Ternary pseudorandom encoding of Fourier transform holograms

Robert W. Cohn and Markus Duelli

The ElectroOptics Research Institute, University of Louisville, Louisville, Kentucky 40292

Received June 1, 1998; revised manuscript received September 10, 1998; accepted September 14, 1998

Pseudorandom encoding is a statistical method for designing Fourier transform holograms by mapping ideal complex-valued modulations onto spatial light modulators that are not fully complex. These algorithms are notable because their computational overhead is low and because the space–bandwidth product of the encoded signal is identical to the number of modulator pixels. All previous pseudorandom-encoding algorithms were developed for analog modulators. A less restrictive algorithm for quantized modulators is derived that permits fully complex ranges to be encoded with as few as three noncollinear modulation values that are separated by more than 180° on the complex plane. © 1999 Optical Society of America [S0740-3232(99)02001-3]

OCIS codes: 230.6120, 090.1760, 030.6600, 070.0070.

1. INTRODUCTION

A. Rationale for Using Pseudorandom Encoding for the Design of Fourier Transform Holograms

The first computer-generated hologram (CGH) solved the problem of representing complex-valued modulations with a binary amplitude-only transmittance.[1] High-quality reconstructions were possible because of the high spatial resolution of large-area plotters (followed by successive photographic reductions). Since then various CGH algorithms have been developed in response to the particular physical properties of the modulating medium—physical properties such as modulating type: amplitude-only, phase-only, or coupled amplitude–phase modulation; modulation levels: continuous or quantized; spatial structure: continuous or discretely sampled; spatial resolution/space–bandwidth product: low to high; and update rate: fixed-pattern to programmable in real time.[2]

CGH algorithms are also shaped by the intended application. For example, for today's fixed-pattern diffractive optic Fourier transform holograms that are replicated *en masse*, there is no major time constraint in employing design algorithms that use numerically intensive optimizations and search strategies. However, if individual custom-designed CGH's are to be used by a large customer base (e.g., a unique CGH for each holder of a national credit card), then the amount of time required to design (and also to fabricate) each CGH should be on the order of 1 s.[2] Computationally intensive design algorithms may also not be appropriate for many optical processors based on spatial light modulators (SLM's); especially, adaptive processors that incorporate new information into newly designed SLM modulations on the fly. It is the later time-critical applications that the algorithms presented in this paper are designed to address.

Methods that we refer to as encoding are especially suitable for fast design of modulations because they calculate the mapping between each desired complex value and each modulator pixel in sequence [see Fig. 1(a)]. Encoding was the principal method of designing CGH's before 1973. With the work of Gallagher and Liu on iterative encoding,[3] there has been a continuing use and refinement of computationally intensive optimization and global search methods to design Fourier transform holograms. For time-critical applications we believe that the advantages of speed and flexibility of encoding are preferable to the performance advantages (especially, diffraction efficiency) of the slower iterative methods.

A second important aspect of encoding onto real-time SLM's is that today's SLM's have far fewer pixels (i.e., space–bandwidth product) than their earlier counterparts, the fixed-pattern CGH pen plots. For the earlier CGH algorithms, it was reasonable to cluster or group pixels together, thereby reducing the space–bandwidth product of the encoded signal by a factor equal to the number of pixels in each group. However, given the low pixel count and the relatively large cost of current SLM's, it is important to utilize as much of the space–bandwidth product of the SLM as possible.

These two considerations on computational speed and bandwidth utilization led to the development of pseudorandom encoding,[4–6] a class of algorithms that encode individual complex values to individual SLM pixels. Since each given value encodes to an individual pixel rather than a group, the space–bandwidth product of the modulation (for periodically sampled SLM's) is identical to the number of pixels in the SLM.

B. Developments in Pseudorandom Encoding Leading toward Ternary Pseudorandom Encoding

The pseudorandom-encoding process represents desired fully complex values (on SLM's that do not produce a complete set of complex values) through the statistical approximation known as the law of large numbers.[7] A unique random distribution of the available pixel modulations is specified for each desired complex value such that the average modulation equals the desired complex value. Under this set of conditions, the resulting Fourier plane intensity pattern will, on average, produce the desired Fourier plane diffraction pattern plus a broadly

 © 1999 Optical Society of America

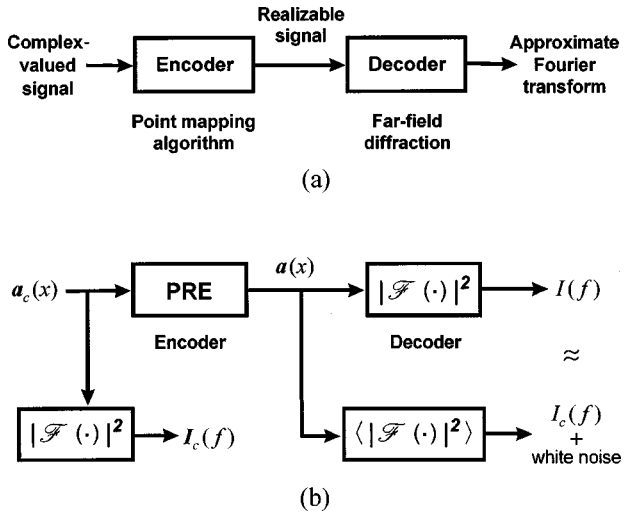


Fig. 1. Systems definition of complex-valued encoding. (a) Systems viewpoint of the Fourier transform computer-generated hologram. A desired complex-valued signal is encoded into a realizable SLM modulation. This signal is decoded through diffraction into a spectrum that approximates the desired complex-valued spectrum. (b) Systems viewpoint specialized for pseudorandom encoding. The desired complex-valued signal $\mathbf{a}_c(x)$ is pseudorandom encoded (PRE) to produce the realizable modulation $\mathbf{a}(x)$. The observed intensity diffraction pattern $I(f)$, which is the squared magnitude of the Fourier transform of $\mathbf{a}(x)$, approximates the desired intensity diffraction $I_c(f)$ in a statistical sense. Specifically, the expected value of $I(f)$, i.e., $\langle I(f)\rangle$, is the desired diffraction pattern $I_c(f)$ on a background of white noise.

spread pedestal that represents the average level of background noise[4] [see Fig. 1(b)]. According to the law of large numbers, the actual diffraction pattern will approximate the average pattern with increasing accuracy as the number of pixels in the SLM is increased.[8] The background noise in the actual pattern is a speckle pattern, which can be either negligible or dominant, depending on the specific complex-valued function that is to be encoded. Some simple metrics calculated from the desired complex function have been described that can be used to provide designers advance knowledge about the quality of each particular encoding.[4,8] Also, in Section 5, a more generally applicable model of signal-to-noise ratio (SNR) is developed.

Until now, pseudorandom-encoding algorithms have always been derived under the assumption that the SLM produces a continuous range of values (e.g., phase-only or coupled amplitude). In one study a continuous (phase-only) modulation has been augmented with a single quantized (zero-amplitude) modulation,[5] but this algorithm does not encode a fully complex range of values if the continuous, unit-amplitude portion of the modulation curve is quantized. However, a key result of this study that is fundamental to the development of pseudorandom encoding on quantized SLM's is the realization that any complex value contained on the line between two complex-valued points can be pseudorandom encoded. By repeatedly using this result for all possible pairs of points on the modulation characteristic, one can identify the encoding range of any given SLM.[6] The complex values identified as encodable by this geometric construction always form a convex set. This analysis procedure was also used to consider the feasibility of pseudorandom encoding on quantized SLM's.[9] The results in Ref. 9 lead to the conclusion that with three properly chosen modulation values a circular region around the origin of the complex plane (i.e., a fully complex range) can be pseudorandom encoded.

This minimal set of restrictions on ternary pseudorandom encoding is of critical importance given the large number of diffractive optics and SLM's that produce only a few quantized levels of modulation.

C. Preliminary Description: Distinctions between Ternary Pseudorandom Encoding and Traditional Computer-Generated Hologram Algorithms

To appreciate better how ternary pseudorandom encoding differs from traditional CGH algorithms, it is worth contrasting it with Burckhardt's method.[10] Burckhardt's method uses a group of three pixels to represent arbitrary complex values. Each pixel is variable in amplitude between zero and unity, and (through delayed sampling) the pixels represent phases of 0°, 120°, and 240°. The addition of these three vector components with various combinations of the three amplitudes permits any value in the hexagonal region shown in Fig. 2(a) to be encoded. The inscribed circle of unity radius is the fully complex set of values that can be encoded by Burckhardt's method.

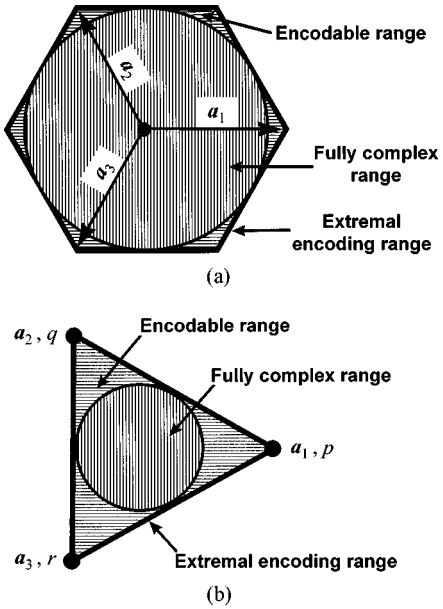


Fig. 2. Distinctions between (a) Burckhardt's ternary encoding method and (b) pseudorandom ternary encoding. In Burckhardt's method the magnitudes of the available complex amplitudes $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ (their loci indicated by the three connected arrows) can be continuously varied between 0 and 1. In pseudorandom encoding, the available complex amplitudes are constant, but the probabilities p, q, and r of selecting $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ can be varied continuously between 0 and 1. The constraint that $p + q + r = 1$ leads to pseudorandom encoding having a different encodable/realizable range from that of Burckhardt's method. The fully complex range is the maximum-diameter circular region that surrounds the origin of the complex plane and that does not exceed the extremal encoding range. The fully complex range is drawn for the specific case that the origin is the center of the circular region.

Ternary pseudorandom encoding differs from Burckhardt's method in two key respects: (1) Any one pixel can be set to one of only three complex amplitudes rather than be continuously varied through a range of values, and (2) one pixel, rather than three pixels, is used to represent a desired complex value. Figure 2(b) illustrates the encoding range and the fully complex range (indicated by the inscribed circle) for unit-amplitude modulation at the three values of phase of 0°, 120°, and 240°. The amplitudes that are physically available in Burckhardt's method are effectively represented by the probabilities p, q, and r in ternary pseudorandom encoding. If one of the three possible values of modulation is randomly selected with relative frequencies of occurrence p, q, and r, then the average value of modulation will be the vector sum of the three available modulation values scaled by the respective values of probability. As shown in Ref. 4, this statistical average does effectively represent the pixel modulation for purposes of designing Fourier transform holograms.

