Reflections on the origin and subsequent course of holography

Emmett N. Leith
The University of Michigan, Ann Arbor, MI, 48109-2122, USA

ABSTRACT

I reflect on the various paths that led to holography, along with various other paths that could have led to holography. Also discussed is how holography, while maintaining a central core, exemplified by display holography, has expanded in scope and has diffused into vast areas of modern technology; in the process, the boundaries between holography and non holography have been blurred**.**

Keywords: holography, origins of, synthetic aperture radar, pulse compression

1. THE PATHS TO HOLOGRAPHY

I reflect on how various researchers, Gabor, Denisyuk and Leith and Upatnieks, came onto holography by three quite different paths, and how others, by yet quite different paths, could have come onto holography. We might think of holography as a destination, which can be arrived at by many, diverse paths, some rather direct, others rather circuitous (Fig. 1).

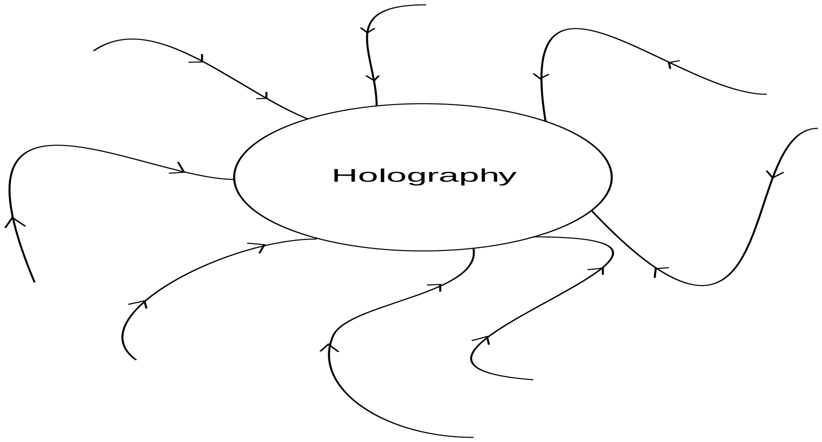

Fig. 1. The many paths to holography

Gabor, of course, arrived there first, by several years. His story is well known; he has described it and it has been recounted many times. It was a rather indirect path, involving such a seemingly unrelated problem of aberration correction in an electron microscope. This was a most ingenious way to arrive at holography, certainly the most ingenious one I can imagine. And I am of the view that the ingenuity of Gabor lies not just in his invention of holography, but also in the way that he arrived at it. It was a path that could have been discovered only by the most ingenious. The path that he chose significantly enhanced his discovery, in my opinion.

Denisyuk relates, in an earlier SPIE publication, a detailed account of his path to holography.[1] It begins with his reading of a science fiction novel, in which beings from outer space visit the earth millions of years ago and leave behind buried in the ground a strange photograph that was in 3-D and had animation. Denisyuk then embarked on a search for a way to invent such a photographic process. At the time of his discovery, was unaware of Gabor's earlier work. Had it

Practical Holography XVII and Holographic Materials IX, Tung H. Jeong, Sylvia H. Stevenson, Editors,
Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5005 (2003)

not been so, the story would have been far less interesting. Denisyuk also relates an all too familiar story--the skepticism of his peers in the Soviet Union, and the subsequent difficulty he had in getting his work published. Part of the difficulty lies in the rather radical ideas embodied in the work–that a light wave could be captured in flight and then regenerated so as to proceed on its previously interrupted course and form an image of the original object. Of course, Gabor's work was little known even to those who worked in optics.

I suppose that a second reason for the skepticism was the lack of highly impressive experimental demonstrations. An Hg arc source of sufficient intensity had coherence length of less than a mm, and his configuration allowed the reference and object beam paths to be equalized to the required degree only by placing the object against the recording plate, which led to an image plane hologram, not the dramatic 3-D imagery of which his paper spoke. Had the laser, with its light of considerable coherence length, been available, the story could have been quite different. The situation is somewhat different both for Gabor, and for Leith-Upatnieks; their systems. Their systems allowed the recording of objects deep in the Fresnel field, or even in the Fraunhofer field, with very little coherence requirement, as long as the objects were transparencies, instead of reflecting objects.

