Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH

J. D. Maynard, E. G. Williams,[a] and Y. Lee
The Pennsylvania State University, Department of Physics, University Park, Pennsylvania 16802

(Received 18 December 1984; accepted for publication 7 June 1985)

Because its underlying principles are so fundamental, holography has been studied and applied in many areas of science. Recently, a technique has been developed which takes the maximum advantage of the fundamental principles and extracts much more information from a hologram than is customarily associated with such a measurement. In this paper the fundamental principles of holography are reviewed, and a sound radiation measurement system, called nearfield acoustic holography (NAH), which fully exploits the fundamental principles, is described.

PACS numbers: 43.20.Ks, 43.20.Rz, 43.20.Ye

INTRODUCTION

Since the time of its conception around 1950, holography[1] has become an increasingly powerful research tool. However, in conventional optical and acoustical holography the full potential of the technique has not been realized. In acoustical holography one can obtain much more information from a hologram than is customarily associated with such a measurement. In this paper we outline the fundamental theory and experimental and signal processing requirements for what we refer to as "generalized holography" which fully exploits the potential of the technique. We also describe the practical application of generalized holography in an actual experimental measurement system called nearfield acoustic holography (NAH).

On a fundamental level, the great utility of holography arises from its high information content; that is, data recorded on a two-dimensional surface (the hologram) may be used to reconstruct an entire three-dimensional wave field, with the well-known result of obtaining three-dimensional images. A popular science magazine once noted that if a picture is worth a thousand words, then a hologram is worth $1000^{3/2}$, or approximately 32 000 words. In the case of digitally processed holograms this statement is literally correct; if a sampled two-dimensional hologram contains 1000 digital words of data, and if the reconstruction is performed for a cubical three-dimensional region, then the resulting reconstruction will contain $1000^{3/2}$ digital words of data. An actual digitally sampled hologram may contain hundreds of thousands of words of data, and the amount of reconstruction data is limited only by the restriction of computation time. In generalized holography, the reconstruction may be expanded in other ways as well. For example, in nearfield acoustic holography, the recording of the sound pressure field on a two-dimensional surface can be used to determine not only the three-dimensional sound pressure field but also the particle velocity field, the acoustic vector intensity field, the surface velocity and intensity of a vibrating source, etc. Furthermore, each data point in the hologram need not be simply phase information from single frequency radiation, but may be a complete time sequence recording from incoherent "white-light" or noise radiation; in this case one may not only reconstruct a three-dimensional field, but may also observe its evolution in time. An interesting application would be the visualization of energy flow from a transient source. Generalized holography also removes the generally assumed limitations of conventional holography, such as the resolution of a reconstructed image being limited by the wavelength of the radiation[2–6] and the limited field of view resulting from conventional recording requirements.

I. EXAMPLE OF APPLICATION

It would be useful to precede discussion of the theory and implementation of generalized holography with a brief description of a fundamental research area in acoustics and the utility of holographic techniques when applied to it.[7] This fundamental research area concerns the radiation of sound into a fluid medium (air or water) by a complex vibrator. A basic goal of research in this area is to correlate the properties of the vibrator (such as structural features, vibration modes, existence of damping material, etc.) with properties of the radiated sound field (such as the total radiated power, the farfield directivity pattern, the vector intensity field, etc.). Though this is a basic field of study, it turns out to be a difficult problem even in the simplest cases, and nearly intractable with more complex sources. Consider the sound field of a plane rectangular vibrator. When the plate is surrounded by an infinite rigid baffle, calculation of the generated sound field is quite easy. However, the same plate when unbaffled produces a sound field which is impossible to determine analytically, and can be approximated only with laborious calculations on a large computer. Naturally, if the vibrator is made more complex with the addition of ribs, etc., the ensuing calculations become all the more ineluctable, thus providing little insight into how the vibrator couples acoustic energy into the medium.

Understanding the relationship between vibrator features and sound field properties is a difficult problem from the standpoint of acoustic experimentation as well. A simple example would be the case of a rectangular plate vibrating in

[a] Present address: Naval Research Laboratory, Washington, DC 20375.

 0001-4966/85/101395-19$00.80

a normal mode with some definite nodal pattern. The plate may have some nominal displacement amplitude producing some nominal particle velocity amplitude v_0 and pressure amplitude P_0 in the medium around the plate. If the plate is vibrating below coincidence (assuming the typical plate dimension is smaller than the acoustic wavelength in the medium),[8] only a small amount of acoustic energy will be radiated into the farfield. A similar velocity amplitude v_0 and pressure P_0 on another plate vibrating above coincidence will result in a large amount of energy being radiated into the farfield. The conclusion to be drawn from this example is that measurement of only vibrator displacements, particle velocity, or sound pressure around a vibrator is insufficient to determine how energy is delivered into the sound field by the vibrator. At high frequencies, well above coincidence, the problem is simpler as areas of the vibrator with large displacement amplitudes are usually the primary energy sources. Consider, however, sound sources that radiate wavelengths that are larger than the typical dimensions of the vibrator's features, as in the case of rotating machinery, musical instruments, etc. Here, we find many instances where areas of large displacements or large pressure amplitudes are not significant energy sources and may even be large sinks of acoustic energy.

The key to understanding how a vibrator radiates sound is the determination of the acoustic intensity vector field $\mathbf{S}(\mathbf{r})$. For radiation of a single frequency, intensity is defined as the product of the in-phase components of the pressure amplitude and the velocity amplitude[9]: $\mathbf{S}(\mathbf{r}) = \frac{1}{2}P_0(\mathbf{r})\mathbf{v}_0(\mathbf{r})\cos\theta$. This quantity specifies at each point in space the rate and direction of acoustic energy flow. When the component of this field normal to the surface of a vibrator is large, this indicates the location of a large energy source. It is this tremendously vital property of the intensity field that has generated so much interest in its measurement, and methods such as the "two-microphone" technique[10] have been developed. Serious limitations in such techniques stem from the fact that they measure only one component of the vector intensity at a single point in space, or in an average over some region in space. This limitation may lead one to mistakenly identify an area as a radiating source when it may in fact be a part of a circulating energy flow pattern.[7] These patterns, which occur frequently, represent real energy flow (as opposed to a reactive out-of-phase kind of energy) in which energy leaves an area of a vibrator, only to quickly (within a fraction of a wavelength) turn around and flow back into another part of the vibrator, and return through the vibrator back to the "source" area. From this it is clear that in order to measure with confidence the sound energy radiation from a complex vibrator, a detailed map of the energy flow field must be obtained. Mapping the three components of the vector intensity field at tens-of-thousands of points in space with a point-by-point probe is impractical. With nearfield acoustic holography, however, such information is readily obtained. The basic features of the actual NAH system, which will be described in more detail later, are as follows:

(1) Only a single, noncontact measurement over a two-dimensional surface is required. This is achieved using a planar grid of regularly spaced microphones.

(2) Comprehensive results are produced in minutes. Such rapid turnaround time results in more time spent analyzing the vibrator/radiation relationship rather than making tedious measurements.

(3) The measurement covers a large area (10 m^2) and subtends a large solid angle from the sources. Multidirectional sources can be examined without missing information.

(4) The measurements have high spatial resolution: Our prototype system can pinpoint energy sources to within ~ 5 cm.

(5) Minutes after measurements, computer graphic displays of the following information are produced:

(a) The sound pressure field, from source to farfield.
(b) The particle velocity field, from source to farfield.
(c) Modal structure of a vibrating surface as determined from the normal particle velocity evaluated at the surface.
(d) The vector intensity field, which maps the energy flow throughout the sound field and pinpoints energy producing sources.
(e) The farfield radiation pattern.
(f) The total radiated power.

In addition to graphic stills and hard copy output, the fields (a)–(e) listed above may be displayed evolving in real time through the use of motion-picture computer graphics. Such graphic visual information permits a researcher to quickly discern salient features (effects of fluid loading, presence of traveling waves, for example) of the otherwise obscure interaction between a complex vibrating structure and the acoustic medium. Just how generalized holography generates so much information efficiently is discussed in the following section.

II. CONVENTIONAL AND GENERALIZED HOLOGRAPHY

Holography in its basic form is really quite straightforward. Measurements of a wave field are made on a two-dimensional surface, and then used to calculate the complete wave field in a three-dimensional space. The sources of the wave field may be scattering (or diffracting) objects or active sources. Measurements are usually made on a planar surface (the hologram plane), and these data are used to reconstruct the three-dimensional field. What makes this explosion of information possible is the fact that a known Green's function (as will be discussed subsequently) can be used, and the fact that the field being measured obeys the wave equation. Holography is unique in acoustic measurement techniques precisely because it takes the maximum advantage of this simple equation.

The paragraph above provides a basic definition of generalized holography. Conventional holography suffers from significant restrictions and limitations. In the majority of applications of conventional holography:

(1) The hologram is recorded with single frequency radiation. No broadband or noise sources are used.

(2) The hologram is recorded with a reference wave and primarily phase information only is retained with a "square-law" detector.

(3) The wavelength of the radiation limits the spatial resolution of the reconstruction.[2-6] This means, for example, that two point sources cannot be resolved if they are separated by less than a wavelength. Naturally, in optics this limitation does not pose any difficulties since the optical wavelengths are so small. However, in acoustics, where a large class of long wavelength radiators (vibrating machinery, musical instruments, etc.) is considerably smaller than the radiated wavelength, this limitation prevents the location of features which might be crucial to understanding the energy radiation.

(4) A hologram which records a specific scalar field can only be used to reconstruct that same field. Thus, in conventional acoustical holography, a measurement of the sound pressure field cannot be used to reconstruct an independent particle velocity field or the vector intensity field, and one is unable to map the source or flow of acoustic intensity. Conventional holography cannot offer the dramatic advantages described in the preceding section.

(5) A conventional hologram must be recorded many wavelengths from the source (i.e., in the Fresnel or Fraunhofer zone).[11] Thus, due to the practical limitation in hologram size, the hologram may subtend a small solid angle from the source. A directional source may not be properly recorded because of this, and important information might be missing.

In the past, as techniques of optical holography were used to form the basis of acoustical holography, these limitations were carried over into acoustical holography. By reexamining the fundamental theory of holography the limitations traditionally associated with long-wavelength acoustic holography can be removed.

III. FUNDAMENTALS OF GENERALIZED HOLOGRAPHY

A. General

As discussed in the preceding sections, generalized holography involves the measurement of a wave field on an appropriate surface and the use of this measurement to uniquely determine the wave field within a three-dimensional region. This description indicates that generalized holography is equivalent to the use of a Dirichlet boundary condition[12] on a surface for which the Green's function is known. One usually imagines boundary value problems as having boundary conditions determined by a source (for example, a vibrating surface in an acoustics problem); such problems are difficult because the source may provide conditions for which there is no known Green's function. In generalized holography, one simply measures a uniform (Dirichlet or Neuman) boundary condition on a surface for which there is a known Green's function. The holographic reconstruction process is then simply the convolution (or deconvolution) of the measured boundary values with the Green's function. In theory this is a straightforward process; in practice some care must be taken in order to identify and avoid the limitations of conventional holography. The causes of the limitations occur in the method of measuring the boundary data, in the formulation of the Green's function, and in the evaluation of the convolution integral. These areas will be discussed in subsequent sections. In later sections the calculation of quantities other than the measured wave field will be discussed. We begin with a description of the formal assumptions required for generalized holography.

The basic assumption is that some sources are creating a wave field $\psi(\mathbf{r},t)$ (a function of position $\mathbf{r}$ in a three-dimensional region of space and time t) which, within a three-dimensional region of interest, satisfies the homogeneous wave equation:

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = 0 . \tag{1}$$

Here, ∇^2 is the Laplacian operator and c is a constant propagation speed. The following is further assumed:

(1) There is a surface S enclosing the three-dimensional region of interest for which there is a known Green's function $G(\mathbf{r}|\mathbf{r}_S)$ satisfying the homogeneous Helmholtz equation for $\mathbf{r}$ inside S and vanishing (or having a vanishing normal derivative) for $\mathbf{r} = \mathbf{r}_S$ on S. Part of S may be at infinity; in practice the part of S not at infinity will be a level surface of some separable coordinate system which is in close contact with the sources.

(2) There is a surface H (the hologram surface) which may coincide with S or have a level surface parallel to S for which $\psi(\mathbf{r}_H,t)$ (or its normal derivative) can be measured or assumed for all $\mathbf{r}_H$ on H and all t.

If the above conditions are met, then $\psi(\mathbf{r},t)$ for $\mathbf{r}$ inside S can be uniquely determined from $\psi(\mathbf{r}_H,t)$ with $\mathbf{r}_H$ on H. The exact procedure and a discussion of the consequences resulting from deviations from the assumptions are presented in the following subsections.