The remainder of this paper is organized as follows. Section 2 gives the essential mathematical background on pseudorandom encoding and its properties needed to derive encoding algorithms for quantized SLM's. Section 3 derives the ternary pseudorandom-encoding algorithm. Section 4 evaluates the encoding error for ternary encoding. Sections 3 and 4 also present geometric interpretations of encoding and encoding error. Section 5 describes how ternary encoding can be used to build up encoding algorithms for quantized modulation characteristics and specifically compares the encoding errors for three, four, and five levels of quantization. Section 5 also develops a new model of SNR in terms of measures of the signal to be encoded and the SLM characteristics. Section 6 encodes the same complex function by the various algorithms and compares the resulting simulated and experimentally produced diffraction patterns.

2. MATHEMATICAL BACKGROUND

A. General Description of Pseudorandom Encoding

All pseudorandom-encoding algorithms specify the modulation of any given pixel in terms of a user-specified 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 bold-face type. The pseudorandom-encoding design statement is, in general, to find a value of the ensemble average

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

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 specified to ensure that the expected value of $\mathbf{a}$ and the desired complex value are identical. 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.

This pseudorandom-encoding prescription is applied to each pixel in sequence to encode the desired spatially varying complex modulations $\mathbf{a}_c$. With i as the spatial coordinate, the spatial samples of the desired complex function, the pdf, and the random modulation can be 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 [see Fig. 1(b)]. 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 of the desired diffraction pattern is

$$I_c(f) = \left| \sum_i \mathbf{A}_{ci} \right|^2 = \left| \mathcal{F}\left\{ \sum_i \mathbf{a}_{ci} \right\} \right|^2, \tag{2}$$

where $\mathcal{F}\{\cdot\}$ is the Fourier transform operator, $\mathbf{A}_{ci}(f)$ is the Fourier transform of the transmittance of the ith pixel located at position i in the modulator plane, and f is the spatial coordinate across the Fourier plane. The expected intensity of the diffraction 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. Under the pseudorandom design condition $\langle \mathbf{a}_i \rangle = \mathbf{a}_{ci}$, the ensemble average pattern is expressed as[4–6]

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

where $\mathbf{A}_i(f)$ 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 corresponds to the average level of background (i.e., speckle) noise that is produced as a result of the randomness of the modulation. It is the error signal referred to in Subsection 1.C. For the case of pixels that are modeled as pointlike apertures, the average background noise is of constant intensity for all frequencies f (i.e., it is white).

B. Encoding Error Defined

Equation (3) identifies individually the noise contribution of each pixel. Therefore insight can be gained by evaluating the noise contribution in the modulation plane. Under the assumption that the pixels are infinitesimally wide apertures, the inverse Fourier transform of a single pixel noise term gives the encoding error[6]

$$\epsilon = \langle |\mathbf{a}|^2 \rangle - |\mathbf{a}_c|^2, \tag{4}$$

where the subscript has been dropped to simplify presentation. (If the pixels are of finite width, then an autocorrelation of the pixel aperture function would also be in-

cluded in the formula.[4] This term is dropped because it adds no essential insight to the current discussion.)

C. Geometric Interpretation of Binary Pseudorandom Encoding

As stated in Subsection 1.B, the pseudorandom selection between two complex values permits any value on the line segment connecting these two points to be realized on average. This geometric construction was used to determine the encoding range of continuous-range SLM's[6,9] and to identify the pseudorandom-encoding range of quantized SLM's.[9] These results are reviewed here and used to develop insights into pseudorandom encoding with three (or more) quantized levels.

Binary encoding is directly developed by using the pdf for the binary distribution in Eq. (1). The binary pdf is

$$p(\mathbf{a}) = p\,\delta(\mathbf{a} - \mathbf{a}_1) + q\,\delta(\mathbf{a} - \mathbf{a}_2), \tag{5}$$

where $\delta(\cdot)$ is the Dirac delta function, $\mathbf{a}_1$ and $\mathbf{a}_2$ are a pair of complex values from the modulation characteristic, and p and $q = 1 - p$ are the probabilities of selecting $\mathbf{a}_1$ and $\mathbf{a}_2$. Since p is a probability, its value is between 0 and 1. [It will be clear from usage when we are referring to the binary probability p and the density function $p(\mathbf{a})$.] Evaluating Eq. (1) with this pdf gives an expression for the effective complex amplitude of

$$\langle \mathbf{a} \rangle = p\,\mathbf{a}_1 + (1 - p)\mathbf{a}_2. \tag{6}$$

Equation (6) is recognized as the expression for a line as a function of the variable p. For $p = 1$, $\mathbf{a}_1$ is encoded; for $p = 0$, $\mathbf{a}_2$ is encoded; and for values of p between 0 and 1, any value lying on the line segment between $\mathbf{a}_1$ and $\mathbf{a}_2$ can be encoded. This geometric interpretation (see Fig. 3) can be brought out further by considering that the desired complex value $\mathbf{a}_c$ can be expressed in terms of the two complex values $\mathbf{a}_1$ and $\mathbf{a}_2$ as

$$\mathbf{a}_c = \frac{l_2}{l}\,\mathbf{a}_1 + \frac{l_1}{l}\,\mathbf{a}_2, \tag{7}$$

where l_1 is the distance between $\mathbf{a}_c$ and $\mathbf{a}_1$, l_2 is the distance between $\mathbf{a}_c$ and $\mathbf{a}_2$, and $l = l_1 + l_2$. Clearly, the lengths can be chosen so that the desired value $\mathbf{a}_c$ can be realized by the average (or effective) value $\langle \mathbf{a} \rangle$. Evaluation of Eq. (4) by using Eqs. (5) and (7) (and some further algebraic manipulation[9]) shows that the encoding error for binary encoding,

$$\epsilon = l_1 l_2 = pql^2, \tag{8}$$

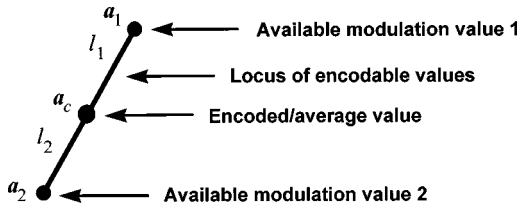


Fig. 3. Geometry of pseudorandom biamplitude encoding. Any desired value $\mathbf{a}_c$ between the two available modulation values $\mathbf{a}_1$ and $\mathbf{a}_2$ can be encoded by pseudorandom encoding. The product of the lengths of the line segments that connect $\mathbf{a}_1$ and $\mathbf{a}_2$ to $\mathbf{a}_c$ is the encoding error $\epsilon = l_1 l_2$.

is simply the products of the distances from $\mathbf{a}_1$ and $\mathbf{a}_2$ to $\mathbf{a}_c$. The maximum encoding error $0.25l^2$ occurs if $\mathbf{a}_c$ is the midpoint of the line segment between $\mathbf{a}_1$ and $\mathbf{a}_2$.

Equations (6) and (7) suggest the following encoding formula for binary SLM's: For a given value of p, the desired complex value $\mathbf{a}_c = \langle \mathbf{a} \rangle$ is represented (i.e., encoded) by a single randomly selected value

$$\mathbf{a} = \begin{cases} \mathbf{a}_1 & \text{if } 0 \leqslant s \leqslant p \\ \mathbf{a}_2 & \text{if } p < s \leqslant 1 \end{cases}, \tag{9}$$

where s is a uniformly distributed random number between 0 and 1.

D. Using Binary Encoding to Evaluate Encoding Range

The analysis and the geometric interpretation of binary pseudorandom encoding [Eqs. (6)–(8)] provides insight into pseudorandom encoding for various continuous and quantized modulation characteristics.[9] One use of this analysis is in determining those complex values that can be pseudorandom encoded for a particular modulation characteristic. As mentioned in Subsection 1.B, the range is found by combining the ranges encoded by each possible pair of values from the SLM characteristic.[6] Because the binary encoding algorithm has the fewest constraints, the maximum possible range of values (a convex set) is found by this procedure. Other constraints [e.g., using a nonbinary pdf in Eq. (1)] are known to reduce this range.[6] Figure 2(b) shows the convex region (triangular shaded) that is bounded by the three possible binary encodings $\mathbf{a}_1 - \mathbf{a}_2$, $\mathbf{a}_2 - \mathbf{a}_3$, and $\mathbf{a}_3 - \mathbf{a}_1$. This is the range of possible complex values that can be realized with three quantized values of modulation. Also, the circular shaded region represents the range over which fully complex-valued functions can be encoded. Section 3 derives ternary encoding and further shows that there is only one unique solution for encoding a given complex value.

3. TERNARY PSEUDORANDOM ENCODING

The ternary pdf for the three modulation values $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ is

$$p(\mathbf{a}) = p\,\delta(\mathbf{a} - \mathbf{a}_1) + q\,\delta(\mathbf{a} - \mathbf{a}_2) + r\,\delta(\mathbf{a} - \mathbf{a}_3), \tag{10}$$

where the three probabilities p, q, and r of selecting $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ satisfy

$$p + q + r = 1. \tag{11}$$

Evaluating Eq. (1) with this pdf gives an expression for the effective complex amplitude of

$$\mathbf{a}_c \equiv \langle \mathbf{a} \rangle = p\,\mathbf{a}_1 + q\,\mathbf{a}_2 + r\,\mathbf{a}_3. \tag{12}$$

There is at most one solution for the values of p, q, and r that encodes $\mathbf{a}_c$. This follows from the fact that there are three linear equations in the three unknowns. The real and imaginary parts of Eq. (12) give two of the equations, and the third expresses that the sum of the three probabilities is unity. These equations written in matrix form are

$$\begin{pmatrix} a_{cr} \\ a_{ci} \\ 1 \end{pmatrix} = \begin{bmatrix} a_{1r} & a_{2r} & a_{3r} \\ a_{1i} & a_{2i} & a_{3i} \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix}, \tag{13}$$

where subscripts r and i indicate the real and imaginary parts of the corresponding complex values $\mathbf{a}_c$, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. As long as the matrix is nonsingular, Eq. (13) has a single solution. However, it is possible that the values found for p, q, and r could be less than zero or greater than unity. Since these values are probabilities, such solutions cannot be pseudorandom encoded. We will show that this situation corresponds to the value of $\mathbf{a}_c$ being outside the convex region formed by $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ [see Fig. 2(b)].

