

Computer Generated Hologram
(and a few odd things people have done in the past)

James Wu
May 11, 2007

Table of Content

Introduction	3
Hardware	4
Proximity Correction	4
Ablation by Excimer Laser	7
Spin-on Glass	10
Calibration of Mask Alignment	12
A circular CGH Writer	13
Two-Photon Four-Level Material	15
Algorithm	16
Fourier Transform and the 4-F System	16
Hough Transform	19
Errors and Noise	21
Application	22
Digital Image Hiding	22
Security Devices	24
Viewing Angle Enhancement with a Superresolution Phase Mask	24
Parallex 2D Encoding CGH	27
A Light-Light Switch	27
Active CGH for Aspheric Testing	29
Encoding Images on the 0 th -order of a Quaternary Phase CGH	31
Derivation	35
Matlab Simulations	41
Conclusion	48
References	49

INTRODUCTION

A hologram is an interference recording of a 3D surface. Upon the proper recording, reconstruction, and viewing conditions, unlike the traditional 2D photography, the image appears to be 3D again. A computer generated hologram or a CGH is a computed interference pattern of such 3D scene. A CGH can be imprinted on a storage media, and the “volumetric” data can later be retrieved and processed. It offers the advantage of an extremely high data density and parallel processing (it displays a volume of data instead of bits of data in series).

This report surveys three components concerning the making of a computer generated hologram – the hardware developments, algorithms, and a few interesting and specific applications utilizing CGH’s. See Figure 1. All three areas are intertwined, and it is impossible to discuss one without examining the others. The resulting CGH systems are very application specific. The goal of this report is to bring up a few “odd” things that people have investigated in the past, so the readers can be aware of their existence.

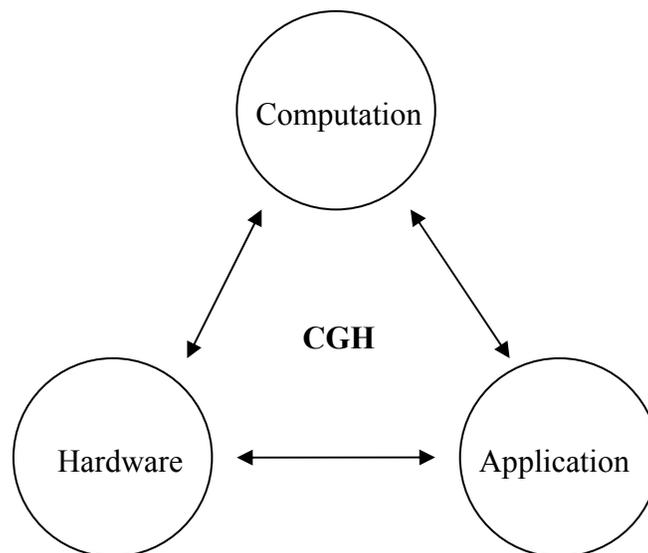


Figure 1: CGH and its relevant components.

The last section of the report is a Matlab code/simulation of a particular type of CGH (quaternary phase CGH) that the author and Professor Dallas worked on in Spring 2007. The simulation shows two images at two different depths of focus, but they are recorded on one common CGH. One of the images is scrambled and hidden in the host image. The hidden image can be reconstructed if a “key” is given; otherwise, it looks like noises.

HARDWARE DEVELOPMENTS

CGH fabrication is primarily based on the micro electronic lithography process. Readers are assumed to be familiar with the lithography process. Following is a list of lithographic tools employed in today's manufacturing processes:

1. Mask vs maskless
2. Projection (contact or imaging) vs e-beam
3. Photo resist (sacrificial material) vs direct printing (laser ablation)
4. Wet vs dry etching (plasma and reactive ion etching, transport or reaction limited)
5. Coherent vs incoherent illumination (compensation methods for smaller and smaller feature size and phase shift technique)
6. Gray tone vs binary
7. Calibration process

