Pseudorandom encoding of complex-valued functions onto amplitude-coupled phase modulators

Robert W. Cohn

The Electro Optics Institute, University of Louisville, Louisville, Kentucky 40902-0001

Pseudorandom encoding is a method of statistically approximating desired complex values with those values that are achievable with a given spatial light modulator. Originally developed for phase-only modulators, pseudorandom encoding is extended to modulators for which amplitude is a function of phase. This is accomplished by transforming the phase statistics to compensate for the amplitude coupling. Example encoding formulas are derived, evaluated, and compared with a noncompensating pseudorandom-encoding algorithm. Compensating algorithms encode a smaller area of the complex plane and can produce more noise than is possible for arbitrary pseudorandom algorithms. However, the encoding formulas have greatly simplified numerical implementations. © 1998 Optical Society of America [S0740-3232(98)00704-2]

OCIS code: 050.1970.

1. INTRODUCTION

It is common to classify spatial light modulators (SLM's) as being either amplitude-only or phase-only, but, in practice, SLM's usually exhibit some degree of coupling between amplitude and phase (e.g., as illustrated in Fig. 1).[1] A notable example is liquid-crystal SLM's, which can be continuously changed from phase-mostly to amplitude-mostly operation by rotation of a wave plate or a polarizer.[2,3] Currently, no commonly available SLM's produce all complex values. Nonetheless, the design of diffractive optics and the implementation of other signal processing functions can often be simplified greatly if there are no constraints on the complex values. When the possible modulation values are constrained in some way, it has become common to employ numerically intensive global searches for functions that are implementable and that meet desired performance criteria. In many real-time applications using programmable modulators, these computational constraints may rule out the use of global searches. Although encoding does not usually match the performance of global searches, it can provide acceptable performance and numerically efficient and direct methods of representing fully complex functions with SLM's that are not fully complex.

The encoding problem considered in this paper is that of the design of Fourier transform holograms for implementation on available SLM's. One of the earliest discussions of this problem is by Brown and Lohmann.[4] Their methods use groups of pixels to represent a single complex value. Thus the space–bandwidth product of the SLM that uses such an algorithm will be reduced by the factor corresponding to the number of pixels in the group. Kirk and Jones introduced a point-oriented method of encoding complex values with a phase-only modulator.[5] The phase is specified to be the product of an amplitude-modulating function and a sinusoidal carrier. For discretely sampled phase-only SLM's, at least

two pixels are required to represent one period of the sinusoid. Thus the space–bandwidth product of the signal is, at best, half that of the SLM. Cohn and Liang developed a method in which any desired complex values can be mapped to a single pixel, thereby using the entire space–bandwidth product of the SLM.[6] The method, referred to as pseudorandom encoding, uses ensemble averages of the values that are achievable with the SLM to represent the desired complex values. The actual modulation produced by the SLM corresponds to a single sample from the ensemble. The diffraction pattern of this random modulation consists of an approximate reconstruction of the desired diffraction pattern and a diffuse, approximately white-noise background. The method has also been interpreted as a carrier-based method.[7] Rather than using a single-frequency carrier, as does the Kirk–Jones method, a carrier of all frequencies is used. This diffracts unwanted light into all spatial frequencies. By a distribution of the unwanted light over the entire spatial bandwidth, the average noise level can often be much lower than the intensity of the desired reconstruction. This permits reconstructions to be formed anywhere over the bandwidth set by the grating frequency of the modulator.

Pseudorandom encoding also has many similarities with the parity sequence method of Chu and Goodman.[8] This method realizes a desired complex value by vectorial addition of two values of transmittance that are separated by $N/2$ pixels in an N-pixel phase-only modulator. This method perfectly reconstructs the desired diffraction pattern, but only at $N/2$ resolvable locations in the diffraction pattern. Between each sample of the desired reconstruction is a sample of the error signal (corresponding to the vector subtraction of the two values of transmittance), referred to as the parity sequence. Chu and Fienup described a version of this encoding method (named the synthetic coefficient method) in which the
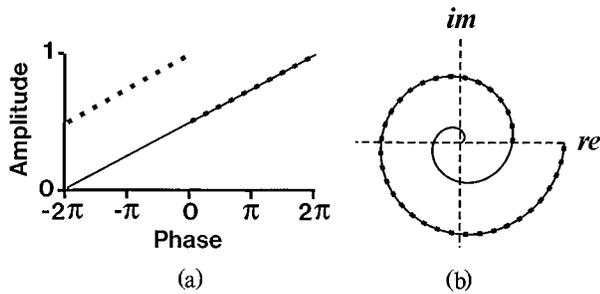
Fig. 1. Amplitude-coupled phase modulation characteristics illustrated in (a) rectangular and (b) polar plots. The characteristics have continuous phase ranges of 4π (solid curves) and 2π (dotted curves). Although amplitude coupling is drawn as a linear function, it is not necessary that it be linear for example 3 in Section 4.

transmittances of a pixel and its nearest neighbor are programmed as a group.[9] In this case the desired reconstruction is centered around the optical axis, and the error signal is centered at the Nyquist frequency $\pm N/2$. The desired reconstruction is most accurate on the optical axis, and its accuracy tends to decrease out to $\pm N/4$, where the error signal tends to dominate. Thus, as with any other group-oriented method, these methods, by their limiting the modulation bandwidth, are unable to use the entire space–bandwidth product of the SLM. However, in that pseudorandom encoding represents a desired complex value with only one pixel transmittance, it is possible to form a desirable reconstruction over the entire spatial bandwidth N for an N-pixel SLM. The error signal also reconstructs over this same spatial bandwidth. However, because pseudorandom encoding diffuses the error signal to (on average) a uniform level over the spatial bandwidth, it is often possible (depending on the desired complex function) for the desired reconstruction to be much brighter than the error. In that Chu and Fienup were studying the encoding of complex-valued functions, there are many other similarities between their methods and pseudorandom encoding. These include issues on trade-offs between diffraction efficiency and reconstruction accuracy, and the selection of phase degrees of freedoms. There are even some mathematical similarities between their encoding formulas and those for pseudorandom encoding. These points will be drawn out at appropriate places in the paper to distinguish the novel features of pseudorandom encoding.

Pseudorandom encoding was originally developed for phase-only SLM's that produce analog phase over a range of 2π[6] and, since then, for SLM's that produce analog phase and two levels of amplitude.[10] The phase-only encoding algorithms have been applied in laboratory demonstrations of beam shaping[11] and spot array generation.[7] This paper generalizes the pseudorandom-encoding concept for coupled modulators and describes new aspects of the encoding algorithms that appear when the modulation characteristics contain amplitude coupling. I will specifically consider that amplitude can be expressed as a function of phase. General derivations are presented together with concrete examples that use the two modulation characteristics illustrated in Fig. 1. Encoding for a variety of other characteristics can also be derived by the

methods presented here. The use of these example characteristics is done only to provide consistency throughout the paper. A notable feature of the pseudorandom-encoding algorithms developed to date is that their implementation requires only a few numerical calculations per pixel. Thus, desired fully complex functions can be mapped to electrically addressable SLM's in real time by using low-end serial processors. A central goal of this investigation is to determine to what degree it is or would be possible to develop numerically efficient encoding algorithms for amplitude-coupled phase modulators.

Although the focus of this paper is on the design of encoding algorithms themselves, I have also included (in Section 8) a simulation of the diffraction patterns resulting from encoding one specific complex-valued function by various encoding algorithms. The results are used to illustrate the performance considerations of the algorithms that are discussed in Sections 5–7, as well as to provide comparisons with existing encoding algorithms (pseudorandom and others) that have already been developed for phase-only SLM's.

2. GENERAL DESCRIPTION OF PSEUDORANDOM-ENCODING AND PREVIOUS ENCODING METHODS

All pseudorandom-encoding algorithms specify the modulation of any given pixel in terms of a random variable. The statistical properties of the random variable are selected in such a way that the expected value, or average, of the random modulation is identical to the desired, but unobtainable, fully complex value. The desired complex-valued modulation is written as $\mathbf{a}_c = (a_c, \psi_c)$, and the resulting modulation by the SLM is $\mathbf{a} = (a, \psi)$, where the ordered pairs are the polar representations of the complex quantities. Complex quantities are indicated by boldface type. The pseudorandom-encoding design statement is, in general, to find a value of the ensemble average

$$\langle \mathbf{a} \rangle = \int \mathbf{a} p(\mathbf{a}) \, d\mathbf{a} \tag{1}$$

of the random variable $\mathbf{a}$ such that $\langle \mathbf{a} \rangle = \mathbf{a}_c$. The statistical properties of $\mathbf{a}$ are determined by its probability density function (pdf) $p(\mathbf{a})$. The pdf is *selected* to ensure that the expected value of $\mathbf{a}$ and the desired complex value are identical. This selection of a pdf corresponds to solving the integral equation (1) for $p(\mathbf{a})$. [The solution is not unique, since the integral in Eq. (1) is a projection from the multidimensional space of $\mathbf{a}$ into a single value $\langle \mathbf{a} \rangle$. Various auxiliary conditions can be imposed on the solution and are considered in this paper.] After an appropriate density function is determined, the desired complex value $\mathbf{a}_c$ is encoded by drawing a single value of $\mathbf{a}$ from a random distribution having the density function $p(\mathbf{a})$. Since the value of $\mathbf{a}$ is found deterministically by computer, rather than from a random process occurring in nature, the procedure has been named pseudorandom encoding.

To this point the discussion has focused on encoding a single complex value. The procedure can be applied to encode spatially varying complex modulations $\mathbf{a}_c$. Spe-

cifically, in this paper I will assume that the SLM is a discretely sampled array of pixels. With the use of i as the spatial coordinate, the spatial samples of the desired complex modulation, the density function, and the random modulation are written as $\mathbf{a}_{ci}$, $p_i(\mathbf{a}_i)$, and $\mathbf{a}_i$. (This indexing scheme can be conveniently applied to one- or two-dimensional arrays, and it is not restricted to equally spaced samples.)

The far-field diffraction pattern of the encoded modulation $\mathbf{a}_i$ approximates the desired diffraction pattern. This can be seen by comparing the intensity of the desired far-field diffraction pattern with the ensemble average diffraction pattern that would result from the encoded modulation. The intensity pattern of the desired diffraction pattern is

$$I_c(f_x) = \left| \sum_i \mathbf{A}_{ci} \right|^2 = \left| \mathscr{F}\left\{ \sum_i \mathbf{a}_{ci} \right\} \right|^2, \qquad (2)$$

where $\mathscr{F}\{ \cdot \}$ is the Fourier transform operator; $\mathbf{A}_{ci}(f_x)$ is the Fourier transform of $\mathbf{a}_{ci}$, the desired complex transmittance of the ith pixel located at position i in the modulator plane; and f_x is the spatial coordinate across the Fourier plane. (Although this equation can also be written as a function of two spatial coordinates, one-dimensional coordinates are used throughout to simplify the presentation.) The expected value of the intensity pattern from the encoded modulation has been derived for the condition that the random variable $\mathbf{a}_i$ for the ith pixel is statistically independent of $\mathbf{a}_j$ for all j not equal to i. The ensemble average pattern is then expressed[6,10] as

$$\langle I(f_x) \rangle = I_c(f_x) + \sum_i \left(\langle |\mathbf{A}_i|^2 \rangle - |\mathbf{A}_{ci}|^2 \right), \qquad (3)$$

where $\mathbf{A}_i(f_x)$ is the Fourier transform of $\mathbf{a}_i$. The expected intensity consists of two terms. The first term is the desired diffraction pattern from Eq. (2). The second term represents the average level of background (i.e., speckle) noise that is produced as a result of the randomness of the modulation. For the case of pixels that are modeled as point sources, the average background noise is of constant intensity for all frequencies f_x (i.e., it is white). Many useful diffraction patterns can be synthesized for which I_c is accurately approximated and the noise level is adequately low.