The analysis is performed by using Eq. (11) to eliminate r from Eq. (12), which yields

$$\mathbf{a}_c - \mathbf{a}_3 = p(\mathbf{a}_1 - \mathbf{a}_3) + q(\mathbf{a}_2 - \mathbf{a}_3). \tag{14}$$

This relationship is illustrated in Fig. 4(a). Using the geometry in Fig. 4(a) and choosing the vector $\mathbf{a}_c - \mathbf{a}_3$ as a coordinate axis, we can write Eq. (14) as

$$\begin{pmatrix} 0 \\ l_{c3} \end{pmatrix} = \begin{bmatrix} l_{13} \sin \theta_{13} & -l_{23} \sin \theta_{23} \\ l_{13} \cos \theta_{13} & -l_{23} \cos \theta_{23} \end{bmatrix} \begin{pmatrix} p \\ q \end{pmatrix}, \tag{15}$$

where l_{13}, l_{23}, and l_{c3} represent the respective lengths from $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_c$ to $\mathbf{a}_3$ and where θ_{13} and θ_{23} are the angles from $\mathbf{a}_c - \mathbf{a}_3$ to $\mathbf{a}_1 - \mathbf{a}_3$ and $\mathbf{a}_2 - \mathbf{a}_3$, respectively. The solution for Eq. (15) is

$$p = \frac{l_{c3}}{l_{13}} \frac{\sin \theta_{23}}{\sin(\theta_{13} + \theta_{23})}, \qquad q = \frac{l_{c3}}{l_{23}} \frac{\sin \theta_{13}}{\sin(\theta_{13} + \theta_{23})}. \tag{16}$$

The geometry in Fig. 4(a) limits p and q to be positive, since the angles θ_{13} and θ_{23} are positive and their sum (since they are part of the same triangle) is less than 180°. Equations (16) do admit values of p and q that exceed unity.

We consider some special cases to appreciate better the relationship between values of p, q, and r and values of $\mathbf{a}_c$. First, consider cases where $q = 0$. The construction in Fig. 4(a) indicates that $\mathbf{a}_c$ lies on the line defined by $\mathbf{a}_1$ and $\mathbf{a}_3$. Thus $\theta_{13} = 0$. Solving Eq. (15) for $q = 0$ then gives $p = l_{c3}/l_{13}$, and $r = 1 - (l_{c3}/l_{13})$. This result indicates that ternary encoding reduces to binary encoding [see Eq. (7)] if $q = 0$. Therefore $\mathbf{a}_c$ is contained between $\mathbf{a}_1$ and $\mathbf{a}_3$ as long as p and r are contained between zero and unity. Similarly, if $p = 0$, these equations reduce to binary encoding between $\mathbf{a}_2$ and $\mathbf{a}_3$. If $r = 0$, then binary encoding is performed between the points $\mathbf{a}_1$ and $\mathbf{a}_2$. This situation is illustrated in Fig. 4(b). It is interesting to note that the geometric construction forms two triangles that are identical to the outer triangle $\mathbf{a}_1 - \mathbf{a}_2 - \mathbf{a}_3 - \mathbf{a}_1$, except that they are scaled in size by p and q. Figure 4(c) generalizes this construction for cases where $r \neq 0$. From Eq. (11) it is clear that each side of the outer triangle is divided into lengths that are proportional to p, q, and r. Now three triangles contained inside $\mathbf{a}_1 - \mathbf{a}_2 - \mathbf{a}_3 - \mathbf{a}_1$ are apparent that are identical except for their scaling by p, q and r. This discussion shows that as long as $\mathbf{a}_c$ is contained on the boundary of $\mathbf{a}_1 - \mathbf{a}_2 - \mathbf{a}_3 - \mathbf{a}_1$ or inside the enclosed area, it can be pseudorandomly encoded. Values that are outside correspond to probabilities that are less than zero or in excess of unity, which cannot be realized by this statistical procedure.

Once p and q are found by solving Eq. (13) or by using Eqs. (16), then ternary pseudorandom encoding of the desired complex value $\mathbf{a}_c = \langle \mathbf{a} \rangle$ is accomplished by randomly selecting

$$\mathbf{a} = \begin{cases} \mathbf{a}_1 & \text{if } 0 \leq s \leq p \\ \mathbf{a}_2 & \text{if } p < s \leq p + q, \\ \mathbf{a}_3 & \text{if } p + q < s \leq 1 \end{cases} \tag{17}$$

where s is a uniformly distributed random number between 0 and 1.

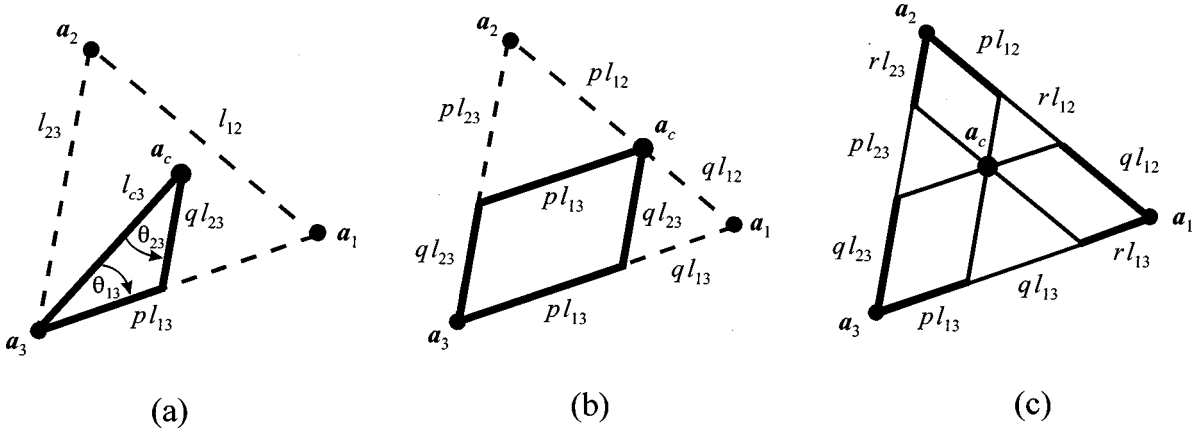


Fig. 4. Geometric relationships for ternary pseudorandom encoding. (a) The three vectors (thick lines) correspond to the three terms in Eq. (14). Each term corresponds to a vector that has $\mathbf{a}_3$ as its origin. (b) Geometry for ternary encoding when the probability $r = 0$. For this condition ternary encoding reduces to biamplitude encoding between $\mathbf{a}_1$ and $\mathbf{a}_2$. This construction also identifies two triangles that are identical except for scaling by p and q. The thick lines indicate the two vectors that add together to produce the desired complex value $\mathbf{a}_c$. (c) Geometry for ternary encoding when the probability r is $0 < r < 1$. This construction shows that there are three triangles that are identical except for scaling by p, q, and r. The thick lines indicate the six line segment lengths that are used in Eq. (22) to calculate the encoding error. The products of the lengths of the three pairs of collinear segments are added together to give the encoding error.

4. ENCODING ERROR AND GEOMETRIC INTERPRETATION

Encoding error provides information on the amount of noise generated by encoding. Since it can be directly calculated from the desired complex value $\mathbf{a}_c$, it can be used to anticipate the quality of the encoding before actually performing the encoding. Therefore pseudorandom encoding has the desirable property that it automatically includes error analysis with the encoding algorithm. In this section the encoding error is evaluated for ternary pseudorandom encoding.

Using Eqs. (10) and (12) in Eq. (4) gives the encoding error for ternary pseudorandom encoding as

$$\epsilon = p|\mathbf{a}_1|^2 + q|\mathbf{a}_2|^2 + r|\mathbf{a}_3|^2 - |p\mathbf{a}_1 + q\mathbf{a}_2 + r\mathbf{a}_3|^2, \tag{18}$$

where the three possible modulation values for a pixel are $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. Equation (18) can be rewritten as

$$\begin{aligned}\epsilon = {} & (p - p^2)a_1{}^2 + (q - q^2)a_2{}^2 + (r - r^2)a_3{}^2 \\ & - 2pqa_1a_2 \cos\phi_{12} - 2pra_1a_3 \cos\phi_{13} \\ & - 2qra_2a_3 \cos\phi_{23},\end{aligned} \tag{19}$$

where a_i is the magnitude of $\mathbf{a}_i$ for $i = 1, 2, 3$ and $\phi_{i,j}$ is the angle between $\mathbf{a}_i$ and $\mathbf{a}_j$ for $j = 2, 3$. (Note that in this section the subscripts refer to one of the three possible modulation values for a pixel rather than to spatial position of a pixel.) This result can be dramatically simplified by repeated use of the law of cosines

$$l_{i,j} \equiv |\mathbf{a}_i - \mathbf{a}_j|^2 = a_i{}^2 + a_j{}^2 - 2a_i\,a_j \cos\phi_{ij} \tag{20}$$

and the use of

$$\begin{aligned} p - p^2 &= pq + pr, \\ q - q^2 &= pq + qr, \\ r - r^2 &= pr + qr, \end{aligned} \tag{21}$$

which follows from Eq. (11). Using Eq. (20) three times (for $\{i, j\} = \{1, 2\}$, $\{1,3\}$, $\{2, 3\}$) in Eq. (19) together with Eqs. (21) gives the simplified expression for encoding error of

$$\epsilon = pql_{12}{}^2 + prl_{13}{}^2 + qrl_{23}{}^2. \tag{22}$$

This result is quite similar to Eq. (8), the encoding error for binary encoding. Since ternary encoding reduces to binary encoding for $r = 0$, then Eq. (22) should also reduce to Eq. (8). Making the identification $l \equiv l_{12}$, we can see that this is indeed the case. Likewise, when $q = 0$ or $p = 0$, the encoding error is identical to the binary-encoding error between $\mathbf{a}_1$ and $\mathbf{a}_3$ or $\mathbf{a}_2$ and $\mathbf{a}_3$, respectively. In general, for p, q, and r not equal to zero, each of the three encoding error terms corresponds to the product of a pair of lengths on the respective line segments $\mathbf{a}_1-\mathbf{a}_2$, $\mathbf{a}_1-\mathbf{a}_3$, and $\mathbf{a}_2-\mathbf{a}_3$. To help visualize the lengths that contribute to encoding error, Fig. 4(c) indicates the lengths as thick lines.

