Invited Paper

Teaching Holography: Holography as a Teaching Tool

Nils H. Abramson

Industrial Metrology, Royal Institute of Technology
S-100 44 STOCKHOLM, Sweden

The holo-diagram is a diagram originally designed as a tool for teaching the making and evaluation of holograms. It was soon found to include much more general uses than first expected. During the 20 years since its introduction we have applied this diagram to explain many different fields such as holography, interferometry, diffraction. Fermat's principle, light-in-flight recordings, three-dimensional measurements with ultrashort lightpulses and the special theory of relativity.

1. HOLOGRAPHY

The set up of Fig. 1 can be used to produce a hologram. When a photographic plate is placed where the two beams intersect the interference fringes will be recorded, but only if the difference in pathlength for the beams is shorter than the coherence length of the laser light. After that the plate is processed it is placed back in the previous position and illuminated by just one of the two beams. By diffraction the recorded interference fringes will deflect part of the beam so that also the other beam is formed.

The described experiment is a simple example of a holographic process. During exposure we can for example designate the left and the right beam the reference respectively the object beam. Later on we reconstruct the object beam by illuminating the hologram plate with the reference beam which is now designated the reconstruction beam.

The reconstruction could perhaps be explained in the following way. As the reconstruction beam passes through the hologram plate it gets shadowed at every place where a dark fringe was formed. The beam can not distinguish these shadows from the interference fringes and accordingly leaves the plate in the form of two beams with exactly that angle which would have produced those dark fringes.

The holographic process works even if there are many object beams that arrive at different angles. The object beam can even consist of all the millions of lightrays arriving from a diffusely reflecting object. In that case the interference fringes will look extremely complicated but still the reconstruction beam can sort them all up, like "unscrambling scrambled eggs". It is, however, important that the reference beam is stronger than the sum of all the object beams because otherwise they start to reconstruct each other.

2. THE HOLO-DIAGRAM

Take a string that is about one meter long and make at one end a knot with a small loop. Make another knot some twenty centimeter from the other end. Nail the two knots, with a horizontal separation of about fifty centimeters, to a vertical drawingboard. Tighten the string with the tip of a horizontal pencil and move the pencil around. It will then draw an ellipse with the two nails as focal points. Fig. 2.

Let the left focal point, which we name A, represent a pointsource of light, while the right one, B, represents the center of a hologram plate. If two small mirrors are placed anywhere on the periphery

 0-8194-0732-1/92/$4.00

of the ellipse, and parallel to it, they will both reflect the light from A to B. What is more, the pathlengths will both be equal to the stringlength and thus the pathlength difference will be close to zero. Therefore interference fringes are formed, and a hologram is produced at B, even if the coherence length of the light is very short.

Now make two more knots, one on each side of the original right knot. The separation of the three knots should be the coherence length, which could be some fifteen centimeters. Put a nail through the right of the three knots and draw with the pencil a new ellipse, which accordingly becomes larger than the original one. Finally nail the other knot and draw a smaller ellipse.

If we remove one of the mirrors from the original ellipse to anywhere on any of the two new ellipses, there will still be produced a hologram at B, because in no case will the pathlength difference be larger than the coherence length. Accordingly any object of any shape can be holographed as long as it can be positioned in between the two outermost ellipses, while the reference mirror is placed at the center ellipse.

If a set of ellipses are drawn with a pathlength difference of one coherence length between adjacent ellipses the holo-diagram of Fig. 3 is formed. To make the situation more clear every second area between the ellipses is painted black. Finally we make the holo-diagram of Fig. 4 where one set of ellipses is drawn and also one set of arcs of circles the meaning of which we will explain later. If the reference mirror is placed on one ellipse, everything within those adjacent ellipses can be recorded.

Finally it should be pointed out that what we have described here is only a two-dimensional section of the real three-dimensional world. If there existed a three-dimensional drawingboard, then we could instead move the pencil in three dimensions all the time keeping the string tight, thus producing rotational symmetric ellipsoids.

Using the method of Fig. 4 we have succeeded to make hologram so objects more than two meters "deep" in spite of a coherence length of only 15 cm. (Ref. 1) As far as we know a world record at the time it was made. Another example Fig. 5 shows how a two meters high vertical milling machine was placed within the coherence ellipsoids. The reference mirror was placed on the middle of three broken lines representing the ellipsoids. The laser and the hologram plate were both placed about five meters to the left of the machine. Fig. 6 shows the resulting hologram. Almost the whole machine is recorded, only those parts are missing that were sticking out of the coherent area as seen in fig. 5. The fringes reveal a deformation as described later.