B. Time dependence

The first step in finding $\psi(\mathbf{r},t)$ from $\psi(\mathbf{r}_H,t)$ is to Fourier transform in time:

$$\tilde{\psi}(\mathbf{r},\omega) = \int_{-\infty}^{\infty} \psi(\mathbf{r},t)e^{i\omega t}\,dt \tag{2}$$

and

$$\tilde{\psi}(\mathbf{r}_H,\omega) = \int_{-\infty}^{\infty} \psi(\mathbf{r}_H,t)e^{i\omega t}\,dt . \tag{3}$$

The symbol $\tilde{}$ indicates a complex field having an amplitude and phase depending on $\mathbf{r}$. The wave equation becomes the Helmholtz equation

$$\nabla^2\tilde{\psi}(\mathbf{r},\omega) + k^2\tilde{\psi}(\mathbf{r},\omega) = 0 , \tag{4}$$

with wavenumber $k = \omega/c$. It should be noted that formally the boundary data $\tilde{\psi}(\mathbf{r}_S,t)$ must be measured for all time $-\infty \leqslant t \leqslant \infty$. Here, $\psi(\mathbf{r}_S,t)$ may be measured within a finite time window of duration T if $\psi(\mathbf{r}_S,t)$ is known to be periodic with period T. For a noise source, one may assume that there exists a time scale T for which statistical averages become stationary within specified limits of fluctuation[13]; in this case also a finite Fourier transform is sufficient. For most noise sources a reasonable T can be used; however, there are exceptions where T may be so large as to preclude the acquisition of a manageable amount of data. An example would be a high-frequency transient in a highly reverberant room. In digital holography, $\psi(\mathbf{r}_H,t)$ is sampled at N discrete points in

time $t_n = t_{r_H} + nT/N$ (noting that the starting time t_{r_H} may be different for different positions on the surface H). It is assumed that the sampling is accomplished at the Nyquist rate to prevent aliasing in the time domain. Expression (3) becomes

$$\tilde{\psi}(\mathbf{r}_H,\omega_m) \approx e^{i\omega_m t_{r_H}} \left(\sum_{n=0}^{N-1} \psi(\mathbf{r}_H,t_n) e^{i2\pi nm/N} \right) \frac{T}{N}, \tag{5}$$

where $\omega_m = 2\pi m/T$ and m is a non-negative integer less than $N/2$. The summation in large parentheses can now be accomplished with a fast Fourier transform (FFT) computer algorithm. The errors associated with the approximation (5) are not unique to holography but are common to all signal processing involving discrete, finite window sampling. Since discussions of these errors can be found in any text on signal processing,[14] we shall not concern ourselves with them here; the more interesting aspects of generalized holography are found in the spatial, rather than the temporal, signal processing. For the purpose of the spatial analysis in the next sections, it can be assumed that the sources are driven at frequencies $\omega_m = 2\pi m/T$ with $m(<N/2)$ some integer and T and N fixed. For most sources it can be assumed that the wave field generated by these harmonic sources does not differ significantly from the actual wave field. If the actual operating frequency of the source is known, then signal processing techniques can be used to correct $\tilde{\psi}(\mathbf{r}_H,\omega_m)$. At any rate, for sources operating at the set of frequencies ω_m, expression (5) becomes exact.

For the spatial analysis we consider a fixed value of ω so that there is a fixed wavenumber $k = \omega/c$ and a single characteristic wavelength $\lambda = 2\pi c/\omega$. The spatial problem is now to find the complex field $\tilde{\psi}(\mathbf{r})$ satisfying the homogeneous Helmholtz equation

$$\nabla^2\tilde{\psi}(\mathbf{r}) + k^2\tilde{\psi}(\mathbf{r}) = 0, \tag{6}$$

for $\mathbf{r}$ within the three-dimensional region of interest, given $\tilde{\psi}(\mathbf{r}_H)$ for $\mathbf{r}_H$ on the hologram surface H.

At this point the source of one of the limitations of optical holography can be discussed. In order to carry out the spatial processing it is necessary to use the complex field $\tilde{\psi}(\mathbf{r}_H)$, amplitude as well as phase, for each temporal frequency. In theory $\tilde{\psi}(\mathbf{r}_H)$ can be found from $\psi(\mathbf{r}_H,t)$; however, in optics there is no detector fast enough to record the real-time development of the wave field. Instead the recorded wave field must contain only a single temporal frequency, the source wave field must be mixed with a reference wave, and the resultant is recorded with a square-law detector.[11] The contributions to this (zero frequency) recording which come from the cross terms in the mixed wave field can be used to obtain some information about $\tilde{\psi}(\mathbf{r}_H)$; however, the amplitude and phase information have become irretrievably intermixed. In practice, optical holograms are measured many wavelengths from the source (in the Fresnel or Fraunhofer zone), where the amplitude information has become unimportant (having a simple spherical wave dependence on distance from the source) and only the phase information is significant. The phase information contained in the optical hologram cross terms can be processed as $\tilde{\psi}(\mathbf{r}_H)$; however, the lack of precise amplitude and phase information and the requirement of recording in the Fresnel or Fraunhofer zone result in the limitations of conventional optical holography as described in Sec. I. These limitations will be discussed further in a later subsection. For generalized holography it is assumed that $\tilde{\psi}(\mathbf{r}_H)$ is known.

In acoustical holography it is possible to record $\psi(\mathbf{r}_H,t)$ with conventional experimental techniques and precisely determine $\tilde{\psi}(\mathbf{r}_H)$. It is interesting to note that early implementations of acoustical holography were copies of optical systems in that reference waves were used and square-law recordings were made in the farfield of the source.[15]

C. Spatial processing

Since it is assumed that the Green's function $G(\mathbf{r}|\mathbf{r}_S)$ satisfying the homogeneous Dirichlet condition on the surface S is known, then the solution $\tilde{\psi}(\mathbf{r})$ for Eq. (6) can be found with a surface integration[12]:

$$\tilde{\psi}(\mathbf{r}) = -\frac{1}{4\pi} \iint \tilde{\psi}(\mathbf{r}_S) \frac{\partial G}{\partial n}(\mathbf{r}|\mathbf{r}_S) d^2 r_S, \tag{7}$$

where $\partial G/\partial n$ is the normal derivative of G with respect to $\mathbf{r}_S$. If the surface S is the same as the surface H, where $\tilde{\psi}(\mathbf{r}_H)$ is measured or assumed, then the determination of $\tilde{\psi}(\mathbf{r})$ is complete. If H lies inside S, then processing proceeds as follows.

In practice, the Green's function G is known provided that the part of S not at infinity is the level surface of a separable coordinate system. We denote the three spatial coordinates of this system as ξ_1, ξ_2, and ξ_3, with the level surface given by $\xi_3 = \xi_3^S$, a constant. According to the assumptions of generalized holography (in Sec. I) the hologram surface is given by $\xi_3 = \xi_3^H$, where the constant $\xi_3^H > \xi_3^S$ describes a surface inside S. In terms of ξ_1, ξ_2, and ξ_3 Eq. (7) becomes

$$\begin{aligned}\tilde{\psi}(\xi_1,\xi_2,\xi_3) &= -\frac{1}{4\pi} \iint \tilde{\psi}(\xi_1',\xi_2',\xi_3^S) \\ &\times \left(\frac{\partial G}{\partial \eta}(\xi_1 - \xi_1', \xi_2 - \xi_2', \eta) \Big|_{\eta = \xi_3 - \xi_3^S} \right) d\xi_1' \, d\xi_2',\end{aligned} \tag{8}$$

but this cannot be evaluated directly because $\tilde{\psi}(\xi_1,\xi_2,\xi_3^H)$ is known instead of $\tilde{\psi}(\xi_1,\xi_2,\xi_3^S)$. If expression (8) is evaluated for $\xi_3 = \xi_3^H$ we obtain

$$\tilde{\psi}(\xi_1,\xi_2,\xi_3^H) = \iint \tilde{\psi}(\xi_1',\xi_2',\xi_3^S) G_{HS}(\xi_1 - \xi_1', \xi_2 - \xi_2') \, d\xi_1 \, d\xi_2, \tag{9}$$

where

$$G_{HS}(\alpha,\beta) = -\frac{1}{4\pi} \frac{\partial G}{\partial \eta}(\alpha,\beta,\eta) \Big|_{\eta = \xi_3^H - \xi_3^S}.$$

The right-hand side of Eq. (9) is a two-dimensional convolution; by using the convolution theorem Eq. (9) can be inverted to obtain $\tilde{\psi}(\xi_1,\xi_2,\xi_3^S)$ in terms of $\tilde{\psi}(\xi_1,\xi_2,\xi_3^H)$. Denoting a two-dimensional spatial Fourier transform by $\hat{}$ and its inverse by $\mathcal{F}^{-1}$, we have from Eq. (9) and the convolution theorem

$$\hat{\tilde{\psi}}(\xi_3^H) = \hat{\tilde{\psi}}(\xi_3^S) \hat{G}_{HS}. \tag{10}$$

Solving for $\tilde{\psi}(\xi_1,\xi_2,\xi_3^S)$ yields

$$\tilde{\psi}(\xi_1,\xi_2,\xi_3^S) = \mathcal{F}^{-1}[\hat{\tilde{\psi}}(\xi_3^H)\hat{G}_{HS}^{-1}] . \tag{11}$$

Once $\tilde{\psi}(\xi_1,\xi_2,\xi_3^S)$ is found from the hologram data $\tilde{\psi}(\xi_1,\xi_2,\xi_3^H)$, then Eq. (8) is used to reconstruct $\tilde{\psi}(\xi_1,\xi_2,\xi_3)$ over the entire three-dimensional region inside S. It should be noted that the two-dimensional Fourier transforms used in Eqs. (10) and (11) may be in the form of decompositions in terms of a complete set of eigenfunctions appropriate for the coordinate system used. In fact, the Green's function is usually only known in terms of such a decomposition. This feature will become evident in subsequent subsections.

If instead of $\tilde{\psi}(\mathbf{r}_S)$ one determines its normal derivative with respect to $\mathbf{r}_S$, $\partial\psi/\partial n(\mathbf{r}_S)$, then Eq. (7) is replaced by[12]

$$\tilde{\psi}(\mathbf{r}) = +\frac{1}{4\pi}\int\int \frac{\partial\psi}{\partial n}(\mathbf{r}_S)G(\mathbf{r}|\mathbf{r}_S)\,d^2r_S , \tag{12}$$

where the Green's function G now must satisfy a homogeneous Neuman condition on S. Processing in terms of a separable coordinate system proceeds as before.

Derivatives of the field $\tilde{\psi}(\mathbf{r})$ with respect to the three spatial coordinates may be transferred to the Green's functions in Eqs. (7) and (12), so that calculations of such quantities simply involve processing with a different kernel.

It is important to note that all of the formulations discussed above [Eqs. (7)–(12)] are exact; there have been no approximations which would lead to resolution limits, etc. Equations (7) and (12) are not approximate expressions of Green's theorem, nor are they approximate solutions to the Helmholtz integral equation; they should not be confused with the approximate formulas used in diffraction problems.[2] The Green's functions in Eqs. (7) and (12) should not be confused with the free-space Green's function even though in some cases it has an identical form. Historically Eqs. (7) and (12) are referred to as the first and second Rayleigh integrals.[16]

D. Plane generalized holography

In conventional holography, holograms are usually recorded on plane surfaces, and in generalized holography the processing of plane holograms is the easiest from a computational point of view. Other hologram surfaces (cylindrical, spherical, etc.) can be used when they more closely conform to the shape of the sources. When the sources have odd shapes which do not conform to the level surface of a separable coordinate system, then plane generalized holography may be used in conjunction with a finite element technique; this will be discussed in Sec. VI. In any case the features of plane generalized holography represent all forms of generalized holography. The discussion of plane holography given below will present the basic equations underlying the actual nearfield acoustic holography computation algorithms, and will illustrate in detail the departures from conventional holography and the sources of problems in real applications of generalized holography.

For plane holography the separable coordinate system is of course the Cartesian system with rectangular coordinates (x, y, z). The surface S (described in Sec. III A) is taken to be the infinite plane defined by $z = z_S$ (a constant) and the infinite hemisphere enclosing the $z > z_S$ half-space. It is assumed that the sources lie in a finite region just below the z_S plane, and that the field which they generate obeys the Sommerfeld radiation condition[2] [i.e., $r(\partial\tilde{\psi}/\partial n - ik\tilde{\psi})$ vanishes on the hemisphere at infinity]. As an aid in understanding, it is useful to assume that the sources are planar, such as vibrating plates, etc., lying in the z_S plane; nonplanar sources and depth resolution below the z_S plane will be discussed later.

For expression (7) relating $\tilde{\psi}(x, y, z)$ to $\tilde{\psi}(x, y, z_S)$, we need the Green's function which satisfies the homogeneous Dirichlet boundary condition on z_S; this is given by[12]

$$G(x, y, z|x', y', z') = \frac{\exp[ik\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}]}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}} - \frac{\exp[ik\sqrt{(x-x')^2+(y-y')^2+(z+z'-2z_S)^2}]}{\sqrt{(x-x')^2+(y-y')^2+(z+z'-2z_s)^2}} . \tag{13}$$

The normal derivative $(\partial/\partial z')$ at $z' = z_S$ is

$$-4\pi G'(x-x', y-y', z-z_S) \equiv \frac{\partial G}{\partial n}(x, y, z|x', y', z_S) = -2\frac{\partial}{\partial\alpha} \times \left(\frac{\exp[ik\sqrt{(x-x')^2+(y-y')^2+\alpha^2}]}{\sqrt{(x-x')^2+(y-y')^2+\alpha^2}}\right)_{\alpha=(z-z_S)} , \tag{14}$$

so that Eq. (7) becomes

$$\tilde{\psi}(x, y, z) = \int\int_{-\infty}^{+\infty} \tilde{\psi}(x', y', z_S)G'(x-x', y-y', z-z_S)dx'\,dy' . \tag{15}$$

It should be noted that expression (13) is not the free-space Green's function[12] which has just one term in the form $\exp(ikR)/R$. Although expression (14) follows this form, the free-space Green's function is not used in this boundary value problem. Equation (15) is not an approximate form of Green's theorem with one of the free-space Green's function terms dropped, as is sometimes mistakenly assumed.

