# Control of volume holograms

## David Brady

*Department of Electrical and Computer Engineering, University of Illinois at Urbana—Champaign,*
*Urbana, Illinois 61801*

## Demetri Psaltis

*Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91105*

The mismatch in the number of degrees of freedom supported by volume holograms and the boundary fields that control them limits the dynamic range of recorded holograms. For holograms controlled by using fractal sampling grids, the maximum dynamic range falls inversely with the minimum number of exposures needed to record the hologram, the rank of the hologram. In adaptive holography, feedback between coupled holograms prevents the dynamic range from decreasing faster than the fundamental limit. If the control problem is overcome, the maximum dynamic range that a hologram can support falls inversely with the square root of the rank. In principle, holograms in which the dynamic range falls inversely with the square root of the rank can be recorded by using cross-spectrally coherent polychromatic pulses.

## 1. INTRODUCTION

Volume holography is of interest for optical data storage,[1,2] artificial neural processing,[3-7] parallel optical interconnections,[8,9] and the active control of cavities. Interest arises because the number of degrees of freedom that may be stored in a volume is large and because holographic control methods permit the parallel dynamic updating of these degrees of freedom. In this paper we analyze the capabilities and limitations of holographic control methods. A hologram is a three-dimensional (3-D) function describing the optical properties of a volume. The hologram can be described in the spatial domain by a vector describing the state of all the resolution elements, or voxels. The hologram can also be described in the spatial-frequency domain by a vector describing the state of all the resolvable Fourier components, or gratings. Holograms are difficult to control because there is no way to reach into a material volume and change one voxel without affecting other voxels. This paper addresses the static problem of specifying the final state of a volume perturbation and the dynamic problem of adaptively programming the state of a volume.

Nonholographic methods for programming the state of a volume include layered growth and multiphoton recording. Layered growth can be accomplished by conventional photolithography or by photopolymer micromachining.[10] In multiphoton recording, control signals consist of beams at different wavelengths that cross in only one voxel or focusing beams that exceed critical intensity in only one voxel.[11] From a control perspective, the difference between holographic control and other techniques is that in holographic systems there is no one-to-one relationship between the state of a specific control signal and the state of a specific voxel. Holographic control signals are fields on the boundary of the volume. The state of a voxel is a function of the fields generated in it over time. The relationship between the final states of the voxels and the control signals is specified by a global transformation in space and in time.

To drive the voxels to some desired state, one must invert this transformation to find the appropriate control signals. While the analysis of holographic systems is more complex, holographic control is easier to implement and adapt than layered growth and is less material and source dependent than critical control.

Holographic control is complementary to optical tomography. Tomographic imaging is achieved by measuring signals scattered from a discrete sequence of probe fields.[12] To a first approximation, holographic control is achieved by inverting this process and recording the interference between the desired signals and a set of reference fields. In practice, a number of problems unique to the control problem arise. The effects of sequential exposures on a holographic volume are not independent, as information obtained by different tomographic probes is. The main coupling mechanism between information recorded in different exposures is response saturation. Less fundamental problems include cross gratings between signal fields, dynamic wave mixing, and geometric constraints on the material response. In this paper we consider fundamental and materials-specific constraints on holographic control of volume media. Our goals are to derive fundamental constraints on the information storage and processing capabilities of volume holograms and to discover control techniques and materials that most nearly approach optimal capabilities.

We focus on the holographic control system sketched in Fig. 1. The system consists of a photosensitive volume, a pair of two-dimensional (2-D) spatial light modulators (SLM's), and passive optical elements. The recorded hologram is a photo-driven perturbation of the optical properties of the volume. We represent this perturbation as $A(\mathbf{r}, t)$. In general, $A(\mathbf{r}, t)$ is a complex-valued tensor function of the frequency of the probing field. The evolution of the perturbation is determined by the fields on the SLM's and the properties of the recording medium. The fields generated in the volume by the SLM's are $\mathbf{E}_1(\mathbf{r}, t)$
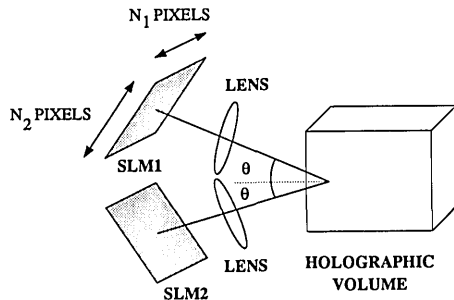
Fig. 1.   System for recording a volume hologram.

and $E_2(r, t)$. Our goal is to select $E_1(r, t)$ and $E_2(r, t)$ such that $A(r, t)$ is driven to a desired state. Since $E_1(r, t)$ and $E_2(r, t)$ are determined by 2-D devices, they lack sufficient degrees of freedom to control $A(r, t)$ instantaneously. Assuming that the fields vary discretely in time, we define the minimum number of time steps, or exposures, necessary to create $A(r, t)$ to be the rank $R$ of the hologram. If $N_1$ and $N_2$ are the numbers of resolvable spatial degrees of freedom in $E_1(r, t)$ and $E_2(r, t)$, respectively, and $N_g$ is the number of spatial degrees of freedom in $A(r, t)$, equating the control and controlled degrees of freedom implies that

$$R \geq \frac{N_g}{N_1 + N_2}. \tag{1}$$

We assume that $A(r, t)$ can assume any state consistent with the total dynamic range of the holographic medium. The number of degrees of freedom in $A(r, t)$, $N_g$, is equal to the number of independent voxels or, equivalently, the number of independent gratings needed to describe it. No matter how $A(r, t)$ is controlled, if the gratings that describe it are statistically independent, then the maximum expected value of the amplitude of an individual grating is $A_0/\sqrt{N_g}$,[5] where $A_0$ is the maximum amplitude the material supports in a single-grating hologram. Substituting for $N_g$ from Eq. (1), we find that, at best, the amplitude of the recorded gratings falls inversely with $\sqrt{R}$. As we show below, this optimal limit is elusive in practice. The more natural result for holographic control techniques is for the mean grating amplitude to fall inversely with $R$.[13,14] However, it is possible to obtain a result closer to the information theoretic limit by choosing a nonlocal recording medium or by using polychromatic recording techniques. In the remainder of this section, we overview the characteristics of holographic materials and summarize our results for each material class and control technique.

Common holographic materials may be grouped into three classes:

1. Materials in which the holographic effect is due to changes in dopants. This class includes a number of photo-organics, such as $\alpha$-diketones and dyes in organic matrices,[15,16] and organic and inorganic photochromics. The supporting matrix is relatively inert in holographic processes in these materials.

2. Materials in which the holographic effect is due to interactions between dopants and the supporting matrix. This class includes photorefractive crystals, photopolymers, and gelatins. In photorefractives the effect is based on reversible charge transport.[17] The effect in photopoly-

mers and gelatins is due to irreversible photo-driven bond formation.[18]

3. Pure materials with weak interactions. Liquid crystals that may be reoriented by optical or optically driven thermal effects are an example for this class.[19,20] The interaction must be weak if the hologram is to be thicker than a few wavelengths.

Holographic effects may be further differentiated into local and nonlocal, reversible and irreversible, latent and nonlatent, and linear and nonlinear phenomena. Effects for which the holographic perturbation at a given point is a function only of the energy absorbed at that point are local. Materials in classes 1 and 3 are generally local. Holograms are reversible if the unperturbed material has lower energy and entropy than the holographically perturbed material. Examples of both reversible and irreversible behavior exist in all three classes of material. A latent effect is a perturbation that does not couple its recording signals. An effect is linear if the perturbation is proportional to the recording intensity. While truly latent and linear media are not available, the concept of such a material is useful in analyzing holographic control.

To a first approximation, the dynamics of holographic recording in most materials can be modeled either by monotonic irreversible growth toward saturation or as exponential growth toward saturation and exponential relaxation back to the unperturbed state. Irreversible saturable behavior is observed when permanent photochemical changes are made, as in silver halide films and photo-driven polymerizations. Reversible exponentially decaying behavior is observed when photochemicals are pumped into unstable states, as in cis–trans isomerization, or into unstable thermal and orientational states, as in liquid crystals or when photo-generated charge is redistributed in the supporting lattice in photorefractives.

Two levels of difficulty are encountered in controlling volume holograms. In the simplest case, we know the perturbation that we wish to record. We are not interested in the dynamics of the recording process as long as the final state of the recorded hologram matches our target. This case corresponds, for example, to the fabrication of volume holographic optical elements and read-only memories. Monochromatic methods for recording target holograms are the subject of Section 2. We simplify our analysis by assuming that the control fields are confined to fractal sampling grids that ensure that the control parameters that address a given grating in $A(r, t)$ are unique. Under these conditions our analysis confirms the result that the dynamic range of $A(r, t)$ is inverse in $R$ for both irreversible and reversible media. However, in nonlocal reversible media the loss in dynamic range that is due to increasing $R$ can be alleviated a factor in the ratio of the real and imaginary parts of the response times. Such media can behave as though there were an asymmetry in the recording and erasing response times. Assuming that this asymmetry is large, it may be possible to approach the information theoretic limit for the dynamic range in these media. The second level of difficulty in holographic control arises when the desired hologram is unknown or when we wish to constrain the dynamic states of $A(r, t)$ rather than simply specify its final state. Adaptive holographic systems such as optical neural networks and dynamically
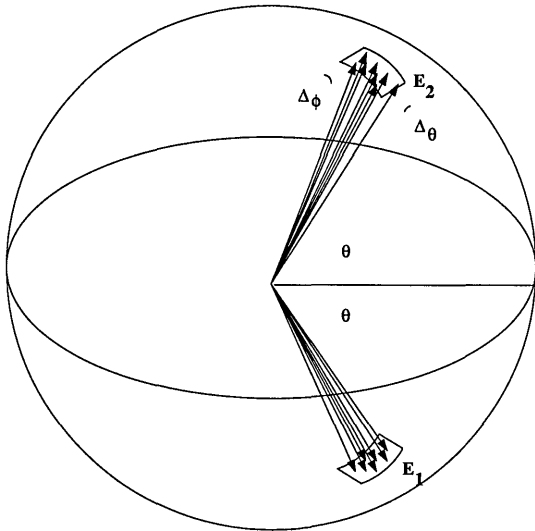
Fig. 2. Geometry of the control signals on the wave normal surface.

controlled cavities are examples of this case. The control of adaptive holograms is discussed in Section 3. In general, an adaptive system will lose dynamic range faster than the $1/R$ falloff of targeted systems. However, feedback between adaptive holograms can maintain the dynamic range at the level of targeted systems. In both Sections 2 and 3 we limit ourselves to monochromatic control. Polychromatic control allows the fields $\mathbf{E}_1$ amd $\mathbf{E}_2$ to support instantaneously as many degrees of freedom as $A(\mathbf{r}, t)$. In Section 4 we find that, while cross-spectrally incoherent polychromatic control achieves no better dynamic range than sequential control, cross-spectrally coherent ultrahigh-speed control may allow us to record holograms at the information theoretic limit of the dynamic range.

## 2. TARGETED RECORDING

Given a target for $A(\mathbf{r}, t)$ at time $t = T_0$, our goal is to find $\mathbf{E}_1(\mathbf{r}, t)$ and $\mathbf{E}_2(\mathbf{r}, t)$ such that the target hologram is recorded with maximal amplitude. We analyze this problem for latent linear media and for media in which the perturbation is proportional to the modulation depth of the exposing energy. We then show that our results apply to simple models of both irreversible and reversible holography. For the reasons discussed below, we assume that the spatial spectrum of the control fields is constrained by fractal sampling grids.

### A. Recording Geometry
A medium is linear if the holographic perturbation is proportional to the recording energy. Neglecting the low-spatial-frequency components, in a linear medium

$$\frac{dA(\mathbf{r}, t)}{dt} = \frac{A_0}{E_0}[\mathbf{E}_1 \cdot \mathbf{E}_2{}^* + \mathbf{E}_1{}^* \cdot \mathbf{E}_2], \qquad (2)$$

where $E_0$ and $A_0$ are constants. It is simpler to analyze this equation in Fourier space. Given a finite recording medium, $A(\mathbf{r}, t)$ is represented by the discrete Fourier series

$$A(\mathbf{r}, t) = \sum_{\mathbf{n}} A_{\mathbf{n}}(t)\exp(j\mathbf{K}_{\mathbf{n}} \cdot \mathbf{r}), \qquad (3)$$

where $\mathbf{K}_{\mathbf{n}} = (2\pi n_x/L_x)\hat{x} + (2\pi n_y/L_y)\hat{y} + (2\pi n_z/L_z)\hat{z}$; $n_x$, $n_y$, and $n_z$ are integers; and $L_x$, $L_y$, and $L_z$ are the extents of the holographic volume along the corresponding axes. Since the control fields are specified by 2-D spatial boundary conditions, their Fourier representations are 2-D for monochromatic fields. In Fig. 1 the control fields are specified by the two SLM's. If the SLM's are pixellated, $\mathbf{E}_1$ and $\mathbf{E}_2$ consist of discrete sums over the modes excited by individual pixels. We assume that the pixels are arranged so that they excite modes corresponding to discrete plane waves with wave vectors spaced uniformly on the wave normal surface. This assumption implies that there is curvature in the rectilinear layout of the pixels on the SLM's. As sketched in Fig. 1, the central axis of the fields incident from each SLM forms an angle $\theta$ with respect to the $z$ axis. Figure 2 shows the arrangement of the components of $\mathbf{E}_1$ and $\mathbf{E}_2$ on the wave-normal surface. The spacing in the plane of the central axes is $\Delta_\theta$, and the spacing normal to this plane is $\Delta_\phi$. $\mathbf{E}_1$ and $\mathbf{E}_2$ are

$$\mathbf{E}_1(\mathbf{r}, t) = \sum_{n_1=-N_1/2}^{N_1/2} \sum_{n_2=-N_2/2}^{N_2/2} \mathscr{E}_{n_1 n_2}^1(t)\exp(-j\omega_0 t)$$
$$\times \exp(j\mathbf{k}_{n_1 n_2}^1 \cdot \mathbf{r}), \qquad (4)$$

$$\mathbf{E}_2(\mathbf{r}, t) = \sum_{n_1=-N_1/2}^{N_1/2} \sum_{n_2=-N_2/2}^{N_2/2} \mathscr{E}_{n_1 n_2}^2(t)\exp(-j\omega_0 t)$$
$$\times \exp(j\mathbf{k}_{n_1 n_2}^2 \cdot \mathbf{r}), \qquad (5)$$

where

$$\mathbf{k}_{n_1 n_2}^1 \approx k_0\{[\cos(\theta) - n_1\Delta_\theta \sin(\theta)]\hat{z}$$
$$+ [\sin(\theta) + n_1\Delta_\theta \cos(\theta)]\hat{x} + n_2\Delta_\phi \sin(\theta)\hat{y}\}, \qquad (6)$$

$$\mathbf{k}_{n_1 n_2}^2 \approx k_0\{[\cos(\theta) - n_1\Delta_\theta \sin(\theta)]\hat{z}$$
$$+ [\sin(\theta) + n_1\Delta_\theta \cos(\theta)]\hat{x} + n_2\Delta_\phi \sin(\theta)\hat{y}\}. \qquad (7)$$

We assume that $(N_1\Delta_\theta)^2$, $(N_2\Delta_\phi)^2 \ll 1$.