5. DESIGN AND EVALUATION OF SPECIFIC ALGORITHMS

Encoding algorithms for modulation characteristics of any degree of quantization can be built up out of elementary ternary pseudorandom-encoding algorithms. Therefore the analysis presented in Sections 3–5 can be specialized and applied to m-ary quantized modulation characteristics. In this section we define specific pseudorandom-encoding algorithms for three, four, and five levels of quantization that each provide a circular encoding range around the origin of the complex plane. The encoding error for each algorithm is evaluated, and these results are compared with the encoding errors for continuous phase-only and biamplitude phase modulation characteristics. We also use the encoding error together with measures of the desired complex-valued signal to define an estimate of SNR of the resulting diffraction pattern. In Section 6 these specific algorithms are demonstrated for a specified function, and the SNR's of the resulting diffraction patterns are compared with our estimated SNR.

The specific pseudorandom-encoding algorithms evaluated in the remainder of this paper are defined with the help of Table 1. The ternary algorithm is defined to use three modulation values that are equally spaced by $2\pi/3$ rad around the unit circle. The m-ary 1 algorithm uses four modulation values that are equally spaced by $\pi/2$ around the unit circle, and the m-ary 2 algorithm uses the same four modulation values as those for m-ary 1, with the addition of the value of zero. The m-ary 1 algorithm is built up out of two ternary encoding algorithms. The modulation values of $\{1, j, -1\}$ are used for desired complex values that lie in the upper half of the complex plane, and $\{1, -j, -1\}$ is used for encoding complex values that lie in the lower half-plane. The m-ary 2 encoding algorithm is composed of four ternary encoding algorithms (listed in the third column of Table 1). Each of the four ternary algorithms corresponds to encoding desired values in one quadrant of the complex plane.

To compare pseudorandom encoding for discrete modulation characteristics with pseudorandom encoding for continuous characteristics, we also consider encoding algorithms for phase-only[4] and biamplitude phase SLM's.[5] The biamplitude encoding algorithm is identical to that

Table 1. Defining Parameters and Metrics for Various Pseudorandom-Encoding Algorithms

SLM Type	SLM Values	Ternary Groups	γ	ϵ
Ternary	$1, \exp(\pm j\,2\pi/3)$	$\{1, \exp(\pm j\,2\pi/3)\}$	1/2	$1 - a_c{}^2$
m-ary 1	$\pm 1, \pm j$	$\{1, j, -1\}, \{1, -j, -1\}$	$\sqrt{1/2}$	$1 - a_c{}^2$
m-ary 2	$0, \pm 1, \pm j$	$\{0, \pm 1, j\}, \{0, 1, \pm j\}$	$\sqrt{1/2}$	$p + q - a_c{}^2$
Phase-only	$\exp(j\psi)$	—	1	$1 - a_c{}^2$
Biamplitude phase	$0, \exp(j\psi)$	—	1	$a_c - a_c{}^2$

reported in Ref. 5. It can be viewed as binary pseudorandom encoding (as described in Section 2) by using the two modulation values of 0 and $\exp(j\psi_c)$, where ψ_c is the phase of the desired complex value $\mathbf{a}_c$ and the binary selection between the two modulation values is controlled by the probability $p = a_c$. The encoding algorithm for phase-only SLM's used in this study is simpler than the algorithm originally reported in Ref. 4. The new algorithm is based on binary encoding between the two values $\pm\exp(j\psi_c)$. In this case the appropriate probability needed to encode the amplitude a_c is $p = (1 + a_c)/2$.

We choose to use this phase-only encoding algorithm over the algorithm in Ref. 4 because it is somewhat simpler to program. However, very similar results are anticipated by either algorithm, since they both produce identical levels of encoding error. This can be appreciated by evaluating the encoding error with the use of Eq. (4). This equation reduces to $\epsilon = 1 - a_c^2$ for any phase-only modulation.[4,6] This is true for both continuous and discrete phase-only modulation characteristics [as can also be shown by evaluating Eq. (18)]. Therefore the phase-only, ternary, and m-ary 1 algorithms all produce identical encoding errors when the same amplitude is encoded. The encoding error for biamplitude phase modulation, $\epsilon = a_c - a_c^2$, is given in Refs. 5 and 6. This result also follows from Eq. (8) if $l = 1$ and $a_c = p$, which is the case for biamplitude encoding. For the m-ary 2 algorithm, analysis of Eq. (4) or (18) by using any one of the ternary groups (given in the third column of Table 1) gives the encoding error $\epsilon = p + q - a_c^2$. The encoding error for each of the five algorithms is summarized in Table 1.

Even though the three algorithms for phase-only SLM's produce identical encoding errors when the same value is encoded, this does not mean that their performance is identical. The reason is that the circular encoding range [see Section 2 and Fig. 2(b)] is less for our discrete modulation characteristics than it is for our continuous characteristics. The scaling of the desired complex-valued function to fit within the maximum circular radius γ of each modulation characteristic can cause significant differences in the amount of encoding error for the various algorithms. For the case of a SLM that produces M uniformly spaced phase-only modulation values around the unit circle, the maximum circular radius can be expressed as

$$\gamma_M = \cos(\pi/M). \tag{23}$$

This result is determined by considering that the fully complex encoding region intersects the chord connecting nearest-neighbor modulation values at the half-angle π/M between them. The values of γ for our specific encoding algorithms are listed in the fourth column of Table 1.

For many pseudorandom algorithms, scaling the complex values to be smaller than the maximum possible radius γ reduces diffraction efficiency and increases SNR.[6] It is also possible to combine pseudorandom encoding with other algorithms, which permits the complex values to be scaled to be larger than the maximum circular radius.[5] This can produce greater diffraction efficiencies and higher SNR's. Considering these additional two possibilities would needlessly complicate this study. In this paper the fully complex values are scaled so that the maximum amplitude for a given encoding algorithm is its maximum circular radius γ.

With this definition of the maximum circular encoding range, it is now possible to make a comparative analysis of the performance of each algorithm. Since the desired fully complex function $\mathbf{a}_c(x)$ is normalized so that its maximum amplitude is γ, it is appropriate to compare the amplitudes a_{c1}/γ_1 and a_{c2}/γ_2, where the subscripts 1 and 2 indicate the values of amplitude and maximum amplitude for two different algorithms. Since encoding error is proportional to intensity rather than amplitude, it is appropriate to compare the normalized encoding errors ϵ_1/γ_1^2 and ϵ_2/γ_2^2. Therefore, rather than describing the absolute encoding error, this normalization presents the error-to-signal ratio, or relative error, for the same value encoded by two different algorithms. Relative error is more informative of the fidelity of the diffraction patterns resulting from pseudorandom encoding than is absolute encoding error.

The encoding errors for the Table 1 algorithms are presented in this way in Fig. 5. Phase-only encoding always produces larger encoding errors than biamplitude phase encoding, as reported in Ref. 5. The m-ary 2 algorithm produces a range of relative errors that are contained between the two m-ary 2 curves in Fig. 5. The lower curve corresponds to the case where $q = 0$, and the upper curve corresponds to the case where $p = q$. Figure 5 shows that the m-ary 2 algorithm always produces more relative error than does biamplitude phase encoding but frequently produces less relative error than does phase-only encoding. If the desired complex-valued function $\mathbf{a}_c(x)$ has many more amplitudes that are less than 1/2, then the total relative error produced by m-ary 2 encoding can be substantially less than that for phase-only encoding. Alternatively, if most of the desired complex values are well above 1/2, then m-ary 2 encoding produces a total relative error that is much greater than that for phase-only encoding.

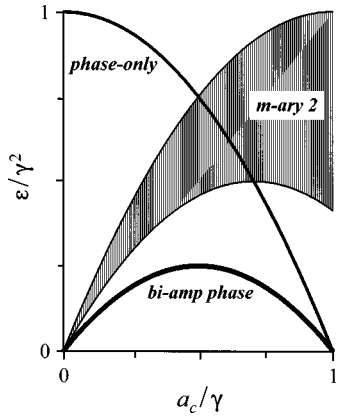


Fig. 5. Relative encoding errors for various pseudorandom-encoding algorithms. The desired magnitude $\mathbf{a}_c$ is normalized by γ, the maximum radius for the fully complex encoding range, and the relative encoding error is $\epsilon_{\text{rel}} = \epsilon/\gamma^2$. The striped region gives all possible encoding errors for the m-ary 2 algorithm. The phase-only curve also represents the relative encoding error for pseudorandom encoding with M phase-only values that are uniformly spaced in angle. In this case the plotted phase-only curve is offset by $\tan^2(\pi/M)$ [see Eq. (24)].

The relative error curves for ternary and m-ary 1 encoding are identical to the relative error curve for phase-only encoding, except that they are offset by amounts that depend on the degree of quantization. This can be seen by considering that for quantized phase-only modulation characteristics for which there are M evenly spaced modulation values on the unit circle, the relative error can be written with the help of Eq. (23) as

$$\epsilon_{\mathrm{rel}} \equiv \epsilon_M/{\gamma_M}^2 = 1 + \tan^2(\pi/M) - (a_c/\gamma_M)^2. \quad (24)$$

For the maximum amplitude of $a_c = \gamma_M$, Eq. (24) gives the minimum relative error $\epsilon_{\mathrm{rel}} = \tan^2(\pi/M)$, and for the minimum amplitude of $a_c = 0$, the relative error is $\epsilon_{\mathrm{rel}} = 1 + \tan^2(\pi/M)$. Between these limiting points the curve has the identical quadratic dependence as that of the relative error curve for the continuous phase-only modulation characteristic. Therefore the quantized characteristics produce additional relative error by the amount $\tan^2(\pi/M)$. For the ternary encoding this offset is 3, or in other words, the relative error for ternary encoding is always larger by a value of 3 than that for phase-only encoding. For m-ary 1 encoding the offset is 1. With increasingly fine quantization the relative error curve approaches the phase-only curve. For instance, the offset would be 0.17 for eight levels and 0.04 for 16 levels of quantization.

Despite the significant amount of relative error produced by ternary pseudorandom encoding, it is possible to use this noisiest of pseudorandom algorithms to encode many desired complex functions with good fidelity. The key factor is that the spatial extent (i.e., the bandwidth B) of the desired diffraction pattern is small enough that the signal is sufficiently greater than the background noise that is due to the sum of the encoding errors from each pixel.