3. HOLOGRAPHIC INTERFEROMETRY

In Fig 4 two adjacent ellipses represent a pathlength difference of one coherence length. If they instead represented one wavelength, then the holo-diagram could be used to visualize and to evaluate interference fringes. If an object point is moved from one ellipsoid half way to an adjacent one, then the phase at B will change by 180 degrees, which means that the interference pattern at B moves half a fringe separation.

The result will be, that if we make a double exposure with such an object motion in between the two exposures, then the fringes will be wiped out at B, and the corresponding object point appears dark. Consequently, if an object is fixed to one end and the other end is moved so that it crosses say five ellipsoids, then the object will be covered by five fringes in the reconstructed holographic image.

A movement parallel to one ellipsoid causes no pathlength difference and therefore no fringes while a movement perpendicular to the ellipsoids causes most fringes. Thus the latter represents the sensitivity direction for the fringeforming process.

We designate the separation as k times the wavelength. The numerical value of k is one over cosine half the angle ACB where C is an object point somewhere in the diagram as seen in Fig. 4. The perpherical angle on a circle segment is constant and therefore also the k-value is constant along arcs of circles that pass through A and B. Thus we have in the diagram printed the k-value where these circles cross the Y-axis. When the position in the holo-diagram is known for a studied object point, then the sensitivity is known both to amplitude and direction. The displacement perpendicular to the ellipsoids is calculated as the number of fringes seen on the object multiplied by k times half the wavelength. In this way the holo-diagram can be used to simplify the planning of the holographic set up and to evaluate the displacement from the number of fringes.

By the use of the holo-diagram we have managed to lower the sensitivity of holographic interferometry so much that a movement of 2 mm caused only two fringes. This value should be compared to the ordinary high sensitivity of one fringe for a movement of 0.3 thousandth of a millimeter. We also made an interferometer the *"interferoscope"* in which the sensitivity could be changed from one micrometer to 5 micrometers per fringe just by changing the k-value. Thus, a greater k-value caused by more grazing incidence of the light rays works just *as if* there was a longer wavelength or a *red-shift* of the light.

4. MOIRÉ ANALOGY TO THE HOLO-DIAGRAM

There exists another very different pathway towards the understanding of the holo-diagram, see Fig. 7. As before, A is the pointsource of coherent light, thus A can be visualized as a point that is the center of concentric equally spaced circles, which move outwards with the speed of light.

As before, B is the point of observation which only reacts to coherent light, thus B can be visualized as a point that is int he center of concentric equally spaced circles, which with the speed of light move inwards, Where the two sets of circles interact, a moiré pattern forms one set of hyperbolas and one set of ellipses. If the spacing between the circles were half a wavelength the hyperbolas would be identical to *Youngs Fringes,* the interference fringe pattern caused by illumination from two pointsources of coherent light. Their spacing would also represent the *diffraction limited resolution* of a lens that we can almost visualize as positioned with its diameter touching A and B.

To make the situation more clear let us study Fig. 3 again where the ellipses are drawn in such a way that the pathlength difference between the middle of a bright fringes respective a dark fringe is half a wavelength.

The ellipses are identical to the ellipses of the holo-diagram and thus represent *the interference limited resolution* of a set up where A is the light source and B is the point of observation.

The ellipsoids will everywhere reflect light from A to B, and, what is very important, the angles and the separations are such that all the reflected lightrays arrive in phase at B. It is said that the *Bragg conditions* are everywhere fulfilled by the ellipsoids.

A glass surface that is introduced anywhere among the ellipsoids will deflect light from A to B by *diffraction,* if it were covered with one grove for each ellipsoid that is intersected. Depending on the position of the glass either a *transmission (E)* or a *reflection (F) grating* is formed.

Fermat´s Principle says that light reflected from a mirror of any shape always chooses the fastest pathway. Let the circle at D in Fig. 3 represent a soapbubble. Apparently the principle is right, a reflection will form where the bubble tangents, and thus is parallel to, an ellipsoid. At the front surface of the bubble this point certainly represents the shortest pathlength because it is the only point reached by the string used to draw tat tangential ellipsoid. But Fermat´s Principle has to be extended, the back surface of the bubble also reflects light to B, and it represents the longest pathlength. Using the holo-diagram it has been possible to introduce another modification so that the Principle besides reflections also includes diffractions. (Ref. 2)

If A and B both are lightsources, then the ellipses move with a speed faster than light, while the hyperbolas will be stationary. This is why we can record fringes and holograms and measure the phase with high accuracy even with slow detectors in spite of the fact that the frequency of light is enormously high, some ten to the 14 cycles per second.