Usually the hologram data is not recorded on the sources ($z = z_S$) but rather on a plane $z = z_H > z_S$ above and parallel to the source plane. Evaluating Eq. (15) with $z = z_H$ yields

$$\tilde{\psi}(x, y, z_H) = \int\int_{-\infty}^{\infty} \tilde{\psi}(x', y', z_S)G'(x-x', y-y', z_H-z_S)dx'\,dy' , \tag{16}$$

where $\tilde{\psi}(x, y, z_H)$ is the hologram data (assumed to be available for all x and y in the z_H plane). Since $z_H - z_S$ is a constant, Eq. (16) is a two-dimensional convolution, and $\tilde{\psi}(x', y', z_S)$ can be found in terms of $\tilde{\psi}(x, y, z_H)$ with the convolution theorem. We denote the two-dimensional spatial Fourier transform as $\hat{\tilde{\psi}}$,

$$\hat{\tilde{\psi}}(k_x,k_y,z_H) = \int\int_{-\infty}^{\infty} \tilde{\psi}(x,y,z_H)e^{-i(k_x x + k_y y)}\,dx\,dy \quad (17)$$

and the inverse transform as $\mathcal{F}^{-1}$. With the convolution theorem we can rewrite Eq. (15) as

$$\tilde{\psi}(x,y,z) = \mathcal{F}^{-1}[\hat{\tilde{\psi}}(k_x,k_y,z_S)\hat{G}'(k_x,k_y,z-z_S)] \quad (18)$$

and Eq. (16) can be written as

$$\hat{\tilde{\psi}}(k_x,k_y,z_H) = \hat{\tilde{\psi}}(k_x,k_y,z_S)\hat{G}'(k_x,k_y,z_H-z_S)\,. \quad (19)$$

Solving Eq. (19) for $\hat{\tilde{\psi}}(k_x,k_y,z_S)$ and substituting in Eq. (18) yields

$$\tilde{\psi}(x,y,z) = \mathcal{F}^{-1}\left[\hat{\tilde{\psi}}(k_x,k_y,z_H)\left(\frac{\hat{G}'(k_x,k_y,z-z_S)}{\hat{G}'(k_x,k_y,z_H-z_S)}\right)\right]. \quad (20)$$

Equation (20) is the expression which gives the holographic reconstruction of the three-dimensional field $\tilde{\psi}(x,y,z)$ in terms of the (Fourier transformed) hologram data $\tilde{\psi}(x,y,z_H)$.

From Eq. (14) the two-dimensional spatial Fourier transform $\hat{G}'$ can be found explicitly:

$$\hat{G}(k_x,k_y,z) = \begin{cases} e^{iz(k^2-k_x^2-k_y^2)^{1/2}}, & k_x^2+k_y^2 \le k^2, \\ e^{-z(k_x^2+k_y^2-k^2)^{1/2}}, & k_x^2+k_y^2 > k^2. \end{cases} \quad (21)$$

The interpretation of $\hat{G}'(k_x k_y, z-z_S)$ and its role in Eq. (18) is as follows.

The source plane at $z=z_S$ is considered as a superposition of surface waves $\exp(ik_x x + ik_y y)$ with amplitudes $\hat{\tilde{\psi}}(k_x,k_y,z_S)$. Since there are no restrictions on the nature of the sources, then $\hat{\tilde{\psi}}(k_x,k_y,z_S)$ can have nonzero values for any point in the two-dimensional k space (k_x,k_y). In fact, if the sources are of finite extent in the z_S plane, then $\hat{\tilde{\psi}}(k_x,k_y,z_S)$ must be nonzero for arbitrarily large values of k_x and k_y.[17] One must then consider both forms of $\hat{G}'(k_x,k_y,z-z_S)$ in Eq. (21) and their role in Eq. (18). When $k_x^2+k_y^2<k^2$, then the surface waves in the z_S plane simply couple to ordinary propagating plane waves in the three-dimensional region $z>z_S$. These plane waves have amplitudes $\hat{\tilde{\psi}}(k_x,k_y,z_S)$, which travel in the direction given by the wave vector $(k_x,k_y,\sqrt{k^2-k_x^2-k_y^2})$, and have wave vector magnitude k so as to satisfy the original Helmholtz equation [see Eq. (6)]. The kernel or "propagator" in Eq. (18),

$$\hat{G}'(k_x,k_y,z-z_S) = \exp\left[i(z-z_S)\sqrt{k^2-k_x^2-k_y^2}\right],$$

simply provides the plane-wave phase change in going from the z_S plane to the z plane. The propagating plane wave emerges from the z_S plane at just such an angle so as to exactly match the surface wave in the z_S plane.

When $k_x^2+k_y^2>k^2$, then there is no way that one can add a real z component to (k_x,k_y) and form a three-dimensional plane wave with wave vector magnitude k. If $k_x^2+k_y^2>k^2$ then the length of the surface wave is shorter than $\lambda=2\pi/k$; having a three-dimensional plane wave (of wavelength λ) emerging from the z_S plane at some angle can only match surface waves which have two-dimensional wavelengths greater than or equal to λ. Surface waves with $k_x^2+k_y^2>k^2$ must be matched with evanescent waves[18] which have imaginary z components in their wave vector, and which exponentially decay in the z direction as $\exp\left[-(z-z_S)\sqrt{k_x^2+k_y^2-k^2}\right]$. This is correctly represented in Eq. (18) with the form of $\hat{G}'$ in Eq. (21). The boundary in k space which separates the propagating plane-wave region from the evanescent wave region is the "radiation circle," defined by $k_x^2+k_y^2=k^2$.

The situation described above is illustrated in Fig. 1. In this figure the "FT" dashed lines represent the two-dimensional forward Fourier transform going from an (x,y) plane in real space to the (k_x,k_y) plane in k space, the "IFT" dashed lines represent the inverse Fourier transform, and $k_z \equiv \sqrt{|k^2-k_x^2-k_y^2|}$. Features of the source in the z_S plane which vary in space more slowly than λ get mapped by the FT to points in k space lying inside the radiation circle; features of the source which vary in space more rapidly than λ get mapped to points in k space lying outside the radiation circle. The wave field in a plane a distance z above the z_S plane is determined in k space by multiplying the amplitudes

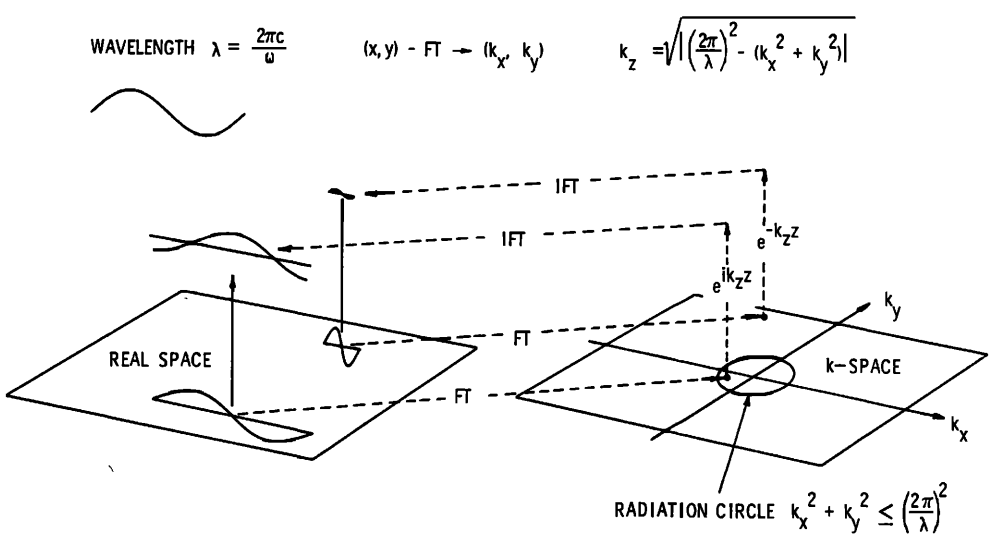

FIG. 1. Schematic representation of the holographic reconstruction process, illustrating propagating and evanescent waves.

$\hat{\tilde{\psi}}(k_x, y_y, z_S)$ inside the radiation circle by $\exp(ik_z z)$ (thus surface waves varying more slowly than λ simply undergo a phase change in moving to a plane away from the sources), and by multiplying the amplitudes outside the radiation circle by $\exp(-k_z z)$ (so that surface waves varying more rapidly than λ suffer an exponential decay in amplitude in moving to a plane away from the sources).

Having discussed the role of the propagator $\hat{G}'$ [Eq. (21)] in radiation from the z_S plane [Eq. (18)], we now consider its action in the expression for holographic reconstruction, Eq. (20). By inserting the expression for $\hat{G}'$ [Eq. (21)] into Eq. (20) we obtain

$$\tilde{\psi}(x, y, z) = \mathcal{F}^{-1}\left(\hat{\tilde{\psi}}(k_x, k_x, z_H) \begin{cases} e^{ik_z(z - z_H)}, & k_x^2 + k_y^2 \leqslant k^2 \\ e^{-k_z(z - z_H)}, & k_x^2 + k_y^2 > k^2 \end{cases}\right). \tag{22}$$

When $z > z_H$, then Eq. (22) is analogous to Eq. (18), and it represents the phase change of the propagating plane-wave components and the exponential decay of the evanescent wave components in going from the z_H plane outward (away from the sources) to the z plane. When $z < z_H$, then the factor $\exp[ik_z(z_H - z)]$ reverses the phase change of the propagating plane waves, and the positive exponential $\exp[+k_z(z_H - z)]$ restores the decayed evanescent wave amplitudes to their original values in the z plane.

It should be noted that z_S does not appear in Eq. (22), nor will it occur explicitly in any final reconstruction expressions. The role of the surface S in the derivation of generalized holography is only to establish rigorously the region of validity of the final expressions. In real applications of generalized holography, the z_S surface is the one parallel to the z_H surface which just touches the physical sources or scattering objects.

If we redefine k_z to be a complex function of k_x and k_y as

$$k_z = \begin{cases} \sqrt{k^2 - k_x^2 - k_y^2}, & k_x^2 + k_y^2 \leqslant k^2, \\ i\sqrt{k_x^2 + k_y^2 - k^2}, & k_x^2 + k_y^2 > k^2, \end{cases} \tag{23}$$

then Eq. (22) becomes, with $\mathcal{F}^{-1}$ explicitly expressed,

$$\tilde{\psi}(x, y, z) = \frac{1}{(2\pi)^2} \int\int_{-\infty}^{\infty} [\hat{\tilde{\psi}}(k_x, k_y, z_H) e^{-ik_z z_H}] \times e^{i(k_x x + k_y y + k_z z)}\, dk_x\, dk_y . \tag{24}$$

Equation (24) is of the form

$$\tilde{\psi}(x, y, z) = \frac{1}{(2\pi)^2} \int\int_{-\infty}^{\infty} \tilde{A}(k_x, k_y) e^{i(k_x x + k_y y + k_z z)}\, dk_x\, dk_y , \tag{25}$$

which is the general solution of the Helmholtz equation (6) which one would obtain using the method of separation of variables in Cartesian coordinates. The two constants of separation are k_x^2 and k_y^2, the mode labels are k_x and k_y, and the product solutions (eigenfunctions) are the propagating plane waves and the evanescent waves:

$$\tilde{u}(k_x, k_y; x, y, z) = \begin{cases} e^{ik_x x} e^{ik_y y} e^{iz\sqrt{k^2 - k_x^2 - k_y^2}}, & k_x^2 + k_y^2 \leqslant k^2, \\ e^{ik_x x} e^{ik_y y} e^{-z\sqrt{k_x^2 + k_y^2 - k^2}}, & k_x^2 + k_y^2 > k^2. \end{cases} \tag{26}$$

The reconstruction expressions of generalized holography may be derived quickly from general solutions such as Eq. (25). One simply evaluates the general solution at the hologram coordinate, $z = z_H$, and then uses the orthogonality of the product solutions to uniquely solve for the coefficients $\tilde{A}(k_x k_y)$ in terms of the hologram data $\tilde{\psi}(x, y, z_H)$. The result is an expression of the form of Eq. (24). This separation of variable and eigenfunction technique will be used to derive the expressions of generalized holography for coordinate systems other than Cartesian; in non-Cartesian coordinate systems the Green's function is known only in terms of an eigenfunction expansion so that no convolution expressions analogous to Eq. (15) are available, and only expressions analogous to Eq. (25) can be used. For the derivation of the expressions of plane generalized holography [Eqs. (18)–(22)], it would have been easier to use separation of variables in Cartesian coordinates and expansions in terms of the eigenfunctions of Eq. (26); however, the use of the real-space Green's function $G'(x - x', y - y', z - z_S)$ [Eq. (14)] in the convolution expression (15) will be necessary in dealing with the problem of a finite hologram aperture in real applications of plane generalized holography.