The general encoding concept presented above makes no assumptions about the properties of the SLM or the specific statistical distributions selected. For coupled modulators in which amplitude $a(\psi)$ is a function of phase ψ, Eq. (1) becomes

$$\langle \mathbf{a} \rangle = \int a(\psi) p(\psi) \exp(j\psi) \mathrm{d}\psi \equiv a_0 \exp(j\psi_0), \qquad (4)$$

where $a_0 \equiv |\langle \mathbf{a} \rangle|$ is the effective amplitude and $\psi_0 \equiv \arg(\langle \mathbf{a} \rangle)$ is the effective phase resulting from the averaging operation. For the specific amplitude coupling $a(\psi) = 1$, Eq. (4) also describes the effective amplitude for phase-only modulators.

An important factor that controls the performance of not only pseudorandom encoding but several other encod-

ing algorithms as well is the absolute magnitude scaling of the desired complex-valued function $\mathbf{a}_c$. I will use the symbol

$$\gamma = \max_i(\mathbf{a}_{ci}) \qquad (5)$$

for this scaling factor. For SLM's that are (usually assumed to be) passive devices, it may not be possible with some algorithms to encode complex values that exceed unity magnitude. As will be shown below, for some algorithms and modulation characteristics, the scaling parameter γ can be constrained to be much less than unity. Low values of γ reduce the amount of energy in the reconstruction and, even worse, can often increase the amount of noise energy. This observation is probably most easily seen for phase-only SLM's.[11] In this case all the energy incident on the modulator is transmitted to the Fourier plane. Thus, as γ decreases, the desired portion of the reconstruction, I_c, becomes increasingly dim and the noise component becomes increasingly bright. Consequently, the desired reconstruction becomes increasingly perturbed by noise and more difficult to see.

Chu and Fienup made quite similar observations for the parity sequence method.[9] Making $\gamma < 1$ ($A = 1/\gamma$ is used for the scaling parameter in their paper) reduces the energy in the desired reconstruction but does not affect accuracy. On the other hand, they also considered cases for $\gamma > 1$. For those desired magnitudes that exceed unity, the SLM transmittance is set to $\exp[j \arg(\mathbf{a}_{ci})]$, which they referred to as a kinoform but which today is more frequently referred to as a phase-only filter. They noted that it is possible to trade off the amount of energy in the diffraction pattern versus reconstruction accuracy for values of the scaling parameter, e.g., $1 \leq \gamma < \infty$. Similar sorts of trade-offs have been identified for some types of pseudorandom encoding in Refs. 7 and 10. In Sections 5–7 below, many new possibilities of blending together various pseudorandom (and also nonrandom) algorithms to improve the quality of the diffraction pattern are given. In Ref. 10 it has been shown that with blending it is even possible to obtain better reconstruction accuracy for $\gamma > 1$. Thus, from Section 5 on, I will frequently describe the encoding range of the various algorithms in terms of γ.

Closely related to γ is the diffraction efficiency of the *desired* (as opposed to the resulting) complex-valued function. Through the use of Parseval's relation, this can be calculated in the modulation plane as

$$\eta = \frac{1}{N} \sum_{i=1}^{N} |\mathbf{a}_{ci}|^2. \qquad (6)$$

This definition of diffraction efficiency has the physical interpretation of the average energy transmittance of the desired fully complex function. This result shows that for two different values of scaling factor, γ_1 and γ_2, the two resulting diffraction efficiencies vary according to $\eta_1/\eta_2 \propto (\gamma_1/\gamma_2)^2$. This relationship, together with Eq. (3), shows why (especially for pseudorandom encoding of phase-only SLM's) it is important to make the diffraction efficiency large. For the coupled SLM's considered in Fig. 1, there is a significant amount of phase modulation,

and, as a result, making diffraction efficiency large usually leads to improved performance. (Simulations illustrating such improvements are presented in Section 8.)

The diffraction efficiency also depends on the desired function as well. This was recognized early in the development of phase-only computer-generated holograms and was also mentioned by Chu and Fienup[9] as well as Brown and Lohmann.[4] This has led to the continued development of code that optimizes the performance of the diffraction pattern under the condition that only the intensity of the diffraction pattern is of concern. The phases (the so-called degrees of freedom) are varied, with the goal of achieving near-unity diffraction efficiency and low reconstruction error. Because of the time-consuming nature of such optimizations, this type of design is not considered in this paper. Instead, a single complex-valued function is selected for the design examples in Section 8. Only the parameter γ is varied as permitted by the various encoding algorithms.

3. REVIEW OF PSEUDORANDOM ENCODING FOR PHASE-ONLY MODULATORS

This section reviews pseudorandom encoding for phase-only modulators and discusses the desirable features sought in developing a specific encoding formula. These results for phase-only encoding are used to motivate the derivations of encoding formulas in Section 4.

Various families of density functions $p(\psi; \langle\psi\rangle, \sigma)$, parameterized in terms of the mean value $\langle\psi\rangle$ and the standard deviation σ of the phase distribution, were evaluated in Eq. (4) of Ref. 6 and shown to produce all complex amplitudes having an amplitude between 0 and 1. In general, a two-dimensional search over $\langle\psi\rangle$ and σ is required to obtain a desired complex value. It was found (though not explicitly stated in Ref. 6) that the solution method could be simplified to a one-dimensional search by using density functions that are symmetric around their means $\langle\psi\rangle$. This led to the specific result that $\psi_0 = \langle\psi\rangle$ and that σ can be found by a one-dimensional search to obtain a desired value of effective amplitude in the range 0–1. Thus parameters that describe the density function (such as mean and variance) are individually associated with a value of effective phase and a value of effective amplitude. These conditions have led to simple expressions that can be evaluated with low numerical overhead for pseudorandom encoding of phase-only modulators. These points are brought out by way of the following example.

Example 1: Derivation of a pseudorandom phase-only encoding method. In Ref. 6 the effective complex amplitude was derived by using the uniform family of density functions [see Fig. 2(a)]

$$p(\psi; \langle\psi\rangle, \nu) = \frac{1}{\nu} \mathrm{rect}\left(\frac{\psi - \langle\psi\rangle}{\nu}\right), \qquad (7)$$

which is specified in terms of the two parameters $\langle\psi\rangle$ and ν, the spread of the density function (where the spread is more conveniently used instead of the standard deviation $\sigma = \nu/\sqrt{12}$ in the case of uniform distributions). Substituting Eq. (7) into Eq. (4) for $a(\psi) = 1$ gives

$$\langle\mathbf{a}\rangle = \mathrm{sinc}(\nu/2\pi)\exp(j\langle\psi\rangle) \equiv a_0 \exp(j\psi_0). \qquad (8)$$

The amplitude is then identified as (solid curve in Fig. 3)

$$a_0 = |\mathrm{sinc}(\nu/2\pi)|, \qquad (9)$$

and the phase is identified as

$$\psi_0 = \begin{cases} \langle\psi\rangle, & \mathrm{sinc}(\nu/2\pi) > 0 \\ \langle\psi\rangle + \pi, & \mathrm{sinc}(\nu/2\pi) < 0 \end{cases}, \qquad (10)$$

where the phase offset of π reflects sign changes of the sinc function. All values of effective amplitude between 0 and 1 can be realized by limiting the maximum spread to 2π. The effective complex amplitude then simplifies to

$$\langle\mathbf{a}\rangle = \mathrm{sinc}(\nu/2\pi)\exp(j\langle\psi\rangle), \quad 0 \leq \nu \leq 2\pi. \qquad (11)$$

Equating Eq. (11) with the desired complex value $\mathbf{a}_c$, it is found that $\langle\psi\rangle = \psi_c$ and that

$$\nu = 2\pi \, \mathrm{sinc}^{-1}(a_c). \qquad (12)$$

The desired value of $\mathbf{a}_c$ is encoded by selecting a random value of phase ψ from the distribution in Eq. (7) for the specified values of $\langle\psi\rangle$ and ν. In practice, this distribution is simulated by transforming the uniform random variable $s \in [0, 1]$, which has the pdf $p_s(s) = \mathrm{rect}(s - 1/2)$, into a random variable ψ that has the required
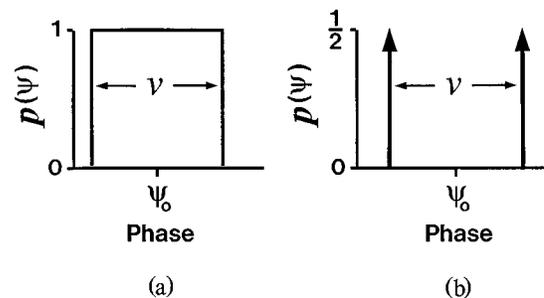

Fig. 2. Probability density functions (pdf's) for (a) uniform and (b) binomial random distributions of phase. In Section 4 effective pdf's of these same forms are sought.
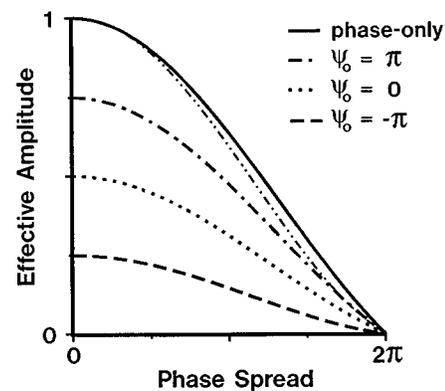

Fig. 3. Effective amplitudes for examples 1 and 2. The amplitude coupling used is identical to the solid curves in Fig. 1. The curve without a legend (dashed–dotted–dotted curve) is the effective amplitude for $\psi_0 = -\pi$ with the maximum effective amplitude normalized to unity. The two other curves even more closely match the sinc (phase-only) curve if they are normalized similarly.

pdf of Eq. (7). The random variable s is simulated by using the standard uniform random-number generator $s_i = \mathrm{ran}(s_{i-1})$. The appropriate random value of phase is then drawn by using the formula

$$\psi = \langle\psi\rangle + \nu[\mathrm{ran}(s) - 1/2].\qquad(13)$$

Equation (13) shows that pseudorandom encoding for phase-only modulation can be accomplished with a small number of mathematical operations. The speed of computation can be extremely fast if ν and $\mathrm{ran}(s)$ in Eqs. (12) and (13) are calculated with the use of lookup tables.