If a is a lightsource and B is a point of observation, then the hyperbolas will move with a speed faster than light, while the ellipses are stationary. This explains why we can make holographic recordings even with long exposure time of the wavefronts themselves in spite of the extremely high velocity of light, some 300 00 km per second.

5. MOIRÉ ANALOGY TO HOLOGRAPHIC INTERFEROMETRY

Now that we have found moiré analogies to interferometry and to the holo-diagram there are good reasons to believe that there should also exist a simple moiré analogy to the interference fringes formed on the image in a double exposed hologram. If the object between the exposures has made a rigid motion and if the point of illumination and the point of observation are close together and far from the studied object, in that case the fringe pattern will be made up of straight lines and thus simple and easy to understand. However, in other situations the pattern might become quite complicated and in that case a moiré analogy would be very useful to visualize the interference process.

It did take us quite a long time until we suddenly understood how very simple the moiré analogy was. Let us make two transparent copies of Fig. 3 and place one on top of the other so that they exactly coincide.

What would the fringepattern be on any object displaced between the two exposures parallel to the x-axis of the holo-diagram? To find out move one of the transparencies slowly in that direction. When the displacement is a quarter of a wavelength all the x-axis outside A and B becomes dark. After a motion of another quarter it becomes bright again and so on. This effect represents the sensitivity of an ordinary interferometer or of holographic interferometry for objects placed at the x-axis. Everywhere else the fringes arrive later because the sensitivity to the introduced movement is lower. The fringes that are formed on the transparencies are everywhere analogous to the fringes formed on objects placed at corresponding position in the space of the holo-diagram. The maximum sensitivity to fringe formation is where the motion is perpendicular to the ellipses and where these are most closely spaced.

6. LIGHT-IN-FLIGHT RECORDING BY HOLOGRAPHY

Now that we have spoken so much about the importance to holography of the ellipsoids it would be interesting to actually see them. There is one rather simple way to do that. If you remember, the holo-diagram was invented to overcome the problems and limitations of a short coherence length. In a hologram there is recorded only what is within the space of two ellipsoids. Therefore, if the coherence is very short and we try to make a holographic image of a flat screen, then the recording would consist of only an elliptical line representing the intersection of the screen by one of the ellipsoids. As before, one focalpoint is the point source of illumination while the other is the point of observation. Thus in this way we really can see intersections of the ellipsoids of the holo-diagram. By moving the eye behind the plate during observation, the elliptic line will move over the screen because the eye represents one focal point of the ellipsoids. (Ref. 3)

We know that light from a point source expands int he form of spherical waves, so why do we see intersections of ellipsoids? One obvious reason is that through one point on the hologram plate we see only that scattered light which has travelled the same time and thus the same pathlength as the reference beam. Another way to explain the same thing is that it does not only take time for the light to move to the screen, *it also takes time for the light to move from the screen to the hologram plate.* If different parts of the screen are at different distances from the plate, then there will be different delays in the observation and consequently different parts of the screen are observed at different points of time . However, if we place the screen at a large distance and perpendicular to our line of sight, then the distance from every point of the screen to the point of observation will be about equal. It can be proved that for that configuration the intersections of the ellipsoids are identical to those of spheres. See Fig. 7. Along circles around B the intersections of the ellipses are identical to the intersections of the circles around A. Therefore the bright line we see on the screen represents the spherical waves themselves without distortions. Because of the already mentioned analogy between light of short coherence length and short pulse length we will in the following only discuss short pulses, and show how the described method can be used to study ultrashort pulses of light as they fly by.

The holographic plate records a hologram only when it is illuminated by the object beam and simultaneously by the reference beam. If therefore the latter consists of a picosecond pulse the object will be recorded only during that picosecond, which represents a lightpath of about 0.3 mm. If the reference pulse illuminates the plate almost parallel to its surface e.g. from the left then the left part of the plate will record what happens first to the object, while the right part records what happens last. The time span recorded will be the time it takes for the light to move from one edge of the plate to the other. Thus the picosecond reference pulse works like a *light-shutter* which with the speed of light, or even faster, moves over the plate. When the plate is studied during reconstruction. a motion of the eye along the plate results in a continuous change of time at the recorded object.