A simple analysis is presented to make this relationship more apparent. Consider that a particular desired function is pseudorandom encoded for an N-pixel SLM. The average encoding error per pixel is ϵ_a, and the average intensity transmittance that is encoded is ${a_{ca}}^2$. The total energy in the encoding error is then $N\epsilon_a$, and the total energy in the encoded signal is $N{a_{ca}}^2$. The encoding error in the diffraction plane transforms into a white spectrum over a bandwidth of N, whereas the desired signal has a designed bandwidth of B. Therefore the desired diffraction pattern will have a directionality gain of N/B over the spectrum of the encoding error. The SNR can then be written for this approximate analysis as

$$\mathrm{SNR} = \frac{N}{B}\frac{{a_{ca}}^2}{\epsilon_a}. \quad (25)$$

To appreciate this analysis better, consider the following numerical example. For a 128×128-pixel phase-only SLM ($N = 16{,}384$), an average encoding error $\epsilon_a = 0.9$, a root-mean-square amplitude transmittance ${a_{ca}}^2 = 1 - \epsilon_a = 0.32$ (from Table 1), and a desired signal-to-noise ratio SNR $\geqslant 100$, Eq. (25) gives the result that the bandwidth of the desired signal needs to be $B \leqslant 18.2$. This analysis shows that even for pseudorandom algorithms that produce the greatest encoding errors, there are many diffraction patterns that can be successfully encoded as long as the SNR is acceptably large and the signal bandwidth is correspondingly low. The trends predicted by Eq. (25) are evident in the computer simulations and the experimental demonstrations that are presented in Section 6.

Although fidelity of diffraction patterns is the focus of this investigation, diffraction efficiency is probably the most widely discussed metric. For this reason a few basic relationships are reviewed that relate diffraction efficiency to SNR. In pseudorandom encoding, the diffraction efficiency has been shown to be identical to the average intensity of the fully complex function that is to be encoded.[8] This is written as

$$\eta = \frac{1}{N}\sum_{i=1}^{N} {a_{ci}}^2 \equiv {a_{ca}}^2. \quad (26)$$

Since, for a particular encoding method, the a_{ci} are scaled so that the maximum amplitude equals the circular encoding radius γ, it is clear from Eq. (26) that the diffraction efficiency is proportional to γ^2.[6] Consequently, the efficiencies of the five algorithms in Table 1 vary by a factor of 4.

For arbitrary modulation characteristics the average intensity transmittance in Eq. (26) can be replaced by the diffraction efficiency η, and for the specific case of phase-only modulation Eq. (26) reduces to

$$\mathrm{SNR} = \frac{N}{B}\frac{\eta}{1-\eta}. \quad (27)$$

However, for the two modulation characteristics in Table 1 that also have a zero amplitude the denominator term ϵ_a is less than $1 - \eta$, which increases the SNR over that possible for the corresponding discrete or continuous phase-only characteristic.

6. DEMONSTRATIONS OF TERNARY ENCODING

A. Specification of the Desired Function To Be Encoded

In this section the same desired function $\mathbf{a}_{ci}$ is encoded by various algorithms, and the resulting simulated and (for some cases) experimentally measured diffraction patterns are presented. The desired function is a 128×128 array of complex values that has a Fourier transform that produces a 7×7 array of uniform-intensity diffraction-limited spots. To better compare pseudorandom encoding with known art, we relate our results to previously published designs. Krackhardt *et al.* have reported the highest possible diffraction efficiency designs for continuous phase-only SLM's.[11] Our complex values are derived from their design for a 1×7 spot array. Their Table III specifies seven phases θ_k associated with seven equally spaced spots. These phases are used to specify a desired fully complex function of the form

$$\mathbf{a}_c(x, y) = \sum_{k=2}^{8} \exp(j\theta_k)\exp(j2\pi kx) \times \sum_{l=2}^{8} \exp(j\theta_l)\exp(j2\pi ly) \quad (28)$$

that is rectangularly separable. This periodic function is sampled to produce a 32×32-unit cell of complex values and a 4×4 array of cells that form the 128×128 desired complex values.

B. Definition of the Encoding Algorithms Used

These values are then encoded by each of the five algorithms in Table 1. (Part of the encoding includes scaling the desired complex values so that the largest amplitude of the complex values is equal to the appropriate value of γ given in Table 1.) The resulting diffraction patterns are compared and evaluated.

Equation (28) can also be interpreted as a desired fully complex function for the study of Krackhardt *et al.* The phases found through their global optimization of the diffraction pattern specify a desired function that is then encoded onto continuous phase-only SLM's by transforming the desired fully complex function into the phase-only function or kinoform[12]

$$\mathbf{a}(x, y) = \exp\{j \arg[\mathbf{a}_c(x, y)]\}. \tag{29}$$

This encoding is performed not only so that the desired function can be implemented by a phase-only modulator but also to maximize diffraction efficiency.[13] The encoding indicated in Eq. (29) is also applied to the 128×128 desired complex values, and the resulting diffraction pattern is compared with those of the five pseudorandom algorithms. We will refer to this algorithm as nonrandom phase-only encoding to help distinguish it from pseudorandom phase-only encoding.

Since the algorithms in which we are most interested are for quantized SLM's, we also quantize the phase to three and four values of phase (uniformly spaced around the unit circle) and evaluate the diffraction patterns for these modified encodings. We refer to these algorithms as nonrandom ternary and nonrandom m-ary 1, respectively. It should be noted in Tables I and II of Ref. 11 that different values of the spot phases θ_k would lead to maximum diffraction efficiency. However, this would require a new optimization to find the phases for each modulator characteristic. Instead, in keeping with the spirit of encoding the same complex function, we have chosen to compare quantized pseudorandom encoding with quantized versions of the maximum-efficiency, rectangularly separable design.

It is also possible to specify a nonrandom version of the pseudorandom m-ary 2 algorithm. In this case a zero-value modulation is selected if the desired complex value is closer to zero than to the four other phase-only modulation values. To compare the pseudorandom and nonrandom m-ary 2 algorithms fairly, the desired complex values $\mathbf{a}_{ci}$ are similarly scaled, so that the maximum amplitude encoded is $\gamma = \sqrt{1/2}$. However, Juday has shown that the quality of an encoding depends on the value of γ.[5,14] For this reason we also perform an iterative search to find the value of γ that optimizes the performance measures that are of most interest to us. The optimum value found in our simulations is $\gamma = 1.3$, which optimizes signal-to-peak-noise ratio (SPR) and uniformity. Subsection 6.C defines these two and the other metrics of interest.

C. Simulation Procedures and Definition of the Performance Metrics

Two metrics, diffraction efficiency η and signal-to-noise ratio SNR, are directly calculated from the desired complex values $\mathbf{a}_{ci}$ for each of the five pseudorandom-encoding algorithms. After $\mathbf{a}_{ci}$ is scaled by the appropriate value γ in Table 1, η is calculated by using Eq. (26) and SNR is calculated by using Eq. (25). In these calculations we use $N = 128^2$ for the number of SLM pixels, and, considering a diffraction-limited spot to have a space–bandwidth product of 1, we use $B = 7^2$ for the space–bandwidth product of the desired signal. These metrics that are based on theory are listed in parentheses in Table 2 beside the values of SNR and η, which are calculated directly from the simulated diffraction patterns.

The far-field diffracted intensity patterns are simulated by fast-Fourier-transforming the encoded values $\mathbf{a}_i$ and then squaring the magnitude for each of the pseudorandom and nonrandom encodings. For all metrics except diffraction efficiency, the 128×128 array is placed in a 512×512 array of zeros that is fast Fourier transformed. The zero padding is used to resolve the features of the diffraction pattern more finely and to produce more realistic gray-scale images. For diffraction efficiency the 128×128 array is fast Fourier transformed directly. For phase-only modulation (for either pseudorandom or nonrandom algorithms), the efficiency is simply the sum of the intensities of the 49 spots divided by the sum of all intensities in the 128×128 diffraction pattern. For the other modulation characteristics that contain a zero value, the energy absorption in the modulator plane also needs to be accounted for.[5] Therefore the ratio of desired energy to total energy in the diffraction pattern is multiplied by the ratio of unit-amplitude pixels (i.e., "on" pixels) to the total number of pixels (i.e., number of on plus off pixels). The SNR is the ratio of the average intensity of the peak values of each of the 49 spots to the average intensity of the 512×512 pattern, excluding the square

Table 2. Performance Measures of the Pseudorandomly Encoded Desired Function

Pseudorandom	η (%)[a]	SNR[b]	SPR	n-unif (%)
Biamplitude phase	44 (44)	685 (710)	38	4
m-ary 2	22 (22)	197 (203)	15	10
Phase-only	43 (44)	254 (262)	17	10
m-ary 1	22 (22)	93 (94)	8	11
Ternary	11 (11)	40 (41)	3	20

[a] Numbers in parentheses are calculated from Eq. (26).
[b] Numbers in parentheses are calculated from Eq. (25).

Table 3. Performance Measures of the Nonrandomly Encoded Desired Function

Nonrandom	η (%)	SNR	SPR	n-unif (%)
m-ary 2 ($\gamma = \sqrt{1/2}$)	7	175	1	60
m-ary 2 ($\gamma = 1.3$)	68	961	5	11
Phase-only	92	1076	77	18
m-ary 1	75	1189	5	18
Ternary	63	376	1	18

