths(b) illustrating likelihood of success of x-ray holographic microscopy. Interference fringes produced with Lloyd's mirror. FIGURES 18(a) and (b) according to F.A.Jenkins and H.E.White, Fundamentals of Optics, McGraw Hill Book Co. New York(1957). (a) Taken with visible light $\lambda = 4358\text{\AA}$ (after White). (b) Taken with x-rays, $\lambda = 8.33\text{\AA}$ (after Kellstrom³⁴).



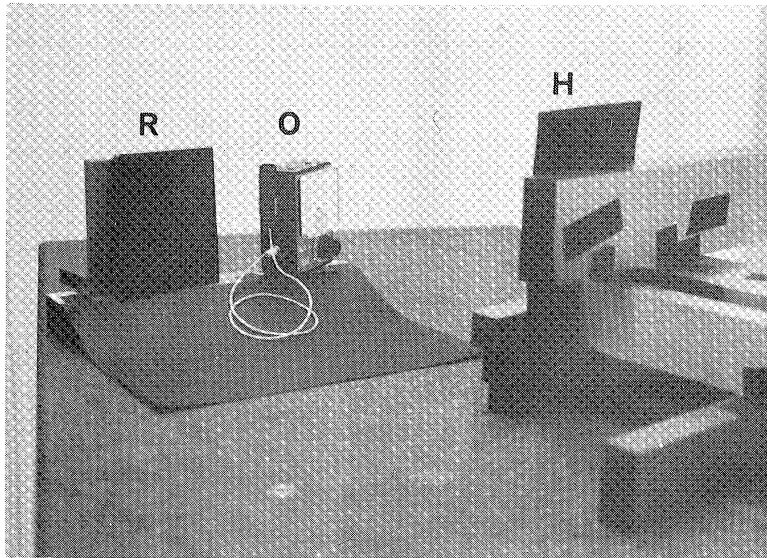
(a)



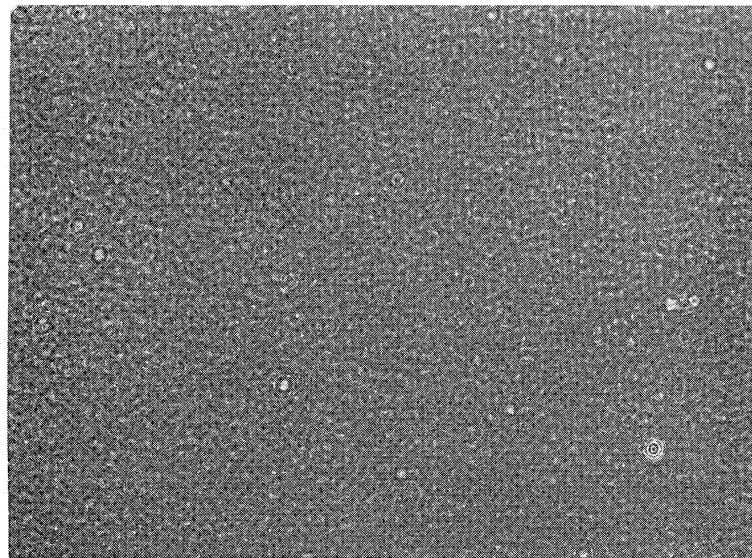
(b)

FIG.19. (a) Images of three-dimensional scene²¹ reconstructed from a hologram according to Leith and Upatnieks²² and Stroke²³.

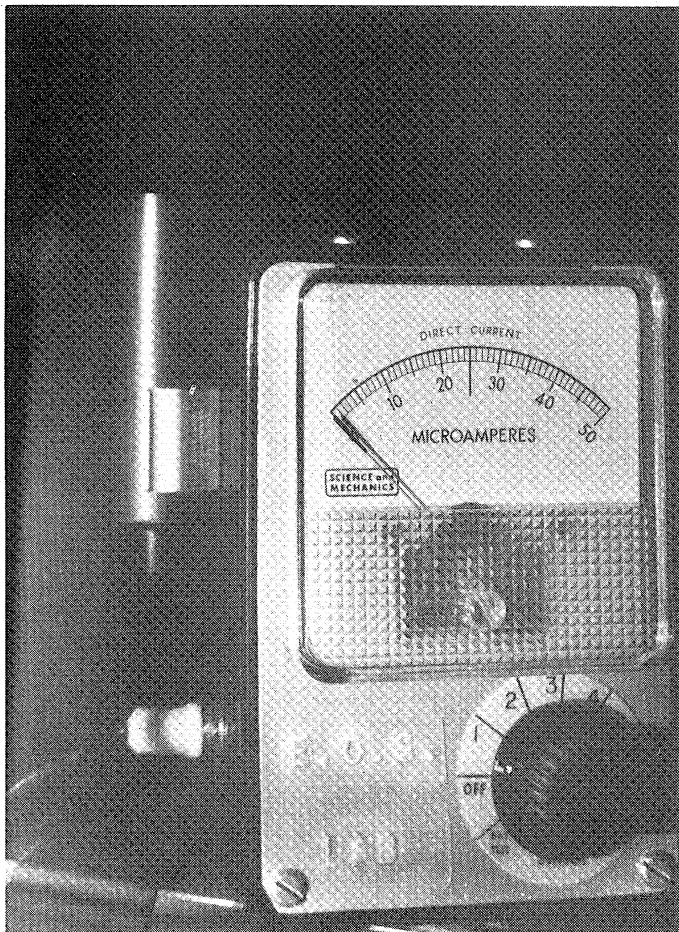
All four images were photographed by means of a camera looking through the hologram, and so focused and inclined to reveal the three-dimensional nature of the reconstruction. The top two images are photographs with a small camera aperture, at two angles, to show parallax. The lower two images are photographs with a large camera aperture, at the same angle with respect to the hologram(b) to show the three-dimensional field depth. (Hologram and reconstructions in 6328Å laser light).