Using this technique it has been possible to produce continuous frameless moving pictures of light reflected by a mirror. Fig. 8, focused by a lens Fig. 9 and many other situations where light is reflected, refracted or diffracted.

When a more general three-dimensional light-in-flight recording is evaluated one has to compensate for the apparent distortions of the wavefronts. For this purpose the following general rule derived from the holo-diagram can be applied:

A wavefront recorded by any ultra highspeed method appears transformed into a mirror surface that would reflect the light from the point of illumination to the eye of the observer.

Thus, a spherical wave appears distorted into one of the ellipsoids of the holo-diagram. A flat wavefront that passes by appears to the first approximation to be tilted fortyfive degrees, at a closer study it is found to be distorted into a paraboloid, the focalpoint of which is the eye of the observer.

The pulse appears to move perpendicular to the ellipsoid or the paraboloid and therefore the pulse length and the pulse speed both appear to be multiplied by the k-value of the holodiagram. As k is larger than one, the velocity of the pulse appears to be faster then light.

This phenomena of superluminous velocity can be studied in quite another field, astronomy. Instead of using ultrashort recording time to study pulses of light, we can use an ordinary time scale and study ultralong distances. Thus, in 1988 there was found visible lightechoes from the explosion of a supernova at a distance of some fifty thousand lightyears. About one thousand lightyears from earth was a dusty cloud, the Large Magellanic Cloud, that scattered the light and therefore the lightpulse from the supernova was seen as a ring around the star that expanded at about ten times the speed of light. Explained by the holodiagram this means that the cloud which was between A and B intersected one ellipsoid where the k-value was about ten.

7. RELATIVITY

The whole reason for the design of the holo-diagram is the separation of the point of observation B from the point of illumination A. Had this separation been zero then all the ellipsoids of Figs. 2, 3 and 4 would have been spheres. The intersections by flat surfaces would have been Fresnel zone-plates and their moiré patterns simply straight fringes. It is the separation of A from B that distorts a spherical wavefront into an ellipsoid, and a flat wavefront into a paraboloid.

Thus all the described distortions are caused by the separation of A from B. When we make measurements with holography or any type of interferometry or radar we would make large errors if we do not compensate for those distortions. One very important fact is that exactly the same compensations are needed whether this separation between A and B is *static*, like in holography, or *dynamic*, e.g. the experimenter sits on a train and makes measurements with A and B being at the same point at the train. (Ref. 4 and 5) If the train is moving fast and the stationary world outside is measured then A and B will in relation to that world be separated because of the motion of the train during the time between illumination and observation. Because the speed of light always appears to be the same independent of the velocity of light source or observer, the compensations have to be identical to those of a static separation.

Thus the holodiagram can be used to understand the theory of *special relativity* and all the first and second order apparent distortions caused by a relative motion between experimenter and object can in a graphical way be derived from the holo-diagram. In this way it is found that the *Lorentz contraction* is just one of a number of apparent elongations and contractions of fast moving objects. Objects moving towards the observer appear elongated, while those moving away appear shortened, only those just passing by appear Lorentz contracted. Time behaves in a similar way. Clocks moving towards us appear faster, blueshifted, those moving away appear slower, redshifted, and those just passing by appear to have the ordinary relativistic *time dilation* or, transversal *redshift*.

8. REFERENCES

1. Nils Abramson, The making and Evaluation of Holograms. Academic Press, 1981.

2. Nils Abramson, Principle of least wave change. Journal of the Optical Society of America. A, Vol. 6. No. 5, 1989.

3. Nils Abramson, Light-in-flight recording: high-speed holographic pictures of ultrafast phenomena. Applied Optics. Vol. 22. No. 2, pages 215-232, 1983.

4. Nils Abramson, Light-in-flight recording 4: Visualizing Optical Relativistic Phenomena. Applied Optics. Vol 24. No. 20 pages 3323-3329, 1985.