The encoding formulas (12) and (13) have a form that is quite similar to the parity sequence method of encoding. For both techniques there is a desired phase (i.e., $\langle\psi\rangle$ for pseudorandom encoding) and an effective-amplitude reduction (corresponding to, in the case of pseudorandom encoding, a random phase offset selected over a spread ν and, in the case of the parity sequence method, a fixed phase deviation between the two transmittances). In pseudorandom encoding, a spread of 2π produces an effective amplitude of zero [see Eq. (11)], whereas in Chu and Goodman's method a nonrandom spread of π effectively encodes a zero. An even closer mathematical correspondence is described for the algorithm derived in example 3 of Section 4.

4. DERIVATION OF PSEUDORANDOM-ENCODING ALGORITHMS FOR COUPLED MODULATORS

In this section the pseudorandom-encoding approach that was described in Section 3 is modified so as to compensate for the effects of amplitude coupling. Following this approach leads, once again, to encoding formulas that have low numerical overhead. The subsequent sections consider the range of realizable values (as measured by the value of γ) that can be encoded on coupled modulators by these algorithms and modifications of the algorithms that extend the encoding range.

The use of symmetric pdf's with amplitude-coupled modulation characteristics usually does not produce decoupling between the effective amplitude and the effective phase. The reason is that in Eq. (4) the coupled amplitude $a(\psi)$ is mathematically identical to a weighting function that biases the average of the random phasor $\exp(j\psi)$ away from $\exp(j\langle\psi\rangle)$. The bias is not usually constant, so that a two-dimensional search is required to find a solution to $(a_0, \psi_0) = (a_c, \psi_c)$. Alternatively, the form of the pdf can be chosen to compensate for amplitude coupling, so that it becomes possible to specify $\psi_0 = \psi_c$ directly. This is seen by defining the term $p_{\mathrm{eff}}(\psi) \equiv a(\psi)p(\psi)$ in Eq. (4), which will be referred to as the effective pdf. Written this way, Eq. (4) has a form identical to that for the effective complex amplitude for phase-only encoding [Eq. (4) with $a(\psi) = 1$]. If the effective density function is symmetric about the amplitude-weighted expected value of phase

$$\langle\psi\rangle_{\mathrm{eff}} = \int \psi p_{\mathrm{eff}}(\psi)\mathrm{d}\,\psi,\qquad(14)$$

then the effective phase ψ_0 will equal $\langle\psi\rangle_{\mathrm{eff}}$ even though ψ_0 is not equal to the average phase $\langle\psi\rangle$. Thus the selec-

tion of the density function $p(\psi)$ that makes the effective density function $p_{\mathrm{eff}}(\psi)$ symmetric permits the desired phase $\psi_c = \psi_0$ to be directly specified as the center of symmetry of the effective density function.

Although a desired value of ψ_0 can be directly specified, this approach does not entirely decouple a_0 from ψ_0. It does, however, produce the desired numerical simplification in that the effective amplitude a_0 becomes a one-dimensional function $a_0(\sigma; \psi_0)$ of σ (or ν) and the fixed parameter ψ_0. The desired value of amplitude corresponds to the value of σ that satisfies $a_0(\sigma; \psi_c) = a_c$. From the standpoint of numerical efficiency, sequential encoding of phase and amplitude is preferable to a simultaneous two-dimensional search for the encoding parameters.

The derivation of encoding formulas for coupled modulators also uses the following two results from probability[12]:

1. The effective pdf and the pdf $p(\psi)$ are not fully specified until a scale factor is determined that ensures that the integrated area of $p(\psi)$ equals unity. This requirement is simply part of the definition of a pdf. Specifically, the probability of the certain event is unity. This requirement can be expressed in terms of the cumulative pdf

$$P(\psi) = \int_{-\infty}^{\psi} p(\phi)\mathrm{d}\phi.\qquad(15)$$

The random variable ψ has total probability $P(\psi) = 1$ for $\psi = \infty$.

2. Although random-number generators for arbitrary random distributions are not usually available, it is possible, by using a suitably chosen function, to transform the statistics of the uniform random variable s into the desired statistics. The function is known to be the inverse of the cumulative distribution function[12]:

$$\psi = P^{-1}(s).\qquad(16)$$

The random variable ψ is then simulated by performing the function in Eq. (16) on the numbers produced by the random-number generator $\mathrm{ran}(s)$.

The procedure of deriving encoding formulas by this compensation approach is illustrated by the following two examples. In the first example (example 2), the amplitude-coupling function is given as an explicit function. In the second example (example 3), a solution is found in closed form without the amplitude-coupling function being given explicitly. This second form would be especially useful for SLM's for which the amplitude coupling can be changed *in situ* (for instance, liquid-crystal SLM's that are combined with rotatable polarizers or wave plates).

Example 2: Derivation for amplitude-coupling an explicit function. The amplitude-coupling function [solid line of Fig. 1(a)] is the linear function of phase

$$a(\psi) = m\psi + b, \qquad \psi \in [-2\pi, 2\pi],\qquad(17)$$

where m is the slope and $b = a(0)$. A family of effective pdf's that is similar in form to Eq. (7) [Fig. 2(a)] is

$$p_{\mathrm{eff}}(\psi) \propto \mathrm{rect}[(\psi - \psi_0)/\nu].\qquad(18)$$

A single pdf is specified by two parameters: spread ν and bias ψ_0. After the correct normalizations are determined so that each pdf of the family $p(\psi; \psi_0, \nu)$ has unit area, the pdf that compensates $a(\psi)$ is identified as

$$p(\psi) = \frac{1}{\psi + b/m}$$
$$\times \left[\ln\left(\frac{\psi_0 + \nu/2 + b/m}{\psi_0 - \nu/2 + b/m} \right) \right]^{-1} \mathrm{rect}\left(\frac{\psi - \psi_0}{\nu} \right).$$
(19)

With the use of Eqs. (15) and (16), the transformation from the uniform random variable s to the random variable ψ is found to be

$$\psi = \frac{(\psi_0 + \nu/2 + b/m)^s}{(\psi_0 - \nu/2 + b/m)^{s-1}} - b/m.$$
(20)

Substitution of Eqs. (17) and (19) into Eq. (4) gives a closed-form expression for the effective complex amplitude of

$$\langle \mathbf{a} \rangle = m\nu \left[\ln\left(\frac{\psi_0 + \nu/2 + b/m}{\psi_0 - \nu/2 + b/m} \right) \right]^{-1} \mathrm{sinc}\left(\frac{\nu}{2\pi} \right) \exp(j\psi_0).$$
(21)

Note the similarity between this equation and Eq. (11) for phase-only pseudorandom encoding. Furthermore, in the limit, as the slope m approaches zero, the effective complex amplitude for the coupled modulation ap-

of ψ_0 over a 2π range. For the phase-only SLM the modulation characteristic is periodic, which eliminates the need for a SLM that produces phase in excess of 2π. Thus the algorithm proposed in example 2 suffers from the practical disadvantage that most SLM's available today barely produce a 2π phase range. A second disadvantage with the approach described in example 2 is that the algorithm needs to be custom designed for each individual modulation characteristic. Both disadvantages of the approach in example 2 are overcome in example 3 by designing a pseudorandom-encoding algorithm with the use of a different class of statistical distributions and by treating the amplitude modulation as a periodic function of phase.

Example 3: Derivation for amplitude coupling that is not an explicit function. The amplitude-coupling function is assumed be periodic. Figure 1(a) (dotted lines) illustrates a specific coupling function. The coupling is linear with phase, but no specific form is required or considered in this derivation. The discrete effective density function [see Fig. 2(b)]

$$p_{\mathrm{eff}}(\psi) \propto \delta(\psi - \psi_0 + \nu/2) + \delta(\psi - \psi_0 - \nu/2), \quad (22)$$

where $\delta(\cdot)$, the Dirac delta function, is especially simple to use, since the amplitude-weighting compensation depends on only the amplitude values at the two points $\psi = \psi_0 \pm \nu/2$. Under these assumptions the analysis performed in example 2 can be repeated to give the pdf

$$p(\psi) = \frac{a(\psi_0 + \nu/2)\delta(\psi - \psi_0 + \nu/2) + a(\psi_0 - \nu/2)\delta(\psi - \psi_0 - \nu/2)}{a(\psi_0 - \nu/2) + a(\psi_0 + \nu/2)} \quad (23)$$

proaches the effective complex amplitude for phase-only modulation.

The prescription for pseudorandom encoding, $\langle \mathbf{a} \rangle = \mathbf{a}_c$, and the relationship among $\langle \mathbf{a} \rangle$, ν, and ψ_0 in Eq. (21) can then be used to specify the following encoding algorithm:

select initial values for s_0, b, and m
For $i = 1$ to N pixels
$\psi_{0i} \leftarrow \psi_{ci}$; $a_{0i} \leftarrow a_{ci}$; $s_i \leftarrow \mathrm{ran}(s_{i-1})$
solve Eq. (21) for ν_i with a_{0i} and ψ_{0i} specified
calculate ψ_i from Eq. (20) with s_i, ψ_{0i}, and ν_i specified.

To achieve the greatest computational speeds, the values of s_i, ν_i, and ψ_i would be precomputed and stored as lookup tables. Thus, by including amplitude compensation, it is possible to pseudorandom-encode coupled modulators in a manner similar to the encoding of phase-only modulators. The algorithms are similar in structure, and the encoding formulas are similar in form to those for pseudorandom phase-only encoding. The procedure illustrated in example 2 can be followed to derive encoding formulas for various other coupling functions and effective density functions.

In example 2 the availability of a 4π phase modulation range was assumed. This ensures that an effective amplitude between $a(\psi_0)$ (for a random phase spread $\nu = 0$) and zero ($\nu = 2\pi$) can be encoded for any value

ρ and the effective complex amplitude

$$\mathbf{a}_c = \langle \mathbf{a} \rangle$$
$$= \frac{2a(\psi_0 + \nu/2)a(\psi_0 - \nu/2)}{a(\psi_0 + \nu/2) + a(\psi_0 - \nu/2)} \cos(\nu/2)\exp(j\psi_0).$$
(24)

The ratio in Eq. (24) contains all the terms that compensate for the amplitude coupling. This term can be seen to be a ratio of the square of the geometric mean divided by the algebraic mean of the two amplitude samples $a(\psi_0 \pm \nu/2)$. This ratio reduces to a constant for phase-only SLM's. The uniform random variable s (from the available random-number generator) can be transformed into the random variable of phase by the simple threshold test

$$\psi = \begin{cases} \psi_0 - \nu/2 & \text{if } s \leq \dfrac{a(\psi_0 + \nu/2)}{a(\psi_0 - \nu/2) + a(\psi_0 + \nu/2)} . \\ \psi_0 + \nu/2 & \text{otherwise} \end{cases}$$
(25)

These results are especially useful in that this closed-form result applies to any function $a(\psi)$ for which ψ has a range of at least 2π. This formula can be applied even when samples are randomly selected from either side of

an amplitude discontinuity [such as the discontinuity at $\psi = 0$ in Fig. 1(a)]. The encoding algorithm is identical to the algorithm following example 2 if Eqs. (24) and (25) are used in place of Eqs. (21) and (20), respectively.