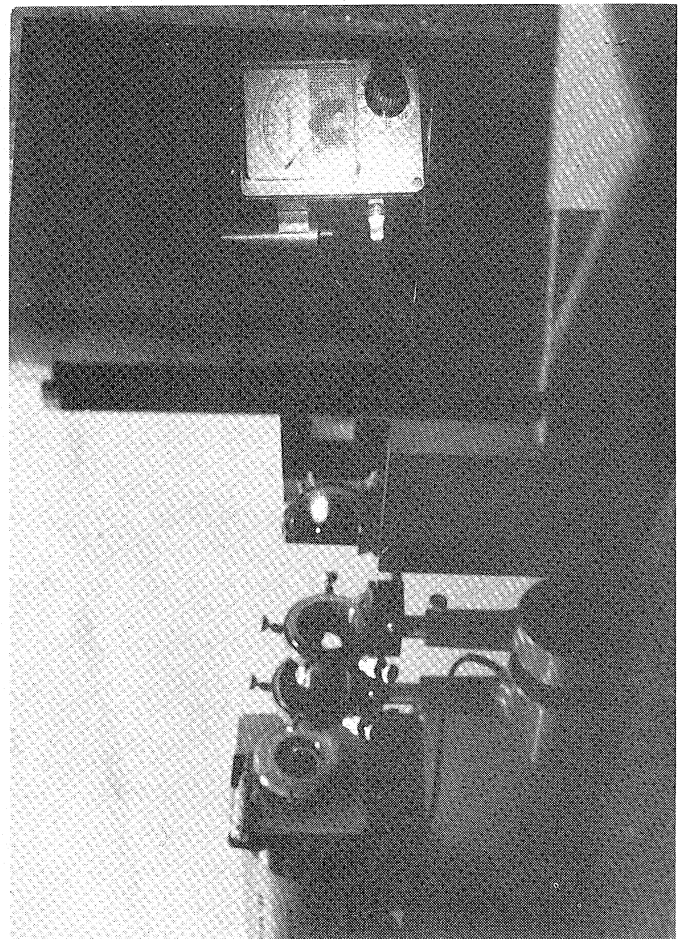
(a)



(b)



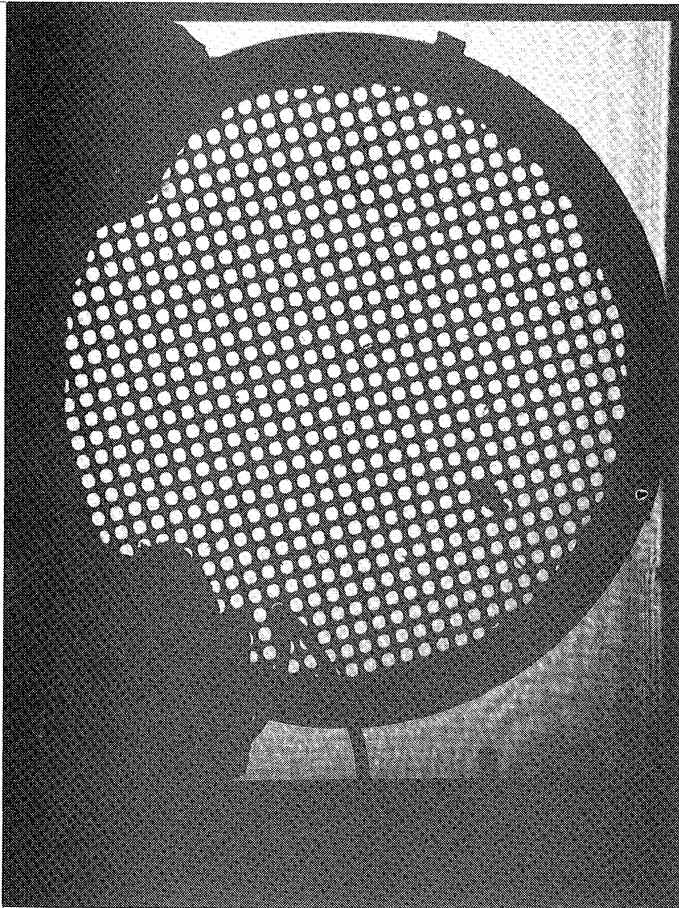
(c)



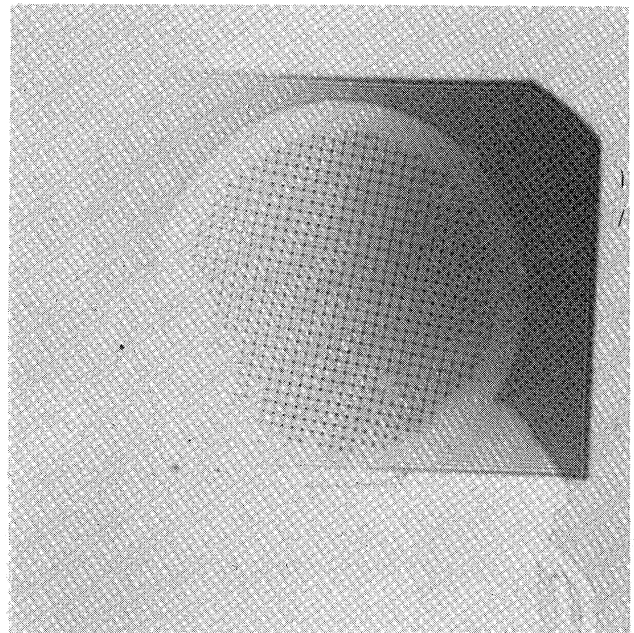
(d)

FIG.20. Holographic recording and reconstruction of three-dimensional object according to the principles first suggested by Stroke²².

- (a) Recording : R=reference mirror, O=object, H=hologram(on Kodak 649 F photo-plate)
 - (b) Hologram
 - (c) Lensless reconstruction of real image
 - (d) Reconstruction of virtual image,as photographed through hologram
- (All holographic work in 6328\AA laser light in Prof.Stroke's Electro-Optical Sciences Laboratory , University of Michigan,Ann Arbor,Michigan).



(a)



(b)

FIG.21. Lensless microscopy.

- (a) Lensless reconstruction of image of a crystal-like grating. The 6 times enlarged image was reconstructed from the hologram(b) by means of the arrangement and apparatus shown in FIG.24. Note the remarkably fine resolution of details (f.ex.the 0.1mm diameter wire supporting the grating).The 6 times enlargement was used here to show the reconstruction in its entirety,but enlargements of well over 100 times have been obtained without any difficulty³².
- (b) Hologram obtained by "projection microscopy" arrangement shown in FIG.22 and in FIG.23 and used in reconstructing image in (a).The entire recording and reconstruction process shown was carried out without the aid of any lenses in 6328Å laser light in Prof.Stroke's Electro-Optical Sciences Laboratory,University of Michigan.An additional magnification of 6328x would result if the recording were carried out in 1Å x-ray light,leading to magnifications in excess of 1 million in x-ray microscopy.