The performance of an optical imaging system can be characterized by the resolvable angular resolution $\Delta\theta \approx \lambda/d$; $\Delta\theta$ is the angular resolution, λ is the wavelength, and d is some critical dimension or the pupil size. Bigger d means better angular resolution $\Delta\theta$. Contrary to traditional imaging systems, an "efficient" CGH needs to diffract light into large diffraction angles. This means a smaller feature size d on a CGH yields a larger diffraction angle $\Delta\theta$; there is an analogy one can make between the finesse of a diffraction grating to the diffraction efficiency of a CGH. With the advances in the micro electronic processes, small feature size required by a CGH can indeed be achieved. Researchers are constantly looking for ways to make smaller features more accurately.

Proximity Correction

Diffraction from otherwise a perfect edge is especially a problem for making small features using projection lithography. The edge diffraction and side-lobe intensity "ripples" overlap and blur the neighboring geometries, thus limit the smallest feature one can possibly project and record on a substrate.

Proximity correction is a compensation technique developed for writing a photo resist with an electron beam having a Gaussian intensity profile.¹⁻³ The technique accounts for the 3D dose absorption of the photo resist, and use an iterative process to solve for the

compensated dose that would develop into the desirable structure. $T(x,y)$ represents the desirable dose distribution, and $I(x,y)$ is the proximity function consisting of two Gaussian intensity profiles with parameters α , β , and η (Eq 1). For the simple compensation dose, $D(x,y) = D_0 = 100\%$ inside the exposure area and exactly 0 outside the area. The $T(x,y)$ expression for such simple compensation scheme is described by Eq 2. For the more accurate 3D compensated dose distribution, the goal is to obtain a dose distribution function $D(x,y)$ for the structures described by $T(x,y)$. Both Eq 2 and 3 are convolution integrals, and Proxy is the software used to do this iterative calculations.

$$\begin{aligned}
 I(\rho) &= \frac{1}{1+\eta} I_1(\rho) + \frac{\eta}{1+\eta} I_2(\rho) \\
 &= \frac{1}{1+\eta} \frac{\exp(-\rho^2/\alpha^2)}{\pi\alpha^2} + \frac{\eta}{1+\eta} \frac{\exp(-\rho^2/\beta^2)}{\pi\beta^2}
 \end{aligned}
 \tag{Eq 1}^{1-3}$$

$$\begin{aligned}
 T(x,y) &= (1+\eta)D(x,y) - \eta \int dx' dy' I_2(x-x', y-y') \\
 &\quad \times D(x', y').
 \end{aligned}
 \tag{Eq 2}^{1-3}$$

$$\begin{aligned}
 T^1(x,y) &= (1+\eta)D(x,y) - \eta \int dx' dy' I_2(x-x', y-y') \\
 &\quad \times D(x', y').
 \end{aligned}$$

$$\begin{aligned}
 T^2(x,y) &= (1+\eta)D(x,y) - \eta \int dx' dy' I_2(x-x', y-y') \\
 &\quad \times T^1(x', y') \\
 T^{k+1}(x,y) &= (1+\eta)D(x,y) \\
 &\quad - \eta \int dx' dy' I_2(x-x', y-y') T^k(x', y').
 \end{aligned}$$

$$\mathbf{T}_2(\mathbf{x},\mathbf{y}) = \mathbf{T}_1(\mathbf{x},\mathbf{y}) + (\mathbf{1} + \eta) * (\mathbf{D}(\mathbf{x},\mathbf{y}) - \mathbf{D}_1(\mathbf{x},\mathbf{y}))
 \tag{Eq 3}^{1-3}$$

Figure 2 is an illustration of the 3D corrected dose exposure for making a Ronchi ruling. It can be seen that the edges of the structure receive more/less dosage than the center portion of the structure, according to the description from Eq 3, to account for the “spillage” from the neighboring geometries. The discrete steps are approximation of the continuous $T(x,y)$ function. After etching, the corrected profile should look like the ideal square wave profile.