In Fourier space Eq. (2) becomes

$$\frac{dA_{\mathbf{n}}}{dt} = \frac{A_0}{E_0}\left[\sum_{n_1, n_2} \mathscr{E}_{n_1, n_2}^1(t)\mathscr{E}_{n_1', n_2'}^{2\,*}(t)\right], \qquad (8)$$

where we have reduced it to scalar form by assuming that $\mathbf{E}_1$ and $\mathbf{E}_2$ are of the same polarization and that $n_1{}'$ and $n_2{}'$ are such that

$$|[\mathbf{K}_{\mathbf{n}} - (\mathbf{k}_{n_1 n_2}^1 - \mathbf{k}_{n_1' n_2'}^2)]i| < \frac{\pi}{L_i}, \qquad (9)$$

where $i$ denotes the $x$, $y$, or $z$ component. Equation (8) describes the dynamics of the Fourier coefficients corresponding to $\mathbf{K}_{\mathbf{n}}$ primarily along the positive $x$ axis; the coefficients for $\mathbf{K}_{\mathbf{n}}$ along the negative $x$ axis are found by conjugating. The range of $\mathbf{n}$ for which $dA_{\mathbf{n}}/dt$ is nonzero is limited by the range of $n_1$ and $n_2$. Inequalities (6) and (7) yield

$$\mathbf{k}_{n_1 n_2} - \mathbf{k}_{n_1' n_2'} = k_0\{(n_1 + n_1')\Delta_\theta \sin \theta \hat{z} + [2 \sin \theta$$
$$+ (n_1 - n_1')\Delta_\theta \cos \theta]\hat{x} + (n_2 - n_2')\Delta_\phi \sin \theta \hat{y}\}. \qquad (10)$$

The number of distinct values of $\mathbf{k}_{n_1 n_2} - \mathbf{k}_{n_1' n_2'}$ corresponds to the number of different triplets $(n_1 - n_1', n_2 - n_2', n_1 + n_1')$. Assuming that the range of $n_1$ and $n_1'$ is $[-N_1/2, N_1/2]$ and that the range of $n_2$ and $n_2'$ is

$[-N_2/2, N_2/2]$, there are $(N_1 + 1)^2(2N_2 + 1)$ such triplets. Because we regard the gratings corresponding to $\mathbf{K_n}$ and $-\mathbf{K_n}$ as equivalent, however, the triplets $(n_1 - n_1', n_2 - n_2', n_1 + n_1')$ and $(n_1' - n_1, n_2' - n_2, -n_1 - n_1')$ are not considered different. The number of different triplets is therefore approximately $N_1^2N_2$. Each triplet corresponds to a distinct component $A_n$ if

$$\Delta_\theta \sin\theta \geq \frac{\lambda}{L_z}, \tag{11}$$

$$\Delta_\theta \cos\theta \geq \frac{\lambda}{L_x}, \tag{12}$$

$$\Delta_\phi \sin\theta \geq \frac{\lambda}{L_y}. \tag{13}$$

These constraints combined with those following inequality (7) imply that the number of distinct Fourier components of the perturbation that can be controlled by this system is much less than $L_y(1/L_z^2 + 1/L_x^2)^{-1} \sin\theta/\lambda^3$, or, assuming that $L_x = L_y = L_z$ and $V = L_xL_yL_z$,

$$N_1^2N_2 \ll \frac{V}{\lambda^3} \sin\theta. \tag{14}$$

When expressions (11)–(13) apply, the sum over $n_1$ in Eq. (8) can be dropped because specifying $\mathbf{n}$ uniquely determines $n_1$ and $n_1'$. ($n_1$ and $n_1'$ correspond to propagation directions in the plane of the optical axes of the SLM's. The uniqueness of $n_1$ and $n_1'$ for a given grating is due to Bragg selectivity for rotations in this plane.) The sum over $n_2$ must be maintained, however, because different values of $n_2$ and $n_2'$ may correspond to the same value of $n_y$. (Degeneracy in $n_2$ and $n_2'$ is due to the degeneracy of the Bragg condition for rotations about the grating wave vector.) Equation (8) can be rewritten as

$$\frac{dA_\mathbf{n}}{dt} = \frac{A_0}{E_0}\left[\sum_{n_2=-N_2/2}^{N_2/2} \mathscr{E}_{n_1,n_2}^1(t)\mathscr{E}_{n_1',n_2'}^{2*}(t)\right], \tag{15}$$

where $n_1 = (n_x + n_z)/2$, $n_1' = (n_z - n_x)/2$, and $n_2' = n_2 - n_y$. If the hologram is controlled by discrete exposures, then $\mathscr{E}_{n_1,n_2}^1(t)$ and $\mathscr{E}_{n_1',n_2'}^2(t)$ may be replaced by the piecewise constant functions

$$\mathscr{E}_{n_1,n_2}^1(t) = \mathscr{E}_{n_1,n_2}^1(i) \quad \text{for} \sum_{j<i} t_j < t < \sum_{j\leq i} t_j, \tag{16}$$

$$\mathscr{E}_{n_1',n_2'}^2(t) = \mathscr{E}_{n_1,n_2}^2(i) \quad \text{for} \sum_{j<i} t_j < t < \sum_{j\leq i} t_j. \tag{17}$$

In this case the solution to Eq. (15) is

$$A_\mathbf{n} = \sum_{i=1}^{N_e}\sum_{n_2=-N_2/2}^{N_2/2} \frac{A_0t_i}{E_0}\mathscr{E}_{n_1,n_2}^1(i)\mathscr{E}_{n_1',n_2'}^{2*}(i) + A_\mathbf{n}(0), \tag{18}$$

where $t_i$ is the exposure time for the $i$th exposure and $N_e$ is the number of exposures made. The rank $R$ of a hologram is the minimum $N_e$ needed to drive the components $A_n$ to the target state. Since there are only $2N_1N_2$ degrees of freedom in the control fields in each exposure, $R$ must be at least $N_1/2$ if all $N_1^2N_2$ components $A_n$ are to be independently specified. In this section we describe the methods for specifying $A_n$ in at most $N_1$ exposures.

Our purpose in recording $A(\mathbf{r})$ is to perform transformations between the input and the output fields. The input fields correspond to probes generated on SLM2 in Fig. 1, and the output fields correspond to diffracted signals in the plane conjugate to SLM1. A geometry for monitoring transformations between the input and the output fields is shown in Fig. 3. Each frequency component of the hologram, $A_n$, is a grating that can connect a point at the input plane with a point at the output. Assuming that the monitoring fields do not substantially change the hologram, the transformations implemented by this system are linear. A linear transformation from a $D_i$ discrete component input space to a $D_0$ discrete component output space may be represented by a $D_0 \times D_i$ matrix. The transformation is arbitrary if each component of the matrix is independent. Just as the $2N_1N_2$ degrees of freedom available on the SLM's for controlling the hologram are insufficient, the $N_1^2N_2$ degrees of freedom in the volume are $N_2/2$ times fewer than the number needed to implement arbitrary linear transforms between the fields on SLM2 and the fields incident upon the detector array conjugate to SLM1. An arbitrary linear transformation is possible only if some of the degrees of freedom available on the SLM's are not used. Sampling patterns that constrain the pixels active at the input SLM and the output detector array such that arbitrary linear transformations are possible have been developed.[6,21,22] These patterns are referred to as fractal sampling grids because they constrain the input and the output fields to fractal dimensions.

The simplest types of fractal grid restrict the product of the number of nodes active at the input, $D_i$, and the number of nodes active at the output, $D_0$, to fewer than $N_1^2N_2$. Fractal grids for which this product is greater than $N_1^2N_2$ have also been developed, but transformations implemented on them, while potentially useful, are not arbitrary in the sense described above. For all $D_i$ and $D_0$ such that $D_iD_0 \leq N_1^2N_2$, sampling grids can be found such that arbitrary linear transformations from one grid to the next are possible. We call grids of this type nondegenerate fractal grids. Although the focus of this paper is the formation rather than the reconstruction of volume holograms, nondegenerate fractal grids are of interest to us because only one pair of control and reconstruction fields is Bragg matched to each component $A_n$ when such grids are used.

## B. Linear and Saturable Media
The control fields generated on the two SLM's are formed on a single pair of nondegenerate fractal grids. A nondegenerate fractal grid limits the dependence of a given $A_n$
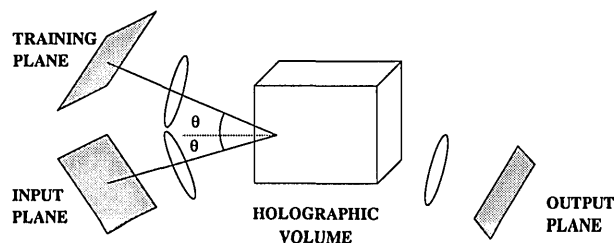


Fig. 3. System for constructing and monitoring holographic linear transformations.

on the control fields to the product of exactly one pair of coefficients $\mathscr{C}^1_{n_1, n_2}$ and $\mathscr{C}^2_{n_1, n_2}$. In other words, the sum over $n_2$ in Eq. (18) is reduced to only one value of $n_2$. If we map the $D_0$ Fourier coefficients $\mathscr{C}^1_{n_1, n_2}$ and the $D_i$ coefficients $\mathscr{C}^2_{n_1, n_2}$ onto vectors $|\mathscr{C}^1\rangle$ and $|\mathscr{C}^2\rangle$ and we map the $D_i \times D_0$ controlled Fourier coefficients $A_n$ onto a matrix $\overline{W}$, Eq. (18) may be expressed as

$$\overline{W} = \sum_{i=1}^{N_e} \frac{A_0 t_i}{E_0} |\mathscr{C}^1(i)\rangle \langle \mathscr{C}^2(i)| + \overline{W}(0).\qquad(19)$$

Given a matrix $\overline{W}$, we can find orthonormal bases $\{|u_n\rangle\}$ and $\{|v_n\rangle\}$, using, for example, singular-value decomposition such that

$$\overline{W} = \Gamma \sum_{n=1}^{R'} \beta_n |u_n\rangle \langle v_n|,\qquad(20)$$

where $R'$ is the rank of $\overline{W}$, $\beta_n$ is a positive real number, and $\Gamma$ is a positive real scaling factor.[23] Including $\Gamma$ allows us to require that

$$\sum_{n=1}^{R'} \beta_n = 1.\qquad(21)$$

$\Gamma$ is used below as a measure of the amplitude of $\overline{W}$. ($\Gamma^2$ is a measure of the diffraction efficiency of the hologram.) Since the number of outer products needed to specify $\overline{W}$ is at least $R'$, the minimum value of $N_e$ with the particular fractal grids selected is $R'$. The rank of a hologram recorded by using fractal grids is the minimum of $R'$ over all the possible grid patterns. For a fixed set of nondegenerate fractal grids, the rank $R$ of a hologram is equal to the rank $R'$ of the linear transformation between the input grid and the output grid that the hologram represents. The maximum value of $R$ is equal to the lesser of $D_i$ and $D_0$. The minimum value of $D_i$ or $D_0$ when $D_i D_0 = N_1^2 N_2$ is $N_1$. Thus the minimum rank for a hologram with $N_1^2 N_2$ independent degrees of freedom is $N_1$. Since any hologram could be recorded by using grids with $D_0 = N_1 N_2$ and $D_i = N_1$, $N_1$ is also the maximum rank for a hologram. In practice, more interesting fractal grids are often used, in which case the effective rank may be higher. A hologram of rank $R$ with one set of fractal grids may have a different rank on grids of another dimension. For example, even though the maximum rank when $D_i = D_0 = N_1 \sqrt{N_2}$ is $N_1 \sqrt{N_2}$, the outer products between $N_e$ pairs of patterns on these grids will probably have a lower rank than the corresponding hologram recorded by using grids with $D_i = N_1$ and $D_0 = N_1 N_2$ if $N_e < N_1$. In most of the analysis in this paper $R$ is the rank $R'$ of the matrix $\overline{W}$ corresponding to given input and output grids rather than the global minimum of $N_e$ with respect to all the possible grids.