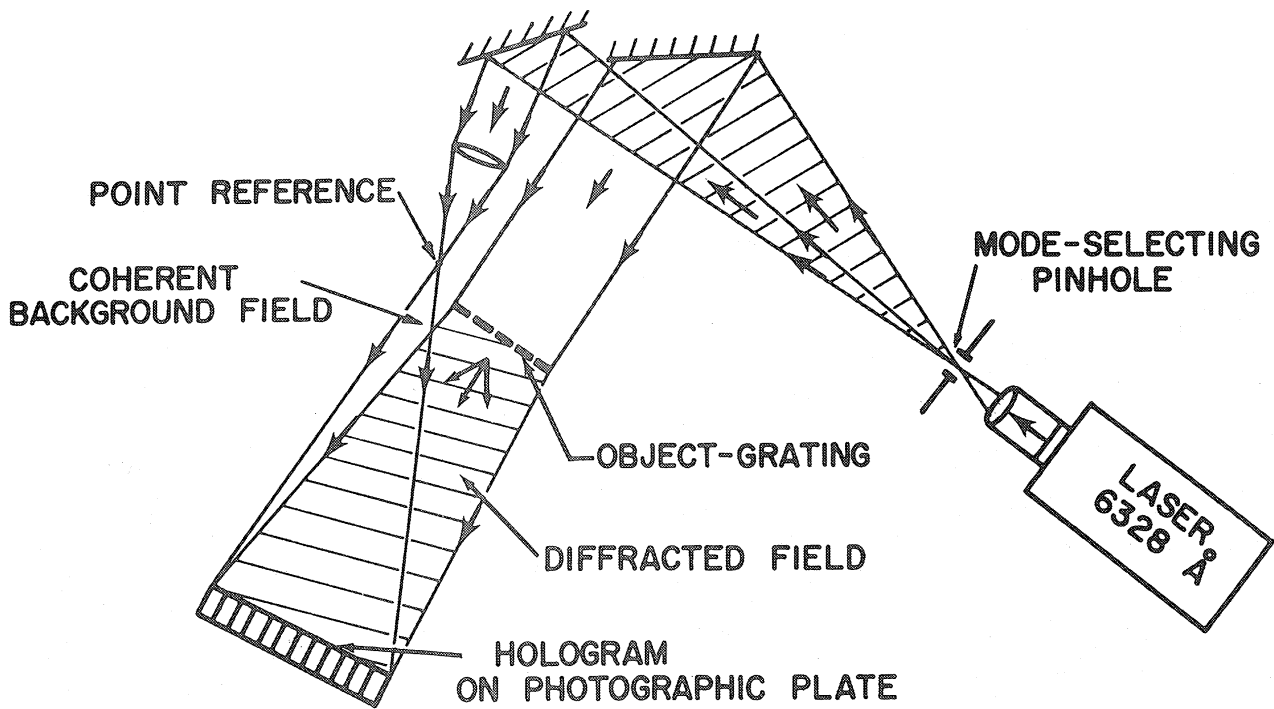


FIG. 22.

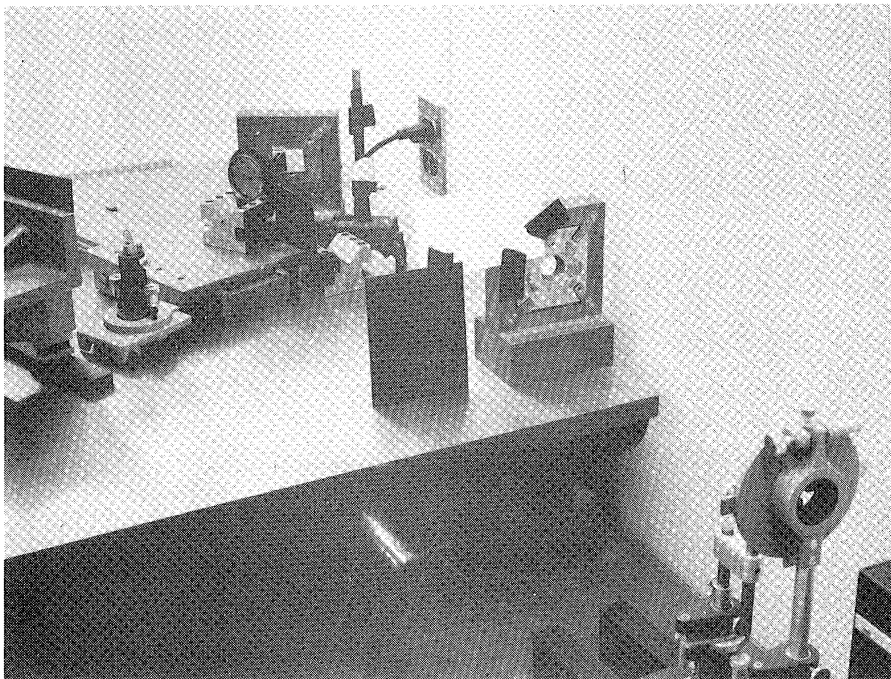
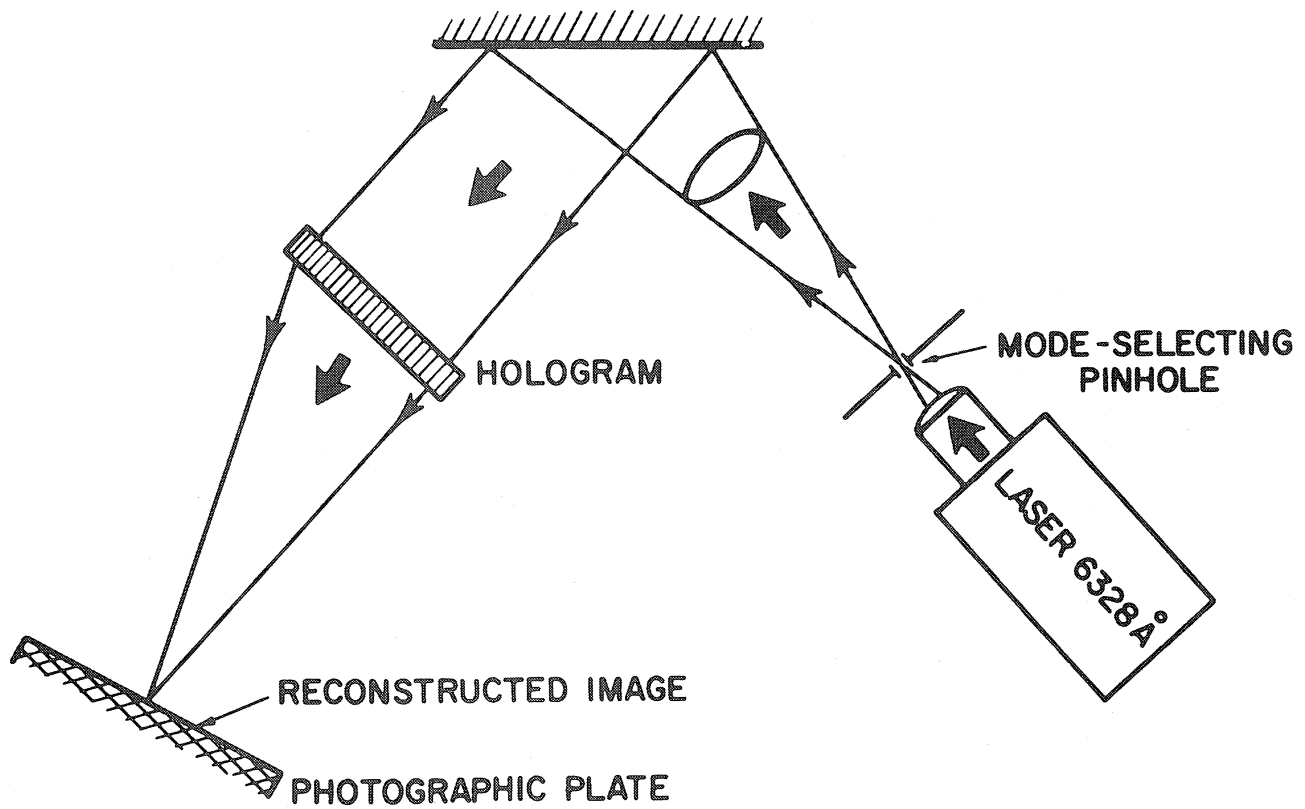