5. Nils Abramson, Minkowski diagram in relativity and holography. Applied Optics. Vol. 27. No. 9, pages 1825-1832, 1988.

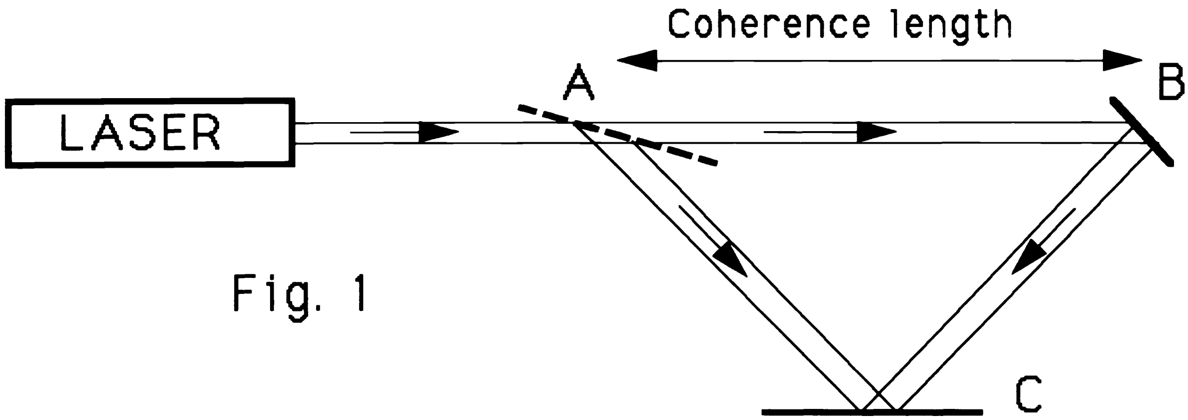

Fig. 1

The laser beam passes through a beamsplitter (A) and half of it is reflected by the mirror (B) towards a screen (C). The other half is reflected by the beamsplitter to the same spot on the screen. Dark interference fringes are seen where the two beams intersect only if the distance between (A) and (B) is shorter than the coherence length of the laser light.

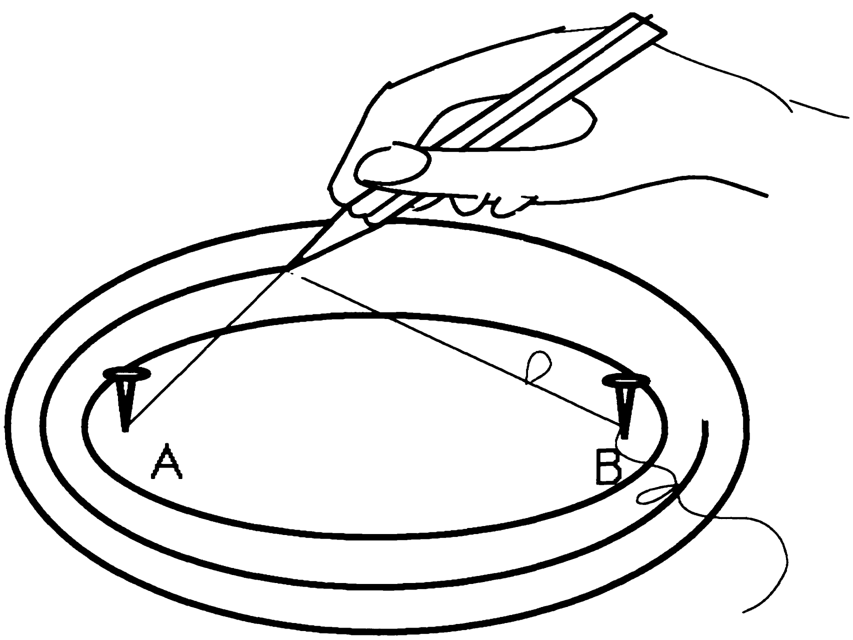

Fig. 2

Two nails are pinned down at (A) and (B), and a string with three knots is fixed to the nails. The separation of the knots is L. A pencil is moved around while the string is kept stretched. Because the stringlength is constant, the result is an ellipse with its focalpoints at the nails. Three ellipses are drawn. Between the making of each ellipse the string length is changed by fixing different knots to the nail (B).
(A) represents the pointsource of light with the coherence length of L, (B) represents the middle of the hologram plate.
If a reference mirror is placed on the middle ellipse every object positioned between the two outer ellipses can be holographed because they will be within the coherence length.