The mathematics for this encoding algorithm are surprisingly similar to those for the parity sequence method.[8] This is most clearly seen for the case in which the coupling function $a(\psi) = 1$ for all values of ψ. The effective amplitude in Eq. (24) then becomes $a_c = \cos(\nu/2)$, and the effective phase is $\psi_c = \psi_0$. These effective values are identical to those that are encoded by the parity sequence method. Furthermore, the actual modulation produced by the pseudorandomly encoded phase-only modulator is either $\mathbf{a} = \exp[j(\psi + \nu/2)]$ or $\mathbf{a} = \exp[j(\psi - \nu/2)]$ with probability 1/2. The only difference with the parity sequence method is that each of these transmittances is produced by a pair of spatially separated pixels. In general, when there is coupling, the probability of selecting one or the other value of $\mathbf{a}$ is adjusted [(according to Eq. (25)] so that the weaker modulation is selected more frequently than the stronger modulation. This compensation effectively attenuates the stronger modulation, so that both values of modulation are effectively of equal strength.

5. NUMERICAL EVALUATION OF THE ENCODING FORMULAS AND IMPLEMENTATION ISSUES

The characteristics and the implementation of the encoding formulas can be appreciated by considering some numerical simulations of effective amplitude. The amplitude-coupling functions shown in Fig. 1 will be used. This corresponds to using a slope of $m = 1/4\pi$ and $b = 1/2$ in the equations in example 2. The effective amplitudes found from Eq. (21) are plotted in Fig. 3. In example 3 the amplitude-coupling function used is similar to Eq. (17). A slope of $m = 1/4\pi$, and $b = 1/2$ for $\psi > 0$ and $b = 1$ for $\psi < 0$, are used to describe the two line segments [the dotted lines in Fig. 1(a)]. Using this coupling function in Eq. (24) produces the effective amplitudes shown in Fig. 4.
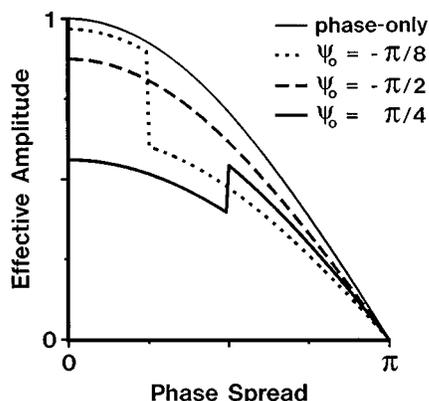

Fig. 4. Effective amplitude curves for example 3. The amplitude coupling is identical to the dotted curve in Fig. 1(b). The effective amplitudes for phase-only encoding with use of the pdf from Fig. 2(b) are included for comparison.

Numerical results for example 2. Figure 3 shows that the range of possible effective amplitudes a_0 depends on the value of the effective phase ψ_0. For a spread of $\nu = 0$, the effective amplitude of each curve is identical to $a(\psi_0)$, and as the spread increases, the effective amplitude decreases monotonically from this point. The shape of each curve is very close to that of a sinc function, as is illustrated by the normalized version of the effective amplitude for the $\psi_0 = -\pi$ curve. The two other effective amplitudes are even closer in shape to that of a sinc. Numerically efficient solutions of ν that are only slightly more involved than those for phase-only encoding can be developed by using this result.

The curves plotted in Fig. 3 correspond to those for which ψ_0 is contained between $-\pi$ and π. For this range of effective phase, all the curves range between $a(\psi_0)$ and 0. For $|\psi_0| > \pi$ the 4π range of the modulator limits the maximum spread to $2(2\pi - |\psi_0|)$. Thus, for these particular values of ψ_0, the spread cannot be made large enough to continuously reduce a curve of effective amplitude to 0. In this subsection I will consider the implications of using only the effective-amplitude curves for which $-\pi \leq \psi_0 \leq \pi$, and I will consider the more general case, for which $-2\pi \leq \psi_0 \leq 2\pi$, in Section 6.

If one chooses to use the Fig. 3 curves to pseudorandom-encode the coupled modulator, the desired complex function $\mathbf{a}_c$ must be scaled by an appropriate choice of γ [see Eq. (5)]. Consider that if $\gamma > 0.25$, then amplitudes for some values of phase cannot be encoded. However, if $\gamma = 0.25$, then any complex value out to a circular radius of γ can be pseudorandom encoded. Since γ can be as large as 1 for pseudorandom encoding on a phase-only modulator, it can be seen that the diffraction efficiency η of the coupled modulator will be 1/16 of the efficiency of the phase-only modulator [see Eq. (6)]. Of even more concern than efficiency is that by scaling the maximum magnitude to $\gamma = 0.25$, large random phase spreads are needed to encode many of the desired complex values. For instance, for the $\psi_0 = \pi$ curve in Fig. 3, all spreads are between approximately 1.5π and 2π. Thus substantial amounts of random noise can be generated by applying this particular algorithm for a SLM that has the characteristic of Fig. 1. This observation has motivated the development of various modified encoding algorithms (presented below) for which γ can be larger. Of course, if the amplitude coupling is not as strong as that considered in this example, then the value of γ can be made correspondingly larger. Certainly, such an algorithm would be well suited to SLM's that are phase mostly (amplitude variation between 0.9 and 1).[2,3] The example 2 encoding algorithm appears to adapt phase-only encoding better to phase-mostly SLM's than to SLM's for which phase is strongly coupled to amplitude.

There are other options for selection of the amplitude degree of freedom. Consider (for the Fig. 3 results) that γ is between 0.25 and 0.75. For the complex values that cannot be pseudorandom encoded, one or more additional encoding methods would be combined with the pseudorandom-encoding formula. One appropriate encoding technique is the deterministic mapping method that maps complex values to the closest point on the modulation curve.[13] Earlier studies on encoding

'complex-valued composite functions onto phase-only[14] and biamplitude phase[10] modulators have already demonstrated that the blending of deterministic and pseudorandom encoding can produce performance that is better than either. The numerical implementation of blended encoding algorithms is not substantially more involved. However, it is more numerically intensive, since, at present, the only way to specify the scaling of the desired complex values that produces the best encoding performance is to perform the encoding repetitively for different scaling factors and then evaluate the performance of the resulting modulation for each encoding.

Summarizing the Fig. 3 results: It has been shown that the encoding of weakly coupled (phase-mostly) SLM's can be performed with a small amount of additional numerical overhead and a slight reduction in the performance as compared with that of phase-only pseudorandom encoding. For strongly coupled SLM's the reduction in performance can be significant and may call for the development of more numerically involved algorithms that blend individual pseudorandom-encoding algorithms with other encoding algorithms.

Numerical results for example 3. The evaluation of the effective amplitude shows that there are three types of effective-amplitude curves. These are for (I) $\pi/2 \leq \psi_0 \leq 3\pi/2$, (II) $0 \leq \psi_0 \leq \pi/2$, and (III) $3\pi/2 \leq \psi_0 \leq 2\pi$. Similar to the results of example 2, the type I curves descend monotonically from $a(\psi_0)$ to 0 as illustrated in Fig. 4 by the curve $\psi_0 = -\pi/2$. (Because of the periodic assumption, this is also referred to as the $\psi_0 = 3\pi/2$ curve.) With the use of a binomial distribution, zero amplitude is realized for a spread of π as opposed to 2π in example 2. For type II curves (e.g., the $\psi_0 = \pi/4$ curve in Fig. 4), a discontinuity in the effective amplitude is found for spread $\nu = 2\psi_0$. This is due to the discontinuous jump between 0.5 and 1 in the value of the term $a(\psi_0 - \nu/2)$ in Eq. (22). The jump produces a range of values for which there are two solutions for a desired effective amplitude. For some values of ψ_0, the effective amplitude can be even larger than the effective amplitude for zero spread. For the type III curves (e.g., the $\psi_0 = -\pi/8$ curve in Fig. 4), there is a discontinuity at $\nu = -2\psi_0$ [for ψ_0 expressed as a negative angle or $\nu = 2(2\pi - \psi_0)$ for ψ_0 expressed as a phase between π and 2π]. The discontinuity in effective amplitude for any given curve shows that some values between $a(\psi_0)$ and 0 cannot be encoded by the formulas in example 3.

The region of the unit disk for which the desired complex values can be encoded is shown in Fig. 5. Complex values in the clear region can be pseudorandom encoded, and values in the striped region cannot. The dotted curves represent the values to each side of the discontinuity for the type II curves. Note that, for ψ_0 between 0 and slightly less than $\pi/4$, effective amplitudes greater than $a(\psi_0)$ can be encoded. The portion of the striped region forming a peninsula that spirals into the origin corresponds to the unrealizable complex values for the type III curves. Its boundaries correspond to the values on each side of the discontinuity of the type III curves.

For the 4π modulator (with use of the example 3 algorithm), it is possible to scale the complex values so that all values can be pseudorandom encoded. For the results in
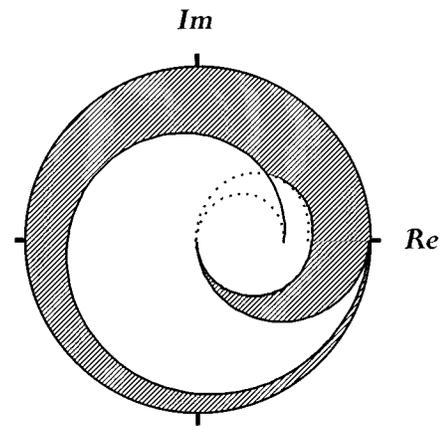

Fig. 5. Map of the effective complex amplitudes that can be pseudorandom encoded (clear region) and those values that cannot be pseudorandom encoded (striped region) on the unit disk by the encoding method from example 3. The amplitude coupling used is the same as the dotted curve in Fig. 1(b), and it is replotted in this illustration. Note that for effective phases between 0 and $\sim\pi/4$ the effective amplitude can be larger than the amplitude of the coupling function. For effective phases between $-\pi/2$ and $\pi/2$, the effective-amplitude curves (Fig. 4) have jump discontinuities. For effective phases between 0 and $\pi/2$, the values on each side of the discontinuity are plotted as dotted curves. For effective phases between $-\pi/2$ and 0, the discontinuities form the boundary of the portion of the unrealizable region that has radii less than the amplitude-coupling function.

this subsection for the 2π modulator, there is no value of γ that will permit the pseudorandom encoding of all desired complex values. This does not mean that the example 3 formulation is less desirable than the formulation in example 2. One advantage of the example 3 method is that it requires a 2π, as opposed to a 4π, modulator. A second advantage is that it encodes a much greater area of the unit disk than the example 2 method. Blending with the deterministic encoding algorithms, as described above, can be used to encode values in the striped peninsular region of Fig. 5. As discussed in Section 2, it is even possible that using deterministic algorithms to encode values outside the unit disk (i.e., $\gamma > 1$) can sometimes produce better performance. Thus deterministic encoding could be used to realize all values on the striped portion of the unit disk, in addition to the complex values that have amplitudes greater than unity.