FIG 23.

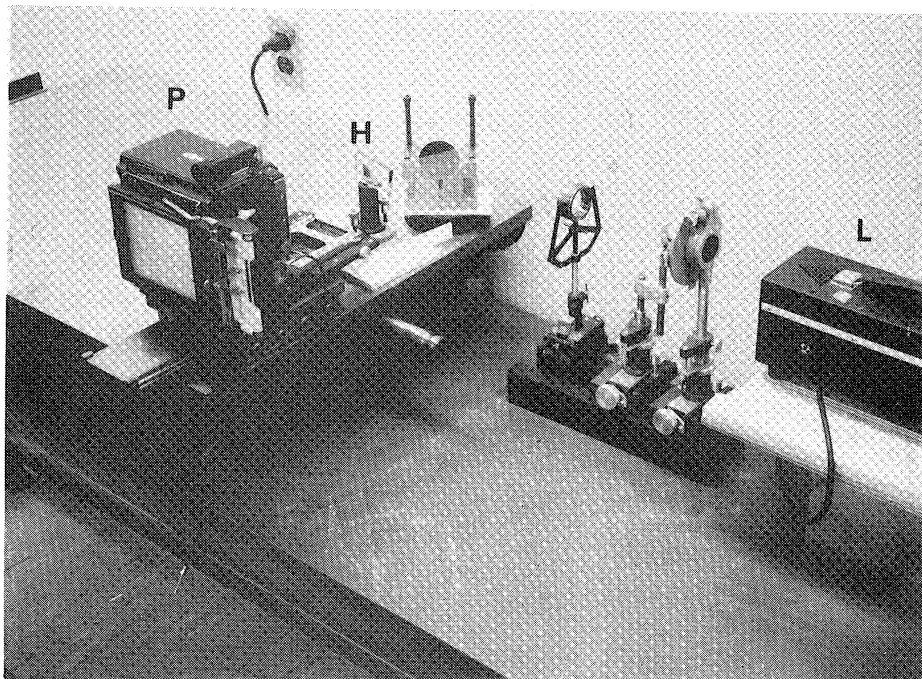
Lensless microscopy

Projection holography arrangement used for "lensless" microscopy in the recording of the hologram of FIG.21. The three-dimensional character of "lensless" microscopy, and the associated great depth of field are of particular interest, not only for x-ray and electron microscopy applications, but for visible and ultraviolet applications as well. The object-grating in the photograph has a diameter of about 14mm.

(Prof.Stroke's Electro-Optical Sciences Lab.University of Michigan)



(a)



(b)

FIG.24. Lensless microscopy.

Apparatus used in the reconstruction of images in "projection microscopy".
P=plate holder, H= hologram, L= laser.

(Prof.Stroke's Electro-Optical Sciences Lab.The University of Michigan)

Lensless reconstruction

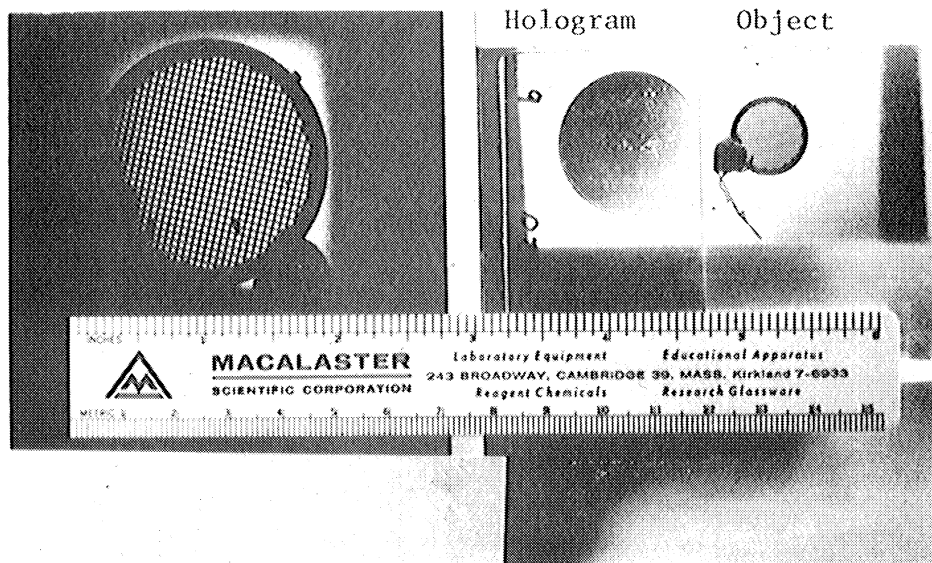


FIG.25. Lensless microscopy using multi-directional illumination and moving scatterers. Shown to scale are ,left to right: image(magnified about 3x), hologram and object.The hologram was obtained in the arrangement shown in FIG.26(photograph in FIG.27) while a diffusing scatterer was actually moved during the exposure. The remarkable quality of the reconstructed image obtained under this unusual condition serves not only to illustrate some important aspects of the degeneracy of the coherence requirements in holography, but the use of a moving scatterer,or mirror, may be a necessity in the attainment of high resolutions in holographic microscopy.
 (Prof.Stroke's Electro-Optical Sciences Lab.Univ.of Michigan)