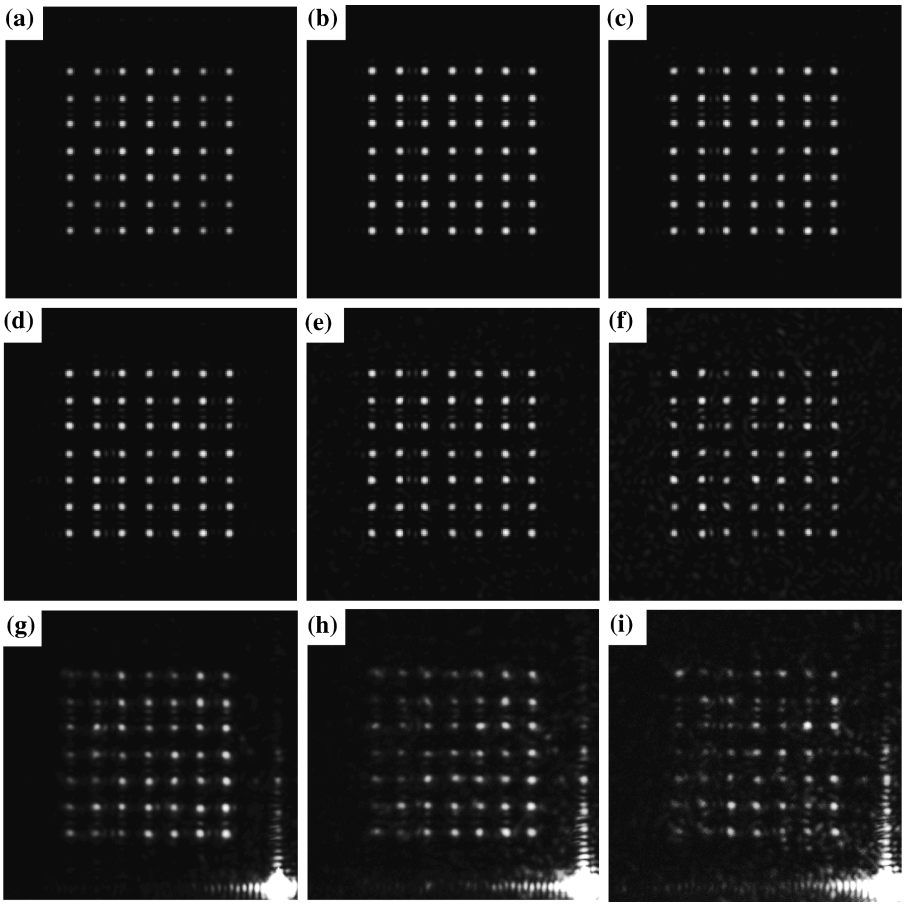


Fig. 6. [(a)–(f)] Simulated and [(g)–(i)] experimental gray-scale images of the diffraction pattern intensity resulting from various encoding algorithms. All encodings are pseudorandom except (a), which is nonrandom phase-only. The simulated pseudorandom encodings are (b) biamplitude phase, (c) m-ary 2, (d) phase-only, (e) m-ary 1, and (f) ternary. The experimental pseudorandom encodings are (g) phase-only, (h) m-ary 1, and (i) ternary.

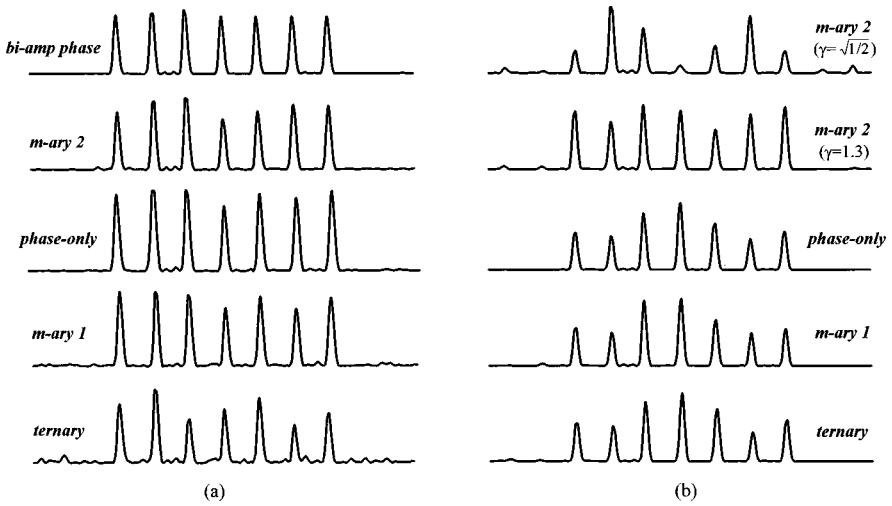


Fig. 7. Simulated cross sections of the diffraction pattern intensity resulting from various encoding algorithms: (a) pseudorandom encodings, (b) nonrandom encodings. Each cross section is along a diagonal that contains the (7,7) order (leftmost spot), the (1,1) order (rightmost spot), and the optical axis (the rightmost side of the curve). For all the pseudorandom curves in (a) and the nonrandom phase-only curve in (b), the cross sections are the intensity values from the diagonal (from upper left to lower right) of the corresponding gray-scale image in Fig. 6. The four other nonrandom encodings in (b) are plotted over an identical range.

region that contains the 7×7 spot array. The SPR is the ratio of the average peak intensity of the spots to the maximum noise peak outside the square region containing the spots. Nonuniformity of the peaks (abbreviated in Tables 2 and 3 as n-unif) is well characterized by the standard deviation of the peak intensities of the 49 spots divided by the average spot intensity. We present the nonuniformity metric this way instead of as the maxi-

mum peak-to-peak fluctuation of the spots because this metric is less susceptible to the random variations that occur for each new set of random numbers used in the encoding algorithm. We did observe that the peak-to-peak fluctuations are 1.8–2.5 times greater than the values of nonuniformity reported in Table 2 and below in Table 3.

D. Simulation Results

Figures 6(a)–6(f) show the same portion of the 512 × 512 diffraction patterns for nonrandom phase-only encoding [Fig. 6(a)] and for each of the five pseudorandom encodings [Figs. 6(b)–6(f)]. Figure 7(a) shows a cross section of intensity pattern [along a diagonal slice containing the optical axis and the (1, 1) and (7, 7) spots] for the five pseudorandom-encoded patterns. Figure 7(b) shows the corresponding cross sections for the nonrandom encodings. The diagonal cross sections tend to accentuate intensity fluctuations between the spots and to exclude most of the sidelobes between the spots that are due to the sinc-squared nature of the diffraction pattern of each spot. The three patterns of Figs. 6(d)–6(f) show increasing levels of background speckle noise and nonuniformity of the spots. The pseudorandom biamplitude phase encoding [Fig. 6(b)] appears to be free of speckle noise, and the quantized biamplitude encoding [i.e., m-ary 2, Fig. 6(c)] appears to have a noise level that is indistinguishable from that of the phase-only encoding [Fig. 6(d)]. Although most of the cross sections in Fig. 7(a) are reasonably uniform, it is clear that the ternary encoding is the least uniform of the five pseudorandom encodings. From the metrics in Table 2, it is possible to make the following observations about pseudorandom encoding:

1. Though not unexpected, Table 2 demonstrates that as the quantization becomes coarser, the performance decreases. This observation also applies to comparisons between phase-only and biamplitude phase encoding.
2. The small observed differences between the gray-scale images for the m-ary 2 [Fig. 6(c)] and phase-only [Fig. 6(d)] encodings are borne out by the relative differences between their metrics in Table 2. The only significant difference is in diffraction efficiency, which is not reflected in the gray-scale images.
3. The theoretical diffraction efficiency of Eq. (26) and the SNR of Eq. (25) are in close agreement with the simulated results.
4. The values of SNR are 12–18 times larger than those of SPR. The much lower SPR is probably due in large part to the background speckle noise being exponentially distributed in intensity.[15] The exponential pdf decreases to zero very slowly with increasing values of intensity. There are also on the order of $N = 16{,}384$ (the number of pixels and also the space–bandwidth product of the speckle noise pattern) independent noise samples in the diffraction pattern. In 16,384 independent Bernoulli trials,[16] there is a probability of approximately 50% that the maximum value is 10× greater than the average value of an exponential distribution, and there is a 10% probability that the maximum value is 12× greater than the average. This leads to the possibility of the maximum-valued noise peak used to calculate SPR being substantially larger than the average value of noise intensity used to calculate SNR. However, since our maximum noise peaks always tend to be somewhat larger than the 50th percentile, there appear to be other contributions to the background that we have not been able to account for.

The model is presented mainly to provide insight into the noise properties of pseudorandom encoding. These properties are noticeably different from those for nonrandom encoding, as we show presently.

The nonrandom-encoding algorithms can all be viewed as a point-by-point nonlinear transformation of the desired complex values. This type of nonlinearity usually produces mixing products that appear as unwanted sum and difference frequencies in the diffraction pattern and as interference in the signal strengths of the desired frequencies.[2] The mixing products are not very evident in the gray-scale image for nonrandom phase-only encoding [Fig. 6(a)]. The interference, which perturbs the uniformity of the spot array, is evident in the intensity curves for phase-only as well as for m-ary 1 and ternary encodings in Fig. 7(b). The m-ary 2 ($\gamma = 1.3$) encoding is the most uniform of the nonrandom encodings shown. Some mixing products at frequencies other than for the desired 7 × 7 spot positions are also apparent in all the intensity curves except for the nonrandom phase-only encoding. However, the mixing products are significantly larger away from the desired spot array, as shown in Fig. 8(a) for the nonrandom ternary encoding. The spot array is designed to lie to the upper left of the optical axis, which gives rise to the strong unwanted harmonics in the lower right corner of the gray-scale image. All the other nonrandom encodings are similar in that the noise is primarily distributed in this same spatial pattern (though with differing intensity levels).

Comparing the simulated nonrandom encodings with pseudorandom encodings, we observe that the pseudorandom encodings all have a speckle/noise pattern of the same average intensity and visual texture over the entire simulated diffraction pattern, as is shown in the close-up views of Figs. 6(b)–6(f). The intensity cross sections for the nonrandom encodings in Fig. 7(b) generally appear less uniform than those for the corresponding pseudorandom curves in Fig. 7(a). The background noise is much more evident on the pseudorandom ternary encoding [Fig. 7(a)] than it is on the nonrandom ternary encoding [Fig. 7(b)]; however, the peak noise of the nonrandom encoding [shown over a wider spatial extent in the cross section in Fig. 8(a)] is significantly larger than the background noise for pseudorandom encoding.

Table 3 provides more detailed information for comparing the individual nonrandom-encoding algorithms with each other and with the results in Table 2 for pseudorandom encoding. The nonrandom diffraction efficiencies are generally much higher than those for pseudorandom encoding. However, the m-ary 2 ($\gamma = \sqrt{1/2}$) encoding has an extremely low diffraction efficiency. This is a direct result of the small value of the maximum complex radius γ, which leads to most of the desired complex values being closer to zero than to a unity magnitude point. This led to our use of the m-ary 2 ($\gamma = 1.3$) encoding, which produces much more uniform spot arrays than does the nonrandom phase-only encoding.