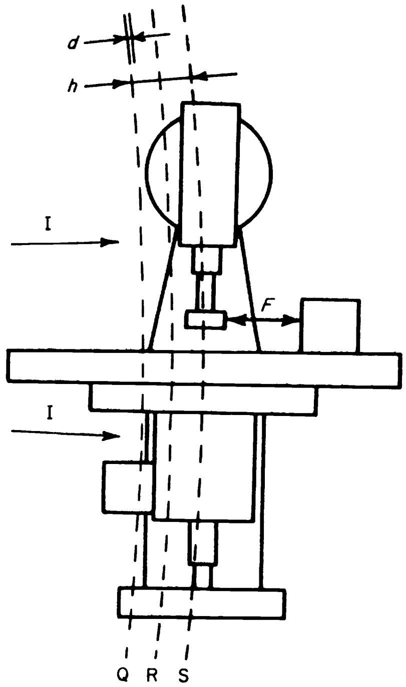

Fig. 5

Fig. 6

Fig. 5
Holographing an almost 2 m high milling machine using a laser with a coherence length of less than 15 cm. Most of the machine could be positioned between the coherence limiting ellipsoidal shells (Q) and (S). The displacement needed to produce one fringe is represented by d (which is of cource greatly exaggerated). h represents the coherence length. R is the ellipse on to which the reference mirror was placed, while (F) represents the static load causing the deformation.

Fig. 6
The milling machine was holographed twice, first without, then with the static load (F) of Fig. 5. Those parts of the machine are missing that were sticking out of the coherence ellipsoids of Fig. 5. Every fringe represents a displacement of about 0,3 mμ normal to the plane of the photograph. Straight fringes represent a tilt around an axis parallel to the fringes. Curved fringes represent deformation. A fixed reference surface is seen at lower right

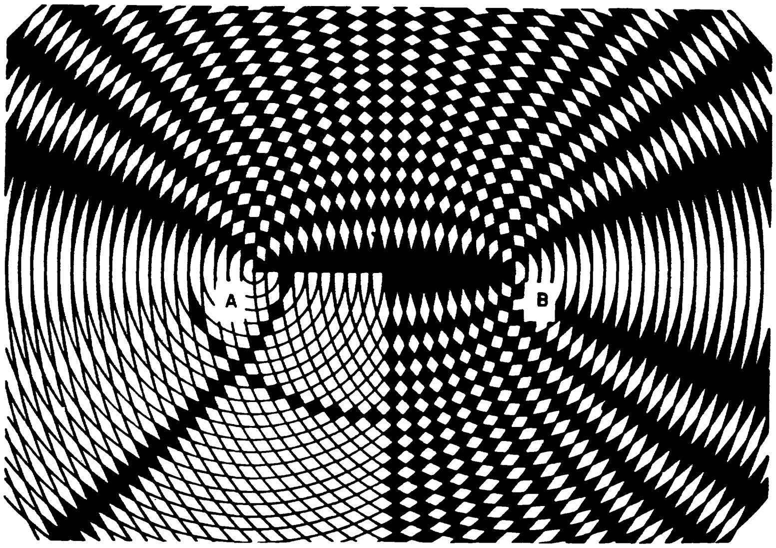

Fig. 7

(A) and (B) are the centres of two sets of concentric circles in a bipolar coordinate system. The moiré fringes form one set of ellipses and one set of hyperbolas. To emphasize these patterns every second rhomboid area has been painted black except for one quarter of the diagram where just one single ellipse and one hyperbola have been marked.