Our invention holography was carried out over a period of several years. The process consisted of two principal parts; the first was in the realm of synthetic aperture radar (SAR). In 1955 I was involved in the theory of coherent optical processing for converting the SAR raw data into a high resolution image. This was about three years before the system was built, so the work was strictly theoretical. To make a fairly long story short, in the process of developing the theory of coherent optical processing of SAR data, and of the way the coherent light beam interacted with the raw data, which we planned to record on photographic film, I was struck by what I thought was a most astonishing observation: when the recorded data was illuminated with coherent light, the transmitted light was an optical regeneration of the original field that would be recorded by the airborne radar antenna as it was carried along the aircraft flight path. The process recreated in miniature the original microwave field, scaled down in both linear dimensions and in wavelength—1000 feet of flight path was scaled down to about 20 mm on the signal record, and the microwave wavelength was scaled down from about a centimeter to about 500 nm., the wavelength of the light illuminating the record. On this observation I formulated a completely new theory of SAR. Prior SAR theory had developed along several different lines. For example, one theory considered the required data processing to be a cross correlation. This was strictly a communication theory viewpoint, which was the natural way for radar scientists and engineers to think about such a process. The new theory was based strictly on physical optics, a theory that paralleled Gabor's original work, which at that time was not known to me. This new viewpoint was not immediately accepted. It was a viewpoint quite strange for those in radar technology. However, a few years later, when the SAR system was built and began acquiring data, the attractiveness of the new view became apparent and soon, by about 1960, it became the standard way of describing the SAR process.

What is most interesting and novel about the SAR development of holography is not the recording of both phase and amplitude; with radio waves this is commonplace and had been done for a long time, perhaps for centuries. What is novel and indeed exciting, is that the spatial structure of the wave could thus be recorded and then regenerated as an identical waveform, except for beam reduction in the spatial scaling and in the wavelength. Had SAR been invented in the age of the computer, the signals impinging on the SAR antenna would have disappeared into a computer, and after much computation would have emerged as a finished product-- a high resolution radar image, and the holographic aspect would have been entirely bypassed. Historically, the optical processing of SAR radar signals and the subsequent holographic interpretation of SAR with coherent optical processing carried SAR technology to heights that would otherwise have had to await the development of the computer.

This new theory was completed in 1956, and was presented in the form of a lengthy memo, which went into the project quarterly report. Shortly afterward I came across an article by Kirkpatrick and El-Sum in the October 1956 issue of the Journal of the Optical Society of America,[2] which described the wavefront reconstruction method of Gabor. So, Gabor had been first, 8 years earlier. My feelings were mixed: some disappointment at not being first, but pleased that the idea was in fact worthy of journal publication. After reading the literature on wavefront reconstruction, i.e., holography, I pondered the possibility of carrying out this new theory of holography with visible light instead of with microwaves. The theory had some improvements over earlier holography; in particular, the use of a spatial carrier to separate the undesired terms, the twin image and what might be called the intermodulation product term, or self interference term, from the desired one. My first impression was that such a separation could not be done with optical holography. What was readily done electronically was impossible to do optically, because electronically we can record the instantaneous signal, whereas with optics only the time averaged signal is recorded. I was ambivalent on this issue for a few years, then finally decided to undertake a serious analysis and optical experimentation. With my colleague Juris Upatnieks we set up optical systems to carry out what is appropriately called carrier frequency wavefront reconstruction, now called off axis holography. The idea in fact worked, and now, for the first time, an optical hologram was produced where, in the reconstruction process, the various emerging waves separated from each other as they propagated way from the hologram, and the desired wave was now totally separated from the various other waves.

2. OTHER PATHS TO HOLOGRAPHY

Now let's consider yet other ways to invent holography—some paths not yet traversed. A hypothetical inventor, Dr. A., looks at the bull's eye pattern produced in an interferometer by the interference of two beams of different curvature, originating from two mutually coherent point sources, displaced in the depth dimension both.. He recognizes that the pattern, when photographically recorded, would produce a transparency remarkably like a Fresnel zone plate. It would, if illuminated with a beam of coherent light, bring a portion of the light to a focus, and another part of the light to a different focus, exactly as would a conventional text book type zone plate. A trivial observation, to be sure. He then asks, what if one of the two beams originated from two point sources instead of just one, but displaced, either laterally or in depth. The photographic record would then consist of two zone plates, forming two image points, whose locations are related to the position of the two point sources. A small leap then suggests three mutually coherent radiating points in one beam, then still more. Now holography has been invented. This would be a rather uninspired way to invent holography. Perhaps someone has traveled this path, without publishing the result.