From the results in this section, it can be seen that some complex values that are encoded by one pseudorandom algorithm may not be encoded by another. This is due to the specification of the form of the pdf's in the derivation of the encoding formulas. Thus, rather than using a deterministic algorithm for these values, it is possible to increase the area of the unit disk that can be encoded by combining two pseudorandom-encoding algorithms. However, as mentioned in Section 2, there are an uncountable number of pdf's that satisfy Eq. (1). This raises the question: What is the total extent of the complex plane that can be encoded by all possible pseudorandom-encoding algorithms? This is answered in Section 6, which also contains a discussion of various approaches for modifying the pseudorandom-encoding algorithms to increase their range.

6. EXTENDING THE RANGE OF PSEUDORANDOM ENCODING

Four methods of extending the complex range are described. For the sake of clarity, the discussion pertains specifically to the examples presented herein, even though the results can be similarly applied to other coupling functions and pdf's. The four methods are (1) for the 4π modulator, using all effective values of phase ψ_0 between -2π and 2π for encoding; (2) for the 4π modulator, blending the example 2 and example 3 pseudorandom algorithms; (3) for either modulator, randomly combining two pseudorandomly encoded values to encode a previously unrealizable value; and (4) for the 2π modulator, using a pseudorandom encoding that does not compensate for amplitude coupling. This last method is also used to evaluate which values can and cannot be implemented by any possible means of pseudorandom encoding.

Method 1. The example 2 pseudorandom encoding algorithm has been applied above to realizing complex amplitudes for effective phases $-\pi \leqslant \psi_0 \leqslant \pi$. The rationale for this is that spread ν can then be varied between 0 and 2π, which allows all values of effective amplitudes between $a(\psi_0)$ and 0 to be encoded for any given effective-amplitude curve (e.g., those shown in Fig. 3). There is a substantially greater area of the complex plane that can be pseudorandomly encoded if the effective complex amplitudes for $\pi \leqslant \psi_0 \leqslant 2\pi$ are also admitted. (Additional solutions for effective phase $-2\pi \leqslant \psi_0 \leqslant -\pi$ are also possible but are not considered here, since these solutions do not increase the area that can be pseudorandomly encoded.) Over this extended range of effective phase, the spread ν can be varied from 0 to a maximum of $2(2\pi - \psi_0)$. The minimum values of effective amplitude calculated by using Eq. (21) are plotted (dashed curve) in Fig. 6. Figure 6 shows that a substantially greater area of the complex plane can be encoded than had originally been considered for the example 2 method. Previously, the complex values had been scaled to a maximum ampli-
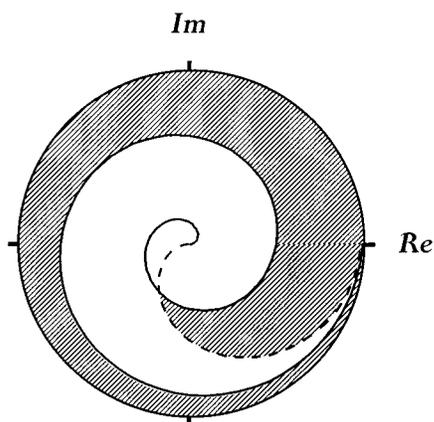

Fig. 6. Map of the effective complex amplitudes that can be pseudorandom encoded (clear region) and those values that cannot be pseudorandom encoded (striped region) on the unit disk by the encoding method from example 2. Additionally, the minimum value of effective amplitude has been calculated for phases between π and 2π by using Eq. (21) (dashed curve). This has increased greatly the number of complex values that can be encoded. The amplitude coupling used is the same as the solid curve in Fig. 1(b), and it is replotted in this illustration.

tude of $\gamma = 0.25$ to use the pseudorandom-encoding algorithm by itself. The point where the dashed curve crosses the effective-amplitude curve $a(\psi)$ is $\psi_0 = 4\pi/3$. Thus the desired amplitudes can be normalized to $\gamma = a(4\pi/3) = 0.33$ instead of $\gamma = 0.25$. Also, a large area (the clear peninsular region) is available for pseudorandom encoding if $\gamma > 0.33$ is used.

Method 2. Although neither the example 2 nor the example 3 encoding algorithm is capable of encoding all complex values, each can encode values that the other cannot. Even though the unrealizable area for the example 2 algorithm is larger than that for the example 3 algorithm, the two algorithms taken together encode an even larger area than the two taken separately. By overlaying Figs. 5 and 6, it can be seen that all complex values having amplitudes of $\gamma = 0.43$ or less can be encoded by combining the two algorithms.

Method 3. Two effective complex amplitudes may be pseudorandomly combined to realize values in the peninsular regions in Figs. 5 and 6. I will show this for the specific condition that both effective amplitudes have the same effective phase ψ_0. This is in keeping with the goal throughout this paper of producing encoding formulas for which a range of effective amplitudes can be specified as a function of spread for any given value of effective phase. In general, the new effective complex amplitude can be written as

$$\langle \mathbf{a} \rangle = d\langle \mathbf{a} \rangle_{\mathbf{u}} + (1 - d)\langle \mathbf{a} \rangle_{\mathbf{l}}, \qquad (26)$$

where the subscripts $\mathbf{u}$ and $\mathbf{l}$ are used to distinguish the upper/larger and lower/smaller effective complex amplitudes. These values could be selected to correspond to the complex amplitudes to each side of the peninsular region at a given effective phase. The value d is the probability of randomly encoding the upper value, and $1 - d$ is the probability of encoding the lower value. Under this set of assumptions, the effective complex amplitude is written as

$$\langle \mathbf{a} \rangle = [da_u + (1 - d)a_l]\exp(j\psi_0), \qquad (27)$$

where a_u and a_l are the individual effective amplitudes. The effective amplitude a_0 for Eq. (27) can be seen to produce any amplitude between a_u and a_l for values of d between 0 and 1. This shows that the method could be used to encode the values in the peninsular region. This method generalizes the biamplitude phase encoding algorithm in Ref. 10 by using effective amplitudes instead of amplitudes from the modulation characteristic. Method 3 can also be viewed as encoding with the use of a different density function. This can be shown by explicitly expressing the effective amplitudes in Eq. (27) as integrals with the use of Eq. (1). The arguments of the integrals can be collected to form a single integral. This permits the identification of the single pdf

$$p(\psi; d) = dp_u(\psi; \psi_0, \nu_u) + (1 - d)p_l(\psi; \psi_0, \nu_l), \qquad (28)$$

where ν_u and ν_l are the two values of spread and p_u and p_l are used to distinguish the two density functions and the new density function p from each other. Although the pdf p is shown to depend on only the parameter d, it also depends on ν_u, ν_l, and ψ_0. These dependencies are not shown because in this encoding method the effective complex amplitudes $\langle \mathbf{a} \rangle_{\mathbf{u}}$ and $\langle \mathbf{a} \rangle_{\mathbf{l}}$ are determined individually, which completely determines p_u and p_l. Then only d needs to be specified to complete the encoding.

If this extension is used, it becomes possible to pseudorandom-encode the entire peninsular regions, in addition to the clear areas in Figs. 5 and 6. If this method is used with the example 2 algorithm, then all complex values of amplitude $\gamma = 0.5$ or less can be encoded by pseudorandom encoding alone. If the example 3 algorithm is used instead, then fully pseudorandom encoding can be used if the complex values are normalized to a maximum amplitude of approximately $\gamma = 0.56$ (on account of the presence of the clear peninsular region in Fig. 5 that covers the phase range of 0 to approximately $\pi/4$).

Method 4. The pseudorandom-encoding range can be extended further by using different pdf's in deriving the encoding formulas. The realizable range can be evaluated by using, as in example 3, a binomial distribution. However, in this evaluation the distribution is not constrained to compensate for bias drift of the effective phase, and the spread is not constrained to be symmetric about the value of effective phase. The evaluation follows from considering the effective complex values that can be encoded by a given pair of complex values $\mathbf{a}_1$ and $\mathbf{a}_2$ on the modulation characteristic. Using the binomial density function in Eq. (4) gives an expression for the effective complex amplitude of

$$\langle \mathbf{a} \rangle = d\mathbf{a}_1 + (1-d)\mathbf{a}_2, \qquad (29)$$

where d is the probability of selecting $\mathbf{a}_1$. The expression is identical to Eq. (26), except that the sample points are constrained to lie on the modulation characteristic. Equation (29) is recognized as the expression for a line as a function of the variable d. Thus any value lying on the line segment between $\mathbf{a}_1$ and $\mathbf{a}_2$ can be encoded. This is illustrated in Fig. 7 for four pairs of points. The three line segments drawn as thin lines all cross at a common point. Thus any one of these segments, or an infinite number of other line segments, could be used to encode this particular complex value. This observation is equally valid for any desired complex value found in the clear region in Fig. 7 that is bounded by the modulation characteristic curve (dashed curve) and the thick line segment. The interior of this boundary is a convex set of all the complex values that can be encoded by the union of all possible pseudorandom-encoding algorithms. Values on the boundary are also realizable, but they have only a single possible solution. Values outside the boundary cannot be realized because the ensemble average of a random phasor never produces a magnitude that is larger than the average magnitude of the phasors in the ensemble.

This simple evaluation is also useful for evaluating the range over which pseudorandom encoding can be applied to modulators that do not produce a full 2π range of phase modulation. For the specific curve and construction in Fig. 7, pseudorandom encoding can be used to encode fully complex functions of any phase and amplitude less than approximately $\gamma = 0.58$. This means that the same function encoded by method 4 would have [by Eqs. (5) and (6)] a diffraction efficiency 5.4 times greater than the example 2 algorithm for which $\gamma = 0.25$. The efficiency of method 4 would also be one third of that for pseudorandom phase-only encoding where $\gamma = 1$.