At this point the wavelength resolution limit of conventional holography should be discussed. The "resolution" of a field refers to how rapidly the field varies in space. It may be quantitatively measured by Fourier transforming the field in some direction (for example, the x direction) and then examining the amplitudes for the different "spatial frequencies" k_x. If in any direction there are no amplitudes larger than some predefined cutoff value for spatial frequencies beyond some value $k_{\max}$, then the minimum distance over which the field varies in space, or the resolution distance, is $R \equiv \pi/k_{\max}$. In generalized holography the resolution is determined by the values of k_x and k_y for which $\tilde{\psi}(k_x, k_y, z_S)$ has a significant magnitude. As already discussed, if there are no limits on the nature of the sources, then $\tilde{\psi}(k_x, k_y, z_S)$ may have finite amplitudes for arbitrarily large values of k_x and k_y. In the reconstruction expressions of generalized holography [e.g., Eq. (24)] the integrals in k space extend over the infinite domain, so that generalized holography has no intrinsic resolution limit; as already stated the reconstruction expressions of generalized holography are exact. The actual resolution limits of practical generalized holography will be discussed in Sec. IV; however the resolution limit of conventional holography may be obtained immediately. In typical conventional holography, holograms are recorded at a distance d in the Fraunhofer or Fresnel zone of the sources (many wavelengths away from the sources, $d \gg \lambda$) so that the hologram represents the Fourier transform of the sources $\tilde{\psi}(kx/d, ky/d, z_S)$[2]. That is, the forward Fourier transform of generalized holography is performed by the field propagation itself. However, what is wrong here (and ignored in most textbooks on holography) is that this procedure does not

work for the evanescent wave components. The reasons that the evanescent waves are ignored is because they decay {by the factor $\exp[ik_z(z-z_S)]$} to an unmeasurable level in the Fraunhofer or Fresnel zone. Taking $2\pi/\lambda$ as a typical value for k_z, and taking 2λ for $(z-z_S)$, we have $\exp[-(2\pi/\lambda)(2\lambda)] = \exp(-4\pi) \approx 10^{-6}$, so that the evanescent waves may decay by six orders of magnitude within only two wavelengths from the source. On the other hand, the propagating wave components maintain their amplitudes and only change phase in traveling to the farfield (thus phase is more important in conventional holography). In conventional holography (optical and acoustical) only the propagating wave components $[\hat{\tilde{\psi}}(k_x,k_y,z_S)$ with $k_x^2+k_y^2 \leqslant k^2]$ are measured, and only these are used in the reconstruction.[2-5] With only these components the maximum spatial frequency is $k_{\max}=k=2\pi/\lambda$, and the resolution distance is $R=\lambda/k_{\max}=\lambda/2$; thus the resolution of conventional holography is limited by the wavelength of the radiation. If better resolution is to be obtained in generalized holography, then the evanescent wave components must be measured; furthermore, the reconstruction expressions (including the Fourier transforms) must be evaluated numerically, since there are no techniques in Fourier optics which can reconstruct the evanescent wave components.

E. Calculation of other quantities

1. Field gradient (particle velocity field)

Once the three-dimensional wave field $\tilde{\psi}(x,y,z)$ has been determined, other quantities such as the field gradient $\nabla\tilde{\psi}$ can be determined. In acoustics, where $\tilde{\psi}$ is the sound pressure field, the particle velocity field can be calculated from

$$\tilde{\mathbf{V}}(\mathbf{r}) = \nabla\tilde{\psi}(\mathbf{r})/i\mu ck\,, \tag{27}$$

where μ is the fluid mass density. By taking the gradient operator inside the integral in Eq. (24), the expressions for the three particle velocity components V_η, $\eta = x,y,z$, become

$$\tilde{V}_\eta(x,y,z) = \frac{1}{4\pi^2\mu c}\int\int_{-\infty}^{\infty} \hat{\tilde{\psi}}(k_x,k_y,z_H) \times\left[(k_\eta/k)e^{ik_z(z-z_H)}\right] e^{i(k_x x+k_y y)}\,dk_x\,dk_y\,. \tag{28}$$

It is important to remember that in expressions such as (24) and (28), k_z is a complex function of k_x and k_y.

At this point it is worth considering solving Eq. (28) for $\hat{\tilde{\psi}}(k_x,k_y,z_H)$ in terms of $\hat{\tilde{V}}_z(x,y,z_S)$ and using this in Eq. (24). The result is

$$\tilde{\psi}(x,y,z) = \frac{\mu c}{4\pi^2}\int\int_{-\infty}^{\infty} \hat{\tilde{V}}_z(k_x,k_y,z_S) \times\left[(\sqrt{k^2-k_x^2-k_y^2})^{-1}\,e^{ik_z(z-z_S)}\right] \times e^{i(k_x x+k_y y)}\,dk_x\,dk_y\,, \tag{29}$$

which is the same result which would be obtained if the original problem had been specified with Neuman instead of Dirichlet boundary conditions. [This is the expression which would be used to predict the radiation from a planar vibrator; the surface velocity $V_z(x,y,z_S)$ might be determined from a structural analysis program.] The important thing to notice here is the appearance of k_z [written out as $\sqrt{k^2-k_x^2-k_y^2}$ in Eq. (29)] in the denominator of the kernel (the term in brackets); on the radiation circle $(k_x^2+k_y^2=k^2)$ the kernel is singular. This singular behavior must be kept in mind when one attempts to evaluate Eq. (29) using conventional computer techniques; this will be discussed further in Sec. IV.

As already mentioned, finite-aperture effects may be more readily handled if one uses real-space convolution expressions rather than the Fourier transform expressions such as Eq. (29). If the convolution theorem is applied to Eq. (29) and the kernel is transformed analytically, then one obtains

$$\tilde{\psi}(x,y,z) = \frac{1}{4\pi}\int\int_{-\infty}^{\infty} i\mu ck\,\tilde{V}_z(x',y',z_S) \times\left(2\,\frac{\exp\left[ik\sqrt{(x-x')^2+(y-y')^2+(z-z_S)^2}\right]}{\sqrt{(x-x')^2+(y-y')^2+(z-z_S)^2}}\right) \times dx'\,dy'\,. \tag{30}$$

Equation (30) is the real-space convolution expression for the solution to the Neuman boundary value problem, i.e., Eq. (12) in Cartesian coordinates; the term in large parentheses is the Green's function evaluated at z_S, and $i\mu ck\,v_z(x',y',z_S) = \partial\tilde{\psi}/\partial z$ on z_S.

2. Farfield directivity pattern

A farfield directivity pattern can be determined if the Cartesian coordinates are written in terms of spherical coordinates r,θ,ϕ, defined by

$$x = r\sin\theta\cos\phi\,, \tag{31a}$$

$$y = r\sin\theta\sin\phi\,, \tag{31b}$$

$$z = z_S = r\cos\theta\,. \tag{31c}$$

A complex directivity function $\tilde{D}(\theta,\phi)$ may be defined by

$$\tilde{\psi}(r\sin\theta\cos\phi\,,\, r\sin\theta\sin\phi\,,\, r\cos\theta) \underset{r\to\infty}{\to} \tilde{D}(\theta,\phi)\exp(ikr)/r\,. \tag{32}$$

If expression (30) is used for $\tilde{\psi}$ with the large r approximation

$$\sqrt{(x-x')^2+(y-y')^2+(z-z_S)^2} \approx r - x'\sin\theta\cos\phi - y'\sin\theta\sin\phi\,, \tag{33}$$

then one obtains

$$\tilde{D}(\theta,\phi) = i\mu ck\,\hat{\tilde{V}}_z(k\sin\theta\cos\phi,\ k\sin\theta\sin\phi, z_S)\,. \tag{34}$$

Using $\hat{\tilde{V}}_z$ from the Fourier transform of Eq. (28) (with $z=z_S$) yields

$$\tilde{D}(\theta,\phi) = k\cos\theta\,\hat{\tilde{\psi}}(k\sin\theta\cos\phi, k\sin\theta\sin\phi, z_H) \times ie^{-ik\cos\theta(z_H-z_S)}\,. \tag{35}$$

Thus the farfield directivity pattern can be found from the Fourier transform of the hologram data. Usually the phase factor is ignored. It is important to note that since $(k\sin\theta\cos\phi)^2+(k\sin\theta\sin\phi)^2 = (k\cos\theta)^2 \leqslant k^2$, then the

farfield directivity pattern depends only on those components $\tilde{\tilde{\psi}}(k_x,k_y,z_H)$ which lie inside the radiation circle.

3. Second-order quantities (acoustic vector intensity field, total power radiated)

From the three-dimensional field ψ and its gradient $\nabla\psi$ second-order products may be determined. A particularly important example is the acoustic vector intensity field, defined by

$$\mathbf{S}(\mathbf{r}) = \frac{1}{T}\int_{t_0}^{t_0+T} \psi(\mathbf{r},t)\mathbf{V}(\mathbf{r},t)\,dt\,, \tag{36}$$

where ψ and $\mathbf{V}$ are the sound pressure and particle velocity fields, and T is a suitable time scale[13] for noise sources or the period for harmonic sources. With the assumption required to make Eq. (5) an equality (i.e., harmonic sources), then Eq. (36) becomes a sum of independent frequency terms, each contributing to the intensity field an amount

$$\mathbf{S}(\mathbf{r}) = \tfrac{1}{2}\,\mathrm{Re}[\tilde{\psi}(\mathbf{r})\tilde{V}^*(\mathbf{r})]\,. \tag{37}$$

This can be calculated from the hologram data using Eqs. (24) and (28).

By integrating the normal component of the intensity field over a suitable surface, the total power radiated may be obtained. For plane holography, the total power radiated into the half-space away from the source plane is

$$P \equiv \frac{1}{2}\,\mathrm{Re}\int\int_{-\infty}^{\infty} \tilde{\psi}(x,y,z)\tilde{V}_z^*(x,y,z)\,dx\,dy\,, \tag{38}$$

for any $z \geqslant z_S$. For some sources it may be known that $\tilde{\psi}$ or $\tilde{V}_z$ vanishes except over some finite region, so that Eq. (38) may be evaluated numerically. In any case, Eq. (38) may be rewritten using the identity

$$\int\int_{-\infty}^{\infty} \tilde{\psi}(x,y,z)\tilde{V}_z^*(x,y,z)\,dx\,dy = \frac{1}{4\pi}\int\int_{-\infty}^{\infty} \tilde{\tilde{\psi}}(k_x,k_y,z)\tilde{\tilde{V}}_z^*(k_x,k_y,z)\,dk_x\,dk_y\,, \tag{39}$$

so that

$$P = \frac{1}{8\pi}\,\mathrm{Re}\int\int_{-\infty}^{\infty} \tilde{\tilde{\psi}}(k_x,k_y,z)\tilde{\tilde{V}}_z^*(k_x,k_y,z)\,dk_x\,dk_y\,. \tag{40}$$

From Eq. (24) we have

$$\tilde{\tilde{\psi}}(k_x,k_y,z) = \tilde{\tilde{\psi}}(k_x,k_y,z_H)\,e^{ik_z(z-z_H)} \tag{41}$$

and from Eq. (28) we have

$$\tilde{\tilde{V}}_z(k_x,k_y,z) = (1/\mu c)\tilde{\tilde{\psi}}(k_x,k_y,z_H)(k_z/k)\,e^{ik_z(z-z_H)}\,. \tag{42}$$

Keeping in mind that $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$ is a complex function, we have

$$\mathrm{Re}[\tilde{\tilde{\psi}}(k_x,k_y,z)\tilde{\tilde{V}}_z^*(k_x,k_y,z)] = \begin{cases} (1/\mu c)|\tilde{\tilde{\psi}}(k_x,k_y,z_H)|^2\sqrt{1-(k_x^2+k_y^2)/k^2}\,, & k_x^2+k_y^2 \leqslant k^2\,, \\ 0\,, & k_x^2+k_y^2 > k^2\,. \end{cases} \tag{43}$$

Now

$$P = \frac{1}{8\pi^2\mu c}\int\int_{k_x^2+k_y^2<k^2} |\tilde{\tilde{\psi}}(k_x,k_y,z_H)|^2 \times\sqrt{1-(k_x^2+k_y^2)/k^2}\,dk_x\,dk_y\,. \tag{44}$$

Like the farfield directivity pattern, the total power radiated depends only on the components $\tilde{\tilde{\psi}}(k_x,k_yz_H)$ which lie inside the radiation circle. When Eqs. (35) and (44) are evaluated numerically using actual hologram data, care must be taken to insure that there is a sufficient density of data points inside the radiation circle. This will be discussed further in Sec. IV.

F. Cylindrical holography

As demonstrated by Eqs. (24) and (25), the expressions of generalized holography may be found by using separation of variables to find the general solution to the Helmholtz equation (6) and then using the hologram data and the orthogonality of the eigenfunctions to find the unique solution. In cylindrical coordinates (ρ,ϕ,z) the general solution (for sources contained just inside the surface S given by $\rho = \rho_S$ and radiating outward) is

$$\tilde{\psi}(\rho,\phi,z) = \sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{A}_m(k_z)e^{im\phi}e^{ik_zz}H_m(k_\rho\,\rho)\,dk_z\,, \tag{45}$$

where m is an integer, $\tilde{A}_m(k_z)$ are the eigenfunction amplitudes to be determined from the hologram data, $k_\rho = \sqrt{k^2-k_z^2}$ is a complex quantity analogous to k_z in Eq. (23), and $H_m(k_\rho\,\rho)$ is the Hankel function (when $k_z \leqslant k$) or modified Hankel function (when $k_z > k$) behaving asymptotically as $\exp(i\sqrt{k^2-k_z^2}\,\rho)$ or $\exp(-\sqrt{k_z^2-k^2}\,\rho)$. The modified Hankel function solutions are analogous to the evanescent wave components of the Cartesian coordinate eigenfunctions.