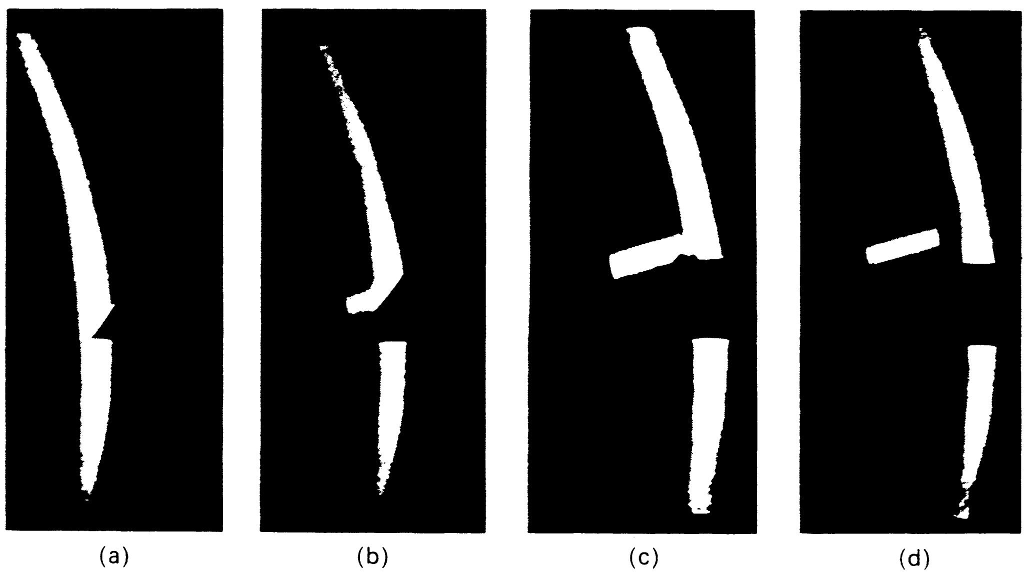

Fig. 8

a) A spherical wavefront from an argon laser enters at the left, illuminating a white painted flat surface at an oblique angel. The lower left end of a tilted mirror is just reached. b) The wavefront has reached the middle of the mirror, the normal of which is inclined at 40 degrees to the horizontal. The light is reflected upwards and to the left. c) All the reflected light is separating from the main wavefront which has just passed the mirror. d) The two components of the light have separated completely, the reflected light leaving a black hole in the spherical wavefront, which exits to the right.

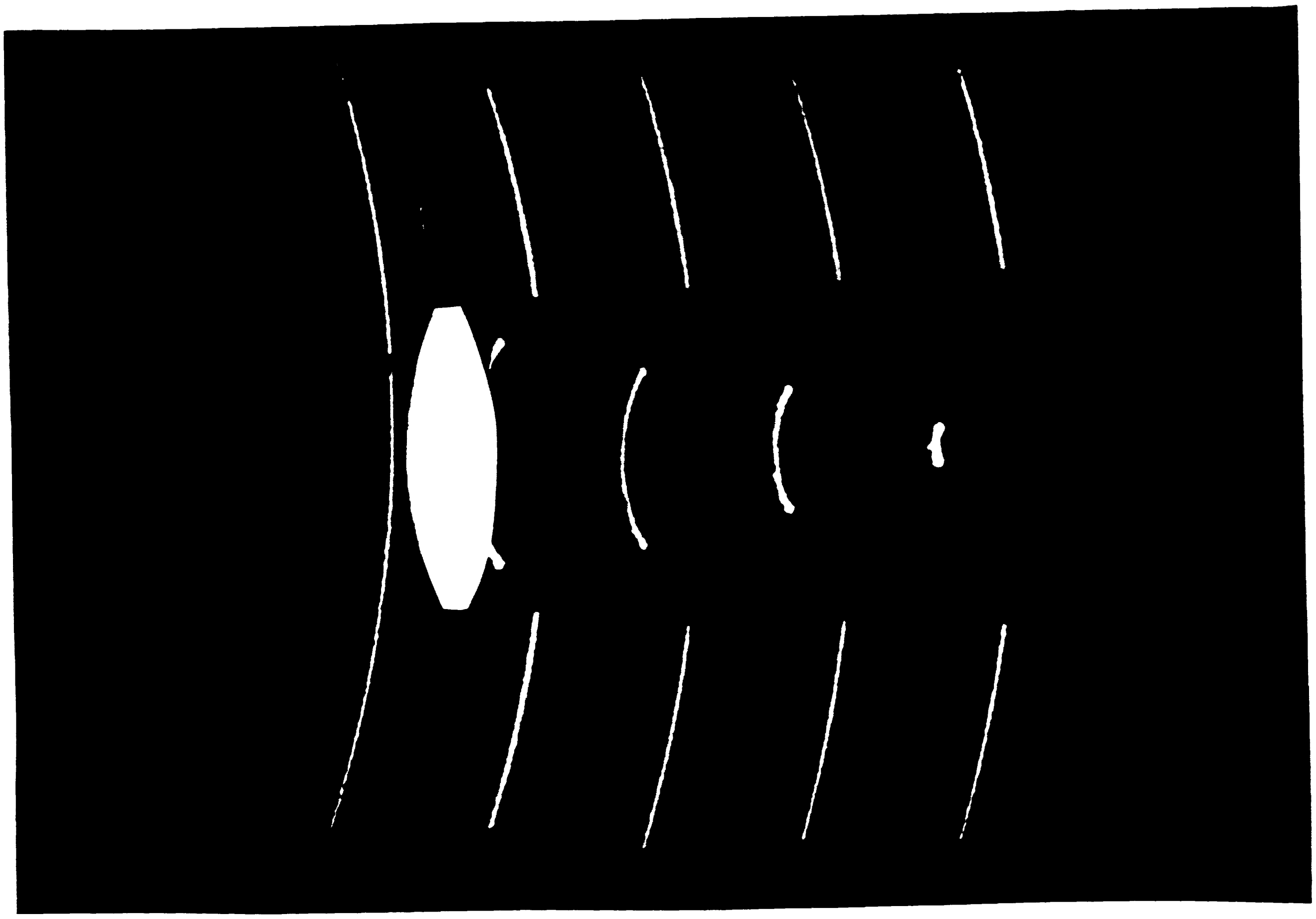

Fig. 9

A spherical wavefront enters from the left. The light that passes through the focusing lens is delayed because the speed of lighht is lower in glass than in air As the lens is thicker in the middle the convex wavefront becomes concave and moves towards focus. The image is a composite of five different photographs taken through different parts of the same hologramn. The pulselength is some 10 picoseconds which is equal to about 3 mm. The time separation between different photographed wavefronts is about 200 picoseconds.