These results answer the question about the maximum range possible. They also raise a new question about which of the possible solutions to the encoding problem is preferable. The approaches considered to this point emphasize the finding of formulas that are simple to implement and that are numerically efficient. For the example 3 algorithm, in which amplitude compensation was used to simplify the implementation, no more than two binomial distributions (i.e., line segments) could be found to realize a desired complex value, and for some desired and realizable complex values no distribution could be found. This section shows that the limitations introduced by these particular assumptions can be substantially reduced by the use of Methods 1–3. The benefit of simple implementations is that they provide great flexibility and easy access to the complex modulating properties of a SLM in an environment that can require real-time programming of the SLM. If the highest levels of optical performance are required and computation time is not a significant concern, then there are many numerically intensive design algorithms that can produce near-optimal optical performance. These considerations have led to my emphasis on simply implemented algorithms. However, it would be possible instead to derive
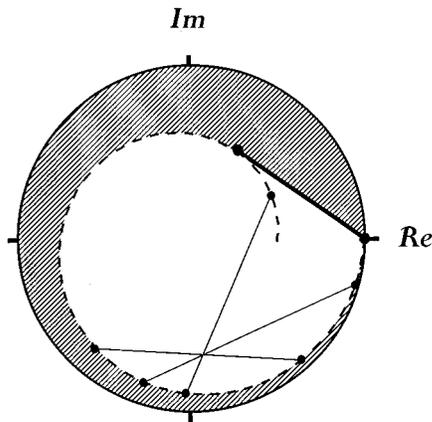

Fig. 7. Map of the effective complex amplitudes that can be pseudorandom encoded (clear region) and those values that cannot be pseudorandom encoded (striped region) on the unit disk by use of all possible binomial distributions. The amplitude coupling (dashed curve) is the same as the dotted curve in Fig. 1 (b). For a given pair of samples on the amplitude-coupling curve, any effective complex amplitude can be realized on the line segment connecting the two points. This is shown for four pairs of samples. Any of the three thin lines could be used to produce the same effective complex amplitude at their common intersection. The thick line segment and the amplitude-coupling curve bound a convex set of all complex values that can be realized by pseudorandom encoding for the particular modulator characteristic.

pseudorandom-encoding algorithms on the basis of minimizing the random errors produced by encoding itself. This is the subject of Section 7. The resulting algorithms appear to be more numerically involved, and, for this reason, it is not my intent to recommend them. Instead, the minimum-error (and also the maximum-error) pseudorandom encoding formulas are used to bound the errors produced by amplitude-compensated algorithms.

7. ERROR ANALYSIS OF PSEUDORANDOM ENCODING AND MINIMUM-ERROR ENCODINGS

Encoding error defined. A general error analysis of pseudorandom encoding, not only for phase-only but also for any SLM modulation characteristic is presented in Ref. 6. Expressions are presented that, in addition to describing the expected intensity, describe the standard deviation of the intensity pattern. For the discussions in this section, I evaluate only the expected intensity, which is given in Eq. (3). The noise, or encoding error, has been identified in Section 2 as the second term in Eq. (3). The ith element of this summation corresponds to the encoding error produced by the ith SLM pixel. The evaluation is most directly performed and the key results are most readily apparent under the assumption that the pixel aperture is an ideal impulse (Ref. 6 can be consulted for further analyses that include finite-width apertures). For this set of assumptions, the error for a single pseudorandom-encoded pixel is

$$\epsilon = \int a^2(\psi) p(\psi) \mathrm{d}\psi - |\mathbf{a}_c|^2. \tag{30}$$

Since the desired complex amplitude is fixed, the integral is the only term that can be minimized. The integral is minimized by individually minimizing the integrand at each value of ψ. Since $a(\psi)$ is a given function, the minimum is produced by selecting $p(\psi)$ to be minimum where $a(\psi)$ is maximum. The minimization is subject to the two constraints that (1) the integrated area of the density function is unity and (2) the value $\mathbf{a}_c = \langle \mathbf{a} \rangle$ in Eq. (4) has a solution. The generality of Eq. (30), together with the constraints, makes it difficult to draw additional conclusions about minimum-error pseudorandom encodings without further specialization of the problem.

Consider the specific problem of determining which binomial distribution produces the least encoding error (as illustrated in Fig. 7). Under this set of constraints, the encoding error in Eq. (30) can be expressed as

$$\epsilon = d(1 - d)[a_1^2 + a_2^2 - 2a_1 a_2 \cos(\psi_1 - \psi_2)]$$
$$\text{subject to } \mathbf{a}_c = d\mathbf{a}_1 + (1 - d)\mathbf{a}_2, \tag{31}$$

where $\mathbf{a}_i \equiv (a_i, \psi_i)$. The error can be minimized by separately minimizing the product of the binomial probabilities and minimizing the term in brackets. The product $d(1 - d)$ is minimized by maximizing the distance between d and $1 - d$. The term in brackets is the familiar formula for the law of cosines. It is minimized by minimizing the difference between the two phases. The constraint in Eq. (31) still makes it difficult to draw a useful conclusion. If this constraint is substituted into the

encoding error in Eq. (31) to eliminate the value of probability d, then this equation reduces to

$$\epsilon = |\mathbf{a}_c - \mathbf{a}_1||\mathbf{a}_c - \mathbf{a}_2|. \tag{32}$$

Written in this form, the pseudorandom-encoding error can be directly interpreted as the product of the lengths of the line segments $\mathbf{a}_1 - \mathbf{a}_c$ and $\mathbf{a}_c - \mathbf{a}_2$. For a fixed length segment $\mathbf{a}_1 - \mathbf{a}_2$, the product is minimized by making one of the two segments as large (or as small) as possible. However, the length of $\mathbf{a}_1 - \mathbf{a}_2$ generally varies, depending on the form of the amplitude-coupling function $a(\psi)$. Thus further analysis is usually required to find the solution that produces a minimum error.

Numerical analysis of encoding errors for binomial distributions. Minimum-error encoding formulas have been determined numerically for a number of desired complex values for the modulation characteristic in Fig. 7. The method of finding the minimum-error encoding formula for a single complex value consists of repeated evaluations of the error ϵ by Eq. (32). The steps are the following: (1) A value of $\mathbf{a}_c$ is specified; (2) a value of phase ψ_1 is specified; (3) the complex amplitude $\mathbf{a}_1 = a(\psi_1)\exp(j\psi_1)$ is calculated; (4) $\mathbf{a}_2$ is found by solving for the point on the line through $\mathbf{a}_1$ and $\mathbf{a}_c$ that intersects $\mathbf{a}_2(\psi_1) = a(\psi_2)\exp(j\psi_2)$; (5) the error is found by using Eq. (32). Steps 2–5 are repeated for all values of ψ_1. From these results the values of ψ_1 and ψ_2 are determined that produce the minimum error for the encoding of the value $\mathbf{a}_c$. Step 4 requires the solution of the nonlinear equation $\mathbf{a}_2(\psi_2) = \mathbf{a}_1 + x(\mathbf{a}_c - \mathbf{a}_1)$. Separately equating the real and imaginary parts of this expression gives two equations in the two unknowns x and ψ_2. These values were solved by using the "find" function in the software package MATHCAD (Mathsoft Inc., Cambridge, Mass. 02142).

The complex value $\mathbf{a}_c = (0.67, 1.6\pi)$ corresponds to the intersection of the three thin line segments in Fig. 7. The minimum-error encoding formula corresponds to a line segment that contains $\mathbf{a}_c$ and the sample point at $(0.5, 0)$. In fact, for this particular coupling characteristic, one of the two sample points is usually $(0.5, 0)$. In all other cases one of the two points is $(1, 2\pi)$. This corresponds to the encoding of those effective amplitudes that exceed $a(\psi_0)$ (i.e., for effective phase between 0 and approximately 0.32π). This procedure can be summarized in the following way: Pick the amplitude of one of the sample points to be as small as possible. This is true even if one sample point is $(1, 2\pi)$. In this case the other sample point is the minimum possible.

Comparison of minimum-error and amplitude-compensated encoding algorithms. The encoding error produced by the example 3 algorithm can be evaluated by solving Eq. (24) for the spread ν for which $\mathbf{a}_c = (a_0, \psi_0)$. Then the complex values on the modulation characteristic, $\mathbf{a}_1$ and $\mathbf{a}_2$, are calculated for the phases $\psi_0 \pm \nu/2$. Using these two sample values in Eq. (32) gives the encoding error ϵ_c for encoding the value $\mathbf{a}_c$ by the compensation method. These errors are compared with minimum encoding error $\epsilon_{\min}$ in Table 1 for selected values of $\mathbf{a}_c$. Errors up to three times larger than the minimum error are produced by the compensation method. However, individual errors can be much closer

in value to the minimum error. For an effective amplitude of 0.68 in the table, the compensated encoding has a spread of $\nu = \pi/10$. Since one sample point is $(1, 2\pi)$, the compensated encoding coincides with the minimum-error encoding. The additional error produced by the compensation method can be appreciated by considering the errors produced for $a_0 = 0$. The phase of a desired complex value need not be specified for zero-valued amplitudes, but, depending on the specification of ψ_0, the encoding error varies from 0.375 for the sampling points $(0.5, 0)$ and $(0.75, \pi)$ to 0.75 for the sampling points $(0.75, \pi)$ and $(1, 2\pi)$.

The search procedure used to identify the minimum encoding error has also been used to identify the maximum possible encoding error $\epsilon_{\max}$. The maximum-error solution usually corresponds to one of the sample points being $(1, 2\pi)$. For the case $a_0 = 0.68$ in Table 1, for which the minimum-error solution uses the $(1, 2\pi)$ sample, the maximum-error solution is found by selecting the smaller of the two sample points to have the smallest amplitude possible. This corresponds to the line segment intersecting $\mathbf{a}_c$ that is tangent to the modulation characteristic. In Table 1 the encoding error for the compensated algorithm, ϵ_c, approaches the upper limit $\epsilon_{\max}$ closely for $\psi_0 = 1.6\pi$ and 1.75π. The solution for the compensated encoding of these two points includes one sample point that is close to $(1, 2\pi)$, whereas a sample point at $(0.5, 0)$ is needed for minimum error.

For comparison I also include in Table 1 the encoding errors for a phase-only SLM. As with phase-only modulation in general, for pseudorandom phase-only modulations there is a conservation of the energy diffracted from the modulation to the Fraunhofer plane.[6] For this reason the encoding error for phase-only modulation is $\epsilon_{\mathrm{po}} = 1 - a_0{}^2$. (This result is independent of the particular phase-only pseudorandom-encoding algorithm. Also note that the square root of this error is identical to the magnitude of the error vector of the parity sequence method.[8]) Since the phase-only characteristic has a greater radius than that of the coupled curve, it is not surprising that the errors for the phase-only characteristic are larger than those for $\epsilon_{\max}$.

Consideration of other coupling functions. It is not immediately obvious if the minimum-error solution requires that one of the two sample points have minimum radius for other types of curves. Various types of coupling functions have been examined numerically to determine if the result can be generalized and if there are counterex-

amples. The solution has been examined for amplitude coupling in the form of a power law $a(\psi) = \alpha[\psi/(2\pi)]^r + \beta$, where $\alpha + \beta = 1$ and $0 \leq \psi \leq 2\pi$. The exponent r was varied between 0.2 and 3. Although this testing is not exhaustive, for each desired complex value the minimum-error encoding once again used the sample point at either $(\beta, 0)$ or $(1, 2\pi)$. For each of these coupling functions, the amplitude increases monotonically with phase.

In search of a counterexample, other coupling functions have been examined that are nonmonotonic. For simulations using these functions, neither of the two sample points is minimum in amplitude for minimum-error encoding. One of these coupling functions that provide a counterexample is $a(\psi) = 0.75 + 0.25 \cos \psi$. Another function is that for the phase-only characteristic. It has the property that every pair of sample points that are collinear with the desired complex value produce the identical amount of encoding error. The geometry is unchanged by a shift of the origin, so that the encoding error as calculated by Eq. (32) remains constant as long as the modulation characteristic is perfectly circular on the complex plane.