The eigenfunction amplitudes $\tilde{A}_m(k_z)$ in Eq. (45) can be found from hologram data measured on the surface $\rho = \rho_H$, where $\rho_H \geqslant \rho_S$. The orthogonality of the eigenfunctions is such that

$$\int_{-\infty}^{\infty} dz\int_0^{2\pi} d\phi\,(e^{im\phi}e^{ik_zz})(e^{im'\phi}e^{ik'_zz})^* = 4\pi^2\delta_{mm}\,,\ \delta(k_z-k'_z)\,. \tag{46}$$

Evaluating Eq. (45) at $\rho = \rho_H$ and using Eq. (46) to solve for $\tilde{A}_m(k_z)$ yields

$$\tilde{A}_m(k_z) = (1/2\pi)\tilde{\tilde{\psi}}_m(k_z,\rho_H)/H_M(k_\rho\,\rho_H)\,, \tag{47}$$

where

$$\tilde{\psi}_m(k_z,\rho_n) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dz\int_0^{2\pi} d\phi\,\tilde{\psi}(\rho_u,\phi,z)e^{-im\phi}e^{ik_zz}\,. \tag{48}$$

Substituting $\tilde{A}_m(k_z)$ in Eq. (45) yields

$$\tilde{\psi}(\rho,\phi,z) = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{\psi}_m(k_z,\rho_H)e^{im\phi}e^{ik_zz} \times\left(\frac{H_m(k_\rho\,\rho)}{H_m(k_\rho\,\rho_H)}\right)dk_z\,, \tag{49}$$

which is the analog of Eq. (24). Again it should be kept in mind that $k_\rho = \sqrt{k^2 - k_z^2}$ is a complex function of k_z. Other quantities, such as the field gradient, etc., may be calculated from $\tilde{\psi}(\rho,\phi,z)$ as for the Cartesian coordinate solution. The solution (49) is valid for $\rho \geqslant \rho_S$, where $\rho = \rho_S$ is the smallest cylindrical surface which just touches the physical sources or scattering surfaces.

G. Spherical holography

In spherical coordinates (r,θ,ϕ), the general solution (for sources contained within a spherical surface S defined by $r = r_S$ and radiating outward) is

$$\tilde{\psi}(r,\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \tilde{A}_{lm} Y_{lm}(\theta,\phi) h_l(kr) , \tag{50}$$

where l and m are integers, the $Y_{lm}(\theta,\phi)$ are the spherical harmonics, and $h_l(kr)$ is the spherical Bessel function behaving asymptotically as $\exp(ikr)$. It is interesting to note that there are no exponentially decaying functions in this solution. The eigenfunction amplitudes $\tilde{A}_{lm}$ can be determined from hologram data on a spherical surface $r = r_H$, with $r_H > r_S$, by using the orthonormalization of the spherical harmonics. One obtains

$$\tilde{A}_{lm} = \hat{\tilde{\psi}}_{lm}(r_H)/h_l(kr_H) , \tag{51}$$

where

$$\hat{\tilde{\psi}}_{lm}(r_H) \equiv \int_0^{\pi} \sin\theta \, d\theta \int_0^{2\pi} d\phi \, \tilde{\psi}(r_H,\theta,\phi) Y^*_{lm}(\theta,\phi) . \tag{52}$$

Substituting $\tilde{A}_{lm}$ into Eq. (48) yields

$$\tilde{\psi}(r,\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{\tilde{\psi}}_{lm}(r_N) Y_{lm}(\theta,\phi) \left(\frac{h_l(kr)}{h_l(kr_H)} \right) , \tag{53}$$

which is the analog of Eqs. (24) and (49).

IV. ACTUAL IMPLEMENTATION

A. General

The implementation of generalized holography in an actual system involves acquisition of the hologram data and evaluation of the various expressions of generalized holography. Because the features of the (hardware) system used for actual data acquisition depend on many extraneous design variables, few general comments may be made about data acquisition. On the other hand, a number of interesting general comments can be made concerning the numerical evaluation of the holography expressions. The following paragraphs discuss the general features, problems, limitations, etc. associated with the actual implementation of generalized holography. In these paragraphs it should be assumed that the comments are about plane holography in particular but may be generalized to other coordinate systems unless otherwise stated. The aspects of a particular hardware system (used for data acquisition and processing) will be described in Sec. V.

B. Data acquisition

Concerning data acquisition, it can be assumed that the major temporal frequency components are sampled at the Nyquist rate or faster, and that any other components at higher frequencies are filtered to a sufficiently small "noise" level. The time-sampled data may then be analyzed to produce the temporal frequency complex amplitudes $\tilde{\psi}(\mathbf{r}_H)$, as discussed in Sec. III A. In theory, the hologram data must be known as a continuous function (i.e., known at all points $\mathbf{r}_H$) over the hologram surface H which may be infinite in extent (spherical holography being one exception). In practice, the hologram data can only be sampled at discrete points on a surface of finite extent (referred to as the hologram aperture). As far as the discrete sampling is concerned, one must be certain that the field $\tilde{\psi}(\mathbf{r}_H)$ is being sampled at the spatial Nyquist rate. It should be recalled that any spatial frequencies of the source which exceed those of the characteristic radiated wavelengths exponentially decay with distance from the source. Thus spatial sampling is provided with a natural filter; as an empirical rule-of-thumb, we find that if the hologram sampling is done at a distance d from the source, then the distance between sampling points should be no larger than d (see Sec. V). Discrete spatial sampling does not result in any unusual problems in generalized holography. On the other hand, the finite hologram aperture does result in fundamental problems which require special processing techniques. Of course, the holography expressions which involve integrals over infinite domains in space (as in plane and cylindrical holography) necessitate that some assumption be made about the hologram (or source) data which lie outside the finite hologram aperture. Practical limitations notwithstanding, it can be assumed that the hologram aperture may be made sufficiently larger than the sources (of finite extent) so that the field on the surface beyond the aperture is not significantly different from zero. This is a reasonable assumption for laboratory studies, but other techniques may be required for field measurements, as discussed in Sec. VI. The special processing required even when the field is zero outside the aperture is discussed in Sec. IV D below.

In addition to being finite and discrete, the actual measured hologram data will contain some intrinsic error including background sound, electronic noise, calibration errors, etc. The error level may be characterized by a dynamic range D defined by

$$D = 20 \log_{10}(M/E) , \tag{54}$$

where M is the maximum field amplitude which is measured and E is the amplitude of the error. It is interesting that this dynamic range plays a role in determining the spatial resolution of generalized holography, as discussed in the next subsection.

C. Resolution

A discussion of the resolution of the reconstructed fields of generalized (plane) and conventional holography was presented in Sec. III D. The minimum resolvable distance is on the order of $R \equiv \pi/k_{\max}$, where $k_{\max}$ is the highest spatial frequency for a measurable Fourier component $\hat{\tilde{\psi}}(k_x,k_y,z_H)$. In conventional optical and acoustical holography no evanescent waves are used in the field reconstructions so that $k_{\max} = k$ and $R = \lambda/2$. In actual implementations of generalized holography, the hologram is uniformly sampled at

discrete points in space; from the Nyquist theorem $k_{max} \leqslant \pi/a$, where a is the distance between the spatial sampling points, so that $R \geqslant a$. The sampling lattice constant a is only a lower limit for R because k_{max} may be further limited by the ability of the hologram recording medium to measure all of the necessary evanescent wave components, as discussed below.

In order for generalized holography to surpass conventional holography in resolution, it is necessary to measure some evanescent wave components so that k_{max} will exceed k. The evanescent wave components decay rapidly with distance from the source, and some of the components, in traversing the distance from the source to the hologram, will decay to a level below the error level E of the hologram recording system. These evanescent wave components cannot be used in the reconstruction, and this sets a limit on k_{max}. In order to quantify this, we assume that the source, at z_S, has propagating and evanescent wave components with equal amplitudes; that is $A \equiv [\text{typical } |\hat{\tilde{\psi}}(k_x, k_y, z_S)|]$ is the same for some $k_x^2 + k_y^2 < k^2$ as for some $k_x^2 + k_y^2 > k^2$. Since the propagating wave components maintain their amplitude in traveling to the hologram plane, then $A \leqslant M$, where M is defined above in Sec. IV B. On the other hand, the evanescent wave components in the hologram plane $z_H > z_S$ will have amplitudes $A \exp\left[-\sqrt{k_x^2 + k_y^2 - k^2}(z_H - z_S)\right]$. In order for these to be used in the reconstruction the amplitude must be above the error level E:

$$A \exp\left[-\sqrt{k_x^2 + k_y^2 - k^2}(z_H - z_S)\right] > E . \tag{55}$$

Using Eq. (54) defining the dynamic range D of the hologram recording system and the relation $A \leqslant M$, we obtain

$$(k_x^2 + k_y^2) < k^2 + [D \ln 10/20(z_H - z_S)]^2 . \tag{56}$$

The expression on the right-hand side of the inequality (56) is the upper limit of usable values of $(k_x^2 + k_y^2)$ and hence is k_{max}^2. The minimum resolvable distance $R = \pi/k_{max}$ is now

$$R = \{4/\lambda^2 + [D \ln 10/20\pi(z_H - z_S)]^2\}^{-1/2} . \tag{57}$$

Since the dynamic range term is usually much larger than $4/\lambda^2$, we have

$$R \approx 20\pi(z_H - z_S)/D \ln 10 . \tag{58}$$

Thus, in actual implementation of generalized holography, good resolution is obtained by having a precise recording system (large dynamic range D) and by measuring as close to the sources as possible (small $z_H - z_S$). Measuring close to the sources is no problem in generalized holography since no use is made of Fourier optics and there is no requirement for recording in the Fraunhofer or Fresnel zone.

D. Finite aperture effects: Wraparound error

As already mentioned, practical data acquisition results in the hologram being finite in size and discretely sampled. The processing of the hologram field must also be finite and discrete in nature; that is, even if the hologram data could be assigned some assumed *a priori* values outside the data acquisition range, the time and space limitations of the data processing hardware would still restrict the hologram field to be finite in size. This finite aperture restriction leads to interesting effects, in particular an error referred to as wraparound,[19] which fortunately can be controlled with proper processing techniques. The wraparound error and the techniques used to avoid it are discussed in this section.

To emphasize that the wraparound error results from improper data processing rather than insufficient data acquisition, we shall assume that the actual field in the hologram plane is negligible for points (x, y) outside the square region defined by $x = \pm L/2$ and $y = \pm L/2$. Thus $\tilde{\psi}(x, y, z_H)$ for (x, y) within the finite $L \times L$ aperture accurately represents the full hologram plane.

The expressions of plane and cylindrical holography involve Fourier transforms which are of course numerically evaluated with finite Fourier transforms and with the FFT computer algorithm in particular. The field $\tilde{\psi}(x, y, z_H)$ inside the $L \times L$ hologram aperture is represented by the discrete series:

$$\tilde{\psi}_L(x, y, z_H) \equiv \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \hat{\tilde{\psi}}_{m,n}(z_H) e^{i(2\pi/L)(mx + ny)} , \tag{59}$$

where $\hat{\tilde{\psi}}_{m,n}(z_H)$ is proportional to $\hat{\tilde{\psi}}(k_x, k_y, z_H)$ with $k_x = (m - N/2)\pi/L$, $k_y = (n - N/2)\pi/L$, and N is an integer limited by reasonable computation times. This series evaluates to $\tilde{\psi}(x, y, z_H)$ exactly at a set of points inside the $L \times L$ aperture, but outside the aperture it represents not the actual (negligibly small) hologram field but rather the periodic extension of the field inside the aperture. This periodic extension is illustrated in Fig. 2(a); the small center square represents the $L \times L$ aperture and the localized hologram field within it, and the set of nine duplicate squares represents a portion of the infinite periodic extension. This extended field looks like a field generated by the actual source and an infinite number of image sources.

Equation (16) shows that propagation of the field away from a plane involves the convolution of the Green's function with the field in that plane. The Green's function, of approximate form $\exp(ikR)/R$ [illustrated in Fig. 2(b)], has infinite extent [indicated by the arrows in Fig. 2(b)]. If this is convolved with the periodic extension of the field [Fig. 2(a)], then contributions from the images outside the $L \times L$ aperture will leak, or "wraparound," into the reconstructed field inside the aperture. That is, the Green's function propagates the field from not only the original source, but from all the image sources as well. If one is reconstructing the field in a plane close to the hologram plane ($|z - z_H| \ll L$), then there is

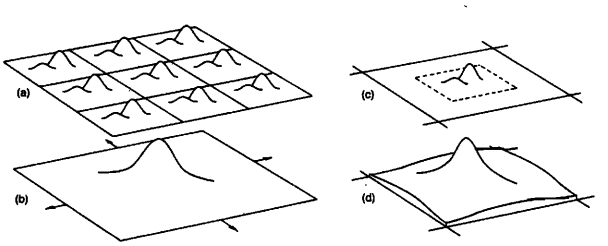

FIG. 2. The wraparound problem. (a) Periodic repetition of the source caused by its representation by a discrete Fourier series; (b) the Green's function to be convolved with the source; (c) the source with a guard band of zeros; (d) the truncated Green's function.

negligible error. However, when $(z - z_H) \sim L$, then considerable wraparound error may result.