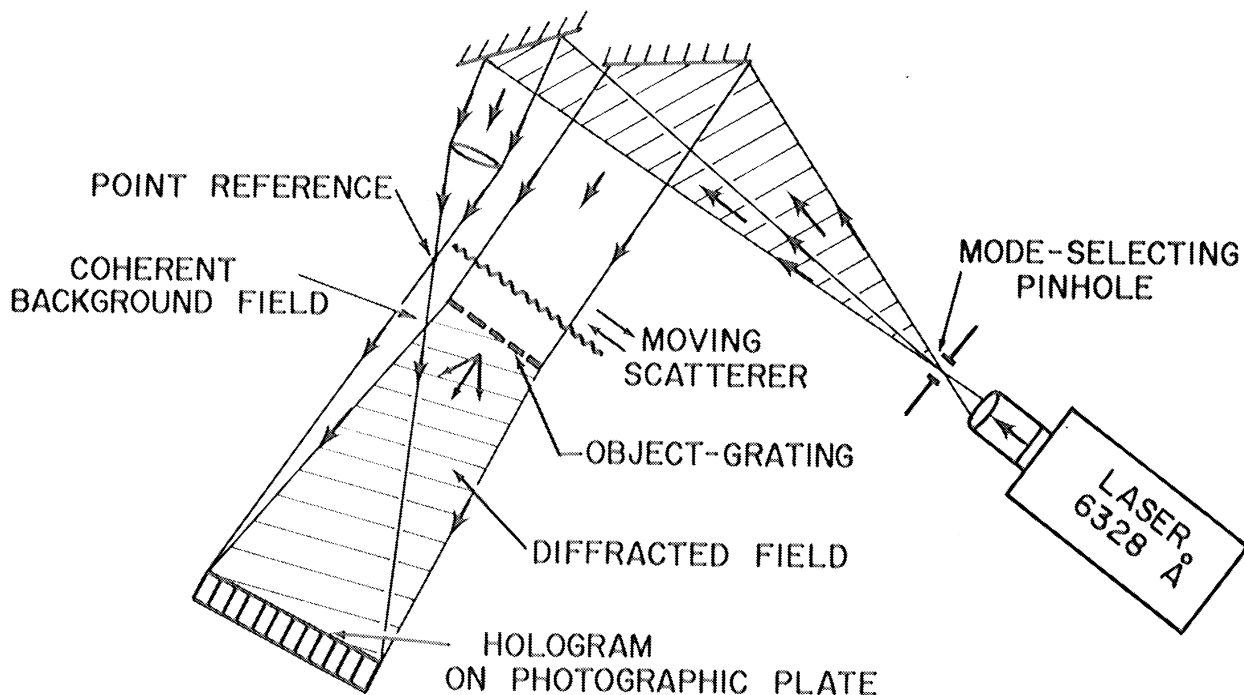


FIG.26. Recording of hologram with "multi-directional" illumination and a moving scatterer. The use of a scatterer in holographic imaging makes it possible not only to obtain multi-directional illumination, but is also a means for "beam-splitting(for instance with the aid of partially diffusing "scatter-plates", such as those already mentioned for interferometry by Newton in his Optics:see also J.M.Burch, Nature 171,889-1953- and J.Strong, Concepts of Classical Optics, article by J.Dyson, p.383(1958).

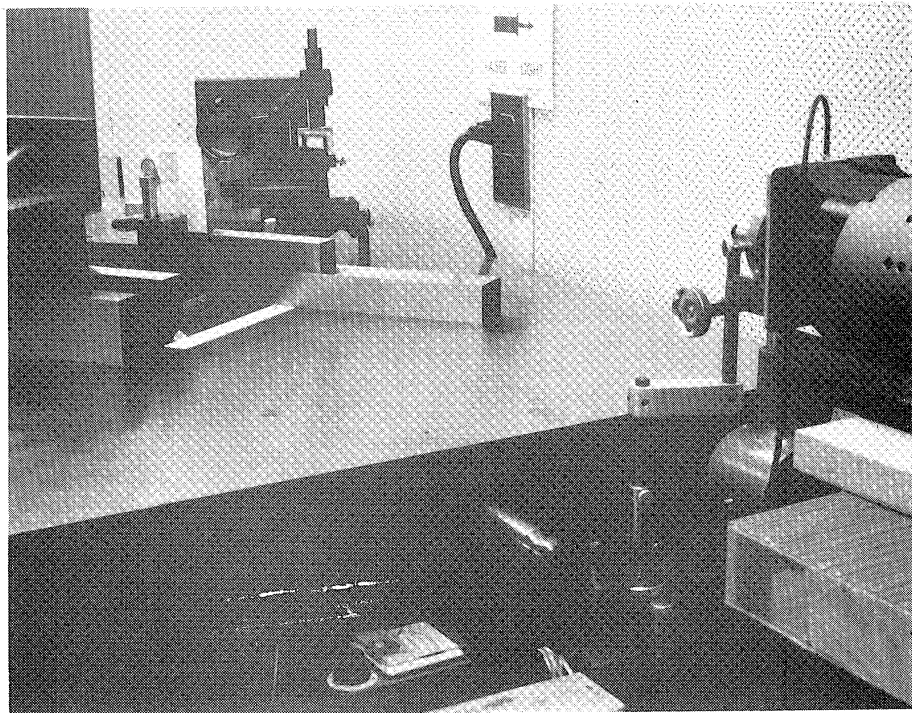


FIG.27. Photograph of apparatus used in holographic microscopy
with "diffused" illumination
(as in FIG.26)

The high-power CW laser used in obtaining the hologram of FIG.25 appears on the right. CW lasers were used in the work shown in the previous pages, but pulsed lasers should help in overcoming vibration difficulties and temperature stability requirements which tend to appear in the interferometers used for the recording of holograms of large-scale objects. The results obtained in the work in Prof. Stroke's Electro-Optical Sciences Laboratory, as assisted by D.G. Falconer, A. Funkhouser and D. Brumm, show that completely satisfactory stability in the interferometers used for holographic recording is achievable for holographic microscopy applications, as well as for comparatively small-scale macroscopic objects.