In a linear medium there is no cost to making multiple exposures. This is not the case when we include saturation in our model. Saturation effects in many media can be accounted for by assuming that $\overline{W}$ is proportional to the modulation depth of the recording energy. Equation (19) is then

$$\overline{W} = \frac{A_0 \sum_{i=1}^{N_e} t_i |\mathscr{C}^1(i)\rangle \langle \mathscr{C}^2(i)|}{\sum_{i'=1}^{N_e} t_i [\langle \mathscr{C}^1(i')|\mathscr{C}^1(i')\rangle + \langle \mathscr{C}^2(i')|\mathscr{C}^2(i')\rangle]},\qquad(22)$$

where we assume that $\overline{W}(0) = 0$. For a single exposure, the $\overline{W}$ of Eq. (22) is proportional to the modulation depth of the recording intensity. To find the amplitude of $\overline{W}$, we cast Eq. (22) in the form of Eq. (20) and find the maximum value of $\Gamma$ (and therefore the maximum diffraction efficiency). The recording vectors $|\mathscr{C}^1(n)\rangle$ and $|\mathscr{C}^2(n)\rangle$ can be expanded in basis vectors of the input and the output spaces, $|u_n\rangle$ and $|v_n\rangle$. These expansions are

$$|\mathscr{C}^1(n)\rangle = \sum_{n'}^{R} s_{nn'} |u_{n'}\rangle,$$

$$|\mathscr{C}^2(n)\rangle = \sum_{n'}^{R} r_{nn'} |v_{n'}\rangle.\qquad(23)$$

The problem of selecting $\mathscr{C}^1_{n_1 n_2}(t)$ and $\mathscr{C}^2_{n_1 n_2}(t)$ corresponds to selecting $s_{nn'}$ and $r_{nn'}$ such that $\overline{W}$ is recorded with maximal $\Gamma$. Substituting Eqs. (23) into Eq. (22) yields

$$\overline{W} = A_0 \sum_{n=1}^{N_e} \sum_{n'}^{R} \sum_{n''}^{R} \frac{t_n s_{nn'} r_{nn''}^* |u_{n'}\rangle \langle v_{n''}^*|}{\sum_{m}^{N_e} \sum_{m'}^{R} t_m [(|r_{mm'}|^2 + |s_{mm'}|^2)]}.\qquad(24)$$

Operating on Eq. (20) with the basis vectors $|u_p\rangle$ and $|v_p\rangle$ yields

$$\langle u_p | \overline{W} | v_p \rangle = \Gamma \beta_p.\qquad(25)$$

Operating on Eq. (24) with the basis vectors $|u_p\rangle$ and $|v_p\rangle$ yields

$$\langle u_p | \overline{W} | v_p \rangle = A_0 \sum_{n=1}^{N_e} \frac{t_n s_{np} r_{np}^*}{\sum_{m}^{N_e} \sum_{m'}^{R} t_m [(|r_{mm'}|^2 + |s_{mm'}|^2)]}.\qquad(26)$$

Equating Eqs. (25) and (26), we find that

$$\beta_p \Gamma = A_0 \sum_{n=1}^{N_e} \frac{t_n s_{np} r_{np}^*}{\sum_{m}^{N_e} \sum_{m'}^{R} t_m [(|r_{mm'}|^2 + |s_{mm'}|^2)]}.\qquad(27)$$

Sum over $p$, and Eq. (27) becomes

$$\Gamma = A_0 \frac{\sum_{m=1}^{N_e} \sum_{m'=1}^{R} t_m s_{mm'} r_{mm'}^*}{\sum_{m}^{N_e} \sum_{m'}^{R} t_m [(|r_{mm'}|^2 + |s_{mm'}|^2)]}.\qquad(28)$$

We know by the triangle inequality that $\Gamma$ is maximal when $s_{mm'} = r_{mm'}$. A solution for suitable recording vectors is $r_{mm'} = s_{mm'} = \delta_{mm'}$, where $\delta_{mm'}$ is the Kronecker delta function. Assuming that $r_{mm'} = s_{mm'}$,

$$\Gamma = \frac{A_0}{2}.\qquad(29)$$

Thus, when the hologram is recorded with maximal $\Gamma$,

$$\overline{W} = \frac{A_0}{2} \sum_{n=1}^{R'} \beta_n |u_n\rangle \langle v_n|,\qquad(30)$$

where we have assumed that $t_n$ is proportional to $\beta_n$.

The diffraction efficiency, i.e., the ratio of the diffracted intensity to the undiffracted intensity, when a pattern $|\psi\rangle$

on the input plane reconstructs the hologram is

$$\eta_\psi = \frac{\langle\psi|\overline{\mathbf{W}}^T\overline{\mathbf{W}}|\psi\rangle}{\langle\psi|\psi\rangle} = \Gamma^2 \sum_n \beta_n^2 \frac{|\langle\psi|v_n\rangle|^2}{\langle\psi|\psi\rangle}.$$  (31)

Assuming that $|\psi\rangle$ has no component in the null space of $\overline{\mathbf{W}}$, $|\psi\rangle$ can be decomposed as $|\psi\rangle = \Sigma_n^R s_n|v_n\rangle$ so that

$$\eta_\psi = \frac{A_0^2}{4} \frac{\sum_n^R \beta_n^2|s_n|^2}{\sum^R |s_n|^2}.$$  (32)

$\eta_\psi$ is maximal when $s_m = 1$, where $\beta_m$ is the largest singular value of $\overline{\mathbf{W}}$. Our purpose in using a volume rather than a thin medium is to store as many different associations as possible. We are not typically interested in holograms in which one singular value is much larger than the others. Assuming that all $\beta_n$ are approximately equal and applying the normalization condition that $\Sigma_n\beta_n = 1$, we find that when $|v_n\rangle$ is active on the input plane the diffraction efficiency is

$$\eta_n = \frac{A_0^2}{4}\beta_n^2 \approx \frac{A_0^2}{4R^2}.$$  (33)

To summarize, we have found that it is possible to store arbitrary matrices $\overline{\mathbf{W}}$ in materials in which the response is described by Eq. (22) but that the diffraction efficiency of the corresponding hologram scales inversely with the square of the rank. We now relate Eq. (22) to the holographic response of real materials.

## C. Irreversible Materials
In irreversible materials the perturbation is linearly proportional to the exposing energy density up to some saturation exposure $E_0$. This response is described by

$$\frac{dA(\mathbf{r}, t)}{dt} = \begin{cases} A_0/E_0|\mathbf{E}_1(\mathbf{r}, t) + \mathbf{E}_2(\mathbf{r}, t)|^2 & \text{for } |A| < A_0 \\ 0 & \text{otherwise} \end{cases}.$$  (34)

If we form the hologram with discrete exposures,

$$A(\mathbf{r}, N_e) = \frac{A_0}{E_0} \sum_{i=1}^{N_e} t_i|\mathbf{E}_1(\mathbf{r}, i) + \mathbf{E}_2(\mathbf{r}, i)|^2,$$  (35)

where $t_i$ is the exposure time for the $i$th exposure and we require that

$$Q = \frac{\sum_{i=1}^{N_e} t_i|\mathbf{E}_1(\mathbf{r}, i) + \mathbf{E}_2(\mathbf{r}, i)|^2}{E_0} \leq 1$$  (36)

for most $\mathbf{r}$. Assuming that the phases of the various Fourier components are uncorrelated, as will tend to be the case if the new information content in each exposure is high, the mean and the variance of $Q$ can both be taken to be equal to

$$\langle Q \rangle = \frac{\sum_{i=1}^{N_e} t_i[\langle\mathscr{E}^1(i)|\mathscr{E}^1(i)\rangle + \langle\mathscr{E}^2(i)|\mathscr{E}^2(i)\rangle]}{E_0}.$$  (37)

We can ensure that $A(\mathbf{r}, N_e)$ does not approach saturation

at almost all values of $r$ by making $t_i$ small enough that $\langle Q \rangle$ does not approach 1. Let $\tau_i$ be such that

$$\sum_{i=1}^{N_e} \tau_i = 1,$$  (38)

and let $t_i = t_0\tau_i$. $\langle Q \rangle$ is less than 1 if

$$t_0 < \frac{E_0}{\sum_{i=1}^{N_e} \tau_i[\langle\mathscr{E}^1(i)|\mathscr{E}^1(i)\rangle + \langle\mathscr{E}^2(i)|\mathscr{E}^2(i)\rangle]}.$$  (39)

Using this inequality and Eq. (35), we find that

$$A(\mathbf{r}, N_e) \leq A_0 \frac{\sum_{i=1}^{N_e} \tau_i|\mathbf{E}_1(\mathbf{r}, i) + \mathbf{E}_2(\mathbf{r}, i)|^2}{\sum_{i=1}^{N_e} \tau_i[\langle\mathscr{E}^1(i)|\mathscr{E}^1(i)\rangle + \langle\mathscr{E}^2(i)|\mathscr{E}^2(i)\rangle]}.$$  (40)

Transforming relation (40) to look at the Fourier components described by $\overline{\mathbf{W}}$, we find that the hologram is described by Eq. (22). This means that the perturbation in a hologram recorded in optimal fashion in an irreversible saturable material is proportional to the modulation depth of the recording energy and therefore that the results derived above [e.g., Eq. (30)] apply.

## D. Reversible Materials
In an irreversible material different exposures affect one another only in the sense that they must share the same overall dynamic range. The situation is more complex for reversible materials, in which each new exposure tends to erase previous ones. Photorefractive crystals are the most fully characterized example in this class. Ignoring for the moment beam-coupling effects and scattering, we see that a band transport analysis of photorefractive hologram formation yields an equation describing the recording of the Fourier components $A_n$ of $A(\mathbf{r}, t)$.[17,24] Since the effect is based on charge transport, the photorefractive effect is nonlocal. The Fourier-domain dynamic equation for this case is

$$\frac{dA_n}{dt} = \frac{A_0}{\alpha}\mathscr{E}_{n_1, n_2}^1(t)\mathscr{E}_{n_1', n_2'}^2{}^*(t) - \frac{A_n}{\tau},$$  (41)

where $\alpha$ and $A_0$ are complex-valued constants and

$$\tau(t) = \frac{\alpha}{[\langle\mathscr{E}^1(t)|\mathscr{E}^1(t)\rangle + \langle\mathscr{E}^2(t)|\mathscr{E}^2(t)\rangle]}.$$  (42)

[Equation (41) is based on an assumption of low modulation depth and no wave mixing.] In matrix form Eq. (41) becomes

$$\frac{d\overline{\mathbf{W}}}{dt} = \frac{A_0}{\alpha}|\mathscr{E}^1(t)\rangle\langle\mathscr{E}^2(t)| - \frac{\overline{\mathbf{W}}(t)}{\tau(t)}.$$  (43)

Assuming discrete exposures, the solution to Eq. (43) is

$$\overline{\mathbf{W}}(N_e) = A_0 \sum_{n=1}^{N_e} \frac{\exp\left[-\sum_{n'>n} t_{n'}/\tau(n')\right]\{1 - \exp[-t_n/\tau(n)]\}}{[\langle\mathscr{E}^1(n)|\mathscr{E}^1(n)\rangle + \langle\mathscr{E}^2(n)|\mathscr{E}^2(n)\rangle]}$$

$$\times |\mathscr{E}^1(n)\rangle\langle\mathscr{E}^2(n)|.$$  (44)

To record a given matrix we select $t_n/\tau(n)$ such that

$$\overline{\mathbf{W}}(R) = A_0 \gamma \sum_{n=1}^{R} \lambda_n |\mathscr{E}^1(n)\rangle \langle \mathscr{E}^2(n)|, \tag{45}$$

where $\lambda_n$ is a real-valued weighting factor and $\gamma$ is a positive scaling constant. $\lambda_n$ are selected to be consistent with Eq. (20). We use $\gamma$ to characterize the amplitude of $\overline{\mathbf{W}}$. If we assume that $\alpha$ is real and that

$$\frac{t_1}{\tau(1)} = -\log(1 - \chi), \tag{46}$$

where $\chi$ is a real constant between 0 and 1, then the appropriate exposure time for $n > 1$ is[5,14,25]

$$\frac{t_n}{\tau(n)} = \log\left(\frac{\lambda_1 I_1 + \chi \sum_{i=2}^{n} \lambda_i I_i}{\lambda_1 I_1 + \chi \sum_{i=2}^{n-1} \lambda_i I_i}\right), \tag{47}$$

where $I_i = [\langle \mathscr{E}^1(i) | \mathscr{E}^1(i)\rangle + \langle \mathscr{E}^2(i) | \mathscr{E}^2(i)\rangle]$. Substituting Eq. (47) into Eq. (44) and comparing Eq. (44) with Eq. (45), we find that

$$\gamma = \frac{\chi}{\lambda_1 I_1 + \chi \sum_{i=2}^{N_e} \lambda_i I_i} \tag{48}$$

for this exposure schedule. In the optimal case, $\chi = 1$, and Eq. (44) reduces to Eq. (22).