How the wraparound error may be eliminated is illustrated in Fig. 2(c) and (d). The first step, shown in Fig. 2(c), is to surround the $L \times L$ aperture with a "guard band" of zeros, forming a $2L \times 2L$ aperture. The discrete series representing this field is

$$\tilde{\psi}_{2L}(x, y, z_H) \equiv \sum_{m=0}^{2N} \sum_{n=0}^{2N} \hat{\tilde{\psi}}_{m,n}(z_H) e^{i\pi/L\,(mx + ny)}, \qquad (60)$$

which has images as $\tilde{\psi}_L$ has, but they are farther apart. However, pushing the images farther away does little to solve the wraparound problem because there are an infinite number of images, and they may constructively interfere inside the reconstruction aperture. What solves the problem is the use of truncated Green's function defined by

$$G_T(x, y, z) = \begin{cases} G'(x, y, z), & \text{if } -L \leqslant x \leqslant L \text{ and } -L \leqslant y \leqslant L, \\ 0, & \text{otherwise}, \end{cases} \qquad (61)$$

which is illustrated in Fig. 2(d). For points (x, y) inside the original $L \times L$ aperture one has

$$\int\int_{-\infty}^{\infty} \tilde{\psi}(x, y, z_H) G'(x - x', y - y', z - z_H)\, dx'\, dy' = \int\int_{-L}^{L} \tilde{\psi}_{2L}(x', y', z_H) G_T(x - x', y - y', z - z_H) \times dx'\, dy'. \qquad (62)$$

With $-L/2 \leqslant x \leqslant L/2$ and $-L/2 \leqslant y \leqslant L/2$, then the truncated Green's function G_T ignores the images of $\tilde{\psi}_{2L}$. Thus calculating the finite convolution on the right-hand side of Eq. (62), which involves the discrete series $\tilde{\psi}_{2L}$, yields exact reconstructions, with no wraparound error, so long as one only reconstructs inside the "duct" enclosing the original $L \times L$ hologram.

In performing actual calculations, the convolution on the right-hand side of Eq. (62) is put into discrete form and evaluated using forward and inverse FFTs. Making the convolution integral discrete involves some approximations and these introduce small errors in the reconstructions. The actual numerical processing of the other quantities which can be determined with generalized holography also involves approximations and small errors. There are a number of different ways of making these approximations and it is found that some procedures result in smaller errors. The development of the techniques to minimize the wraparound and other errors, and the optimization of their computer algorithms, have been accomplished by graduate student W. A. Veronesi[20] and will be published in a second paper.

E. Zoom Imaging

As discussed in the previous section, the wraparound error can be avoided if the reconstruction volume is confined within a duct enclosing the $L \times L$ aperture. For reconstructions in the nearfield of the sources the size of this area is usually more than adequate. However, for reconstructions out to the farfield a much larger aperture would be desirable. Furthermore, having a larger aperture means that there is a higher density of discrete points in k space (the distance between points in k space is π/L), and this may be necessary for calculating quantities such as the farfield directivity and the total power radiated. It should be recalled that these quantities involved $\hat{\tilde{\psi}}(k_x, k_y, z_H)$ at points only inside the radiation circle. For low-frequency sources, the hologram aperture may be only a few wavelengths in size, and this means that there may be only a few discrete (k_x, k_y) points inside the radiation circle, as illustrated in Fig. 3(a); such a low density of points inside the radiation circle may be inadequate for calculating the directivity pattern and the total power radiated.

In order to increase the aperture size one could increase the size of the guard band of zeros, making the effective aperture size $KL \times KL$. Equivalently, one could convolve $\hat{\tilde{\psi}}(k_x, k_y, z_H)$ in k space with a $\sin \alpha/\alpha$ type function[14] in order to intersperse discrete points in k space. Unfortunately, the first technique would require a two-dimensional FFT on a very large data set, and the second would require multiplication by an even larger matrix; both would necessitate prohibitively long computation times.

However, it should be noted that the calculations which require a larger aperture (or higher density of points in k space) only require a higher density of k-space points inside the radiation circle (since reconstructions beyond a few wavelengths contain virtually no evanescent waves). It is possible to reformulate the k-space convolution technique so that it only intersperses data points within the radiation circle, as shown in Fig. 3(b). In a reasonable amount of computation time all of the original N^2 k-space points [as in Fig. 3(a)] may be relocated inside the radiation circle [as in Fig. 3(b)]. With this high density of k-space points, a much larger aperture may be obtained beyond the nearfield. The procedure for enlarging the aperture size is referred to as zoom

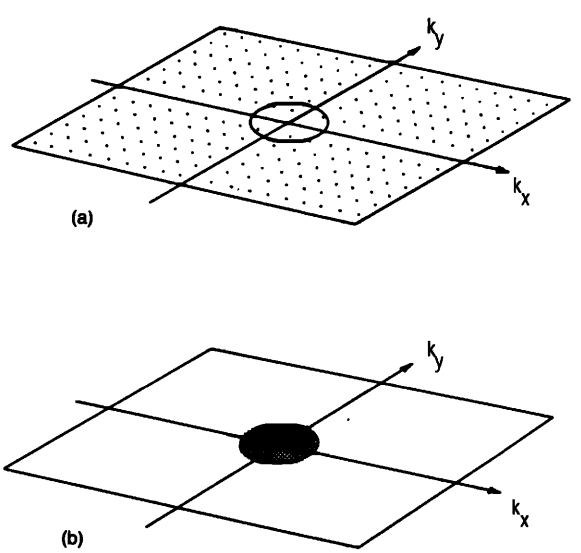

FIG. 3. Zoom imaging. (a) The finite hologram aperture may result in only a few data points falling within the radiation circle. (b) For calculation of farfield quantities, a higher density of data points may be mapped into the radiation circle.

imaging[21]; a paper describing the computer algorithm for this process is in preparation.

V. EXAMPLE OF IMPLEMENTATION: NEARFIELD ACOUSTIC HOLOGRAPHY

A. Introduction

We shall describe in this section a sound radiation measurement system, called nearfield acoustic holography (NAH), which is developed from the principles of generalized holography. At The Pennsylvania State University two NAH systems, one for airborne and the other for underwater sound, are being used for a wide variety of research studies. Several other NAH systems are being, or have been, built at other laboratories by graduate degree candidates trained at Penn State: Earl Williams has developed an underwater system at the Naval Research Laboratories in Washington, DC; Bill Strong is constructing an airborne system at Steinway Piano; and Toshi Mitzutani is developing a facility at the Technics Division of Matsushita, Inc., in Japan. Though each system has its own data acquisition features and innovations, all share the common feature of digital reconstruction based on the principles of generalized holography. Generalized holography in spherical coordinates is used in another system developed independently by G. A. Weinreich[22] at the University of Michigan, Ann Arbor. FFT methods for modeling sound radiation in Cartesian and cylindrical geometries have been explored by Stephanishen and Chen.[23]

B. The NAH system for airborne sound

A large two-dimensional open array of microphones is employed in the Penn State NAH system for airborne sound. An early system, developed with the assistance of research associate E. G. Williams and graduate students W. Y. Strong, T. B. Beyer, and D. J. Bowen, utilized data acquisition electronics which were capable of only single frequency sound radiation measurements and had slow and relatively inefficient data processing algorithms. Recently, the entire data acquisition and processing system has been totally rebuilt.[24] Now real-time measurements on wideband noise sources can be performed. With the present system, one may study the effects of noise source temporal and spatial coherence on such parameters as the acoustic vector intensity field (energy flow patterns). Applications of this work would be found in panel design, radiation cancellation, etc.

A schematic representation of the NAH airborne sound system is shown in Fig. 4. The microphone array is constructed with a square aluminum I-beam frame on which is attached a 16×16 latticework of 0.8-mm-steel wires. The microphones are located at the intersection of the wires, resulting in a total of 256 transducers. Since the microphone size (about 1 cm) is much smaller than the typical 1-m wavelengths studied, and since they are positioned with a unit to

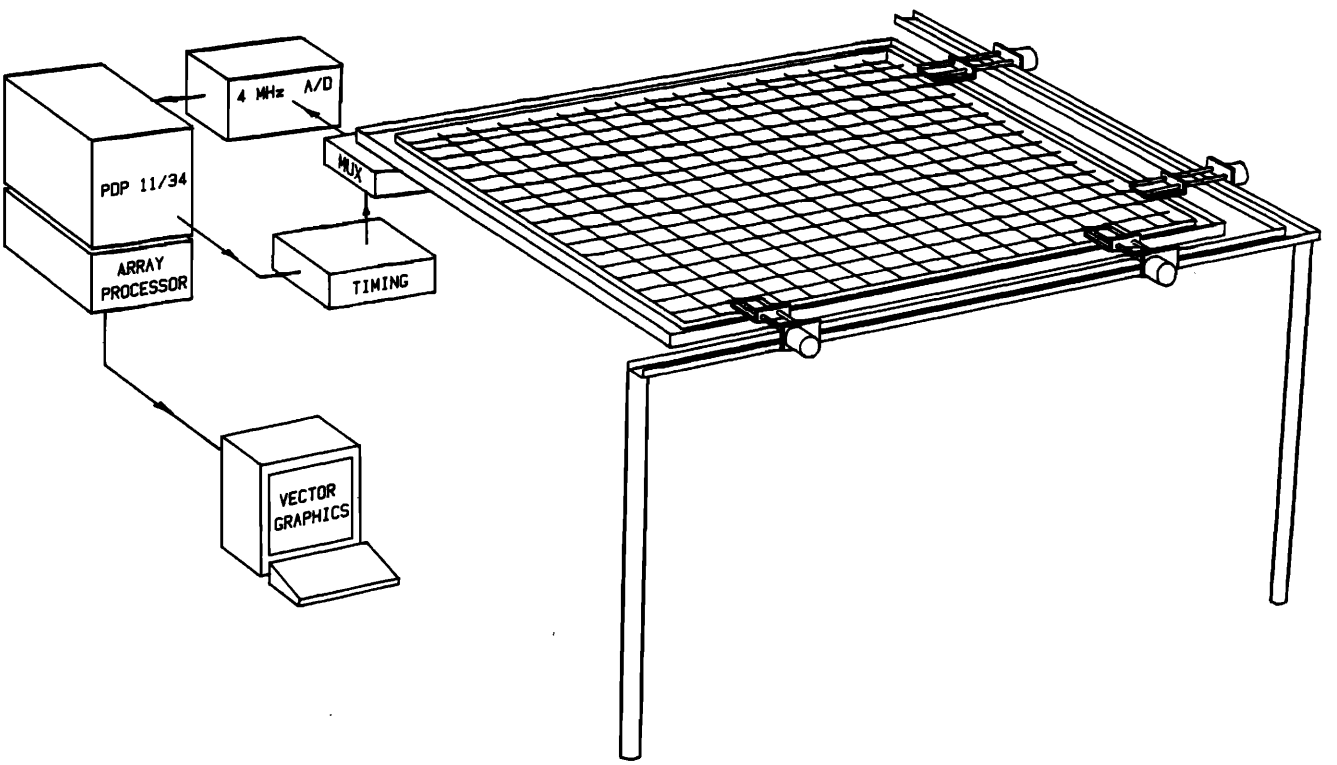

FIG. 4. The Penn State nearfield acoustic holography system for airborne sound radiation research. Illustrated are the 256-microphone array, the data acquisition electronics, and the data processing and display equipment.

unit spacing of 0.18 m, a suitable open (transparent) array is formed. In order for the evanescent wave components to be measured, the sound source to be studied is positioned just a few centimeters below the microphones, so that the array is in the extreme nearfield of the source. Typical sources have dimensions on the order of 1 m, so that the 3-m array, at a distance of about 2 cm, subtends a very large solid angle (nearly 2π sr). Consequently source directivity and total power radiated into the half-space above the source can be properly measured.

Signals from the microphones are switched through a multiplexer to a 4-MHz analog-to-digital converter. Digitized data, consisting of 256 time sequences recorded simultaneously at the 256 microphone sites, are stored in a high-speed buffer memory. At the completion of data taking (which takes but a fraction of a second) the contents of the buffer memory are transferred through an on-line minicomputer to an attached array processor. At this point a temporal FFT is performed on each of the time sequences. This yields a two-dimensional (over the hologram plane) complex data set for each of the temporal frequencies. The 2-D spatial data sets are each processed as described in the theory sections. The array processor speeds along the process by reducing the time required for 2-D complex FFTs, as well as calculations of the Green's functions, window functions, and three-dimensional graphics hidden-line plots. A vector graphics system (having a higher resolution and drawing rate than an conventional rasterscan system) receives the graphics output. The graphics system is capable of redrawing hidden-line plots every 1/60 of a second, permitting motion-picture studies of the holographic reconstructions. These motion pictures allow an experimenter to examine (in slow motion and with greatly exaggerated displacements) subtle aspects of the calculated results as they evolve in time. This feature is especially useful for examining the motion of a vibrating structure and correlating this motion with features of the holographically reconstructed radiated sound field.[24]

The density of the data points can be greatly increased for most sound sources by spatially interspersing measurements. This means that after recording one set of data, the entire microphone array is translated 1/8 of the distance between the microphones, at which point data are again recorded. This process repeats in an 8×8 pattern (within the square areas between the microphones) until a 128×128 data set with sample points every 2.2 cm is obtained. The on-line minicomputer controls the entire translation process using the system of servo motors illustrated in Fig. 4, a system developed by graduate student Toshi Mitzutani.

In order to permit data acquisition and data processing to be carried out in parallel, the timing of the signal multiplexing and digitizing is controlled by separate electronics rather than by the on-line minicomputer, as depicted in Fig. 4.

The signals from the microphones must be digitized simultaneously (on the acoustic time scale of $\sim$0.3 ms) in order to record a hologram for a wideband noise source. Installing 256 separate cables connected with 256 analog-to-digital (A/D) convertors would be both cumbersome and impractical. Instead, a multiplexing system is used in which each microphone unit has its own 60-dB amplifier, remotely selectable 20-dB attenuator (to prevent clipping with especially loud sources), low-pass filter (to prevent time domain aliasing beyond the Nyquist frequency), and an output multiplexing switch. In the array, there are 16 signal lines, with 16 microphones connected in parallel in each line. Orthogonally to the signal lines run 16 gate lines, which enable each multiplexing switch on the 16 microphone units on that gate line. Thus, when one of the gate lines is activated, the signals from the 16 switched microphones along the gate line appear on the 16 signal lines, and are ready to be digitized. Additional multiplexing is required to connect the 16 signal lines to the four inputs of the analog-to-digital convertor, as shown in Fig. 4.