Summary. This section has described a method for encoding based on minimizing pseudorandom encoding error. The method has been specialized so that only binomial distributions are admitted. An even more involved problem would be the search for an arbitrary density function that minimizes the encoding error. This possibility has yet to be considered. Also, the higher-order moments (e.g., the variance of the intensity[6]) provide additional information on the noise and the errors in the diffraction pattern. This information could be used to decide between two encodings that generate the same amount of noise ϵ (as is the case for phase-only and other circular characteristics). Although much more analysis is possible, the analysis procedure presented here does permit comparisons of encoding error for the compensated algorithms with the minimum- and maximum-error limits. The compensation method, although it does produce more error, is more conducive to quick implementation. Furthermore, for modulation characteristics that are closer to circular, the amount of error produced by the amplitude compensation method would approach the minimum error even more closely.

8. DEMONSTRATIONS OF THE ENCODING ALGORITHMS

A better appreciation of the usefulness and the performance of the pseudorandom-encoding algorithms for coupled modulators can be gained by comparing their ability to encode a desired complex-valued function. This is done by way of computer simulation. The identical desired function is used for each encoding algorithm. The diffraction pattern of the desired function and of each encoded function is simulated and plotted on comparable scales. Three of the encodings use algorithms for coupled SLM's that are developed in this paper, and two encodings are previous methods that are included for comparison.

Table 1. Encoding Errors for Minimum-Error, Compensated, Maximum-Error, and Phase-only Pseudorandom Encoding for Various $\mathbf{a}_c$

a_0	ψ_0/π	ν/π	$\epsilon_{\min}$	ϵ_c	$\epsilon_{\max}$	ϵ_{po}
0.80	1.75	0.35	0.09	0.24	0.24	0.36
0.68	0.05	0.10	0.06	0.06	0.18	0.54
0.67	1.60	0.46	0.19	0.36	0.37	0.55
0.60	0.75	0.32	0.10	0.11	0.15	0.64
0.40	0.00	0.62	0.12	0.37	0.70	0.84
0.05	0.75	0.95	0.38	0.46	0.73	1.00
0.00	—	1.00	3/8	3/8 to 3/4	3/4	1.00

Procedure. For this study the SLM is assumed to be composed of 128 × 128 pixels. It is modeled as an array of 128 × 128 samples. The desired fully complex function is formed by adding together subarrays of sizes 128 × 32, 64 × 64, and 32 × 128. Each subarray has amplitude that is constant and phase that varies linearly with position. Thus the Fourier transform of each subarray would be a two-dimensional sinc function centered at a point that is determined by the slope of the phase. The phase slopes are chosen so that diffracted spots will be centered on the horizontal axis. The subarrays are positioned so that only two subarrays overlap at any position in the modulator plane. This gives the resulting ampli-
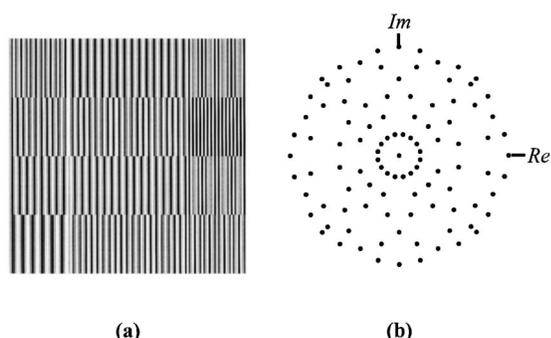
tude modulation in Fig. 8(a) the appearance of vertical interference fringes. The fringe spacing varies, depending on which pair of subarrays overlap. It is desired that each spot have an identical peak intensity, and for this reason the magnitudes of each subarray are identical. This leads to modulation amplitudes that vary from 0 to γ. [Also note that for $\gamma = 1$ the diffraction efficiency, with the use of Eq. (6), is 0.5.]

The resulting diffraction pattern is simulated by placing the 128 × 128 array of complex numbers in a 512 × 512 array of zeros and then performing the fast Fourier transform. A gray-scale rendition of intensity is shown in Fig. 9(a) for the central 65 × 512 samples of the fast Fourier transform, and an intensity plot is shown in Fig. 10(a) for the central 1 × 512 samples. The gray-scale rendition is saturated so that fully white corresponds to a level that is 15% of the peak intensity. Although the Fourier transform of each of the subarrays is a sinc function, some interference from sidelobes is also evident. Nonetheless, the eight spots are nearly identical in intensity.

Three pseudorandom-encoding algorithms for coupled SLM's are implemented by assuming the same two SLM coupling characteristics as those used in Sections 5–7. The first algorithm is the example 2 algorithm with use of the 4π coupling characteristic (the solid curves in Fig. 1). The magnitude of the desired function is scaled so that $\gamma = 0.25$. The second algorithm is the example 3 algorithm with use of the 2π coupling characteristic (the dotted curves in Fig. 1). The algorithm is augmented with method 3 to handle values that lie in the unrealizable

(a)   **(b)**

Fig. 8. Plots of the desired fully complex function $\mathbf{a}_c$ used in the simulation study: (a) desired magnitudes a_c as they would appear on the 128 × 128-pixel SLM, (b) desired values $\mathbf{a}_c$ shown on the complex plane. Each value shown occurs numerous times.
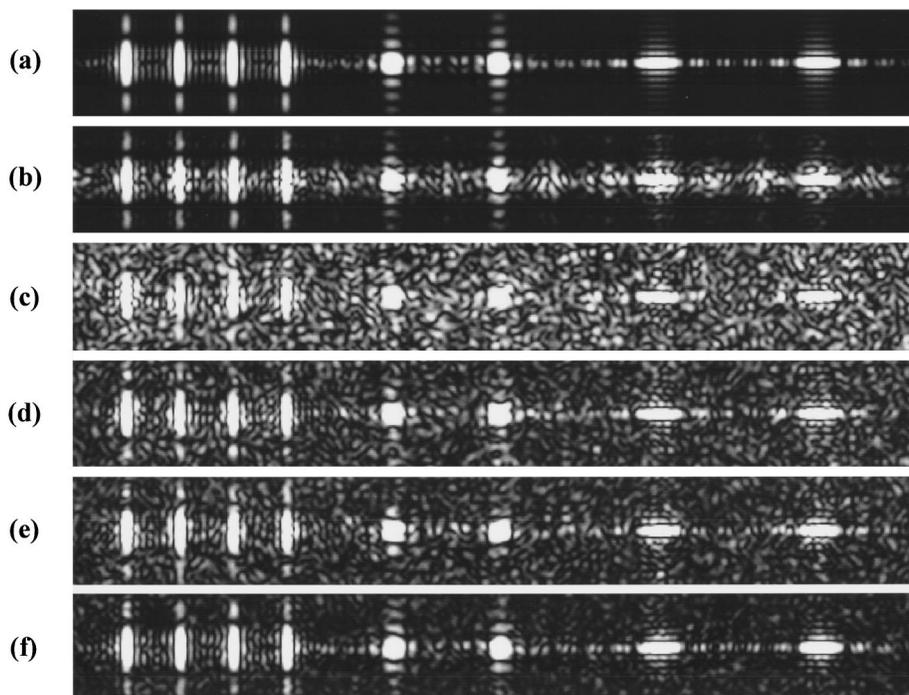

Fig. 9. Gray-scale plots of the intensity of the diffraction patterns resulting from (a) the desired fully complex function and (b)–(f) the various algorithms b–f. The algorithms are described in Section 8. Each image shows the central 65 × 512 samples of the simulated 512 × 512-sample diffraction pattern. In each image a fully white gray scale corresponds to an intensity that is 15% of the maximum spot intensity plus the minimum spot intensity divided by 2 [(max + min)/2].
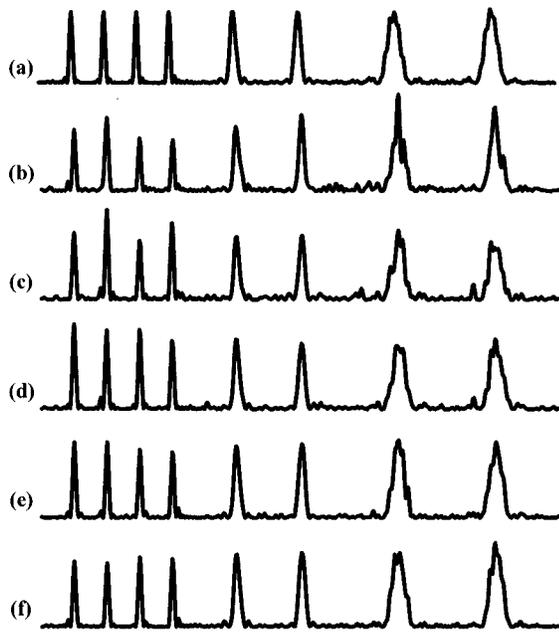
Fig. 10. Cross sections of the diffraction patterns resulting from (a) the desired function and (b)–(f) the various algorithms b–f. Each cross section is the central 1×512 samples of the simulated 512×512-sample diffraction pattern. Each trace is normalized so that the maximum spot intensity plus the minimum spot intensity divided by 2 [(max + min)/2] is of identical length on the vertical scale of each plot.

peninsular region of Fig. 5. The two possible values, a_l and a_u, that are used in Eq. (27) are chosen to be the upper and lower boundaries of the peninsula for a specified value of ψ_c. One or the other of these two boundary values (depending on the value of a random number and d) is then encoded by example 3. The desired function is scaled so that $\gamma = 0.55$. The third algorithm is the minimum-error encoding for binomially selecting one of two possible points. This algorithm is equivalent to method 4 with use of the minimum-error selection criteria given in Section 7. The amplitude coupling is also the 2π coupling characteristic (dotted curves of Fig. 1), and the scaling is $\gamma = 0.49$. For convenience I will refer to these three algorithms as c, d, and e, respectively, which are the same labels as those used for the corresponding results in Figs. 9 and 10.

There are two other algorithms that have been implemented for purposes of comparison. These are a deterministic algorithm that is somewhat similar to the phase-only filter, except that it is applied to the 2π characteristic of Fig. 1, and a pseudorandom encoding for a phase-only characteristic. These algorithms will be referred to as b and f. As with the phase-only algorithm, the nonrandom algorithm sets the actual phase ψ to the desired phase ψ_c. Thus the implemented modulation is $\mathbf{a} = a(\psi_c)\exp(j\psi_c)$. The pseudorandom phase-only encoding is identical to example 3 with the coupling characteristic set to $a(\psi) = 1$ for all values of ψ. The relationship between effective amplitude and spread [Eq. (24)] is plotted in Fig. 4 (thin solid curve). The spread is directly calculated as $\nu = \arccos(2a_c)$. For the phase-only algorithm the desired complex function is scaled so that $\gamma = 1$.