As a variant on this, we suppose that a two-beam interference pattern has been recorded, in which at least one of the beams contains some noise, possibly some point scatterers. These produce a collection of bull's eye patterns on the interferogram. Now if the interferogram were illuminated with a coherent light source, each of the bull's eye patterns would intercept part of the passing light and bring it to a point focus—the holographic reconstruction process. Holography has again been invented. Lohmann has investigated such holograms of the remote past, in the form of published photographic records of interference patterns, and has carried out the reconstruction process on photographic transparencies of the published interferograms.

Or consider Dr. B. who photographs interference patterns comprising uniform fringes. He recognizes that these are in fact diffraction gratings. Indeed Burch had proposed and demonstrated this as a way of producing commercial grade diffraction gratings, back in the 1940's. Of course, a two beam uniform fringe pattern recording would not qualify as a hologram, unless of course, the fringe pattern were modulated, thus giving a pattern more complex than just uniformly spaced fringes. Next, we suppose the optical system producing the pattern to have aberrations, resulting in an imperfect grating. When the flawed grating is illuminated with a beam of monochromatic light, the diffracted orders come to an irregular focus. The fringe-pattern grating thus has information about the optical path traversed by the beams, and information recorded on the pattern has been read out. Since the wave diffracted by the grating is a replica of the recorded wave, the wavefronts produced by the grating can be said to be a reconstruction of an original object field. We are now within a small epsilon of having holography. The nature of the aberration process can be studied from this flawed grating. Indeed, any off-axis hologram can be alternatively viewed as a flawed diffraction grating.

Continuing on a different tack, we consider the recording of an electrical signal. It could be a line trace on a CRT, or the sound track on photographic film, tape or a CD. There is a profound difference between recording of an electrical signal and a light signal. In the former case, the recording process records the instantaneous value of the signal, for example, it records a voltage $v = a(t) \cos [2\pi f_o t + \theta(t)]$, which inherently includes both phase and amplitude, whereas a similar optical signal cannot be recorded on an instantaneous basis, since the frequency is too high. The instantaneous signal inherently embodies both the phase and amplitude; indeed, it is difficult not to record the phase. Hence, almost all recorded electrical signals include both phase and amplitude, whereas recorded optical signals do not, except when interferometric means is used, as in holography.

This view could be developed into a holographic theory for electrical signals. It might go like this: An electric signal, s(t) is recorded on a appropriated medium, say, photographic film, a cathode ray tube, a storage tube, a phonographic disk, or a video or audio disk. The record would inevitably contain both the amplitude and phase, and it would be a wholly trivial task to regenerate the electrical signal from the record. As a slight variant the signal could originally be a sound wave, which is then changed into an electrical signal by a microphone, then recorded, then regenerated as an electrical signal, then changed back to a sound field by means of a speaker. Throughout the process, both the amplitude and phase are retained. The problem with this path to holography is that it is well worn, at least in part; what was not done was to put it formally into a holographic framework.. This path to holography would be rather mundane indeed.

These alternative paths to holography vary greatly in their ingenuity and excitement Might there be other yet undiscovered paths to holography, that are more clever, more exciting? Perhaps, but the requirement that we have placed on them; that they be non obvious and clever, automatically ensures that they will be difficult to find. I suggest another one, that I think has some cleverness to it. It has been known known, at least for over a century, that a way to put an optical signal $|s|^2$ on a spatial carrier, $|s(x,y)|^2 \rightarrow |s|^2 (l + \cos 2\pi f_o x)$ is to image it through a two beam interferometer so that each branch carries the image Here s is the amplitude distribution of the image, and its absolute value squared is the intensity image. This fact has been used by Lippmann for his color process, with the two beams counter-propagating, i.e, oriented 180^0 and entering the recording plate from opposite surfaces.. Similarly, Ives had a color process where the two beams propagated at much lesser angles, and entered the recording plate from the same side. Now, in either case if the signal s traverses only one of the two beams on its way to the recording plate, the system becomes linear in amplitude instead of intensity, and the phase of the signal, as well as its amplitude, is recorded. Stated alternatively, the recorded image becomes a hologram, or more precisely, an image plane hologram. If the light is spatially incoherent, only the image plane of a depth-extended object is recorded, but if the light has some spatial coherence, then the object is recorded in depth, as in typical holography. This suggested process is simple, just passing the signal through one beam instead of two would have yielded a wholly new and remarkable process that was not invented until fifty years later. This would have been an advance that may seem trivial in hindsight, but before the fact, it is anything but trivial. Indeed it appears that this path was never traversed.