While no closed-form solution for $t_n/\tau(n)$ exists when $\alpha$ is complex, it is still possible to develop a discrete exposure procedure that records the desired matrix. Let

$$f_n = \frac{\exp(j\xi_n)\exp\left[-\sum_{n'>n} t_{n'}/\tau(n')\right]\{1 - \exp[-t_n/\tau(n)]\}}{[\langle \mathscr{E}^1(n) | \mathscr{E}^1(n)\rangle + \langle \mathscr{E}^2(n) | \mathscr{E}^2(n)\rangle]}. \tag{49}$$

$\xi_n$ is the global phase difference between $\mathbf{E}_1$ and $\mathbf{E}_2$ for the $n$th exposure. With this notation Eq. (44) becomes

$$\overline{\mathbf{W}} = A_0 \sum_{n=1}^{N_e} f_n |\mathscr{E}^1(n)\rangle \langle \mathscr{E}^2(n)|. \tag{50}$$

To store the desired matrix we must find $\xi_n$ and $t_n$ such that Eq. (50) is equal to Eq. (45) or, equivalently, $f_n = \gamma\lambda_n$. As above, $\gamma$ characterizes the amplitude of the recorded matrix. Given arbitrary $t_1$ and $\xi_1$, $t_n$ and $\xi_n$ may be discovered by using a recursive process based on the requirement that

$$\frac{f_{n+1}}{f_n} = \frac{\lambda_{n+1}}{\lambda_n}. \tag{51}$$

Separating the real and imaginary parts of Eq. (51) yields two transcendental equations that can be solved numerically for $t_{n+1}$ and $\xi_{n+1}$.

The imaginary component of $\alpha$, $\mathfrak{I}(\alpha)$, affects the amplitude of $f_n$ through $\tau(n)$. A finite value for $\mathfrak{I}(\alpha)$ increases the rate at which $|1 - \exp[-t_n/\tau(n)]|$ grows. In the expo-

nential factors of $f_n$ in $\tau(n' \neq n)$, $\mathfrak{I}(\alpha)$ results only in a phase shift. (The purpose of $\xi_n$ is to negate this phase shift.) Since $\mathfrak{I}(\alpha)$ increases the rate at which the perturbation grows but does not change the rate at which it decays, we expect that a finite-valued $\mathfrak{I}(\alpha)$ will yield a greater value in $\gamma$ than that obtained for real $\alpha$. To evaluate the effect of a complex-valued $\alpha$, we estimate $t_n$ for large $n$. Numerical results provide some guidance. Figure 4 is a plot of $\mathfrak{R}[t_n/\tau(n)]$ versus $n$ for various values of $\mathfrak{I}(\alpha)/\mathfrak{R}(\alpha)$. $\lambda_n$ is assumed to be a constant for the cases shown. As is shown in the figure, all the curves tend to approach asymptotically the curve found for the case $\mathfrak{I}(\alpha) = 0$. The asymptotic result becomes approximately valid when $\mathfrak{I}[t_n/\tau(n)] \ll 1$. As we saw in Eq. (47), the solution for $t_n$ when $\mathfrak{I}(\alpha) = 0$ is

$$\mathfrak{R}\left[\frac{t_n}{\tau(n)}\right] = \log\left(\frac{\lambda_1 I_1 + \chi \sum_{i=2}^{n} \lambda_i I_i}{\lambda_1 I_1 + \chi \sum_{i=2}^{n-1} \lambda_i I_i}\right). \tag{52}$$

For these values for $t_n$,

$$\sum_{n'=n+1}^{N_e} \mathfrak{R}\left[\frac{t_{n'}}{\tau(n')}\right] = \log\left(\frac{\lambda_1 I_1 + \chi \sum_{i=2}^{N_e} \lambda_i I_i}{\lambda_1 I_1 + \chi \sum_{i=2}^{n} \lambda_i I_i}\right). \tag{53}$$

Substituting these values for $t_n$ into Eq. (49), we find that

$$f_n = \frac{\exp(j\xi_n)}{I_n} \exp\left\{-j \sum_{n'>n} \mathfrak{I}\left[\frac{t_{n'}}{\tau(n')}\right]\right\}$$
$$\times \frac{\left(\lambda_1 I_1 + \chi \sum_{i=2}^{n} \lambda_i I_i\right)\{1 - \exp[-t_n/\tau(n)]\}}{\lambda_1 I_i + \chi \sum_{i=2}^{N_e} \lambda_i I_i}. \tag{54}$$
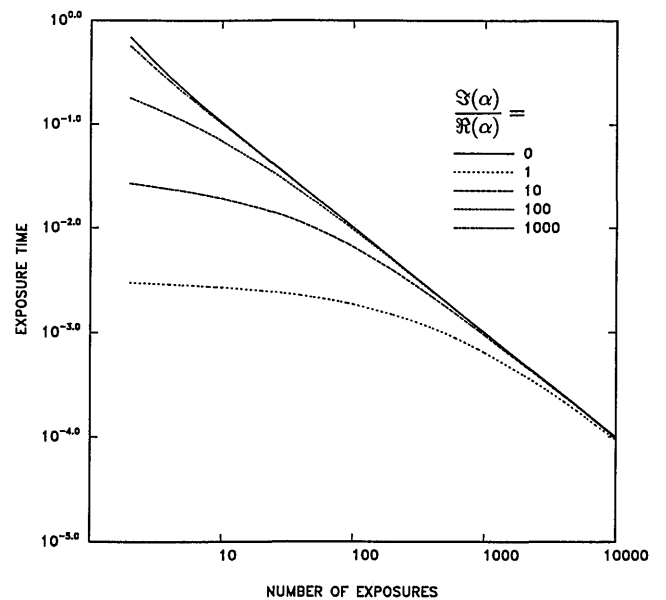


Fig. 4. Plot of $\mathfrak{R}[t_n/\tau(n)]$ versus exposure number. Plots are shown for values of $\mathfrak{I}(\alpha)/\mathfrak{R}(\alpha)$ from 0 to 1000. The data were calculated by using a numerical solution to Eq. (51).

Note that

$$\frac{t_n}{\tau(n)} = \exp\left\{-j\,\tan^{-1}\left[\frac{\mathfrak{F}(\alpha)}{\mathfrak{R}(\alpha)}\right]\right\}\left\{1 + \left[\frac{\mathfrak{F}(\alpha)}{\mathfrak{R}(\alpha)}\right]^2\right\}^{1/2}$$

$$\times \log\left|\frac{\lambda_1 I_1 + \chi\sum_{i=2}^{n}\lambda_i I_i}{\lambda_1 I_1 + \chi\sum_{i=2}^{n-1}\lambda_i I_i}\right| \tag{55}$$

and that, for $n \gg \mathfrak{F}(\alpha)/\mathfrak{R}(\alpha)$, $|t_n/\tau(n)| \ll 1$. In this case, $\{1 - \exp[-t_n/\tau(n)]\} \approx t_n/\tau(n)$, and

$$|f_n| \approx \left\{1 + \left[\frac{\mathfrak{F}(\alpha)}{\mathfrak{R}(\alpha)}\right]^2\right\}^{1/2}\frac{\chi\lambda_n}{\lambda_1 I_1 + \chi\sum_{i=2}^{N_e}\lambda_i I_i}, \tag{56}$$

$$\xi_n = \frac{\mathfrak{F}(\alpha)}{\mathfrak{R}(\alpha)}\log\left|\frac{\lambda_1 I_1 + \chi\sum_{i=2}^{n}\lambda_i I_i}{\lambda_1 I_1 + \chi\sum_{i=2}^{n-1}\lambda_i I_i}\right| + \tan^{-1}\left[\frac{\mathfrak{F}(\alpha)}{\mathfrak{R}(\alpha)}\right]. \tag{57}$$

Since $|f_n|$ is proportional to $\lambda_n$, Eq. (52) represents an asymptotic solution for $t_n$ for large $n$. This asymptotic solution allows us to calculate $\gamma$ for complex $\alpha$:

$$\gamma \approx \left\{1 + \left[\frac{\mathfrak{R}(\alpha)}{\mathfrak{F}(\alpha)}\right]^2\right\}^{1/2}\frac{\chi}{\lambda_1 I_1 + \chi\sum_{i=2}^{R}\lambda_i I_i}. \tag{58}$$

Comparing relation (58) with Eq. (48), we see that the effect of a complex-valued time constant is to increase the amplitude of the recorded hologram by the factor $\{1 + [\mathfrak{F}(\alpha)/\mathfrak{R}(\alpha)]^2\}^{1/2}$. Since both Eq. (48) and relation (58) scale with the modulation depth of the exposing energy, we see that the analysis leading to Eq. (30) applies to materials with an exponential response.

A complex $\alpha$ corresponds to an overdriven exponentially damped system. Damped harmonic responses of this sort have been observed in several photorefractive crystals. The effect is enhanced by applying a strong dc electric field across the crystal during recording. Under the assumption that such a field is applied, the amplitude of $\mathfrak{F}(\alpha)/\mathfrak{R}(\alpha)$ can be as large as

$$\frac{\mathfrak{F}(\alpha)}{\mathfrak{R}(\alpha)} = \frac{1}{2}\left|\left(\frac{\gamma_r \epsilon}{e\mu}\right)^{1/2} - \left(\frac{e\mu}{\gamma_r \epsilon}\right)^{1/2}\right|, \tag{59}$$

where $\gamma_r$ is the carrier–trap recombination rate, $\epsilon$ is the static permittivity, $e$ is the electron charge, and $\mu$ is the carrier mobility.[26] This ratio can be as large as 50–100 in highly polar materials such as $BaTiO_3$ and $Sr_{0.6}Ba_{0.4}Nb_2O_6$ (SBN). Experimental evidence of large values of $\mathfrak{F}(\alpha)/\mathfrak{R}(\alpha)$ is presented in Ref. 27. If $\mathfrak{F}(\alpha)/\mathfrak{R}(\alpha)$ approaches $R$, then it is possible to use the full dynamic range in a multiply exposed hologram.

### E.  Practical Implications

Experiments using the schedule of Eq. (47) to record sequences of exposures have been reported.[28–30] We have recorded more than 100 exposures in a SBN sample with this technique. In our experiments the recording intensity was 32 mW/cm² for all the exposures. The character-

istic response time $\tau$ at this intensity was 37.5 s. Each exposure recorded a single grating between plane-wave signal and reference beams. The same reference was used for all the exposures. New signal beams were generating by shifting the angle of incidence for the signal plane wave. The intensity of the reference beam was 10 times greater than the intensity of the signal beams, and the saturation diffraction efficiency for a single grating was approximately $10^{-1}$. The SBN sample was 2 mm thick. We continuously monitored the diffraction efficiency of all the recorded gratings, using the phase conjugate of the reference beam. The experimental apparatus was identical to the system described in Ref. 30. The diffracted signals were focused onto a CCD camera to yield a spot of light corresponding to each grating. The power in each spot was proportional to the diffraction efficiency of the corresponding grating. Figure 5 is a photograph of the video signal from the CCD when 111 gratings were recorded. The variance in diffracted power is ~20% of the mean for these gratings. Figure 6 is a log–log plot of the relative diffraction efficiency per grating versus the number of exposures for the sequence of exposures shown in Fig. 5. The slope of the least-mean-squares fit to the data points in Fig. 6 is −2.2. In these experiments all the exposure times were calculated a priori from Eq. (47), and no attempt was made to adjust exposure times to ensure uniformity. Given the uncertainty in our knowledge of $\tau$, more uniform results can be obtained by using monitoring techniques to adjust exposure times. This approach was taken by Mok et al., who recorded more than 1000 expo-
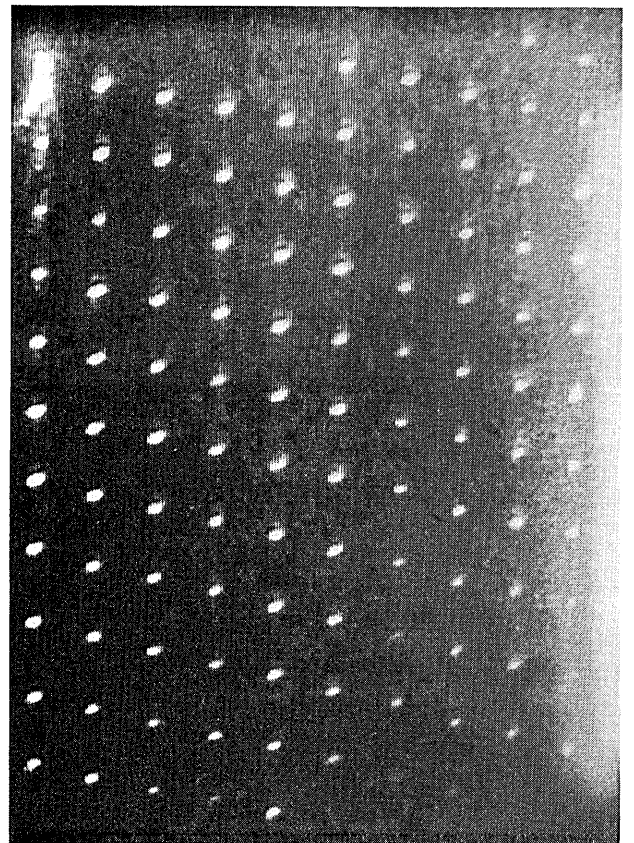


Fig. 5.  Photograph of light diffracted from each of 111 gratings recorded in sequence. Each spot corresponds to the signal diffracted by a single grating.
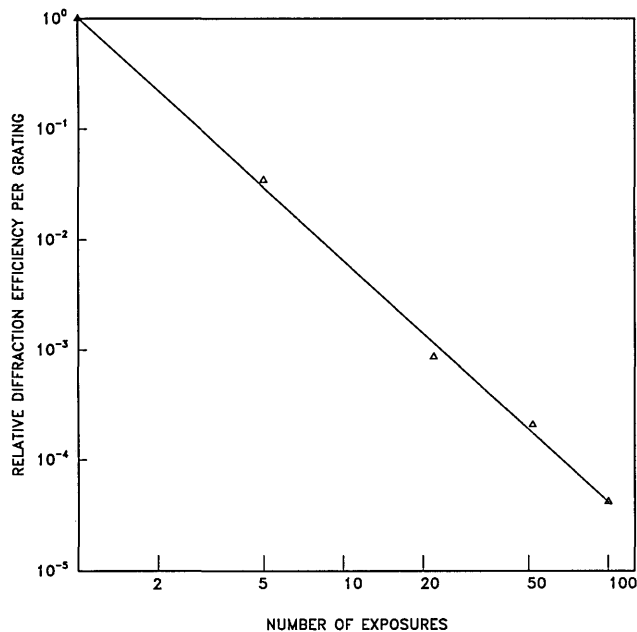
Fig. 6. Relative diffraction efficiency per grating versus the number of exposures for the data of Fig. 5.

sures in $LiNbO_3$ with impressive uniformity.[28,29] The basic falloff of the mean diffraction efficiency with the square of the number of exposures is also confirmed in the results of Mok et al.