Sampling the entire array in $\sim$0.3 ms requires that the microphone signals be switched and settle within 1 μs. In designing such a complex multiplexing system a great deal of effort must be expended to prevent switching transients from inducing large errors in the microphone signals. Digital and analog signals are completely isolated from each other whenever possible, and independent grounds are established for each. After considerable effort Penn State graduate student Donald Bowen[25] was able to develop a data acquisition system which settled to within the digitizing precision (0.1%) within the 1-μs time limit.

Calibration is another important factor in the microphone array operation. Microphones must be calibrated to have a high relative precision in amplitude and phase since when performing reconstructions back towards the radiating source, the positive exponential factor of the inverse Green's function will magnify any calibration error in the hologram data. Using a Bruel and Kjaer 1/4-in. microphone coupled to an array microphone within a small airtight bellows activated chamber, each microphone unit in the array was calibrated *in situ*, using the actual data recording paths. During data acquisition, calibration data are used to normalize each hologram data set so that at each of the 256 data sites it appears as though the signals were recorded with identical Bruel and Kjaer microphones. The careful calibration, and the testing of the calibration using a theoretically tractable (a rigid oscillating piston) source, were the thesis project of graduate student Todd Beyer.[26]

With the system described above, consisting of the high-speed data acquisition electronics and computer controlled array translation system, it takes but a few minutes for the Penn State NAH system to record, transform (from the time domain to the frequency domain), calibrate, and store a 64×64 point hologram. The speed and precision of the data acquisition are matched by fast and efficient computer programs developed for the digital holographic reconstructions. Graduate student Bill Veronesi[20] explored all the possible methods of sampling the Green's functions, and determined which methods were most suitable on a case by case basis. Research associate Yongchun Lee[27] wrote the computer programs which utilize the array processor to its maximum potential in all phases of the computations.

Detailed papers covering all the features of the NAH system discussed above (array construction, data acquisi-

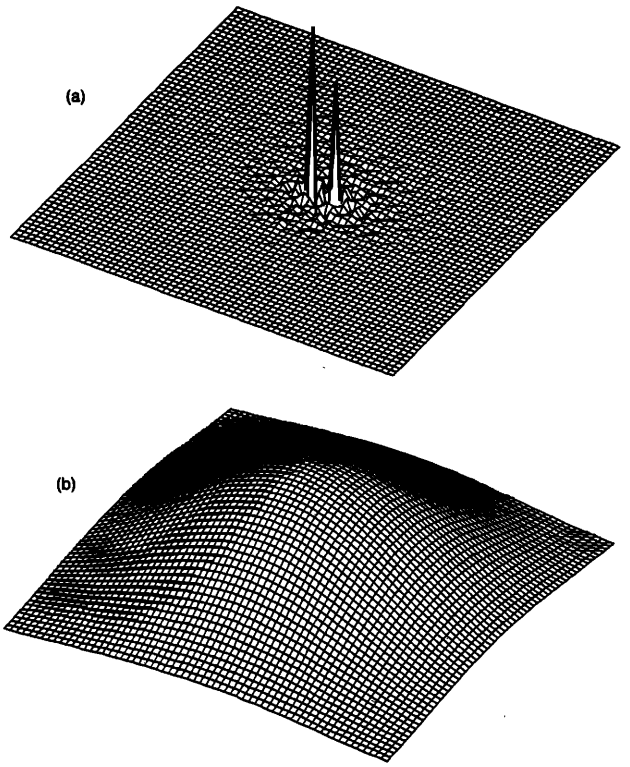

FIG. 5. Holographic reconstruction of the intensity of two point sources. (a) Nearfield holography; (b) conventional holography.

tion, array calibration and testing, and computer algorithms) will be published in the near future.

C. Examples of NAH reconstructions

The reconstruction of two point sources, simulated by the low-frequency radiation from the ends of two 2.5-cm-diam pipes driven at resonance, is shown in Fig. 5(a). The wavelength of the radiation was 3 m, while the spacing between the ends of the pipes was only 6 cm. Figure 5(b) shows the reconstruction of the two point sources using conventional holography (without evanescent waves). Notice that the reconstruction shows only a single broad maximum,

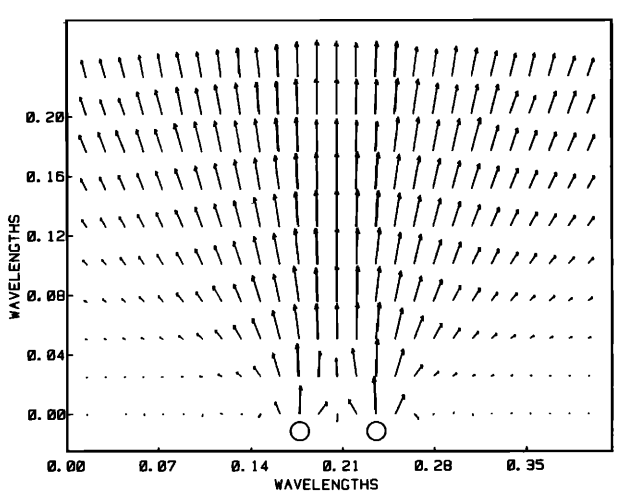

FIG. 6. Acoustic intensity vector field from two point sources.

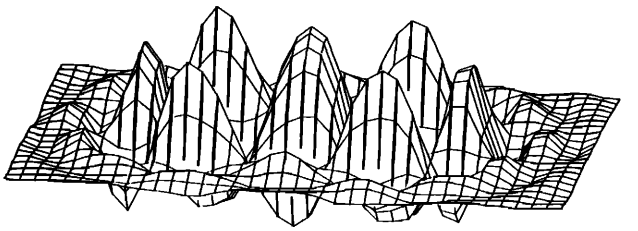

FIG. 7. Nearfield acoustic holography reconstruction of the surface velocity of a rectangular plate vibrating in a (4,2) mode.

roughly a single wavelength (3 m) in diameter. Figure 5(a) shows the NAH reconstruction; the two point sources are clearly defined. The NAH resolution is improved by a factor of ~ 50. Because the point source is such a simple source, one might assume that reconstruction is a simple matter. However, if one recalls that the intensity of an ideal point source is a product of an infinite pressure amplitude, the velocity amplitude, and the cosine of the 90° phase difference (i.e., zero), giving a finite result, the results of Fig. 5(a) are indeed impressive. It is clear the hologram data must be precisely measured, calibrated, and processed in order to so accurately process this computationally challenging wave field.

A plot of a projection of the vector intensity field in a plane containing the two point sources and perpendicular to the hologram plane is shown in Fig. 6. The acoustic energy from the two point sources flows together within a distance of only 0.05 wavelengths.

Figure 7 depicts the reconstructed surface velocity of a rectangular plate vibrating in a normal mode which has four nodal lines traversing the width of the plate and two nodal lines traversing the length. When observing the vector graphics display, subtle shifts in phase occurring between the various sections of the plate can be readily detected. When a plate is vibrating below coincidence, such phase shifts will produce dramatic changes in the vector intensity pattern above the plate surface.

In Fig. 8 a top view of the vibrating plate of Fig. 7 is shown. Different symbols indicating the surface velocity amplitude and approximate relative phase ($\cos \Delta\phi \simeq \pm 1$)

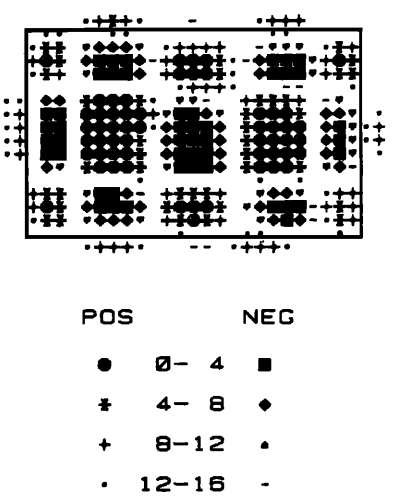

FIG. 8. Top view of the plate in Fig. 7. The different symbols indicate approximate amplitude and phase.

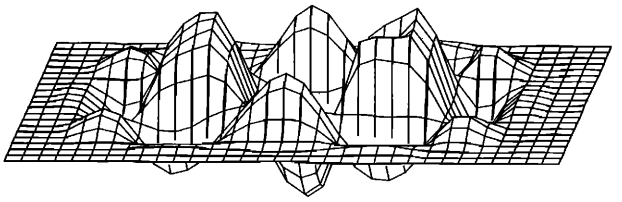

FIG. 9. Normal component of the acoustic intensity at the surface of the plate of Fig. 7.

clearly display the (4,2) nodal line pattern.

Figure 9 plots the normal component of the acoustic intensity at the surface of the plate of Figs. 7 and 8. The total radiated power is greatly diminished by the presence of regions of negative intensity which act to cancel the positive intensity regions. This is shown more clearly in Fig. 10, which gives the projected acoustic intensity in a plane through the centerline of the plate and normal to the plane of the plate. The positive and negative intensity regions are seen to be portions of a circulating energy flow pattern, as discussed in Sec. I. The circulating energy flow, which is always present for plates vibrating in a normal mode below the coincidence frequency, occurs within a fraction of a wavelength of the vibrating plate surface.

The holographically reconstructed surface intensity of a plate vibrating in a (2,2) mode is shown in Fig. 11(b). The plate was below coincidence, so regions of positive and negative intensity are evident. The theoretically predicted surface intensity for the (2,2) mode of a plate (with free edges) vibrating in an infinite rigid baffle is shown in Fig. 11(a). The surface intensity for the same plate under unbaffled conditions, as calculated by an iterative computer algorithm developed by E. G. Williams,[28] shows good agreement with the experimental results of Fig. 11(b). A comparison of Fig. 11(c) with 11(a) illustrates the significant difference in the surface intensity between a baffled and unbaffled vibrating plate.

Figure 12 shows the surface intensity for a different plate, also vibrating in the (4,2) mode, but with a much

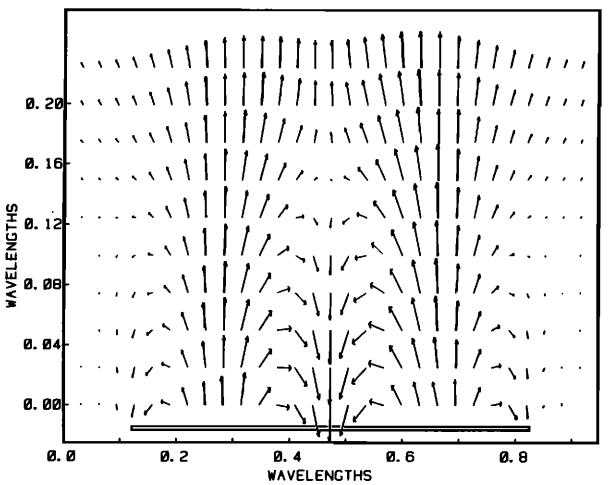

FIG. 10. Projected acoustic intensity in a plane through the centerline of the plate of Fig. 7.

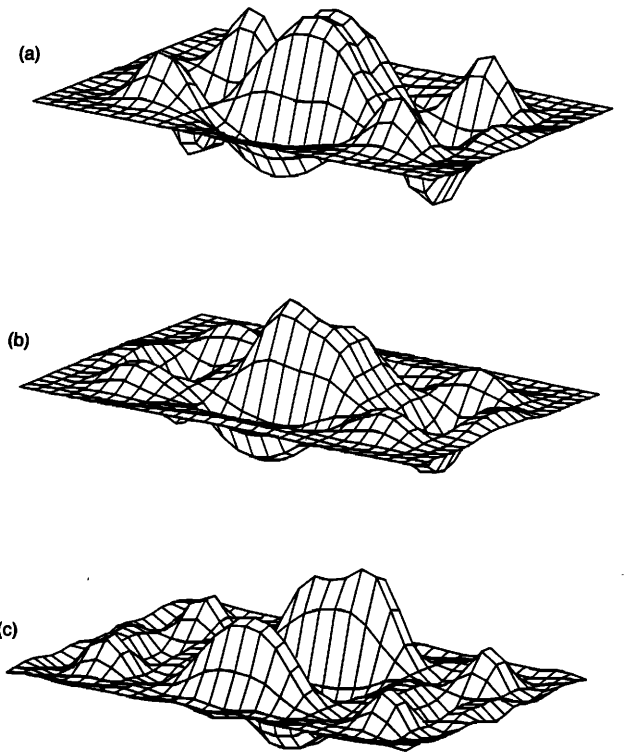

FIG. 11. Surface intensity of a plate vibrating in a (2,2) mode. (a) Theoretical unbaffled plate; (b) experimental result from nearfield acoustic holography (unbaffled plate); (c) theoretical baffled plate.

greater thickness than that of the previous plates. The high-intensity regions of Fig. 12 are all positive, as is characteristic of the higher radiation efficiency of a plate vibrating above coincidence, and contrasts sharply with the surface intensity shown in Fig. 9.