After each of the five encodings is performed, the data are zero padded and fast Fourier transformed by the identical procedure used for the desired fully complex function. Each gray-scale image in Fig. 9 is saturated to approximately 15% of its peak intensity. The maximum (max) and minimum (min) intensities of the eight peaks in the corresponding trace in Fig. 10 are found. Then the peak gray scale for the gray-scale image in Fig. 9 is set to 15% of (max + min)/2. Figure 10 is plotted so that the vertical scale is identical for the quantity (max + min)/2. The peak fluctuations are used to calculate a uniformity measure as well. This is $u = (\text{max} + \text{min})/(2 \times \text{max})$.

A few comments on the numerical implementation of the algorithms may be of interest. The algorithms are all implemented by using MATHCAD whiteboarding software running on a 100-MHz Pentium personal computer with 32 Mbytes of memory and a Windows 95 operating system. The time required to encode the 128^2 desired complex values takes of the order of 2 s for algorithms b, c, and f and of the order of 3 min for algorithms d and e. No special effort has been made to optimize the implementations for computational speed or real-time implementation. Furthermore, since MATHCAD is an interpreter and a graphical interface, rather than a compiled language, the speeds reported here are not representative of what is possible with compiled code. To ensure consistency in comparing the various encodings, the identical 128^2 random numbers are used for each type of encoding. Algorithms b and f are directly implemented with standard functions, so no further discussion of their implementation is given.

Algorithm c requires that the amplitude of Eq. (21) be inverted to specify ν. This can be done by using a one-dimensional nonlinear equation solver or by developing a lookup table. Based on the smoothness of the curves in Fig. 3, the latter option has been pursued. A two-dimensional spline-fitting function is available in MATHCAD. It requires that the sample coordinates be rectangularly spaced in two dimensions. The way that this condition has been obtained is the following: (1) For a fixed value of desired phase ψ_c, calculate the magnitude of Eq. (21) divided by $a(\psi_c)$ for 11 values of ν between 0 and 2π. This produces values of $x = a_c/a(\psi_c)$ that are normalized between 0 and 1. (2) This calculation is performed for 11 values of ψ_c between $-\pi$ and π to yield 121 values. (3) The 11 pairs of values for each fixed value of ψ_c are fitted with a one-dimensional spline to produce 11 spline fits of ν as a function of x. (4) Each spline fit is then interpolated at identical values of x. This produces 121 values of ν on rectangular coordinates of x and ψ_c. (5) These values are then fitted by a two-dimensional spline. The appropriate value of ν is then found by supplying values of ψ_c and $x = a_c/a(\psi_c)$ to the MATHCAD two-dimensional spline interpolation routine. Such routines actually use a lookup table to localize the value of the function followed by interpolation to refine the accuracy of the value.

The development of a two-dimensional spline fit proved too difficult to apply to algorithm d. One problem is that desired/effective amplitude a_c is not always a monotonic function of ν (see Fig. 4). Thus it is not possible to ex-

change the ordinate and the abscissa, and then spline-fit ν against a_c over the full π range. A nonlinear equation solver has also been tried, but the possibility of two solutions for ν and the sharp discontinuity in the curves frequently lead to nonconvergence to a solution. A method that always seems to produce a solution is to limit the fit (and subsequently to interpolate) only over ranges for which a_c is a monotonic function of ν. These various ranges have been evaluated in Section 5 and are identified in Fig. 5. The description of the logical selection of the ranges is somewhat tedious. I will describe only one case for illustration. If ψ_c is a fixed value between $3\pi/2$ and 2π, a_c is a monotonic function for all values of ν between 0 and $2(2\pi - \psi_c)$. (The maximum value of a_c corresponds to the lower bound of the unrealizable peninsular region in Fig. 5.) A spline fit of ν as a function of a_c can be made over this range. Although it is numerically inefficient, the fit is repeated for each new value of $\mathbf{a}_c$ as follows: (1) For a given value of $\mathbf{a}_c$, the appropriate range of ν is identified; (2) 11 values of a_c are calculated over the range of ν for the given value ψ_c; (3) a spline fit of ν as a function of the a_c is performed; and (4) a spline interpolation is performed to calculate the required value of ν.

For algorithm e a nonlinear equation solver is employed to find the two sample points that pseudorandom-encode the desired complex value with a minimum error. The result from Section 7 is used that one of the two sample points of the minimum-error solution is usually (0.5, 0). By limiting the scale factor so that γ is less than 0.5 (as has been done), it becomes possible to avoid consideration of the other possibility [that one sample point can sometimes be (1, 0)] and always use (0.5, 0). Under this condition the equation solver converges for all 128^2 complex values. For γ of 0.5 or somewhat greater, the equation solver did not always converge. The solution is essentially identical to the result of the procedure given in Section 7. Given the points $\mathbf{a}_c$ and (0.5, 0), the solver finds the intersection point of the common line through these two points and the modulation characteristic.

Discussion of results. Figures 9 and 10 show that all five methods produce diffraction patterns that are similar to the desired diffraction pattern. Algorithm b, the deterministic algorithm, is the least uniform, having a uniformity of $u = 0.76$. The pseudorandom algorithms c, d, e, and f have greater uniformities, of $u = 0.82, 0.88, 0.92$, and 0.88, respectively. The uniformity of algorithm e may be overstated, since the peak intensity of its diffraction pattern (unlike the other curves) is on a line other than the one plotted in Fig. 10. If this value had been used instead, then u would be calculated to be 0.89. The key result to note is that the pseudorandom algorithms lead to more uniform or accurate reconstructions of the desired function than does the deterministic method. This correspondence between pseudorandom and nonrandom methods has been seen to a greater[10] or lesser[14] degree depending on the specific functions that are being encoded. A second observation is that, of the pseudorandom encodings c–e, c, which uses a low value of γ, is much less uniform than d and e, which use larger values of γ.

The algorithms are also compared on the basis of back-

ground noise. Algorithm b generates more intense noise than algorithms d and e and levels that are comparable with those of algorithm c. However, away from this axis the noise for algorithm b is substantially weaker than that for any of the other algorithms. More severe background noise has been noted in Ref. 10 for two-dimensional spot arrays produced by deterministic encoding. In the pseudorandom algorithms the background noise of algorithm c, as shown in Fig. 9, is much lower than that for algorithms d and e. Algorithm f has even lower levels of background noise. The background noise in algorithms c–f has the appearance of speckle, as expected. The higher levels of speckle noise in algorithm c are due to the small value of γ, which leads to a significant number of large phase spreads in the design algorithm.

The most pleasing result of this simulation study is that algorithms d and e, which are applied to SLM's with significant degrees of amplitude coupling, perform nearly as well as pseudorandom encoding (algorithm f) on phase-only SLM's. This result shows that pseudorandom encoding is of more than mathematical interest and that it has its uses in control of today's SLM's.

9.   SUMMARY AND CONCLUDING REMARKS

The concept of extending pseudorandom encoding from phase-only to amplitude-coupled phase modulators has been explored. Including amplitude compensation in the probability density function has been used to derive solutions over a continuous range of effective amplitudes for a given value of effective phase. The encoding formulas found are only slightly more involved than those previously found for phase-only modulators. The example 3 algorithm, which uses the binomial distribution, is useful in that it is adaptable to a wide variety of amplitude-coupling characteristics. In developing an encoding algorithm, it is necessary to determine the range of values that can be encoded. Modifications (e.g., blending of multiple encoding algorithms) that extend this range can lead to improved performance. The maximum range that can be encoded by all possible pseudorandom algorithms has been identified as a convex region that is bounded (in part) by the modulation characteristic. This region has been identified by use of the properties of the binomial distribution. This analysis with the binomial has also been used to illustrate how various distributions can encode the same value, though with differing levels of error. The topics covered in this paper provide a framework for the development of pseudorandom-encoding algorithms for various coupled modulators.

Compared with previous encoding algorithms, pseudorandom encoding is novel in that it permits the direct encoding of complex-valued information to one pixel at a time with only a limited consideration of the settings of neighboring pixels. This has great utility for programming SLM's with fairly arbitrary complex-valued functions in real time. The idea that a limited set of modulation values can represent a continuum on the complex plane is in some sense parallel to Shannon's concepts on communication in the presence of noise.[15]   In Shannon's

theory the information content of a signal space is determined by the dimensionality of the signal space (i.e., the number of ways that a signal can be represented) divided by the volume occupied in this space by noise. In optical processing, the modulation characteristic of SLM's limits the area of the complex-valued space that can be addressed. With pseudorandom encoding, a continuum of the complex space can be addressed, but, as a result, white noise is generated. The proposed improvements involving the blending of various pseudorandom and other algorithms appear to be aimed at further increasing the available signal space.

ACKNOWLEDGMENTS

REFERENCES

1. R. D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," Appl. Opt. **28**, 4865–4869 (1989).
2. C. Soutar and S. E. Monroe, Jr., "Selection of operating curves of twisted-nematic liquid crystal televisions," in *Advances in Optical Information Processing VI*, D. R. Pape, ed., Proc. SPIE **2240**, 280–291 (1994).
3. L. G. Neto, D. Roberge, and Y. Sheng, "Programmable optical phase-mostly holograms with coupled-mode modulation liquid crystal television," Appl. Opt. **34**, 1944–1950 (1995).
4. B. R. Brown and A. W. Lohmann, "Complex spatial filter," Appl. Opt. **5**, 967 (1966).
5. J. P. Kirk and A. L. Jones, "Phase-only complex-valued spatial filter," J. Opt. Soc. Am. **61**, 1023–1028 (1971).
6. R. W. Cohn and M. Liang, "Approximating fully complex spatial modulation with pseudo-random phase-only modulation," Appl. Opt. **33**, 4406–4415 (1994).
7. R. W. Cohn, A. A. Vasiliev, W. Liu, and D. L. Hill, "Fully complex diffractive optics by means of patterned diffuser arrays," J. Opt. Soc. Am. A **14**, 1110–1123 (1997).
8. D. C. Chu and J. W. Goodman, "Spectrum shaping with parity sequences," Appl. Opt. **11**, 1716–1724 (1972).
9. D. C. Chu and J. R. Fienup, "Recent approaches to computer-generated holograms," Opt. Eng. **13**, 189–195 (1974).
10. R. W. Cohn and W. Liu, "Pseudorandom encoding of fully complex modulation to bi-amplitude phase modulators," in *Diffractive Optics and Microoptics*, Vol. 5 of 1996 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1996), pp. 237–240.
11. R. W. Cohn and M. Liang, "Pseudorandom phase-only encoding of real-time spatial light modulators," Appl. Opt. **35**, 2488–2498 (1996).
12. A. Papoulis, *Probability, Random Variables, and Stochastic Process*, 3rd ed. (McGraw-Hill, New York, 1991), Chap. 5, pp. 101–102 and Chap. 8, pp. 226–229.
13. R. D. Juday, "Optimal realizable filters and the minimum Euclidean distance principle," Appl. Opt. **32**, 5100–5111 (1993).
14. L. G. Hassebrook, M. E. Lhamon, R. C. Daley, R. W. Cohn, and M. Liang, "Random phase encoding of composite fully-complex filters," Opt. Lett. **21**, 272–274 (1996).
15. C. E. Shannon, "Communication in the presence of noise," Proc. IRE **37**, 10–21 (1949).