At 1-$\mu$m resolution, $N_1$ is $10^4$/cm of aperture. Since the maximum rank is approximately $N_1$, a 1-$cm^3$ hologram formed of ~10,000 exposures would be nearing its maximum information capacity. Even without taking advantage of complex time constants, the state of the art is not far off this pace. By using materials with greater sensitivity than $LiNbO_3$, one may expect 10,000-exposure holograms to be created soon. If $\eta_{min}$ is the minimal acceptable diffraction efficiency when $\beta_n = 1/R$, then the number of exposures that can be recorded and detected is $A_0/\sqrt{\eta_{min}}$, or $\{1 + [\Re(\alpha)/\Im(\alpha)]^2\}^{1/2}A_0/2\sqrt{\eta_{min}}$. Under the assumption that $\eta_{min} \approx 10^{-4}$, $A_0/2\eta_{min}$ has been shown to reach $10^3$ in $BaTiO_3$.[31] Even more promising results have been obtained in photo-organics.[32] Further increases in the dynamic range available in multiply exposed holograms by enhancing complex recording time constants are relatively unexplored but may be expected to lift the number of reasonable exposures above the limit of 10,000.

Several of the assumptions that we have made have important consequences. The principal assumptions were as follows:

1. That the control fields are limited to a single matched pair of nondegenerate fractal grids during the entire exposure sequence. This assumption limits the generality of our results because cases exist in which the rank of the recorded hologram may be reduced either by using unrestricted 2-D images or by changing the dimension of the fractal grids from one exposure to the next. We make the assumption in spite of this because practical factors, especially the difficulty of inverting Eq. (18) to find appropriate control fields and the difficulty of maintaining sufficient phase coherence between different con-

trol pixels that simultaneously address the same grating, lead us to believe that fractal control is necessary in realistic systems when $R$ is large. (There are some cases in which transformations between degenerate fractal grids are desirable, but in most such cases the relative phases of degenerate recording beams would not be controlled, and the rank would therefore not be reduced.)

2. That cross gratings between different plane-wave components of $E_1$ or $E_2$ can be ignored. In some materials, such as diffusion-driven photorefractive crystals, the significance of cross gratings is limited by the fact that the response of the material is weak at low spatial frequencies. In the more general case this assumption tends to be valid because cross gratings build incoherently from one exposure to the next. Fields (eigenvectors) that diffract strongly off $\overline{W}$ will not tend to have high diffraction efficiencies from the cross gratings. A problem may arise if there is gain for diffraction off cross gratings, as in materials with strong scattering. The best hope for these materials is probably to use adaptive feedback to suppress scattering, leading to an effective increase in the rank.

3. That wave-mixing effects can be ignored. This assumption may be approached in several ways. First, although strong wave-mixing effects are often observed, they may be suppressed by controlling polarization, spatial stability, or diffraction efficiency. In multiply exposed holograms, in which the diffraction efficiency for any one exposure is small, the wave-mixing effects are often not strong. Second, wave mixing may sometimes be accounted for without fundamental changes in the exposing process. For example, wave mixing sometimes produces changes in the structure of the recording fields but not in the rank or the exposure schedule. This occurs when the saturation dynamics of hologram formation remain essentially exponential.[33] The third and most challenging approach is to use wave-mixing effects to advance holographic control. While it cannot reduce the fundamental rank of a hologram, wave mixing might provide a means for increasing the effective dynamic range in high-rank holograms. For example, it might be possible to use beams generated internally to sustain previously recorded information as new information is stored. Fully analyzing this approach is exceedingly difficult. While the results presented in this section show that if nonlinear effects are suppressed a volume hologram is a controllable system, these results do not preclude the possibility that accentuating nonlinearities might lead to better control techniques.

## 3. ADAPTIVE HOLOGRAMS

In this section we consider the effect of adaptive recording schemes on holographic control and discuss copying techniques for adaptive holograms. In adaptive holography $\overline{W}$ and $N_e$ are not known a priori. New exposing patterns are selected by a training algorithm based on the current states of $\overline{W}$ and the input and the output fields. In some applications, such as the recognition of patterns drawn from a fixed set, the training algorithm converges in a finite number of exposures to a permanent state. In other applications, such as the adaptive control of dynamic systems, the training algorithm requires an indefinite num-

ber of exposures. In either case, the fact that $N_e$ is not known in advance renders irreversible holographic media unsuitable because, as we saw in expression (39), we must know $N_e$ to calculate the optimal exposure times. We do not need to know $N_e$ in advance to calculate exposure times for reversible media, however, as we saw in Eq. (47). We assume for the rest of this section that we are dealing with reversible media.

Let $\Gamma$, as defined in Eq. (28), represent the amplitude of the corresponding hologram. The effect of many adaptive exposures on $\Gamma$ depends on the details of the training algorithm implemented. There are, however, a few points that can be made without considering the details. First, even if $N_e \gg R$ there exist cases in which $\Gamma$ does not scale inversely with $N_e$ or $R$. A trivial example of this occurs when the sequence of exposures involves repetitive cycling through the basis vectors, i.e., when $|\mathscr{E}^1(n)\rangle = |u_{n'}\rangle$ and $|\mathscr{E}^2(n)\rangle = |v_{n'}\rangle$, where $n' = (n \bmod R) + 1$. In this case $\Gamma$ is again $A_0/2$. The importance of this example is that it shows that the amplitude of the recorded hologram can remain finite even in holograms recorded with large numbers of exposures. The perceptron is an example of a learning algorithm for which one might expect this sort of behavior.[5,34–36] In perceptron learning, associated pairs of input–output training vectors are assumed to represent or to be close to pseudoeigenvectors. A second general comment is that there are cases in which $\Gamma$ does scale inversely with $N_e$ even when $N_e > R$. Calculating the effect of many exposures on $\Gamma$ is difficult because all the exposures control the same set of gratings (Fourier components $A_n$). The growth of the amplitude of a given grating depends on whether the contributions to that grating from one exposure to the next add in phase. By unfortunate selection of a training algorithm it is possible to record $\overline{\mathbf{W}}$ with arbitrarily small $\Gamma$. One example that allows us to isolate the effects that are due to incoherent addition of grating amplitudes and those that are due to unlucky selection of recording vectors occurs if we assume that each exposure records exactly one grating. The exposing fields for this case consist of one plane wave from the input field and one plane wave from the training field. This approach allows us to specify independently the relative value of each component in $\overline{\mathbf{W}}$. Let the eigenvectors $|\varphi_n\rangle$ and $|v_n\rangle$ represent the plane-wave components of the output and the input fields. The desired amplitude for the $ij$th component of $\overline{\mathbf{W}}$ is

$$w_{ij} = \left| \sum_n \beta_n \langle \varphi_i | u_n \rangle \langle v_n | v_j \rangle \right|.  \qquad (60)$$

If we record the components $w_{ij}$ sequentially, then Eq. (22) yields

$$\overline{\mathbf{W}} = \frac{A_0 \sum\limits_{n=1}^{D_0} \sum\limits_{n'=1}^{D_i} w_{nn'} |\varphi_n\rangle \langle v_{n'}|}{\sum\limits_{m=1}^{D_0} \sum\limits_{m'=1}^{D_i} 2w_{mm'}},  \qquad (61)$$

where we assume that the relative phases of $|\varphi_n\rangle$ and $\langle v_{n'}|$ are appropriately adjusted for each exposure. For this matrix,

$$\Gamma = \frac{A_0}{\sum\limits_{m=1}^{D_0} \sum\limits_{m'=1}^{D_i} 2w_{mm'}}$$

$$= \frac{A_0}{2} \left[ \sum_{m=1}^{D_0} \sum_{m'=1}^{D_i} \sum_{n=1}^{r} \beta_n |\langle \varphi_i | u_n \rangle| |\langle v_n | v_j \rangle| \right]^{-1}.  \qquad (62)$$

Using the identities

$$\sum_{n'=1}^{D_0} |\varphi_{n'}\rangle \langle \varphi_{n'}| = \mathbf{I},  \qquad (63)$$

$$\sum_{n'=1}^{D_i} |v_{n'}\rangle \langle v_{n'}| = \mathbf{I},  \qquad (64)$$

we can show that

$$\sum_{n'}^{D_0} \langle u_n | \varphi_{n'} \rangle \langle \varphi_{n'} | u_n \rangle = 1,$$

$$\sum_{n'}^{D_i} \langle v_n | v_{n'} \rangle \langle v_{n'} | v_n \rangle = 1  \qquad (65)$$

for all $n$. In view of this, we estimate $|\langle \varphi_i | u_n \rangle|$ and $|\langle v_n | v_j \rangle|$ as $1/\sqrt{D_0}$ and $1/\sqrt{D_i}$, respectively. Then

$$\Gamma \approx \frac{A_0}{2\sqrt{D_0 D_i}} = \frac{A_0}{2\sqrt{N_e}}.  \qquad (66)$$

This example shows that when completely uncorrelated exposures are recorded the dynamic range can be reduced by a factor as large as $N_e^{-1/2}$. This factor is in addition to the $R^{-1}$ factor from the normalization of $\beta_n$. This phenomenon was demonstrated for rank 1 holograms in Ref. 30, where a periodic refreshing technique to eliminate the $N_e^{-1/2}$ factor is described.

One can compensate for decreases in $\Gamma$ caused by adaptive recording by periodically refreshing the hologram. Three techniques are of interest. First, when the current value of $\Gamma$ becomes unacceptably small we might interrupt the adaptive recording process, probe the hologram to determine $\overline{\mathbf{W}}$, decompose $\overline{\mathbf{W}}$ to determine the optimal recording sequence, and copy $\overline{\mathbf{W}}$ back onto the original material with maximal $\Gamma$. While this approach permits adaptive recording with maximal dynamic range, it is too tedious to be of use in continuously adapting systems. A second approach is to use a fixed simple algorithm to restore the dynamic range of $\overline{\mathbf{W}}$ periodically. In this case no care is taken to determine the optimal exposure sequence for refreshing $\overline{\mathbf{W}}$, thus greatly reducing the time and computation required per recording cycle. As an example of this approach, we consider periodic copying of holograms from one volume to another and back again. A sketch of a system for doing this is shown in Fig. 7. In the simplest case the adaptive hologram, H1, is probed sequentially by $N_1$ different plane waves. For fractal patterns with $N_1$ inputs and $N_2$ outputs, these plane waves represent the basis vectors for the input space $|v_n\rangle$. The plane wave $|v_p\rangle$ is diffracted into a 2-D output image that is the corresponding basis vector for the output space $|u_p\rangle$. The matrix that describes the recorded hologram is

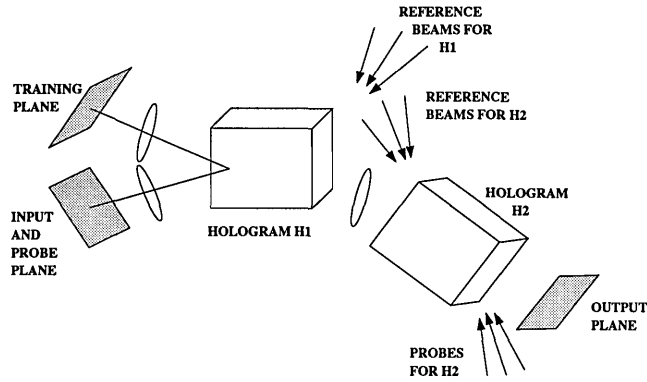$$\overline{\mathbf{W}}_0 = \Gamma_0 \sum_{n=1}^{R} \beta_n |u_n\rangle \langle v_n|.  \qquad (67)$$

Fig. 7. System for periodically refreshing adaptive volume holograms.