Figure 13 displays the intensity field projected onto a plane through the centerline of a rib-stiffened vibrating plate. A heavy rib is bonded across the width of this plate as shown by the edge view of the plate illustrated in Fig. 13. The rib is located at 1/4 of the length of the plate. The plate, driven at a point in the center, is vibrating at a resonance below coincidence. The intensity map shows that acoustic energy is transmitted to the air at the driver (due to the distortion of the plate by the point force) and near the rib (where the bare plate motion is disturbed, and thus promotes radiation by a reduction in cancellation). A side view of the same intensity field is shown in Fig. 14. The end view of the plate, projected in the plane containing the rib, shows the rib ex-

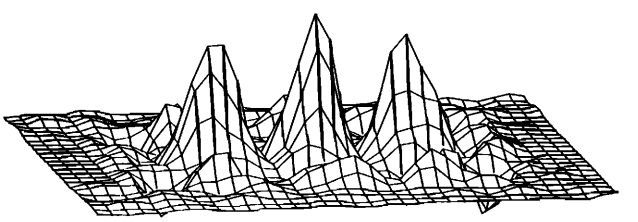

FIG. 12. Surface intensity of a thick plate vibrating in a (4,2) mode above coincidence.

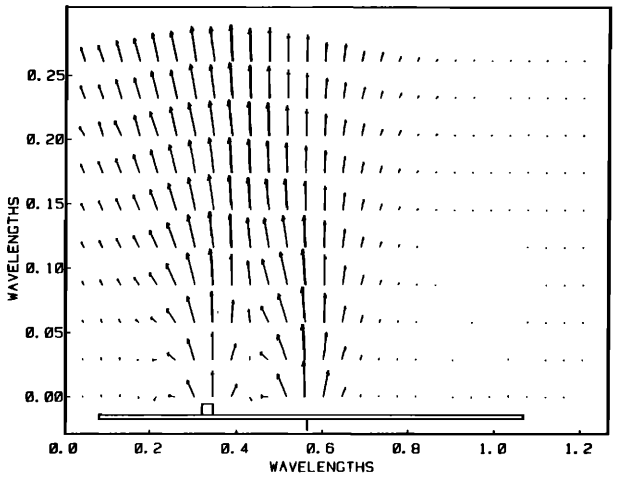

FIG. 13. Acoustic intensity in a plane through the centerline of a center-driven, rib-stiffened plate.

tending across the length of the plate. At the right-hand side, one notices a peculiar circulating energy flow, where a uniform intensity pattern might be expected. This mystery was solved upon a close examination of this plate which revealed that the epoxy bonding the rib to the plate had come loose along the right-hand section. The circulation flow pattern resulting from the faulty epoxy bond was present at all driving frequencies. It is important to stress that the spatial resolution represented in the intensity field plots of Figs. 13 and 14 is much smaller than the radiated wavelength. Thus, low-frequency nondestructive testing is possible, using the NAH system to pinpoint defects.

VI. FURTHER DEVELOPMENTS

A. General

Generalized holography can be a powerful measurement tool when applied to radiation sources or scattering objects which conform to the requirements imposed by a practical data acquisition system. There are numerous types of sources for which the limitations of finite aperture size and dynamic range produce negligible error. There are, of course, sources and environments for which the application of a generalized holography system is impossible or impractical, and other techniques must be used for these sources. However, there are some sources which do not conform to the ideal conditions of generalized holography, but for which the technique may be extended so as to make useful measurements possible. Such sources and environments may occur when generalized holography is used in the field and/or applied to sources or scatters which have a low-symmetry shape, not readily conforming to the level surface of a separable coordinate system. These situations and possible extensions of the technique will be briefly discussed in the next subsections.

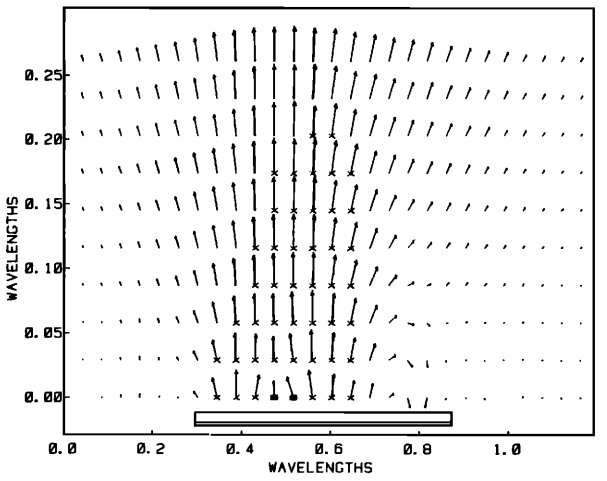

FIG. 14. Acoustic intensity in a plane through the rib of the plate of Fig. 13. The circulating energy flow pattern locates faulty bonding of the rib.

B. Development of a field measurement tool

With current technologies, a reasonably sized microphone array and portable data acquisition electronics may be constructed and used for field measurements of noise sources. Problems which might be encountered in field measurements include:

(a) Sources having a physical extent exceeding the size of the array.

(b) Reverberent environment.

(c) Sources having low-symmetry shape.

For sources which extend beyond the hologram aperature, it may be possible to make several measurements and combine them in a mosaic. Some assumptions must be made concerning the spatial and temporal coherence of the source. The mitigating factor would be that each measurement will be made in the extreme nearfield of the local sources, so that the direct radiation from these sources will dominate each reconstructed field. Once major sources have been quantified, this information can be used as *a priori* knowledge in artificially extending the aperture of a single hologram. In general, *a priori* knowledge of the physical configuration of a radiation source may be used to artificially extend the aperture. Computer simulations of this technique are currently in progress.

Field measurements in a reverberent environment may be possible if one has some knowledge of the environmental characteristics. For example, if measurements are made on a source near a rigid wall, then the hologram data might be processed with eigenfunctions $\cos k_z(z - z_{\text{wall}})$ rather than $\exp(ik_z z)$. Also a system of images may be used in an iterative or variational processing technique in order to find a field solution consistent with the hologram measurement and the environment. Finally, hologram measurements may be made on two closely spaced parallel surfaces, so that the processing can distinguish between outgoing waves from the source and incoming waves from reflections. This technique has already been successfully employed by G. Weinreich.[22]

Although the discussion above may give the impression that field measurements with holography are difficult, they are no more so than with other techniques. Nearfield holography has the advantage that measurements are made as close to the source as possible, so that the direct field of the source dominates the data. In this case holography provides the fastest, most thorough amount of radiation information available per unit effort.

C. Holography for low-symmetry objects; depth resolution

As discussed in Sec. III C, generalized holography can be used to reconstruct the wave field in a three-dimensional region bounded by the level surface S of a separable coordinate system, where S just touches the physical source. Ideally, one would want the surface of the physical source to coincide with the level surface S, so that generalized holography can be used to reconstruct the normal velocity and intensity directly at the source; in this case a detailed correlation between the properties of the source and the radiated field could be obtained. However, many sources (certainly sources encountered in the field) have shapes which do not conform to level surfaces and which may have important radiating surfaces beyond the level surface S. For example, in plane generalized holography one cannot reconstruct an "image" of the source beyond the plane (parallel to the hologram plane) which just touches the physical source region. It would of course be advantageous to obtain "depth resolution" for holographically reconstructing beyond the surface S.

The reason that generalized holography is restricted to level surfaces of separable coordinate systems is because the Green's functions (for the boundary value problems) are known only for such systems. In order to reconstruct down to the surface of an odd-shaped or low-symmetry source, one must find the Green's function satisfying a homogeneous condition on the source surface. Since in actual implementations of holography the measured and reconstructed fields are discrete and finite, then one may use linear analysis to determine a Green's function matrix. If finite element techniques[29] are combined with the principles of generalized holography, then the size of the matrices and the time required for computations can be significantly reduced. A hybrid holography/finite element technique is currently being developed by graduate student Bill Veronesi.[30] An outline of the technique follows.

For sound radiation research (as opposed to acoustic imaging) the physical shape of the source is assumed known. The surface of the source S' is divided into a two-dimensional network of finite elements and nodes, as in the conventional technique. Shape functions for both the surface velocity $v(\mathbf{r}_{S'})$ and surface pressure $P(\mathbf{r}_{S'})$ are then constructed using coordinates conforming to the local curvature of the source surface. Linear expansions of the surface velocity and pressure in terms of the shape functions are then inserted into the surface Helmholtz integral equation[31]:

$$P(\mathbf{r}_{S'}) = \frac{i\mu c}{2\pi k}\int\int_{S'} v(\mathbf{r}'_{S'})G_0(\mathbf{r}_{S'} - \mathbf{r}'_{S'})d^2r'_S - \frac{1}{2\pi}\int\int'_{S'} P(\mathbf{r}'_{S'})\frac{\partial G_0}{\partial n}(\mathbf{r}_{S'} - \mathbf{r}'_{S'})\,d^2r'_S, \quad (63)$$

where G_0 is the free-space Green's function and the prime on the second integral indicates the principal value. Evaluating the integrals in Eq. (63) over the shape functions produces a set of linear equations relating the surface velocity and pressure expansion coefficients.

The next step is to use Green's surface integral to find the pressure at a field point in terms of the surface velocity and pressure[12]:

$$P(\mathbf{r}) = \frac{1}{4\pi}\int\int_{S'}\left(\frac{i\mu c}{k}v(\mathbf{r}_{S'})G_0(\mathbf{r} - \mathbf{r}_{S'}) - P(\mathbf{r}_{S'})\frac{\partial G_0}{\partial n}(\mathbf{r} - \mathbf{r}_{S'})\,d^2r_{S'}\right). \quad (64)$$

This expression is evaluated at hologram data points, $\mathbf{r} = \mathbf{r}_H$, and the integrals are evaluated over the shape functions. Requiring that the resulting expressions for $P(\mathbf{r}_H)$ least-squares fit the hologram data produces another set of linear equations for the expansion coefficients. The two sets of equations can be solved for the coefficients which then give the source surface velocity and pressure. All other quantities of interest can be calculated from these.

ACKNOWLEDGMENTS

This work is supported by the Office of Naval Research, Physics Division, and the National Aeronautics and Space Administration.

[1]For a brief review, see, for example, H. M. Smith, *Principles of Holography* (Wiley-Interscience, New York, 1969).
[2]J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
[3]G. C. Sherman, J. Opt. Soc. Am. **57**, 1160 (1967).
[4]J. R. Shewell and E. Wolf, J. Opt. Soc. Am. **58**, 1596 (1968).
[5]E. Lalor, J. Math. Phys. **9**, 2001 (1968).
[6]H. G. Schmidt-Weinmar, in *Inverse Source Problems*, edited by H. P. Baltes (Springer, Berlin, 1978), p. 83.
[7]J. D. Maynard and E. G. Williams, *Proceedings NOISE-CON 81*, edited by L. H. Royster, F. D. Hart, and N. D. Stewart (Noise Control Foundation, New York, 1981), pp. 19–24.
[8]M. C. Junger and D. Feit, *Sound, Structures, and their Interaction* (MIT, Cambridge, MA, 1972).
[9]A. D. Pierce, *Acoustics, an Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981).
[10]F. J. Fahy, J. Acoust. Soc. Am. **62**, 1057–1059 (1977); J. Chung and J. Pope, GM Research Lab, GMR-2654 (1978).
[11]F. T. S. Yu, *Introduction to Diffraction, Information Processing, and Holography* (MIT, Cambridge, MA, 1973).
[12]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. I, pp. 803–833.
[13]J. S. Bendat and A. G. Piersol, *Measurement and Analysis of Random Data* (Wiley, New York, 1966).
[14]A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[15]See, for example, *Acoustical Holography* (Plenum, New York, 1967–1976), Vols. 1–7.
[16]Lord Rayleigh, Philos. Mag. **43**, 259 (1897).
[17]R. N. Bracewell, *The Fourier Transform and its Applications* (McGraw-Hill, New York, 1978), p. 143.
[18]M. Born and E. Wolf, *Principles of Optics* (Oxford, England, 1965), p. 563.
[19]J. Powers, in *Acoustical Holography*, edited by L. W. Kessler (Plenum, New York, 1976), Vol. 7, p. 193; D. L. Van Rooy, IBM Publications No. 320.2402, Houston Scientific Center (1971).
[20]W. A. Veronesi and J. D. Maynard, J. Acoust. Soc. Am. Suppl. 1 **75**, S71 (1984).
[21]Y. Huang, W. A. Veronesi, and J. D. Maynard, J. Acoust. Soc. Am. Suppl. 1 **75**, S71 (1984).
[22]G. Weinreich and E. B. Arnold, J. Acoust. Soc. Am. **68**, 404–411 (1980).
[23]P. R. Stepanishen and K. C. Benjamin, J. Acoust. Soc. Am. **71**, 803–812 (1982); P. Stephanishen and H. W. Chen, J. Acoust. Soc. Am. Suppl. 1 **74**, S24 (1983).
[24]J. D. Maynard, J. Acoust. Soc. Am. Suppl. 1 **74**, S37 (1983).
[25]D. J. Bowen and J. D. Maynard, J. Acoust. Soc. Am. Suppl. 1 **75**, S71 (1984).
[26]T. B. Beyer, "Test of the Nearfield Acoustical Holography Technique Using an Unbaffled, Uniformly Oscillating Disk," M. S. thesis, The Pennsylvania State University (1984).

[27]Y. Lee and J. D.Maynard, J. Acoust. Soc. Am. Suppl. 1 **75**, S71 (1984).
[28]E. G. Williams, J. Acoust. Soc. Am. **74**, 343–347 (1983).
[29]O. C. Zienkiewicz, *The Finite Element Method* (McGraw-Hill, London, 1977).
[30]W. A. Veronesi and J. D. Maynard, J. Acoust. Soc. Am. Suppl. 1 **75**, S71 (1984).
[31]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II, p. 526.