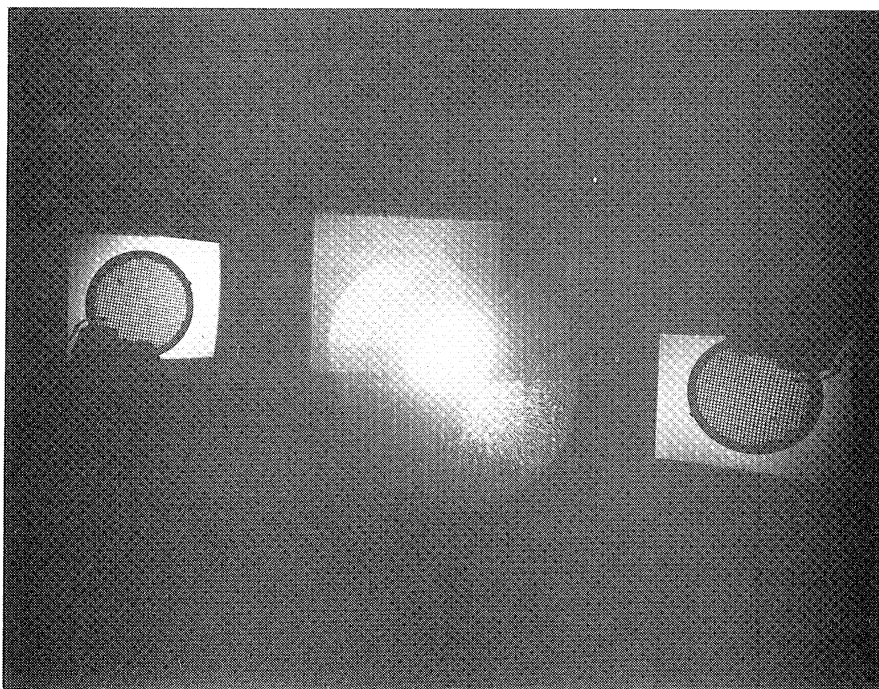


FIG.28.(c)

Reconstruction of the two side-band images from "lensless Fourier-transform hologram" obtained in arrangement of Fig.28.(b) ,according to G.W.Stroke and D.G.Falconer. The figure shows a single exposure of the two side-band images,as well as the unattenuated central image.

(Prof.Stroke's Electro-Optical Sciences Lab.,Un.of Michigan)

VII.6 Conclusion

The determination of the phase of a scattered wave has long been a significant and formidable problem for the x-ray crystallographer. Without some knowledge of both the magnitude and the phase of the scattered waves it does not readily appear possible to obtain a complete, well-resolved 'image' of the crystal specimen. The first attempts to solve the phase problem appear to have been made in 1939 by Bragg^{35,36} and by Buerger³⁷⁻³⁹, who noted an early suggestion by Boersch.⁴⁰

Buerger and Bragg demonstrated that an 'image' of a crystal could be obtained by placing suitably manufactured optical phase plates over the various diffraction points in the reciprocal lattice, and then optically Fourier transforming the composite arrangement in what amounts to an optical image-synthesizer. The technique has been very successful, but it clearly requires some a priori knowledge of the phases of the scattered waves, such as that which is available with centro-symmetrical crystals, for example³⁵⁻³⁹. More recently, Kendrew^{41,42} and co-workers have successfully "synthesized" images of crystals by electronic-computer Fourier-transformation of x-ray diffraction patterns, in particular in those cases where the "heavy-atom" isomorphous replacement technique was applicable, for instance in protein crystals such as the myoglobin molecule⁴³. A general applicability to "x-ray microscopy" of either the Kendrew or of the Buerger and Bragg methods appears to have been restricted, in their present forms, by the difficulty of ascertaining the phases of the various diffraction spots in the reciprocal lattice, in the general case.

A new approach towards the solution to the "phase problem" in x-ray microscopy has recently been proposed by Stroke and Falconer:⁴⁴ it is based on the general principles of wavefront-reconstruction imaging first proposed by Gabor in 1948.¹⁻⁵ Gabor proposed that a

coherent background be superposed onto the diffraction pattern, so as to provide a reference wavefront for the recording of both the amplitude and the phase of the scattered waves. Success in Gabor's method of holographic microscopy in the visible domain was immediately demonstrated by Gabor himself¹⁻⁵ and has been verified by many others since that time. However, it quickly appeared that high resolutions at x-ray wavelengths would be unattainable, in the application of conventional projection holography, because of the difficulties associated with film-resolution and source-dimension discussed in section E above. Indeed, resolutions of only 5,000Å to 10,000Å, rather than 1Å appeared attainable by means of "conventional" projection micro-holography. In noting the basic similarity between the Buerger and Kendrew image-synthesizing methods, on one hand, and micro-holographic image reconstruction, on the other, Stroke and Falconer have argued³² that high resolutions in micro-holography should be attainable by a suitable modification of the early principles of holography. In particular, Stroke and Falconer have shown that "Fourier-transform holograms" permit one to overcome the "source-emulsion problem" characteristic of conventional projection holography. Stroke and Falconer have also shown⁴⁵ how to obtain Fourier-transform holograms by lensless Fourier transformation, thus preserving the original advantages of "lensless" photography first suggested by Gabor. Scattered-light illumination and "structured" source apertures, based on interferometric resolution-luminosity-coherence considerations, are likely to have a primary role in the complete solution of the "source-aperture" problem^{22,33,44,45} in micro-holography, as are "lensless" Fourier-transform recordings of holograms.^{44,45}

REFERENCES

1. GABOR, D., "A New Microscopic Principle," Nature, 161, 777 (1948)
2. GABOR, D., "Microscopy by Reconstructed Wavefronts, I" Proc. Roy. Soc. (London), A 197, 454 (1949).
3. GABOR, D., "Microscopy by Reconstructed Wavefronts II", Proc. Roy. Soc. (London), B64, 449 (1951).
4. GABOR, D., "Research, 4, 107 (1951).
5. GABOR, D., "Generalized Schemes of Diffraction Microscopy," Proc. Congres de Microscopie Electronique, (1952).
6. ROGERS, G., "Gabor Diffraction Microscopy: the Hologram as a Generalized Zone-Plate," Nature, 166, 237 (1950)
7. ROGERS, G., "Experiments in Diffraction Microscopy", Proc. Roy. Soc. (Edinburgh), A58, 193 (1950-51).
8. ROGERS, G., "Artificial Holograms and Astigmatism", Proc. Roy. Soc. (Edinburgh), A63, 193 (1952).
9. BRAGG, W, and G. ROGERS, "Elimination of the Unwanted Image in Diffraction Microscopy", Nature 167, 190 (1951).
10. HAINE, M., and J. DYSON, "A Modification to Gabor's Diffract Microscope," Nature, 166, 315 (1950).
11. HAINE, M. and Mulvy, "The Formation of the Diffraction Image with Electrons in the Gabor Diffraction Microscope," J. Opt. Soc. Am., 42, 763 (1952).
12. BAEZ, A.V., "A Study in Diffraction Microscopy with Special Reference to X-Rays", J. Opt. Soc. Am. 42, 756 (1952).
13. BAEZ, A.V., and H. M. A. EL-SUM, "Effects of Finite Source Size Radiation Bandwidth and Object Transmission in Microscopy by Reconstructed Wavefronts," in X-ray Microscopy and Microradiography, Academic Press, Inc., New York, 1957, pp. 347-366.
14. EL-SUM, H. M. A., Reconstructed Wave-Front Microscopy, Ph.D. Thesis, Stanford University, November 1952; available from University Microfilms, Inc., Ann Arbor (Dissertation Abstracts 4663, 1953).
15. KIRKPATRICK, P., and H. EL-SUM, "Image Formation by Reconstructed Wave-fronts," J. Opt. Soc. Am., 46, 825 (1956).
16. U. S. Patent No. 3,083,615, Granted 2 April 1963 (Filed 16 May 1960 by H. M. A. El-Sum, assignor to Lockheed Aircraft Corp.), Optical Apparatus for Making a Reconstructing Holograms.
17. EL-SUM, H. M. A., "Information Retrieval from Phase Modulating Media," in Optical Processing of Information, Donald K. Pollack, Charles J. Koester, and James T. Tippett, Eds., Spartan Books, Inc. Baltimore, Md., 1963, pp. 85-97.
18. LOHMANN, A., "Optische Eiseitenbandudertragung angewandt auf das Gabor-Mikroskop," Opt. Acta, 3, 97-99 (1956).
19. LEITH, E. and J. UPATNIEKS, "Reconstructed Wavefronts and Communication Theory," J. Opt. Soc. Am. 52, 1123 (1962).
20. LEITH, E., and J. UPATNIEKS, "Wavefront Reconstruction with Continuous-Tone-Objects," J. Opt. Soc. Am. 53, 1377 (1963).