We would like to calculate how the amplitude of the recorded hologram may be increased by copying $\overline{\mathbf{W}}_0$ from H1 to hologram H2 and then H2 back onto H1. We call the value of $\Gamma$ for the hologram that is eventually formed in H1 $\Gamma_c$. Let the amplitude of the probing field be $C$. Diffracting the plane wave $C|\nu_p\rangle$ off H1 yields a diffracted field

$$|\mathcal{E}_p^2\rangle = C\overline{\mathbf{W}}_0|\nu_p\rangle = C\Gamma_0\beta_p|u_p\rangle. \tag{68}$$

If we interfere $|\mathcal{E}_p^2\rangle$ with $|\nu_p\rangle$ at the second hologram and use an exposure schedule that weights all $N_1$ exposures equally, then the hologram recorded in H2 after all $N_1$ plane waves have probed H1 is described by

$$\overline{\mathbf{W}}_c = \frac{CA_0\Gamma_0 \sum\limits_{n=1}^{R} \beta_n|u_n\rangle\langle\nu_n|}{N_1 + C^2\Gamma_0^2 \sum\limits_{n=1}^{R} \beta_n^2} = \Gamma_c \sum\limits_{n=1}^{R} \beta_n|u_n\rangle\langle\nu_n|. \tag{69}$$

The maximum value for $\Gamma_c$ is

$$\Gamma_c = \frac{A_0}{2\left(N_1 \sum\limits_{n=1}^{N_e} \beta_n^2\right)^{1/2}}, \tag{70}$$

which occurs when

$$C = \frac{1}{\Gamma_0}\left(\frac{N_1}{\sum\limits_{n=1}^{N_e} \beta_n^2}\right)^{1/2}. \tag{71}$$

Note that $\Gamma_c$ is independent of $\Gamma_0$. Reversing our method to copy back from H2 to H1 yields a refreshed hologram described by the same $\overline{\mathbf{W}}_c$ in H1. The maximum amplitude of $\overline{\mathbf{W}}_c$ is again $\Gamma_c$. The minimum value of $\Gamma_c$ occurs when $\beta_n = 1$ for some $n$ (and $\beta_{n'} = 0$ for $n' \neq n$). This case, which corresponds to a rank 1 hologram that we copied in $N_1$ exposures, yields $\Gamma_c = A_0/2\sqrt{N_1}$. This result means that at worst we can limit the decay in $\Gamma$ that is due to adaptive recording to $1/\sqrt{N_1}$. For holograms of high rank the maximum of $\Gamma_c$ is more likely. The maximum occurs when $\Sigma_n\beta_n^2 \approx 1/R$, in which case

$$\Gamma_c \approx \frac{A_0}{2}\left(\frac{R}{N_1}\right)^{1/2}. \tag{72}$$

For holograms of rank $R = N_1$ the maximum value of $\Gamma$ is preserved indefinitely by this periodic refreshing technique. Of course, if fractal grids with $D_0$, $D_i > N_1$ are used for the input and the output fields, the rank for these grids may exceed $N_1$, and the copying technique may actually increase $\Gamma$ above the maximum that could be obtained with the training grids.

A major problem with the copying technique arises when one considers the decay of stored holograms during the probing process. Because the diffraction efficiency must be less than 1, the probe intensity incident upon H1 is greater than the intensity available at H2 to record the copied hologram. If the energy required to erase H1 is comparable with the energy required to record H2, then the asymmetry in the power available means that H1 will be erased faster than H2 will be recorded and it will not be possible to improve the amplitude of the hologram by copying. This problem might be overcome by using an asymmetry in the write–erase time constants, as occurs when the time constant is complex, by providing optical gain between H1 and H2 or by using fixing techniques, as was done previously for H2.[30] It might also be overcome by using real-time feedback between H1 and H2 to stabilize the hologram recorded in both materials. This approach is our third refreshment technique.

Figure 8 is a sketch of a system for stabilizing $\Gamma$ in an adaptively recorded hologram, using real-time feedback. The idea behind this system is to use nonlinear feedback through the hologram to find the eigenmodes of the resonator formed by phase-conjugate mirrors (PCM's) 2 at the input and 3 at the output.[37] These eigenmodes would correspond to the optimal recording vectors $|u_n\rangle$ and $|\nu_n\rangle$. For this system to work the PCM's must exhibit gain and mode competition so that one and only one associated pair $|u_n\rangle$ and $|\nu_n\rangle$ oscillates. There must also be a mechanism, such as seeding through the training planes, to force the system to hop from one mode to the next. If these conditions are met, the system will tend to cycle through $|u_n\rangle$ and $|\nu_n\rangle$ and continuously record $\overline{\mathbf{W}}$ optimally, independently of the training vectors. While real-time feedback systems are clearly interesting for holographic control, a detailed and realistic analysis of what would happen in this system when the actual dynamics and nonlinearities of the PCM's are accounted for is beyond the scope of this paper. In Section 4 we discuss a related feedback system that is both easier to analyze and better suited to the task of linear holographic control.
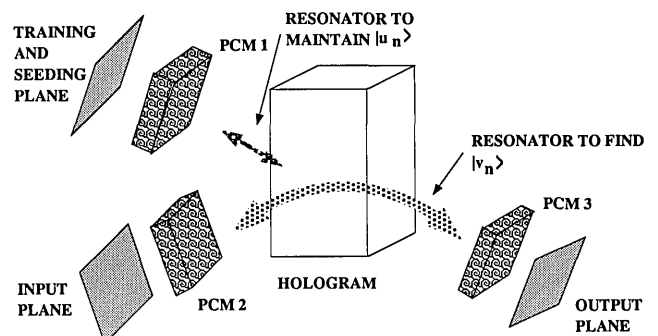


Fig. 8. System for maintaining adaptively recorded holograms, using continuous feedback.

## 4.  POLYCHROMATIC RECORDING METHODS

High-rank holograms may be controlled in a single exposure by using polychromatic light. Since the number of degrees of freedom in a single wavelength component of the polychromatic field cannot exceed $2N_1N_2$, at least $N_1/2$ distinct wavelengths are needed in order to control an arbitrary hologram. In this section we show that a hologram recorded in $N_e$ monochromatic exposures can be recorded in one exposure by using signals carried on $N_e$ independent wavelengths. Polychromatic control is potentially superior to monochromatic control because coherence between different wavelength components can yield greater modulation depth and thus better dynamic range in the perturbation.

Consider the system of Fig. 1. Under polychromatic control, the fields incident from SLM1 and SLM2 must have independent spectral content at each spatial frequency. Letting $\mathbf{E}_1(\mathbf{r}, t)$ and $\mathbf{E}_2(\mathbf{r}, t)$ represent these fields, as before, we represent the components of $\mathbf{E}_1$ and $\mathbf{E}_2$ at the $n$th wavelength as $\mathbf{E}_{1n}(\mathbf{r}, t)\exp(-j\omega_n t)$ and $\mathbf{E}_{2n}(\mathbf{r}, t)\exp(-j\omega_n t)$ so that

$$\mathbf{E}_1(\mathbf{r}, t) = \sum_{n=1}^{N_e} \mathbf{E}_{1n}(\mathbf{r}, t)\exp(-j\omega_n t),$$

$$\mathbf{E}_2(\mathbf{r}, t) = \sum_{n=2}^{N_e} \mathbf{E}_{2n}(\mathbf{r}, t)\exp(-j\omega_n t). \tag{73}$$

SLM1 and SLM2 must be replaced by devices capable of independently controlling all $N_1N_2$ degrees of freedom in each of the $N_e$ independent components of $\mathbf{E}_1$ and $\mathbf{E}_2$. The straightforward approach to controlling this large set of variables is to use $2N_e$ distinct SLM's, each illuminated by a different wavelength, to generate $\mathbf{E}_{1n}$ and $\mathbf{E}_{2n}$. Since we would like $N_e$ to be in the range $10^3-10^4$, however, this approach is not practical. A second approach would be to replace SLM1 and SLM2 with volume holograms H1 and H2. As we saw above, a volume hologram can store $N_1$ independent $N_1 \times N_2$ images. In the monochromatic case different images are reconstructed at different probe angles. Under polychromatic illumination different images can be reconstructed at the same incidence angle by different colors. A Fourier-space diagram of this process is shown in Fig. 9. As long as the separation between wavelengths is greater than $\approx \lambda^2/L_z$ the gratings that diffract different wavelengths are independent. We assume that H1 and H2 diffract probe-signal frequency $\omega_n$ into $\mathbf{E}_{1n}$ and $\mathbf{E}_{2n}$. Assuming that H1 and H2 can be constructed, the field in the holographic volume when the stored images are simultaneously reconstructed by a polychromatic plane wave containing all $N_e$ recording wavelengths is $\mathbf{E}_1(\mathbf{r}, t) + \mathbf{E}_2(\mathbf{r}, t)$. Because all the degrees of freedom of a hologram can be controlled in a single exposure under polychromatic control, there is no need to develop exposure schedules, and the distinction between reversible and irreversible materials outlined in Section 2 is unimportant. For both types of material the perturbation is proportional to the exposing intensity. We model the perturbation as

$$A(\mathbf{r}, T) = \frac{A_0}{E_0} \int_0^T |\mathbf{E}_1(\mathbf{r}, t) + \mathbf{E}_2(\mathbf{r}, t)|^2 \mathrm{d}t. \tag{74}$$

If $T \gg (\omega_n - \omega_{n'})^{-1}$ for all $n' \neq n$, then contributions to $A(\mathbf{r})$ owing to interference between fields at different wavelengths vanish, and

$$A(\mathbf{r}) = \frac{A_0 T}{E_0} \sum_{n=1}^{N_e} |\mathbf{E}_{1n}(\mathbf{r}) + \mathbf{E}_{2n}(\mathbf{r})|^2, \tag{75}$$

where we assume that the recording images are constant in time. This equation is the same as Eq. (35) except that $n$ now corresponds to the $n$th wavelength rather than to the $n$th exposure. Solving for $T$ allows us to show that $A(\mathbf{r})$ is described by relation (40) with the proviso that $\tau_n$ is constant. This means that hologram formation using $N_e$ independently controlled wavelengths with no contribution from interference between different wavelengths yields exactly the perturbation described by Eq. (22) and that, in particular, the dependence of the amplitude of the perturbation on the holographic rank is exactly the same as for the control using multiple exposures. No advantage in dynamic range is obtained for polychromatic control in this case.

In Sections 2 and 3 we used $\overline{\mathbf{W}}$ to characterize the perturbation $A(\mathbf{r})$. This matrix was associated with the fractal grids used to control it. Although we do not use fractal grids under polychromatic control, it is still useful to map the Fourier components $A_n$ onto a matrix $\overline{\mathbf{W}}$ to compare the amplitude of polychromatically controlled holograms with the amplitude of the same perturbation under monochromatic control. As above, the amplitude of $\overline{\mathbf{W}}$ is described by $\Gamma$. We now show that an increase in $\Gamma$ can be obtained if the integration time is short enough that interference can occur between different wavelengths. The simplest means for decreasing the effective integration time is to pulse the recording fields. We assume therefore that

$$\mathbf{E}_{1n}(\mathbf{r}, t) = f(t)\mathbf{E}_{1n}(\mathbf{r}),$$

$$\mathbf{E}_{2n}(\mathbf{r}, t) = f(t)\mathbf{E}_{2n}(\mathbf{r}), \tag{76}$$

where

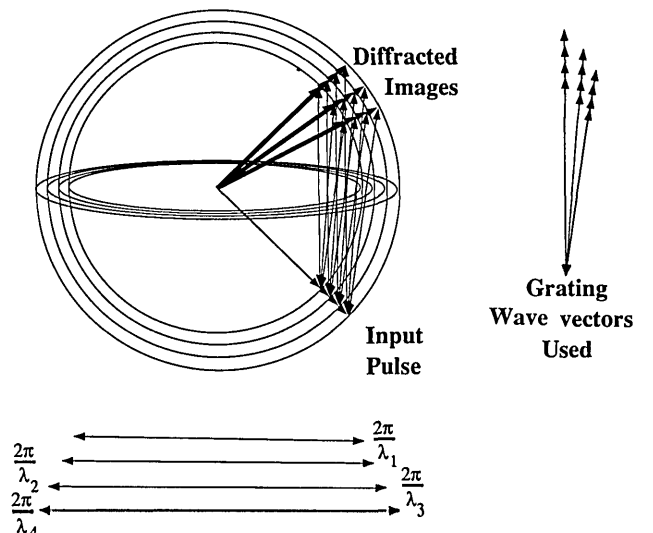$$f(t) = \sum_{n=1}^{\infty} p(t - nt_r) \tag{77}$$



Fig. 9. Reconstruction of multiple independent images, using a unidirectional polychromatic probe.

and $p(t)$ is a short pulsing function of width $t_\delta$. Repetitive pulsing avoids the need for high-power exposing fields. This pulsing differs from the monochromatic multiple exposures because the control fields do not change from one exposure to the next. Suppose that the frequencies of the recording fields are uniformly spaced such that

$$\omega_n = \omega_0 + n\Delta\omega. \tag{78}$$

Substituting Eq. (78) into Eq. (74), we find that

$$A(\mathbf{r},T) = \frac{A_0}{E_0} \sum_{n=1}^{N_e} \sum_{n'=1}^{N_e} [\mathbf{E}_{1n}(\mathbf{r}) + \mathbf{E}_{2n}(\mathbf{r})]$$
$$\cdot [\mathbf{E}_{1n'}{}^*(\mathbf{r}) + \mathbf{E}_{2n'}{}^*(\mathbf{r})]$$
$$\times \int_0^T \exp[-j(n - n')\Delta\omega t]|f(t)|^2 \mathrm{d}t. \tag{79}$$

If $t_\delta > T \gg \Delta\omega^{-1}$, then $|f(t)|^2$ is constant and Eq. (79) reduces to Eq. (75). If, however, $N_e\Delta\omega t_\delta \ll 1$, then

$$A(\mathbf{r},T) = \frac{A_0 T t_\delta}{E_0 t_r} \left| \sum_{n=1}^{N_e} \mathbf{E}_{1n}(\mathbf{r}) + \mathbf{E}_{2n}(\mathbf{r}) \right|^2 . \tag{80}$$

We have implicitly assumed that $\omega_0 t_\delta > 1$ in using the time-averaged form of the optical intensity. To avoid saturation we pick $T$ so that the exposing energy does not exceed $E_0$, which is equivalent to requiring that $|A(\mathbf{r},T)| \le |A_0|$. Assuming that the $2N_e N_1 N_2$ control signals are uncorrelated, this implies that

$$T < \frac{E_0 t_r}{t_\delta} \frac{1}{\sum\limits_{n=1}^{N_e} \langle|\mathbf{E}_{1n}(\mathbf{r})|^2\rangle + \langle|\mathbf{E}_{2n}(\mathbf{r})|^2\rangle}, \tag{81}$$

where $\langle|\mathbf{E}_{1n}(\mathbf{r})|^2\rangle$ is the spatial average of $|\mathbf{E}_{1n}(\mathbf{r})|^2$. Using this value for $T$, we find that

$$A(\mathbf{r},T) = A_0 \frac{\left| \sum\limits_{n=1}^{N_e} \mathbf{E}_{1n}(\mathbf{r}) + \mathbf{E}_{2n}(\mathbf{r}) \right|^2}{\sum\limits_{n=1}^{N_e} \langle|\mathbf{E}_{1n}(\mathbf{r})|^2\rangle + \langle|\mathbf{E}_{2n}(\mathbf{r})|^2\rangle}. \tag{82}$$

Equation (82) is the analog of relation (40) for the cross-spectrally coherent polychromatic case.