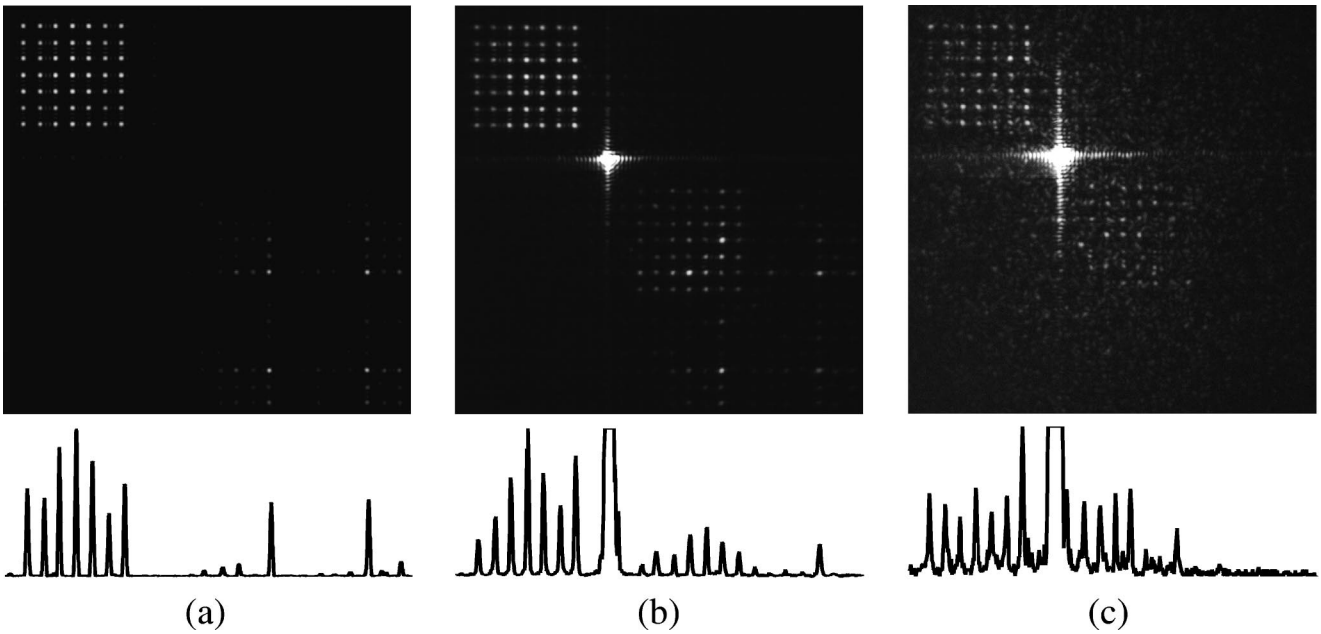


Fig. 8. Delineation of nonlinear effects on encoding: (a) simulated and (b) experimental diffraction pattern intensity for nonrandom ternary encoding, (c) experimental diffraction pattern for pseudorandom ternary encoding. These patterns show a larger view of the diffraction pattern than those in Figs. 6 and 7. Each intensity cross section is along the diagonal of the corresponding gray-scale image. In (a) and (b) the nonrandom ternary encoding produces mixing products, as evident in the lower left corner of each gray-scale image. Although speckle noise is evident in this same region for pseudorandom ternary encoding [(c)], it is much lower in intensity than the mixing products for (b). The saturated spot (centered on the optical axis) in (b) and (c) is primarily a result of the SLM cover glass not being antireflection coated. The most severe effect of the SLM's limited resolution is the appearance, to the lower left of the optical axis, of a duplicate 7×7 spot array in (b) and (c).

For completeness Table 3 reports SNR. This number has little practical use, since much of the noise energy appears in a relatively few noise spikes. This leads to the ratio of SNR to SPR being greater than 100 for four of the five nonrandom algorithms, with the one exception being the maximum-efficiency, phase-only encoding.

The nonrandom phase-only encoding by far produced the highest diffraction efficiency and SPR of any encoding algorithm. In fact, so much energy is in the desired spots that there is little energy left to contribute to noise. However, the spot array is less uniform than for the pseudorandom phase-only encoding. It should be noted that Krackhardt *et al.* reported that by reoptimizing the design, it is possible to lower the diffraction efficiency somewhat, which results in a more uniform array of spots.[11] Nonetheless, pseudorandom phase-only encoding produces reasonably uniform spot arrays and large values of SPR, and it does this without the added computational effort of reoptimizing. This becomes clearer when it is seen that the SPR of the quantized modulation characteristics is much lower for nonrandom encoding (Table 3) than it is for pseudorandom encoding (Table 2). This can be interpreted that pseudorandom encoding is less severely affected (i.e., it is more robust) than nonrandom encoding by quantization effects. The results in Krackhardt *et al.* for binary modulation suggest that reoptimization would be needed for each new type of quantization to minimize these effects. Even with its lower diffraction efficiency, the pseudorandom encoding appears to produce a more faithful reconstruction, with less computational effort, than do the nonrandom encodings. Furthermore, based on the studies in Ref. 5 on biamplitude encoding, it appears likely that there is a way to blend random and nonrandom algorithms for discrete-value modulation such that by adjustment of the value of γ (similar to our optimization of nonrandom m-ary 2 encoding) the uniformity and the SPR are improved over those possible with either random or nonrandom encoding individually. This iteration increases computation, but since only one parameter is adjusted, the computation time should be significantly less than that for the global optimization approaches used in, e.g., Refs. 3 and 11.

E. Procedures Used for the Experiments

We also have attempted to modulate a Hughes birefringent liquid-crystal light valve (set up in a phase-only mode) with the encoded phase values. Cohn *et al.* previously reported experiments on phase-only encoding for this light valve.[15] The current setup differs from the previous setup in that (1) a 488-nm laser is now used in place of a 633-nm laser, (2) the pixel array is now a 128×128 array instead of a 100×100 array [used for Fig. 3(c) in Ref. 15], and (3) the video signal that drives the write light monitor (an Electrohome EDP58XL monochrome monitor with a Hughes high-brightness red tube) is now derived from a Coreco video display card ($S3$ chip set) set to a resolution of 800×600 noninterlaced pixels. Previously, a National Television System Committee (NTSC) signal was the video source. As in Ref. 15, a SLM pixel corresponds to three video lines or 3×3 pixels from the video display card. We have characterized the transfer function from gray-scale values in digital

memory to the phase modulation of the light valve. With the light valve drive voltage set to 27 V p-p (2 kHz) and with proper adjustment of the brightness and the contrast of the monitor, we have realized a nearly linear transfer function in which a gray-scale value of 80 corresponds to zero phase shift and a gray-scale value of 255 corresponds to a phase shift of 2π. The monitor magnification is set to minimum in the horizontal direction. The monitor is reimaged with a 1.9× reduction onto the write side of the light valve. The resulting image is 21 mm × 21 mm, which we have determined to be the maximum input aperture size that allows us to produce diffraction-limited optical Fourier transforms. This area is illuminated on the read side of the light valve with the light polarized along the extraordinary axis of the liquid crystal. The beam converges to a focal point approximately 2 m from the light valve. A 2033 × 2044-pixel cooled CCD camera placed at the focal point records the resulting diffraction patterns.

F. Experimental Results

Figures 6(g), 6(h), and 6(i) show the diffraction patterns for pseudorandom phase-only, m-ary 1, and ternary encoding, respectively. The images differ from the simulated patterns in Figs. 6(d), 6(e), and 6(f), respectively, in that there is a bright spot centered on the optical axis (lower right corner of each image) and that the intensity rolls off/decreases with distance from the optical axis. The bright spot is due primarily to the cover glass of the light valve not being antireflection coated. The roll-off is due to the limited spatial frequency response of the phase of the SLM. The filtering of the phase also produces nonlinear mixing products that contribute energy to the on-axis spot and to an unwanted mirror image of the desired 7 × 7 spot to the lower left of the optical axis in Figs. 8(b) and 8(c). By comparison of Figs. 8(a)–8(c), it can be seen that nonrandom encoding itself produces strong nonlinear mixing products at a few frequencies [at the lower right of Figs. 8(a) and 8(b)]. Instead of producing mixing orders, pseudorandom encoding [Fig. 8(c)] produces an average low level of speckle noise over the entire spatial extent of the diffraction pattern. For either nonrandom- or pseudorandom-encoding experiments, additional nonlinear terms (the mirror images) are present in Figs. 8(b) and 8(c) as a result of the limited phase resolution of the SLM. This distortion, which is due to the limitations of the SLM rather than the encoding method, makes it difficult to make meaningful comparisons between simulation and experiment. Nonetheless, qualitative agreement between Figs. 6(c) and 6(g), 6(d) and 6(h), and 6(e) and 6(i) is seen in that the noise level increases as the quantization becomes coarser.

Potentially much closer agreement between theory and experiment is anticipated by using electrically addressed SLM's. Most of the current devices have individually defined electrodes for each pixel, which would minimize resolution loss that is due to electrostatic fringing fields across the liquid-crystal layer. These SLM's are available only as research-grade or custom devices. In our case this has led to significant delays in obtaining a fully functional device. We have previously observed very close agreement between simulation and theory for pseudorandom phase-only encoded designs that were implemented as diffractive optical elements.[17,18] These results assure us that given an adequately ideal phase-only (or coupled amplitude–phase) SLM, it would be possible to encode fully complex functions, even if the SLM is capable of producing only a few discrete levels of modulation.

7. SUMMARY

In this paper we have derived a statistically based algorithm that with as few as three discrete modulator values encodes a desired complex value to a single pixel in an average sense. This pseudorandom ternary algorithm can be applied directly to SLM's that produce only three values. For SLM's that produce several discrete values, multiple groups of three values can be used to subdivide the complex plane into smaller areas that are ternary encoded, which consequently produces smaller amounts of encoding error. The effect of quantization on pseudorandom encoding is well characterized in a simple model of SNR that depends on only four parameters: two that depend on the signal to be encoded (signal bandwidth and signal diffraction efficiency), one that depends on the modulator characteristic (number of SLM pixels), and one (average encoding error per pixel) that depends on both the signal to be encoded and the modulator characteristic curve. We demonstrated in our simulations that this metric accurately describes the SNR of spot array generators.

To better appreciate the performance of pseudorandom-encoding algorithms, we have compared these algorithms with currently used algorithms in which a desired fully complex function is mapped into modulation values in a systematic and nonrandom way. This function (selected by a global optimization procedure), when mapped to a continuous, phase-only modulation characteristic, produces a diffraction pattern that has the highest diffraction efficiency and the highest SPR of all encoding algorithms studied herein. However, it is less uniform than four of the five pseudorandom algorithms. Furthermore, the SPR of the nonrandom algorithms becomes much worse than that of the pseudorandom algorithms that use similarly quantized modulation characteristics.

Even though the pseudorandom algorithms are less diffraction efficient than nonrandom algorithms, these results nonetheless indicate that pseudorandom algorithms offer significant advantages in terms of fidelity of the diffraction pattern. The advantages of complex-valued encoding techniques, despite their lower diffraction efficiencies, are further amplified by Kettunen *et al.*[19] These advantages, coupled with the low computational overhead of the encoding algorithm and its ability to place a signal anywhere in the available space–bandwidth product of the SLM, make it especially useful for today's low-pixel-count SLM's. We have also observed that the performance advantages of pseudorandom encoding over nonrandom encoding are even more pronounced when it is not possible to maximize the diffraction efficiency of the desired fully complex function, such as in many real-time and time-critical applications.