I suggest yet another path to holography; it is one that I have already traveled, but only after traveling another path earlier. This path is based on the analog between the chirp signal and Fresnel diffraction. The process of Fresnel diffraction can be described as the correlation of an object field with a point spread function of the form $h = \exp[i\pi(x^2 + y^2)/\lambda z]$, where λ is the wavelength of the light and z is the distance from the object lane to the plane of Fresnel diffraction. If we have and an object distribution s(x,y), its Fresnel diffraction field can be described as the convolution of the object with the point spread function, or $s * h$. Similarly, the chirp waveform is a time domain pulse whose frequency changes over the pulse duration, or $f(t) = \mathrm{rect}(t/T) \exp [i(2\pi f_o t + \alpha t^2)]$, where f_o is the mean pulse frequency. The chirp pulse is thus the same as the point spread function of the Fresnel diffraction process, except that its variation is a time function, not a spatial function. This similarity played a major role in the development of coherent optical processing of radar data back in the 1950's and has more recently played an important role in ultrafast optics..

In a chirp radar system, the transmitter radiates a chirp pulse instead of the conventional pulse with constant frequency across its width. The chirp pulse is radiated and the signal reflected back to the receiver is thus given as the convolution of the object field s with the chirp pulse, or $s * h$. The chirp radar thus presents the information about the object field not as an ordinary image, but as a signal having the form of the Fresnel diffraction pattern of the object field. In the 1950's, one way of carrying out the required pulse compression process was to record the radar return on a medium that could be optically read out, such as photographic film or an ultrasonic cell. In the recording process, the time signal was converted into a spatial signal, and the signal was then processed by treating it as a Fresnel hologram. The chirp pulse return from a point object then becomes a Fresnel zone plate, and the pulse compression process is carried out by shining a coherent light beam through the signal record and viewing the plane where the zone plate focuses the diffracted light. The recorded data is now exactly a Fresnel hologram, and the reconstruction process is identical to the hologram reconstruction process. This is a path to holography that was actually traversed, in the 1957-1958 period. The SAR process was combined with the chirp process by radiating a chirp pulse and recording the uncompressed pulses on the SAR data record. In this manner, the two ordinarily separate processes of SAR and pulse compression become a single, two-dimensional operation. This SAR, chirp radar system with optical processing was a highly successful system, and for several years became the basic SAR imaging radar system.

On a final note here, I point out that our 1962 paper that introduced the off-axis technique of holography analyzed the holographic process in the analytical framework of a chirp radar system, a point that apparently no one has ever noticed.

3. THE MANY ARMS OF HOLOGRAPHY

Holography as I see it has two distinctly different aspects. First, there is the core of holography, which encompasses display holography, holography for security purposes and other forms of holography where the end product is a hologram intended for viewing, and which displays an image that has observable 3D. It is the holography that is reported in the Holography Newsletter of Kontnik and Lancaster. It is the business of holography, the stuff of high resolution recording media, art holograms, technically dazzling holograms, holograms displayed in museums and art galleries, of shims and hot stamping. It is a fairly close-knit business which is reported by the Holography Newsletter to have a world-wide value of the order of a billion dollars.

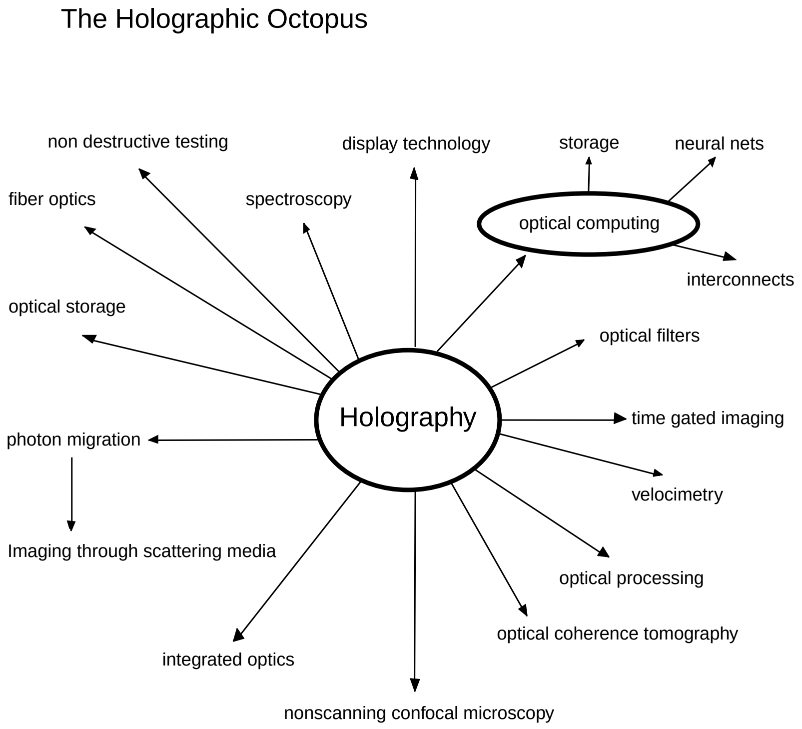

Fig. 2. The tentacled holographic entity.