Equation (82) describes a 3-D perturbation with amplitude proportional to modulation depth. The modulation depth is maximal when one of the control fields, say, $\mathbf{E}_1(\mathbf{r},t)$, consists of a single strong reference beam. We may assume, for example, that

$$\mathbf{E}_1(\mathbf{r},t) = \mathscr{E}_r f(t)\exp[j(\mathbf{k}_r \cdot \mathbf{r} - \omega_r t)]. \tag{83}$$

We then pick $\mathbf{E}_2(\mathbf{r},t)$ so that Eq. (82) yields the desired perturbation. It is possible to find an appropriate $\mathbf{E}_2(\mathbf{r},t)$ for recording the 3-D structure of $A(\mathbf{r})$ in this situation because we are using interference between fields with propagation vectors on different wave normal surfaces. A sketch of the wave-normal space for recording with a single reference is shown in Fig. 10. (The frequencies of pulsed plane waves are not well defined. Figure 10 shows the center wave vectors of the corresponding plane waves.) A grating with wave vector $\mathbf{K}_g$ is controlled by the interference pattern formed by the reference and the signal field separated from the reference by $\mathbf{K}_g$ in the 3-D control

space. Since the amplitude and the phase of the reference are fixed, the relative phase and amplitude of the grating correspond to the phase and the amplitude of the signal field. The signal fields in this case correspond to Fourier components of $\mathbf{E}_2(\mathbf{r},t)$. The Fourier decomposition of $\mathbf{E}_2(\mathbf{r},t)$ is

$$\mathbf{E}_2(\mathbf{r},t) = \sum_{n=1}^{N_e} \sum_{n_1=-N_1/2}^{N_1/2} \sum_{n_2=-N_2/2}^{N_2/2} \mathscr{E}_{n_1 n_2}^{2n} f(t)\exp(-j\omega_n t)$$
$$\times \exp(j\mathbf{k}_{n_1 n_2}^{2n} \cdot \mathbf{r}), \tag{84}$$

where, after Eq. (7),

$$\mathbf{k}_{n_1 n_2}^{2n} \approx \omega_n \sqrt{\mu_0 \epsilon}\{[\cos(\theta) - n_1\Delta_\theta \sin(\theta)]\hat{z}$$
$$+ [\sin(\theta) + n_1\Delta_\theta \cos(\theta)]\hat{x} + n_2\Delta_\phi \sin(\theta)\hat{y}\}. \tag{85}$$

Substituting relation (85) into Eq. (82) and ignoring cross gratings, we find that

$$A_{\mathbf{n}} = A_0 \frac{\mathscr{E}_r \cdot \mathscr{E}_{p_1 p_2}^{2p}{}^*}{|\mathscr{E}_r|^2 + \sum\limits_{n=1}^{N_e} \sum\limits_{n_1=-N_1/2}^{N_1/2} \sum\limits_{n_2=-N_2/2}^{N_2/2} |\mathscr{E}_{n_1 n_2}^{2n}|^2}, \tag{86}$$

where $p_1, p_2,$ and $p$ are such that

$$\mathbf{K}_{\mathbf{n}} = \mathbf{k}_r - \mathbf{k}_{p_1 p_2}^{2p}. \tag{87}$$

$|A_{\mathbf{n}}|$ is maximal when

$$|\mathscr{E}_r|^2 = \sum_{n=1}^{N_e} \sum_{n_1=-N_1/2}^{N_1/2} \sum_{n_2=-N_2/2}^{N_2/2} |\mathscr{E}_{n_1 n_2}^{2n}|^2, \tag{88}$$

in which case

$$|A_{\mathbf{n}}| = A_0 \frac{|\mathscr{E}_{p_1 p_2}^{2p}|}{2\left( \sum\limits_{n=1}^{N_e} \sum\limits_{n_1=-N_1/2}^{N_1/2} \sum\limits_{n_2=-N_2/2}^{N_2/2} |\mathscr{E}_{n_1 n_2}^{2n}|^2 \right)^{1/2}}. \tag{89}$$

The expectation value $\langle|A_{\mathbf{n}}|^2\rangle$, averaged over $\mathbf{n}$, is $A_0{}^2/4N_g$, where $N_g$ is the number of gratings that we wish to control in $A(\mathbf{r})$. $N_g$ is the number of control amplitudes $|\mathscr{E}_{n_1 n_2}^{2n}|$ that are nonzero. For a full-rank hologram, $N_g = N_e N_1 N_2$. Recall that $A_{\mathbf{n}}$ corresponds to a component, $w_{ij}$, of a transformation matrix for monochromatic fields, $\mathbf{W}$. To calculate a value of $\Gamma$ that allows us to compare polychromatic control with monochromatic control, we assume that $\mathbf{W}$ is
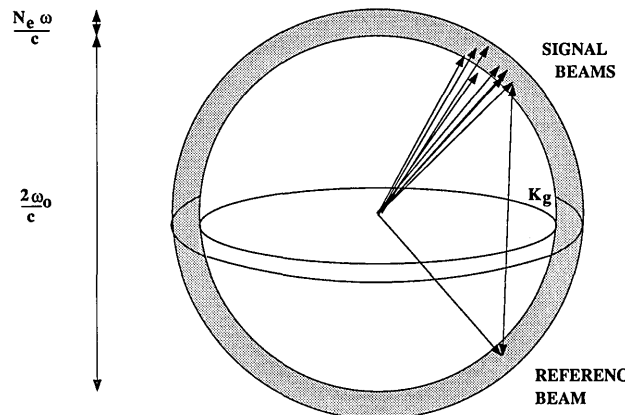


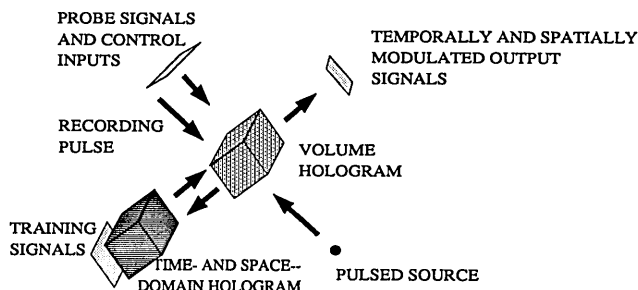Fig. 10. 3-D wave-normal shell for polychromatically controlled volume holograms.

Fig. 11.  System for periodically refreshing volume holograms, using polychromatic light.

the representation of $A(\mathbf{r})$ with minimal rank.  Using the plane-wave vectors $|\varphi_i\rangle$ and $|v_j\rangle$ defined above in Eq. (60) and the general form for $\overline{\mathbf{W}}$ given in Eq. (20), we find that

$$w_{ij} = \Gamma \sum_{n=1}^{R'} \beta_n \langle \varphi_i | u_n \rangle \langle v_n | v_j \rangle. \tag{90}$$

Squaring and summing over the input and the output spaces, we find that

$$\sum_{i=1}^{D_0} \sum_{j=1}^{D_i} |w_{ij}|^2 = \Gamma^2 \sum_{i=1}^{D_0} \sum_{j=1}^{D_i} \sum_{n'=1}^{R} \sum_{n=1}^{R} \beta_n \beta_{n'} \langle u_{n'} | \varphi_i \rangle \\ \times \langle \varphi_i | u_n \rangle \langle v_n | v_j \rangle \langle v_j | v_{n'} \rangle. \tag{91}$$

Applying Eqs. (63) and (64) yields

$$\langle |w_{ij}|^2 \rangle = \Gamma^2 \frac{\displaystyle\sum_{n=1}^{R} \beta_n^2}{D_i D_0}. \tag{92}$$

We now assume that $N_g$ is equal to the number of components in $\overline{\mathbf{W}}$, $D_i D_0$, and that $\sum_{n=1}^{R} \beta_n^2 \approx 1/R$.  Requiring that $\langle |w_{ij}|^2 \rangle = \langle |A_n|^2 \rangle$, we find that

$$\Gamma = \frac{A_0}{2} \sqrt{R} \tag{93}$$

for a cross-spectrally coherent, polychromatically controlled hologram.  Comparing Eq. (93) with Eq. (29), we see that the effect of polychromatic control is to increase the amplitude of the recorded hologram by $\sqrt{R}$.  The mean diffraction efficiency shows a corresponding increase by the factor $R$.

While strong advantages may be obtained through cross-spectrally coherent polychromatic control, the practical aspects of implementing it are daunting.  In assuming that $|\mathscr{E}_{n_1 n_2}^{2n}|$ and $|\mathscr{E}_{n_1' n_2'}^{2n'}|$ address distinct gratings, we have implicitly assumed that

$$|\mathbf{k}_{n_1 n_2}^{2n} - \mathbf{k}_{n_1' n_2'}^{2n'}| > \frac{\pi}{L}, \tag{94}$$

where we assume that $L_x = L_y = L_z = L$.  This constraint implies that

$$\Delta\omega > \frac{\pi\omega}{k_0 L}, \tag{95}$$

$$t_\delta \ll \frac{1}{N_e \Delta\omega} < \frac{L}{N_e c\pi}, \tag{96}$$

where $c$ is the speed of light in the holographic material.

Assuming that $N_e = 100$ and $L = 1$ cm, we find that

$$t_\delta \ll 10^{-13} \text{ s}. \tag{97}$$

One method of limiting $t_\delta$ to this range would be to use a holographic medium that is excited in a two-step process. If the hologram were sensitized by a short pulse during recording with the signals of Eq. (82), then the integration time could be suitably limited.  An alternative approach that is not material dependent is to assume that the recording signals themselves are riding on such a pulse.  An approach to generating such coded pulses is shown in Fig. 11, which is a variation on the real-time feedback system considered in Section 3.  In this system a volume hologram is recorded by using the sequential monochromatic exposures as derived in previous sections.  Periodically the volume is probed by a short pulse.  Since the pulse can simultaneously Bragg match all the gratings in the volume, the field into which it diffracts is 3-D and contains all the information stored in the volume.  By phase conjugating the diffracted field and the pulse, it is possible to copy the volume back onto itself in a single exposure.  The phase conjugation of the temporally and spatially modulated diffracted field would occur in a spectrally sensitive material, such as those used for time and space holography.[38]  While a number of challenges, such as the need to achieve gain in the phase-conjugated fields, lie in the path to demonstrating this technique, the potential rewards for generating high spatial and temporal bandwidth fields make the pursuit worthwhile.

## 5.  CONCLUSION

While the inverse scaling of diffraction efficiency with rank squared when the perturbation is inversely proportional to the total exposing energy applies to two important classes of hologram, those with saturable responses and those with exponential responses, this relationship does not ultimately dim the potential of volume holographic systems.  The dynamic range of holographic perturbations is ultimately limited by the number of active sites per voxel, which for doped materials is likely to be on the scale of $10^6$–$10^9$.  This limit is so far in excess of the $10^3$–$10^5$ range that one might expect for $R$ that the problem of overcoming the $R^{-2}$ loss in diffraction efficiency consists simply of finding a sufficiently sensitive material.  As mentioned at the end of Section 2, such materials are already available if one is willing to accept a value for $\eta_{min}$ of $10^{-4}$.  $\eta_{min}$ in this case is readily detectable but not particularly energy efficient.  However, there is no reason to expect that the materials cited in Section 2 are optimal.  In fact, the prospects for materials with improved sensitivity are, in our opinion, quite good.  In particular, techniques for enhancing the complex response time in certain photorefractive holograms are promising methods for increasing the effective sensitivity for high-rank holograms.

It is important to emphasize that the limit that we have derived on the dynamic range of multiply exposed holograms pertains to the rank of the hologram and not to the actual number of exposures used to record it.  Cases abound, particularly in adaptive holography, in which the number of exposures greatly exceeds the rank without overloading the dynamic range.  This is not to say that

such cases are the norm, however. As described in Section 3, several options exist for maintaining the dynamic range of adaptively recorded holograms such that arbitrarily many exposures may be recorded. These techniques are of particular importance when one considers that the one truly unique aspect of volume holograms is the ability of these holograms continuously and fundamentally to alter the data or interconnections that they store.

While high-speed holography has a fairly substantial following, interactions between polychromatic signals and volume holograms for the purpose of performing transformations between spatial and temporal spectra is a relatively unexplored area. In Section 4 of this paper we described a strong motivating factor for using polychromatic light for holographic control, namely, that the dynamic range of the controlled hologram might be increased by the factor $\sqrt{R}$. This is not, however, the only potential advantage to such systems. The temporal signals that would be created by the system shown in Fig. 11 may in fact be more interesting and useful than the volume hologram itself.