ACKNOWLEDGMENTS

This study was supported by NASA cooperative agreement NCC5-222 and Office of Naval Research grant N00014-96-1-1296.

Send all correspondence to Robert W. Cohn, The ElectroOptics Research Institute, Room 442, Academic Building, University of Louisville, Louisville, Kentucky 40292; tel, 502-852-7077; fax, 502-852-1577; e-mail, rwcohn01 @ulkyvm.louisville.edu.

REFERENCES

1. B. R. Brown and A. W. Lohmann, "Complex spatial filter," Appl. Opt. **5**, 967–969 (1966).
2. R. W. Cohn and L. G. Hassebrook, "Representations of fully complex functions on real-time spatial light modulators," in *Optical Information Processing*, F. T. S. Yu and S. Jutamulia, eds. (Cambridge U. Press, Cambridge, UK, 1998), Chap. 15, pp. 396–432.
3. N. C. Gallagher and B. Liu, "Method for computing kinoforms that reduces image reconstruction error," Appl. Opt. **12**, 2328–2335 (1973).
4. R. W. Cohn and M. Liang, "Approximating fully complex spatial modulation with pseudorandom phase-only modulation," Appl. Opt. **33**, 4406–4415 (1994).
5. 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.
6. R. W. Cohn, "Pseudorandom encoding of complex-valued functions onto amplitude-coupled phase modulators," J. Opt. Soc. Am. A **15**, 868–883 (1998).
7. A. Papoulis, *Probability, Random Variables and Stochastic Process*, 3rd ed. (McGraw-Hill, New York, 1991), pp. 53–55 and 211–212.
8. R. W. Cohn and M. Liang, "Pseudorandom phase-only encoding of real-time spatial light modulators," Appl. Opt. **35**, 2488–2498 (1996).
9. R. W. Cohn, "Analyzing the encoding range of amplitude–phase coupled spatial light modulators," in *Spatial Light Modulators*, R. L. Sutherland, ed., Proc. SPIE **3297**, 122–128 (1998).
10. C. B. Burckhardt, "A simplification of Lee's method of generating holograms by computer," Appl. Opt. **9**, 1949 (1970).
11. U. Krackhardt, J. N. Mait, and N. Streibl, "Upper bound on the diffraction efficiency of phase-only fanout elements," Appl. Opt. **31**, 27–37 (1992).
12. L. B. Lesem, P. M. Hirsch, and J. A. Jordon, Jr., "The kinoform: a new wavefront reconstruction device," IBM J. Res. Dev. **13**, 150–155 (1969).
13. F. Wyrowski, "Upper bound of the diffraction efficiency of diffractive phase elements," Opt. Lett. **16**, 1915–1917 (1991).
14. R. D. Juday, "Optimal realizable filters and the minimum Euclidean distance principle," Appl. Opt. **32**, 5100–5111 (1993).
15. R. W. Cohn, A. A. Vasiliev, W. Liu, and D. L. Hill, "Fully complex diffractive optics by means of patterned diffuser arrays: encoding concept and implications for fabrication," J. Opt. Soc. Am. A **14**, 1110–1123 (1997).
16. A. Papoulis, *Probability, Random Variables and Stochastic Process*, 3rd ed. (McGraw-Hill, New York, 1991), pp. 43–56.
17. R. W. Cohn and M. Liang, "Spot array generator designed by the method of pseudorandom encoding," presented at the Annual Meeting of the Optical Society of America, Rochester, New York, October 20–24, 1996.
18. M. Duelli, D. L. Hill, and R. W. Cohn, "Frequency swept measurements of coherent diffraction patterns," Eng. Lab. Notes **9**, 3–5 (1998).
19. V. Kettunen, P. Vahimaa, and J. Turunen, "Zeroth-order coding of complex amplitude in two dimensions," J. Opt. Soc. Am. A **14**, 808–815 (1997).

ERRATA

Ternary pseudorandom encoding of Fourier transform holograms: errata

Robert W. Cohn and Markus Duelli

The ElectroOptics Research Institute, University of Louisville, Louisville, Kentucky 40292

[S0740-3232(99)01405-2]
OCIS codes: 230.6120, 090.1760, 030.6600, 070.0070.

Owing to the printing process the gray-level values below 60 (out of 256 levels) appear as black in Figs. 6 and 8 of Ref. 1. This makes it difficult in Fig. 8 to delineate between nonlinear effects of encoding and of the SLM. Figure 8 is reproduced here on a glossy paper and with the gray scale scaled by a factor of 2 (and also clipped for gray levels above 255). The inherent nonlinearity in nonrandom encoding produces large undesired diffraction orders that appear in the lower right corner of Fig. 8(a). Applying this encoding to a low-resolution phase-only light valve produces additional diffraction orders, including a bright spot on the optical axis and a set of orders at mirror locations to the desired spot array, as shown in Fig. 8(b). Applying ternary pseudorandom encoding to the same modulator produces the pattern in Fig. 8(c). This figure does not contain the undesired orders that are associated with the nonrandom algorithm of Fig. 8(a) but instead has a broadly spread, low-level background of speckle. Figure 6(i) in Ref. 1 is a closeup of Fig. 8(c). The speckle level is higher in Fig. 6(i) for three levels of quantization than in Fig. 6(h) for four levels of quantization. Figure 6(g) for five levels of quantization has an even lower level of speckle background. The corresponding Figs. 6(d)–6(i) for the simulated encodings show the same trends in background speckle levels. Readers who wish to view a version of Fig. 6 that has higher dynamic range can download the electronic version of the paper from *JOSA A* online or contact the authors for reprints.

The performance measures for phase-only nonrandom encoding were incorrectly reported in Table 3. The signal-to-noise ratio (SNR) was too small and the signal-to-peak-noise ratio (SPR) was too large. The correct

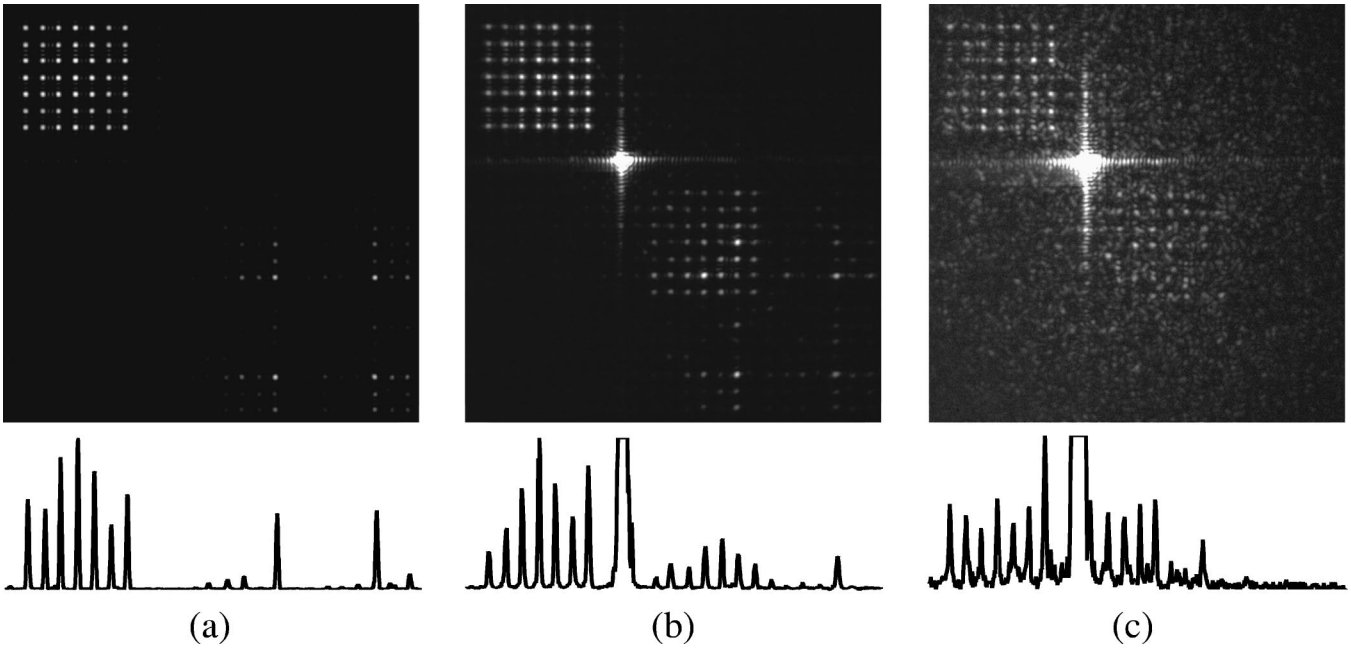


Fig. 8. Delineation of nonlinear effects on encoding: (a) simulated and (b) experimental diffraction pattern intensity for nonrandom ternary encoding, (c) experimental diffraction pattern for pseudorandom ternary encoding. These patterns show a larger view of the diffraction pattern than those in Figs. 6 and 7. Each intensity cross section is along the diagonal of the corresponding gray-scale image. In (a) and (b) the nonrandom ternary encoding produces mixing products, as evident in the lower left corner of each gray-scale image. Although speckle noise is evident in this same region for pseudorandom ternary encoding [(c)], it is much lower in intensity than the mixing products for (b). The saturated spot (centered on the optical axis) in (b) and (c) is primarily a result of the SLM cover glass not being antireflection coated. The most severe effect of the SLM's limited resolution is the appearance, to the lower left of the optical axis, of a duplicate 7×7 spot array in (b) and (c).

0740-3232/99/051089-02$15.00 © 1999 Optical Society of America

numbers are SNR = 5400 and SPR = 17. This indicates that the phase-only pseudorandom encoding also produces a more faithful reconstruction than the nonrandom encoding since each encoding has identical SPR but the nonuniformity of the spot array for pseudorandom encoding is nearly half of that for nonrandom encoding.

Send all correspondence to Robert W. Cohn, The ElectroOptics Research Institute, Room 442, Lutz Building, University of Louisville, Lousiville, Kentucky 40292; tel, 502-852-7077; fax, 502-852-1577; e-mail, rwcohn01 @ulkyvm.louisville.edu.

REFERENCE

1. R. W. Cohn and M. Duelli, "Ternary pseudorandom encoding of Fourier transform holograms," J. Opt. Soc. Am. A. **16**, 71–84 (1999).