There is another holography, only loosely tied to the other, one that is diffuse, that reaches out into many diverse fields, making interesting, novel and often important contributions. I have depicted the holographic entity as a sort of octopus, as shown in Fig. 2. Its many tentacles reach out into a huge realm of other optical technologies, as well as into many other branches of science, technology and engineering. Since it has many more than 8 tentacles, my depiction should perhaps be called a multopus (multipus?)

The many diverse areas into which holography reaches are vast. Of course, there are the areas closely related to basic physical optics, such as optical data storage, optical computing, optical switching, fiber optics telecommunications and imaging through scattering media, and optical sectioning (of which confocal imaging is the best known example). There are more exotic areas, such as the holographic brain, i.e., the view that the brain operates on the principle of holography. John Caulfield, has dealt with this area in a rather intriguing (or devastating) way .[3] There are other areas of optics that are important in their own right, but which may be regarded as offshoots of holography or at least have been closely associated with holography in their formative years. Wavefront correction by phase conjugation and diffractive optical elements are good examples.

All of the earliest published papers in the use of phase conjugation for correcting the wavefront defects produced by propagation through an irregular medium described holographic methods for producing the phase conjugation. Now the area of phase conjugation has grown and other methods for producing the conjugation, both quasi holographic and non holographic, have evolved. Similarly with diffractive optics,. In its infancy the area was called holographic optical elements (HOES). Then it was realized that all that diffracts is not necessarily holography. So the area came to be called diffractive optical elements, or DOES, and holography remains today a substantial part of the area.

But holography has entered into areas remote from basic holography, areas that those in holography and the related area of information optics, would likely never notice, as for example, gravitational holography and the cosmological constant.[4] . To survey this vast area of holographic infusion would be time consuming, but quite revealing, and, I think, worthwhile. It would by itself make a worthy paper.

4. OPTICAL SECTIONING AND THE REDUNDANCY OF INCOHERENT LIGHT

Microscopists, particularly those in biology, have considerable interest in the process of optical sectioning. Optical sectioning is a process in which a single plane within a thick specimen can be imaged without the image being degraded by light scattered from object distributions lying within the specimen but outside the plane of interest. The result is an image that is crisper and clearer than if the sectioning methods were not applied.

Confocal imaging, as with a confocal microscope, is the conventional way of accomplishing this sectioning. This is a double scanning process, in which a point light source is focused onto the plane of interest, and this spot is then focused onto a detector. The point scans across all points in the plane, until the entire plane has been scanned and imaged. The imaging process is characterized by a point spread function (PSF) $h = h_1h_2$, where h_1 is the PSF of the system that images the point onto the plane of interest, and h_2 is the PSF of the system that images the illuminated object plane spot of light onto a point detector. Although the resolution is improved slightly, since h_1h_2 is generally narrower than either h_1 or h_2 alone, the principal value of the process is that light scattered from object points outside the plane of interest is largely suppressed.

There are other ways of achieving optical sectioning, based on coherence. In particular, we described a technique of image plane holography with light that is either spatially or temporally incoherent, or both. We have described this in various SPIE and OSA publications. The suppression of light scattered from scattering centers outside the image plane is rather dramatic Here we further report on this topic.

First, there is the issue of coherent vs incoherent light. Coherent light has great versatility. It can perform some important functions that incoherent light normally does not do, such as complex spatial filtering and holography. The spatial filtering includes phase contrast imaging, a function of crucial importance to microscopists. On the other hand, incoherent light can do some things extremely well, for example, imaging.

Images formed with incoherent light typically have signal to noise ratios (SNRs) much higher than those formed with coherent light. Incoherent systems have, in fact, an enormous resistance to noise,. The basis for this resistance lies in the redundancy of incoherent imaging systems. One might think of an optical system illuminated with a broad source as having many different channels for conveying the object distribution to the image plane; in a loose way it can be said that each source element constitutes a separate channel, with noise that is uncorrelated, or at least less than completely correlated with the light from other channels. This is the root of the noise insensitivity. And what is the source of the noise? It is scattered light from scattering centers located outside the image.