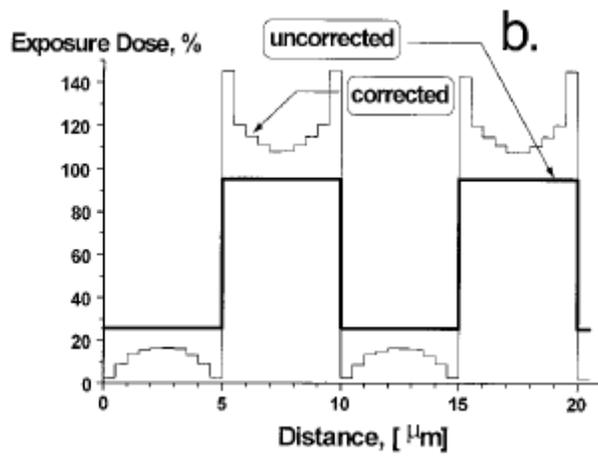


Figure 2: Uncorrected vs corrected Ronchi ruling and patterning the photo resist ¹⁻³

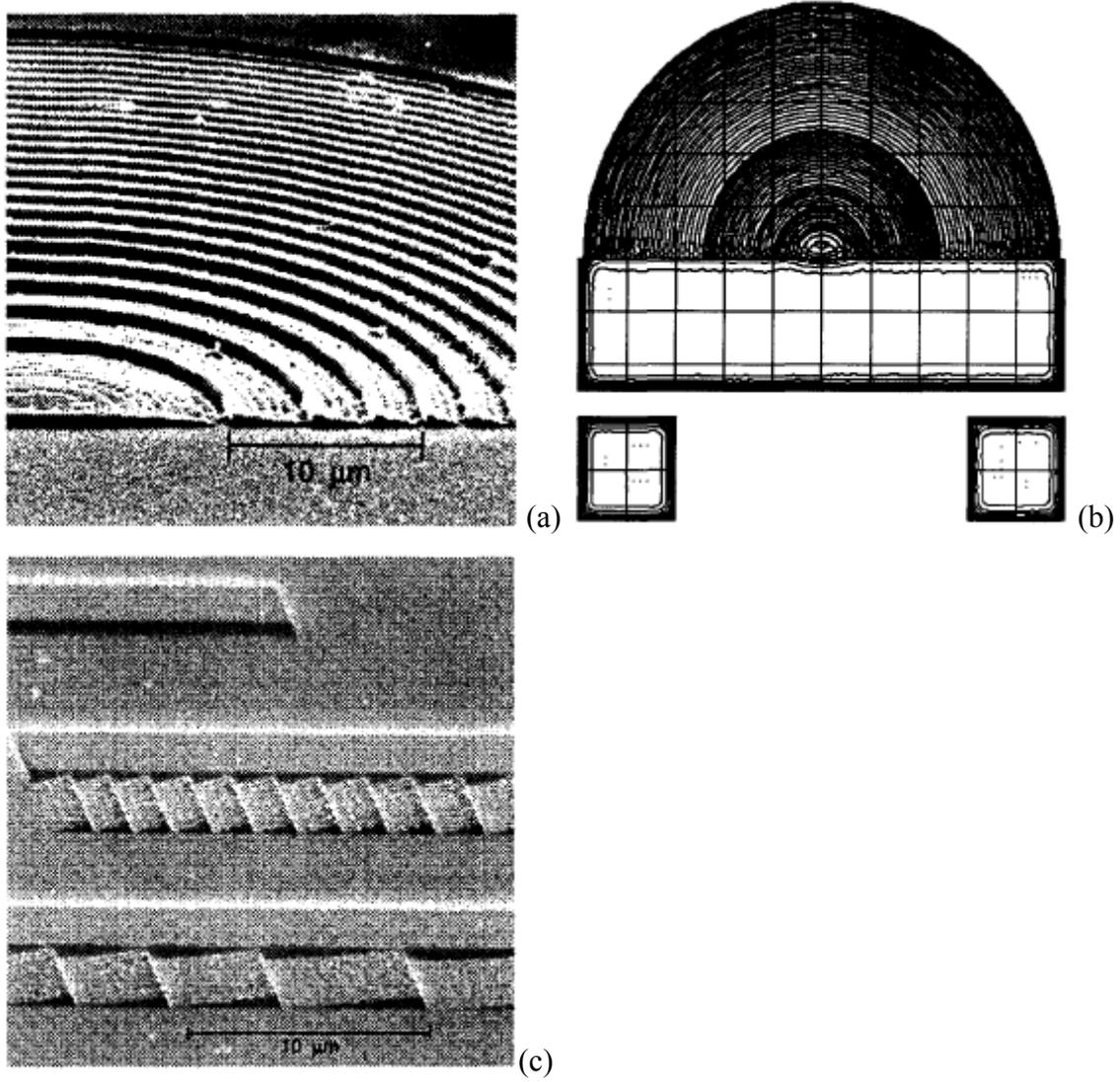


Figure 3: (a) Fresnel “half” zone plate patterned by 3D proximity method, (b) the predicted dosage profile, and (c) calibration geometry patterned on PMMA. ¹⁻³

Figure 3a is an SEM image of a “half” Fresnel zone plate patterned on PMMA. This “half” plate is not a cut-away of a circular Fresnel zone plate, but rather a geometry directly patterned on the PMMA using the 3D proximity calculation. Figure 3b is the predicted dose profile that makes up the “half” plate shown in 3a. And from Figure 3c, one can see the impressive corner sharpness between the wedges patterned by the 3D proximity correction.

There is a group from MIT – Research Laboratory of Electronics that was developing a similar technique for compensating the edge/corner diffraction effects due to the coherent laser illumination and projection (about five years ago). The calculation takes account of the diffraction pattern about the edges and corners of a specific mask pattern; it iteratively calculates the “deformed” pattern that best represents the desirable net shape, and it takes into account of the contrast of the photo resist. The detail of the calculation model employed is not known, but it probably involves physical wave propagation codes, which always complicates the problem when dealing with coherent illumination.

Ablation by Excimer Laser (and calibration of source intensity)

Excimer laser is a UV chemical laser. Typically, it has two components – an inert gas (such as Argon, Krypton, or Xenon) and a reactive gas (such as Fluorine or Chlorine). Lasing occurs under the right excitation condition, which produces an “*excited dimmer* particle”. Most organic and plastic materials exhibit excellent UV absorption, and an excimer laser can be used to remove these materials by ablation (not by burning the materials). During ablation process, a pulse of energy is “injected” into a material. The duration of the pulse is much shorter than the thermal time constant of the material; the amount of thermal energy “injected” or absorbed is so large, that the localized thermal expansion of the material exceeds the rupture strength of the material. Therefore, the localized material is “ejected” or ablated from the surface.

Excimer laser ablation process is inherently a mask patterning process; the poor beam quality from the excimer laser prohibits the direct printing of small features. Energy throughput is a challenge with this mask and ablation process. The discussion here centers on half tone printing. In the half tone mask, the binary feature size is much

smaller than the resolution of the imaging system. The mask can have equally spaced lines/dots at varying width to achieve some gray levels, or it can have equal width of lines/dots at varying spacing. However, these critical dimensions must be carefully design, so only the 0th order diffraction is printed onto the photo resist and the other orders have diffraction angles large enough to miss the printing area.