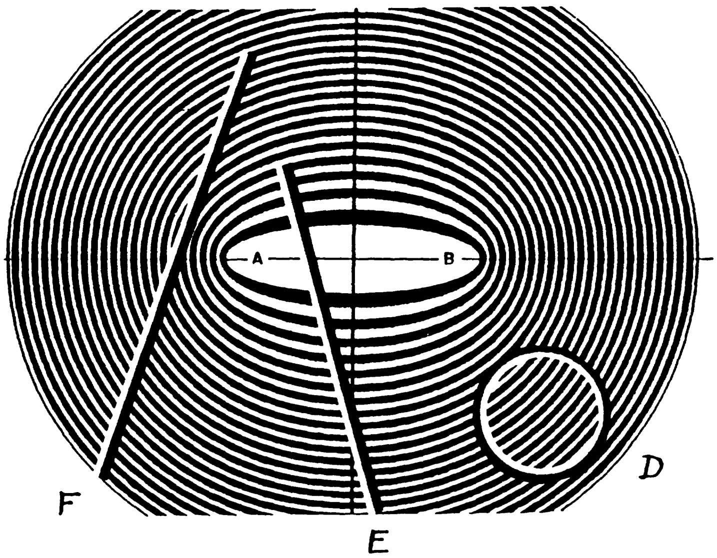

Fig. 3

A number of ellipses have been drawn with a pathlength difference representing half a wave-length. Evey second area between ellipses is painted black. Any glass surfacee.g. E or F of any shape with a groove for each intersection by an ellipse will diffract light from (A) to (B). The circle at (D) is a soap bubble which reflects light from (A) to (B) where it tangents the ellipses. The number of fringes seen between a fixed point and a displaced point studied by holographic inter- ferometry is a measure of the number of ellipses crossed by the displaced point.

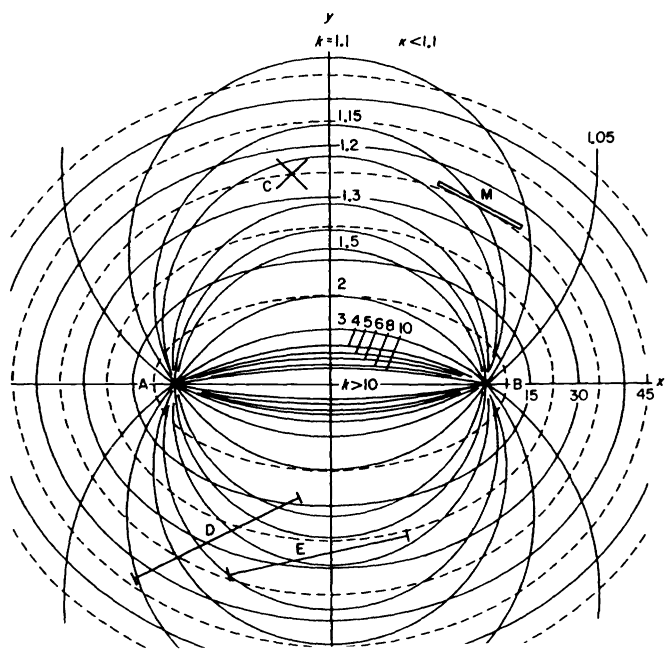

Fig. 4

The holo-diagram used tq utilize the coherence length. If the object (C) is positioned on the same ellipse as the reference mirror (M) almost no coherence length is needed. If the coherence length of the light is 15 cm, the object at (D) could not be recorded because the pathlength difference is more than 15 cm. A slight tilt of the object to the position (E) de-creases the difference to less than 15 cm so that a holographic recording can be made. If the pathlength difference between adjacent ellipses represents one wavelength, one interference fringe would form for an object motion from one ellipse to the next. The separation of the ellipses varies over the diagram, but is constant along arcs of circles named k-circles.