21. LEITH, E., and J. UPATNIEKS, "Wavefront Reconstruction with Diffused Illumination and Three-Dimensional Objects," *J. Opt. Soc. Am.* 54, 1295 (1964).
22. STROKE, G. W., An Introduction to Optics of Coherent and Non-Coherent Electromagnetic Radiations, Engineering Summer Conferences Text, The University of Michigan, Ann Arbor, Michigan, May, 1964, 77 pages.
23. STROKE, G.W., "Diffraction Gratings," in Handbuch der Physik, Vol. 29, S. Flügge, Ed., Springer Verlag, Berlin and Heidelberg, (in print).
24. ZERNIKE, F., "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasen Kontrast Methode," *Physica*, Haag, 1, 43 (1934).
25. ZERNIKE, F. *Z. Tech. Phys.* 16, 454 (1935).
26. ZERNIKE, F., "Das Phasenkontrastverfahren bei der mikroskopischen Beobachtung," *Physik Z.* 36, 848 (1935).
27. WOLTER, H., "Schlieren-, Phasenkontrast- und Lichtschnittverfahren," in Handbuch der Physik, 24, S. Flugge, Ed., Springer Verlag, Berlin, 1956, pp. 555-645.
28. MARECHAL, A., et P. CROCE, "Un filtre de fréquences spatiales pour l'amélioration du contraste des images optiques," *Compt. Rend.*, 237, 607 (1953).
29. ELIAS, P., "Optics and Communication Theory," *J. Opt. Soc. Am.* 43, 229-232 (1953).
30. O'NEILL, E. L., "Selected Topics in Optics and Communication Theory", Optical Research Laboratory, Boston University, Technical Note 133, October 1957.
31. CUTRONA, L. J., E. N. LEITH, C. J. PALERMO, AND L. J. PORCELLO, "Optical Data Processing and Filtering Systems," *IRE Trans. Inform. Theory*, IT-6, no. 3, 386-400, (June 1960).
32. STROKE, G. W., and D. G. FALCONER, "Attainment of High Resolutions in Wavefront-Reconstruction Imaging", *Physics Letters* 13, 306, (1964).
33. STROKE, G. W., Private Communications to E. N. Leith and Associates, August 1963 to November 1964.
34. KELLSTROM, G., "Experimentelle Untersuchungen über Interferenz- und Beugungserscheinungen bei langwelligen Röntgenstrahlen," *Nov. Act. Reg. Soc. Sci. Uppsaliensis*, 8, No. 5, (1932).
35. BRAGG, W. L., "A New Type of X-Ray Microscope," *Nature* 149, 470 (1942).
36. W. Bragg, "A New Type of X-ray Microscope", *Nature* 149, 470 (1942)
37. BUERGER, M. J., "Optically Reciprocal Gratings and their Application to the Synthesis of Fourier Series", *Proc. Nat. Acad. Sci.* 27, 117 (1941).
38. BUERGER, M., "Generalized Microscopy and the Two-Wavelength Microscope", *J. Appl. Phys.* 21, 909 (1950).

39. BUERGER, M. J., "The Photography of Atoms in Crystals," Proc. Nat'l. Acad. Sci., 36, 330-335 (1950).
40. BOERSCH, H. "Zur Bilderzeugung im Mikroskop", Z. Tech. Physik, 337-338 (1938) especially footnote 3, p. 338.
41. KENDREW, J. C., G. BODO, H. M. DINITZ, R. G. PARRISH, W. WYCKOFF and D. C. PHILLIPS, "3-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis," Nature, 181, 662 (1958).
42. KENDREW, J. C., "Three-Dimensional Structure of Globular Proteins", in Biophysical Science--A Study Program, Ed., J. L. Oncley (John Wiley and Sons, New York, 1959), p. 94.
43. PERUTZ, M., "The Hemoglobin Molecule", Scientific Am. 211, 64 (1964).
44. STROKE, G. W. and D. G. FALCONER, "Attainment of High Resolutions in Holography by Multi-Directional Illumination and Moving Scatterers", Physics Letters, 15, 238 (1965).
45. STROKE, G. W. and D. G. FALCONER, "Lensless Fourier-Transform Method for Optical Holography", (in print).

The University of MichiganElectrical Engineering Department

Lectures: M-W-F 10-11

Prof. George W. Stroke

East Eng 3515

Conference hours: Monday 11-12

75(170) Optics of Coherent and Incoherent Electromagnetic Radiations

Prerequisites: Elec. Eng. 220 and Math. 216 (3)

Elective subject introducing the field of electro-optical science. Topics include a review of the properties of light and of basic geometrical and physical optics; operational Fourier-transform and matrix treatment of optical image-forming processes diffraction, spectroscopy and communications; optical filtering and computing; relativistic statistical and coherence properties of light; coherent light generation amplification and control with optical masers; wave-propagation, frequency multiplication and shifts in non-linear media. Similarities and relations with microwave and electronic systems techniques.