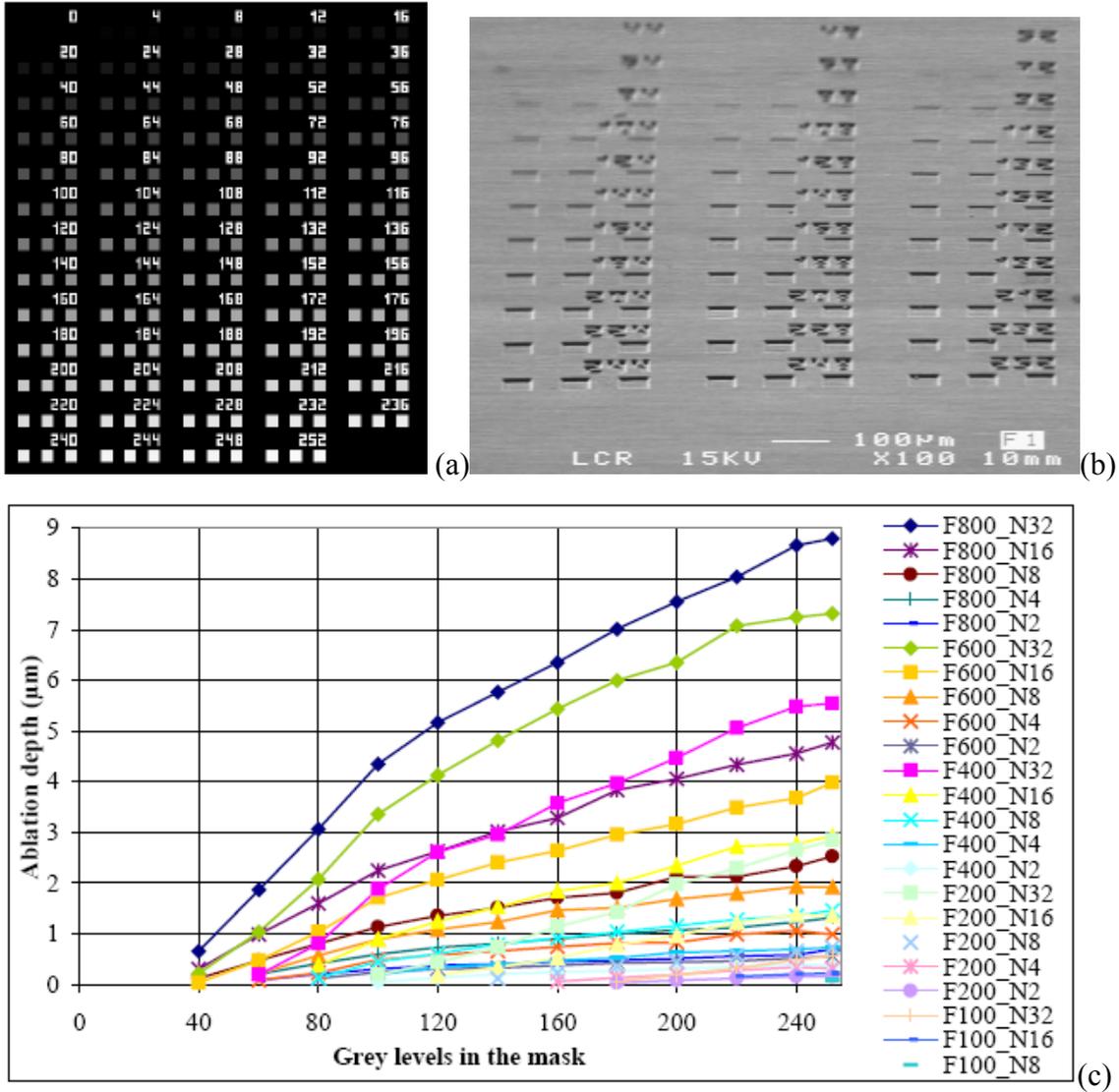


Figure 4: (a) a calibration layout having 64 levels of “grayness”, (b) a calibration sample using half tone mask with $400\text{mJ}/\text{cm}^2$ laser fluence and 32 pulses, and (c) a calibration plot showing the ablation depth as a function of the fluence level and number of pulses. ⁴