Thus there is inherent an optical sectioning capability with incoherent light as opposed to coherent light. But this optical sectioning capability is incomplete, since the defocusing process smooths the field but the scattered light remains, producing a background,. The confocal process carries the optical sectioning a step further by removing the scattered light altogether. Similarly, the coherence imaging methods achieve the same result. This is achieved by image plane holography with incoherent light by rendering the out of plane scattered light incoherent with the reference beam; thus, in the reconstruction process, the scattered light is not part of the reconstruction process and does not appear in the holographic image, even as an ambient background.

Following the considerable success of coherent optical processing in the 1950’s and 1960’s, researchers in optical processing began a quest to achieve similar results with incoherent optics. If this could be done, then one would have the advantages of coherent optics along with those of incoherent optics, viz., the versatility of one along with the high SNR of the other. Thus, research from the field of optical processing has implications for optical sectioning, since the problems are basically similar: improvement of the SNR. The noise in the two cases arises from the same cause, which is scattered light from outside the image plane.

Some papers from the incoherent optical processing era have direct relation to the current interest in optical sectioning; they in fact are papers on optical sectioning. Papers by Leith and Swanson[5] and by Leith and Yang[6] demonstrated the enormous SNR improvement that could be achieved with incoherent light image plane holography. Contributions from out-of-plane scattering centers were thoroughly suppressed, and the contrast of the image remained high. Further papers demonstrated variants on the original techniques.[7,8]

The basic system involves an object s in one branch of the interferometer. Analysis shows that the amplitude PSF of the system is $h = h_1h_2^*$, where h_1 is the PSF of one branch of the interferometer and h_2 that of the other branch. If $h_1 = h_2$, we have $h = |h_1|^2$, and the intensity of the image is $| |s * |h_1|^2 | ^2$. If on the other hand, the signal s is in both branches of the interferometer, the system becomes linear in intensity, giving $|s|^2 * |h_1|^2$. In the one case, the PSF $|h_1|^2$ acts on the amplitude signal s, ; in the other the PSF acts on the intensity signal $|s|^2$ The second configuration can be simplified; since the object and imaging systems in each branch are identical, they can be withdrawn from the interferometer and placed in front of it or after it, which vastly simplifies the system. No longer is a matching pair of lenses needed for the two imaging process; only a single imaging system is needed.

We consider the PSF $|h_1(z)|^2$, a function of z, the distance of an out-of-plane image distribution from the plane of interest. This function describes how rapidly the sectioning process reduces the contributions from out-of-plane object positions. We make the hypothesis that of the various ways to accomplish optical sectioning, including the original confocal process, all, to within a first order approximation, obey either the coherent process or the incoherent process described above.

ACKNOWLEDGEMENT

Part of this work (section 4), was supported by The Army Research Office under grant DAAD19-01-1-0561.

REFERENCES

1. Yu. N. Denisyuk, ”Certain features of the development of display holography in the USSR,” *International Symposium on Display Holography*,” ed. T. H. Jeong, .Proc. SPIE vol 1600**,** 376-386**,** 1992.

2. P. Kirkpatrick and H,M.A. El-Sum, “Image formation by reconstructed wavefront. I. Physical principles and methods of refinement,*” J. Opt. Soc. Am*., **46** , 825-831, 1956.

3. H. John Caulfield, “Holographic brain: a good analogy gone bad,” *Holography: A tribute to Yuri Denisyuk and Emmett Leith*, ed. H. John Caulfield, Proc SPIE vol. 4737, 124-130, 2002.

4. S. Thomas, “Holography stabilizes the vacuum energy,” *Phys. Rev. Lett.* **, 19** Aug. 2002.

5. E. Leith and G. Swanson, “Recording of phase-amplitude images,*”* *Appl. Opt*. **20**, 3081-3084, 1981.

6. E. Leith and G. C. Yang, “Interferometric spatial carrier formation with an extended source,” *Appl. Opt.*

20, 3819-3821, 1981.

7. E. Leith and G. Swanson, “Generalization of some incoherent spatial filtering techniques,” *Appl. Opt***.25**, 499 502, 1986.

8. E. Leith, D. Angell, S. Leon, L. Shentu and C. P. Kuei, “Optical processing with incoherent light,“ *Conference on holography and non destructive testing, Quebec, June, 1986*, Proc. SPIE vol. 661, 1986.