## APPENDIX A: NOTATION

The notation used in this paper is defined as follows:

$\alpha$, Materials- and geometry-specific factor in the characteristic response time for reversible holograms.

$\beta_n$, Normalized singular value for the $n$th constituent outer product of $\overline{\mathbf{W}}$.

$\Gamma, \Gamma_0, \Gamma_c$, Positive real scaling factors for $\overline{\mathbf{W}}$, $\overline{\mathbf{W}}_0$, and $\overline{\mathbf{W}}_c$ when these matrices are written in normalized singular-value decomposed form. $\Gamma$ is used to measure the amplitude of $\overline{\mathbf{W}}$.

$\gamma$, Positive real measure of the amplitude of a hologram recorded in a reversible medium.

$\gamma_r$, Carrier–trap recombination rate in a photorefractive crystal.

$\Delta_\theta, \Delta_\phi$, Spacing between incidence angles in the plane of the optic axes and normal to the optic axes of the plane-wave components of the control fields.

$\delta_{ij}$, Kronecker delta function.

$\epsilon$, Static permittivity.

$\theta$, Angle between the optic axes of the SLM's and the $z$ axis.

$\lambda$, Center wavelength of the optical fields.

$\lambda_n$, Weighting factor for the $n$th exposure. For holograms recorded with basis vectors, $\lambda_n$ is proportional to $\beta_n$.

$\mu$, Carrier mobility in a photorefractive.

$\xi_n$, Global phase difference between $\mathbf{E}_1$ and $\mathbf{E}_2$ during the $n$th exposure.

$\tau_i$, Dimensionless scaling constant for the $i$th exposure time.

$\tau(n)$, Characteristic time constant during the $n$th exposure.

$|\nu_n\rangle, |\varphi_n\rangle$, Vectors describing single plane-wave components of $|\mathscr{E}^1\rangle$ and $|\mathscr{E}^2\rangle$.

$\phi$, Angle between the wave vector of a control-field plane-wave component and the plane containing the optic axes.

$\chi$, Fraction of saturation to which the first exposure in a sequence is recorded.

$|\psi\rangle$, Vector describing an arbitrary input field.

$\Delta\omega$, Spacing between frequencies of a polychromatic field.

$\omega_0$, Carrier frequency for the optical fields.

$\omega_n$, Frequency of the $n$th spectral component of a polychromatic field.

$A_0$, Saturation amplitude of $A(\mathbf{r}, t)$.

$A(\mathbf{r}, t)$, Holographically controlled volume perturbation.

$A_n$, Amplitude of a Fourier component of $A(\mathbf{r}, t)$.

$C$, Amplitude of the probing field for coping holograms.

$c$, Speed of light in the holographic material.

$D_i, D_0$, Number of pixels active on the input and output SLM's. For nondegenerate fractal grids, $D_i D_0 \le N_1^2 N_2$.

$E_0$, Saturation exposing energy flux.

$\mathbf{E}_1(\mathbf{r}, t), \mathbf{E}_2(\mathbf{r}, t)$, Optical control fields generated on SLM1 and SLM2 or by H1 and H2.

$\mathbf{E}_{1n}(\mathbf{r}, t), \mathbf{E}_{2n}(\mathbf{r}, t)$, Components of $\mathbf{E}_1$ and $\mathbf{E}_2$ at frequency $\omega_n$.

$\mathscr{E}_{n_1 n_2}^1, \mathscr{E}_{n_1 n_2}^2$, Amplitudes of the spatial Fourier components of $\mathbf{E}_1$ and $\mathbf{E}_2$.

$\mathscr{E}_{n_1 n_2}^{2n}$, Amplitude of a spatial and temporal Fourier component of $\mathbf{E}_2$ in the polychromatic case.

$\mathscr{E}_r$, Amplitude of the reference for polychromatic recording.

$|\mathscr{E}^1\rangle, |\mathscr{E}^2\rangle$, Vectors with components $\mathscr{E}_{n_1 n_2}^1, \mathscr{E}_{n_1 n_2}^2$.

$|\mathscr{E}^1(n)\rangle, |\mathscr{E}^2(n)\rangle, |\mathscr{E}^1\rangle$ and $|\mathscr{E}^2\rangle$ during the $n$th exposure.

$e$, Electron charge.

$f(t)$, Repetitive pulsing function for polychromatic recording.

$f_n$, Weighting factor for the $n$th exposure in a material with exponential response and complex $\alpha$.

$I_n$, Spatial mean of the recording intensity during the $n$th exposure.

$\mathbf{K}_n$, Wave vector for the $n$th Fourier component of $A(\mathbf{r}, t)$.

$k_0$, $2\pi/\lambda$.

$\mathbf{k}_{n_1 n_2}^1, \mathbf{k}_{n_1 n_2}^2$, Wave vectors for the Fourier components of $\mathbf{E}_1$ and $\mathbf{E}_2$.

$\mathbf{k}_{n_1 n_2}^{2n}$, Spatial wave vector for a Fourier component of $\mathbf{E}_2$ in the polychromatic case.

$\mathbf{k}_r$, Spatial wave vector of the reference for polychromatic recording.

$L_x, L_y, L_z$, Spatial extents of the controlled volume along the corresponding axes.

$N_1$, Number of plane-wave components in each control field along the $\Delta_\theta$ direction for fixed $\phi$.

$N_2$, Number of plane-wave components in each control field along the $\Delta_\theta$ direction for fixed $\theta$.

$N_e$, Total number of exposures made in a recording sequence or, in the polychromatic case, total number of wavelengths used to record a hologram.

$N_g$, Total number of gratings controlled by a single reference polychromatic recording system.

$\mathbf{n}$, Number vector describing the order of spatial harmonics.

$n_1 n_2$, Indices on the ranges $[-N_1/2, N_1/2]$ and $[-N_2/2, N_2/2]$.

$p(t)$, Pulse of width $t_\delta$.

$Q$, Ratio of total exposure energy to $E_0$ for saturable materials.

$R$, Holographic rank, the minimum value of $N_e$ for a given hologram.

$R'$, Rank of $\overline{\mathbf{W}}$. It is shown that $R' = R$.

$r_{nn'}, s_{nn'}$, Coefficients for expanding $|\mathscr{E}^1(n)\rangle$ and $|\mathscr{E}^2(n)\rangle$ in terms of $|u_{n'}\rangle$ and $|v_{n'}\rangle$.

$T$, Recording time for polychromatic holograms.

$t_\delta$, Width of the impulse $p(t)$.

$t_n$, Exposure time for the $n$th exposure.

$t_0$, Total recording time for saturable holograms.

$t_r$, Pulse repetition period in $f(t)$.

$|u_n\rangle, |v_n\rangle$, Orthonormal basis vectors in the spaces spanned by $|\mathscr{E}^1\rangle$ and $|\mathscr{E}^2\rangle$.

$\mathbf{\underline{W}}$, Matrix with components $A_n$.

## ACKNOWLEDGMENTS

## REFERENCES AND NOTES

1. P. J. van Heerden, "Theory of optical information storage in solids," Appl. Opt. 2, 393–400 (1963).
2. A. Kozma and E. S. Barrekette, eds., Special issue on optical storage of digital data, Appl. Opt. 13, (4) (1974).
3. N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, "Optical implementation of the Hopfield model," Appl. Opt. 24, 1469–1475 (1985).
4. D. Z. Anderson, "Coherent optical eigenstate memory," Opt. Lett. 11, 56–58 (1986).
5. D. Psaltis, D. J. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals," Appl. Opt. 27, 1752–1759 (1988).
6. D. Psaltis, D. J. Brady, X.-G. Gu, and S. Lin, "Holography in artificial neural networks," Nature (London) 343, 325–330 (1990).
7. G. Carpenter and S. Grossberg, eds., special issue on neural networks, Appl. Opt. 26(23) (1987).
8. A. Marrakchi, W. M. Hubbard, S. F. Habiby, and J. S. Patel, "Dynamic holographic interconnects with analog weights in photorefractive crystals," Opt. Eng. 29, 215–224 (1990).
9. E. S. Maniloff and K. M. Johnson, "Dynamic holographic interconnects using static holograms," Opt. Eng. 29, 225–229 (1990).
10. M. Cabrera, J. Y. Jezequel, and J. C. Andre, "Three-dimensional machining by laser photopolymerization," in Lasers in Polymer Science Technology (CRC, Boca Raton, Fla., 1990), pp. 73–95.
11. S. Hunter, F. Kiamilev, S. Esener, D. A. Parthenopoulos, and P. M. Rentzepis, "Potentials of 2-photon based 3-D optical memories for high-performance computing," Appl. Opt. 29, 2058–2066 (1990).
12. E. Wolf, "Three-dimensional structure determination of semi-transparent objects from holographic data," Opt. Commun. 1, 153–156 (1969).
13. D. L. Staebler, Second International Conference on Electro-Photography (Society for Imaging Science and Technology, Chicago, Ill., 1974), p. 210.
14. K. Blotekjaer, "Limitations on holographic storage capacity of photochromic and photorefractive media," Appl. Opt. 18, 57–67 (1979).
15. R. A. Bartolini, A. Bloom, and J. S. Escher, "Volume holographic recording characteristics of an organic medium," Appl. Opt. 15, 1261–1265 (1976).
16. T. Todorov, L. Nikolova, N. Tomova, and V. Dragostinova, "Photoinduced anisotropy in rigid dye solutions for transient polarization holography," IEEE J. Quantum Electron. QE-22, 1262–1267 (1986).

17. P. Gunter and J.-P. Huignard, eds., Photorefractive Materials and Their Applications I and II (Springer-Verlag, Berlin, 1989), Vols. I and II.
18. W. J. Tomlinson and E. A. Chandross, "Organic photochemical refractive-index image recording systems," Adv. Photochem. 12, 201–281 (1980).
19. I.-C. Khoo and R. Nomarndin, "The mechanism and dynamics of transient thermal gradient diffraction in pneumatic liquid crystal films," IEEE J. Quantum Electron. QE-21, 329–335 (1985).
20. J. H. Wendorff and M. Eich, "Nonlinear optical phenomena in liquid crystalline side chain polymers," Mol. Cryst. Liq. Cryst. 169, 133–166 (1989).
21. H. Lee, X. G. Gu, and D. Psaltis, "Volume holographic interconnections with maximal capacity and minimal crosstalk," J. Appl. Phys. 65, 2191–2194 (1989).
22. D. Psaltis, X. G. Gu, and D. Brady, "Holographic implementations of neural networks," in An Introduction to Neural and Electronic Networks, S. F. Zornetzer, J. L. Davis, and C. Lau, eds. (Academic, New York, 1989), pp. 339–348.
23. See, for example, C.-T. Chen, Linear System Theory and Design (Holt, Rinehart & Winston, New York, 1984), p. 569.
24. N. V. Kuktarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electro-optic crystals. I. Steady state," Ferroelectrics 22, 949–956 (1979).
25. W. J. Burke and P. Sheng, "Crosstalk noise from multiple thick phase holograms," J. Appl. Phys. 48, 681–685 (1977).
26. D. J. Brady, "Photorefractive volume holography in artificial neural networks," Ph.D. dissertation (California Institute of Technology, Pasadena, Calif., 1990).
27. J. E. Ford, S. H. Lee, and Y. Fainman, "Application of photorefractive crystals to optical interconnection," in Digital Optical Computing II, R. Arrathoon, ed., Proc. Soc. Photo-Opt. Instrum. Eng. 1215, 155–163 (1990).
28. F. Mok, M. Tackitt, and H. M. Stoll, "Massively parallel optical template matcher/correlator," in Optical Society of America 1990 Annual Meeting, Vol. 15 of 1990 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1990), p. 211 (A).
29. F. Mok, M. Tackitt, and H. M. Stoll, "Storage of 500 high-resolution holograms in a LiNbO3 crystal," Opt. Lett. 16, 605–608 (1991).
30. D. J. Brady, K. Hsu, and D. Psaltis, "Periodically refreshed multiply exposed photorefractive holograms," Opt. Lett. 15, 817–820 (1990).
31. J. Hong, P. Yeh, D. Psaltis, and D. J. Brady, "Diffraction efficiency of strong volume holograms," Opt. Lett. 15, 344–346 (1990).
32. R. A. Bartolini, A. Bloom, and J. S. Escher, "Multiple storage of holograms in an organic medium," Appl. Phys. Lett. 28, 506–508 (1976).
33. J. M. Heaton and L. Solymar, "Transient energy transfer during hologram formation in photorefractive crystals," Opt. Acta 32, 397–400 (1985).
34. D. J. Brady and D. Psaltis, "Perceptron learning in optical neural computers," in Optical Computing: Proceedings of the 34th Scottish Universities Summer School in Physics, B. S. Wherrett and F. A. P. Tooley, eds. (Edinburgh U. Press, Edinburgh, Scotland, 1988), pp. 251–265.
35. H. Yoshinaga, K. Kitayama, and T. Hara, "Experimental learning in an optical perceptronlike neural network," Opt. Lett. 14, 716–718 (1989).
36. E. G. Paek, J. R. Wullert II, and J. S. Patel, "Holographic implementation of a learning machine based on a multicategory perceptron algorithm," Opt. Lett. 14, 1303–1305 (1989).
37. D. Z. Anderson and M. C. Erie, "Resonator memories and optical novelty filters," Opt. Eng. 26, 434–444 (1987).
38. P. Saari, R. Kaarli, and A. Rebane, "Picosecond time- and space-domain holography by photochemical hole burning," J. Opt. Soc. Am. B 3, 527–533 (1986).