Half tone printing has the advantage of using only one mask, so no precision motion stage or mask aligner is needed. However, it is crucial to well calibrate the pulse energy,

pulse duration, and the numbers of pulses in order to fabricate a set of desirable gray levels. A look-up table is often required. In addition, the material re-deposition can also cause non-linearity; there is some probability that the ablated material will settle back on the substrate surface. An Exitech 8000 KrF excimer laser⁴ was used for the following results. It has a wavelength of 248nm, typical pulse energy of 0.5J per pulse, a pulse duration of 20nsec, and max repetition rate of 100Hz. Beam forming optics and homogenizer were implemented to overcome the inherent poor beam quality. Figure 4 shows the calibration methodology and results.

From Figure 4c, it is evident that the ablation depth is non-linear as a function of the desirable gray levels. This is especially true for the high fluence + pulse # combination. It was shown that this non-linearity is at least partially caused by the debris re-depositing back onto the surface; when dragging a contact profilometer across the pockets to measure the ablation height, it was observed that the tip of the profilometer left a track of plowing mark through the debris deposited on the surface. See Figure 5. However, low fluence + pulse # scenarios do appear to be linear. To make a quaternary CGH, one needs at least $(n_{\text{refraction}}-1)\lambda$ of ablation depth in transmission, and the low fluence + pulse # combination seems to be suitable.

Figure 5 shows the SEM images of a polycarbonate (PC) vs benzocyclobutene (BCB) surfaces. Side wall angle and micro surface roughness are compared. It is evident that the PC surface is much smoother than the BCB surface. However, the side wall steepness is better on BCB. Both the side wall angle and the micro surface roughness will impact the diffraction efficiency of the CGH. The micro surface roughness is measured to be about 30nm R_a (interpreted as PV not RMS), and the side wall angle is about 10°.

The research group⁴ produced a 4-level phase CGH using this fabrication technique, by using the Gerchberg-Saxton algorithm to calculate the phase function, and the reconstructions are shown in Figure 6. It is not clear why the contrast of the experimental result of the reconstructed image 6(c) is so different from the simulated reconstruction 6(b); it is possible that the simulation was adjusted on a log-scale to enhance the result, or the environment for recording the reconstruction was not ideal or was too bright.