- . Properties of light:
 - a. emission, propagation, detection
 - b. quantum, electromagnetic-wave & relativistic characters
- . Geometrical optics (lumped parameter treatment):
 - a. image formation;
 - b. luminosity
- . Physical optics: interference, diffraction, polarization
- . Matrix formulation of geometrical optics
- . Fourier-transform approximation of electromagnetic diffraction problems in optics: Aperture-Diffraction pattern relationship
- . Object-image and aperture-image relationships in coherent & incoherent light (Convolution-integral and Fourier-transform expressions)
 - spatial filtering and spatial frequency response
 - image contrast
 - two dimensional character of optical imaging
 - resolution & resolving power
- . Effects of optical element imperfections: aberrations
- . Special problems in optical image-processing & spectroscopy
 - phase contrast
 - apodization
 - interferometry
 - infrared
 - coherent background detection
 - (cf. superheterodyning)
 - optical filtering
 - optical computing
 - grating & interferometric spectroscopy
 - small phase difference measurement
- . Coherence character of electromagnetic radiations
 - incoherence, coherence and partial coherence; monochromaticity
 - applications to interferometry, radio astronomy, microscopy, image formation
- . Coherent light generation with optical masers (lasers)
- . Applications of lasers and "optical electronic" I.: Analogies and relations with microwave electronics. - communications ranging (Doppler, etc.), optical radar; - image processing; - optical heterodyning; - modulation and demodulation
- . Applications of lasers and "optical electronics" II.: Wave-propagation in non-linear media - high-field intensity properties of dielectrics; - suppressed carrier modulation; - frequency multiplication and frequency shifts of optical radiations; - parametric oscillators
- . Special topics in statistical optics coherence theory and polarization
 - fluctuations in light beams; - optical information theory; - noise;
 - bunching effects; photoelectric coincidence and mixing effects
- . Special problems in the computing, design, manufacture & testing of optical instruments

te: In general an introductory knowledge of unfamiliar mathematics will be sufficient for the comprehension of the course. Mathematical background, reviews and exercises will be introduced as required.

TABLE OF CONTENTS

I.	INTRODUCTION.....	2
I.1	Emergence of Modern Optics as a Branch of Electrical Engineering.....	2
I.2	Mathematical Character of Electro-Optical Engineering.....	3
I.3	Mathematical Methods of Modern Optics.....	4
I.4	Some Limitations of Optical Formulation of Optical Image Formation and the Need for Boundary-Value Solutions in the Study of "Diffraction" of Electromagnetic Waves in Optics.....	5
I.5	Grating Equation, as Example of Boundary-Value Solution of a Diffraction Problem.....	7
I.6	Gratings as Information-carriers in Optics. (Application to "Wave-front Reconstruction").....	15
I.7	Optics and Communication Theory.....	27
II.	DIFFRACTION THEORY.....	29
II.1	The Two Aspects of the Diffraction of Light.....	29
II.2	Theoretical Calculation of Energy Distribution in Diffraction and of Spectral Diffraction Patterns.....	29
	1. Electromagnetic Boundary-Value Solutions.....	29
	2. Image Formation Solutions Using "Huygens' Principle".....	34
II.3	Image Formation Theory and Optical Signal Processing in Fourier-Transform Formulation.....	36
III.	IMAGE FORMATION IN NON-COHERENT LIGHT.....	38
III.1	Image of Point Source.....	38
III.2	Summation of Light from Several Source Points Reaching One Image Point.....	39
III.3	Spread Function.....	40
III.4	Image of Extended Source in Non-Coherent Light...	40
III.5	Analysis of Image Formation in Terms of Spatial Harmonics.....	43
III.6	Physical Significance of "Spread" Function and of Spatial Harmonic Analysis of Image Formation.....	44
IV.	COHERENCE CHARACTERISTICS OF LIGHT.....	47
IV.1	Introduction.....	47
IV.2	Characterization of Time-Coherence.....	47
	1. Correlation Method.....	47
	2. Alternate Way of Looking at Time Coherence...	48

TABLE OF CONTENTS

Continued

	a. Monochromatic (single frequency) Light...	48
	b. Polychromatic Light.....	49
	3. Compare Correlation Method with Interferometric Method.....	50
	4. Narrow Spectrum.....	51
	5. Photoelectric Interferometry with Gaussian Line-shapes.....	52
	6. Physical Significance of Power Spectra.....	53
	a. Case of Single-Frequency Signal.....	53
	7. Heterodyne Analysis of Signals, Beat Frequencies, etc.....	56
	8. Spatial Coherence.....	56
	9. Partial Coherence with Extended Non-Coherent Source.....	60
	10. Intensity Correlations in Partially Coherent Fields.....	65
	11. Intensity Interferometers.....	66
V.	IMAGE FORMATION IN COHERENT LIGHT.....	69
	V.1 Introduction.....	69
	V.2 "Coherent" Illumination.....	69
	V.3 Image Formation in Coherent Light, Considered as Double Diffraction.....	70
	V.4 Abbe Resolution Criterion.....	71
	V.5 Transfer Functions in Coherent and in Non-Coherent Light.....	73
	V.6 Phase-Contrast Filtering.....	74
	V.7 Optical Filtering with Interferometrically Matched Spatial Filters.....	76
	V.8 Optical Computing Correlating and Signal Processing.....	82
	1. Spectrum Analyzers.....	82
	2. Optical Cross-Correlators.....	83
	V.9 Note.....	84
VI.	CONVOLUTIONS, SPECTRAL ANALYSIS AND THE THEORY OF DISTRIBUTIONS.....	85
	VI.1 Introduction.....	85
	VI.2 Distributions.....	85
	VI.3 Impulse Response of Linear System.....	85

TABLE OF CONTENTS

Continued

VI. 4	Response of Linear System to Any Input Function.....	87
VI. 5	Spectral Analysis.....	87
VII.	OPTICAL HOLOGRAPHY (WAVEFRONT-RECONSTRUCTION IMAGING).....	89
	Introduction.....	89
VII. 1	Background and Experimental Foundations.....	90
VII. 2	Theoretical Foundations.....	92
	A. The Recording Process.....	95
	B. The Reconstruction Process.....	97
	C. Physics of the Method.....	98
	D. Magnification.....	100
	E. Resolution.....	102
	F. Temporal Coherence Requirements.....	104
	G. Steady-State Coherence Requirements.....	104
VII. 3	Summary and Results.....	105
VII. 4	Attainment of High Resolutions.....	107
VII. 5	Electron-Microscopy and X-Ray Hologram Microscopy.....	108
	Figures.....	109-127
VII. 6	Conclusion.....	128
	References.....	130
EE475 -	COURSE ON "OPTICS OF COHERENT AND NON-COHERENT ELECTROMAGNETIC RADIATIONS, As taught at the Univer- sity of Michigan.....	133