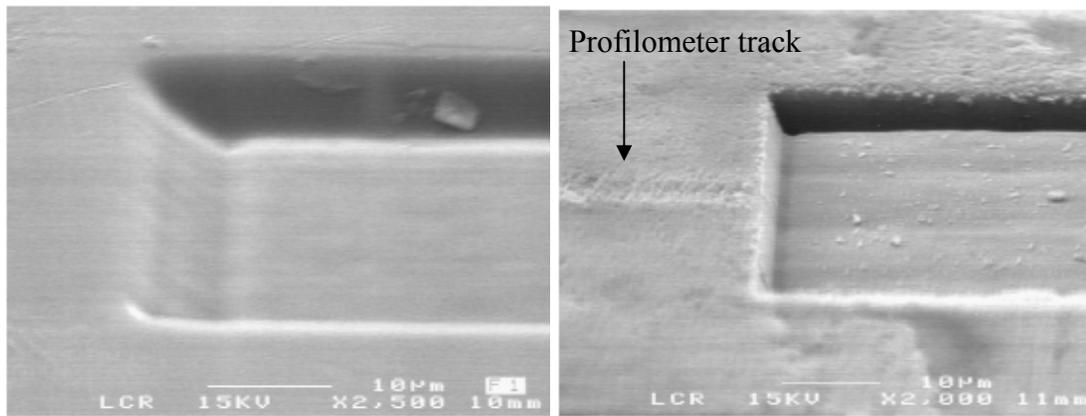


Figure 5: SEM images of PC (left) and BCB (right) patterned surfaces ⁴

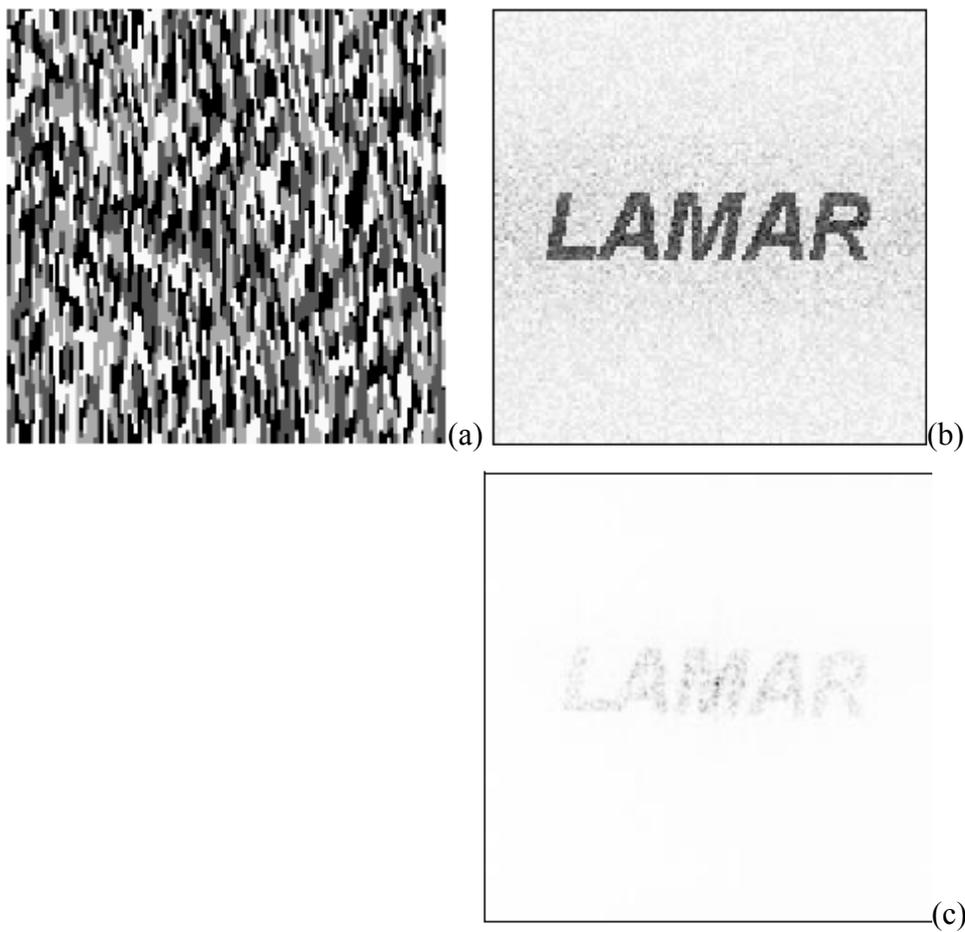


Figure 6: (a) the phase of the Gerchberg-Saxton CGH, (b) simulated reconstruction of the image, and (c) experimental result of the reconstructed image. ⁴

Spin-on-glass (and calibration of source intensity)

Gray scale is similar to half tone, in which they are both capable of producing continuous gray levels but need “dosage” calibration. Generally, a gray scale photo mask employs a

thin layer of silver ion coating, and the transmission of the mask is controlled by the local variations of the silver ion concentration. However, in this research⁵, a very economic spin-on-glass coating is used to modulate the transmission. The spin-on-glass (SOG) changes its transmission characteristics upon exposure to UV light. The exposed SOG film is then permanently fixed by a thermal treatment. The mask then has a continuous gray level without the need of etching steps. This technique was borrowed from the phase-shifting mask technology, where a π shift is used to reduce the side-lobe intensity (by destructive interference principle) during the projection lithography process.

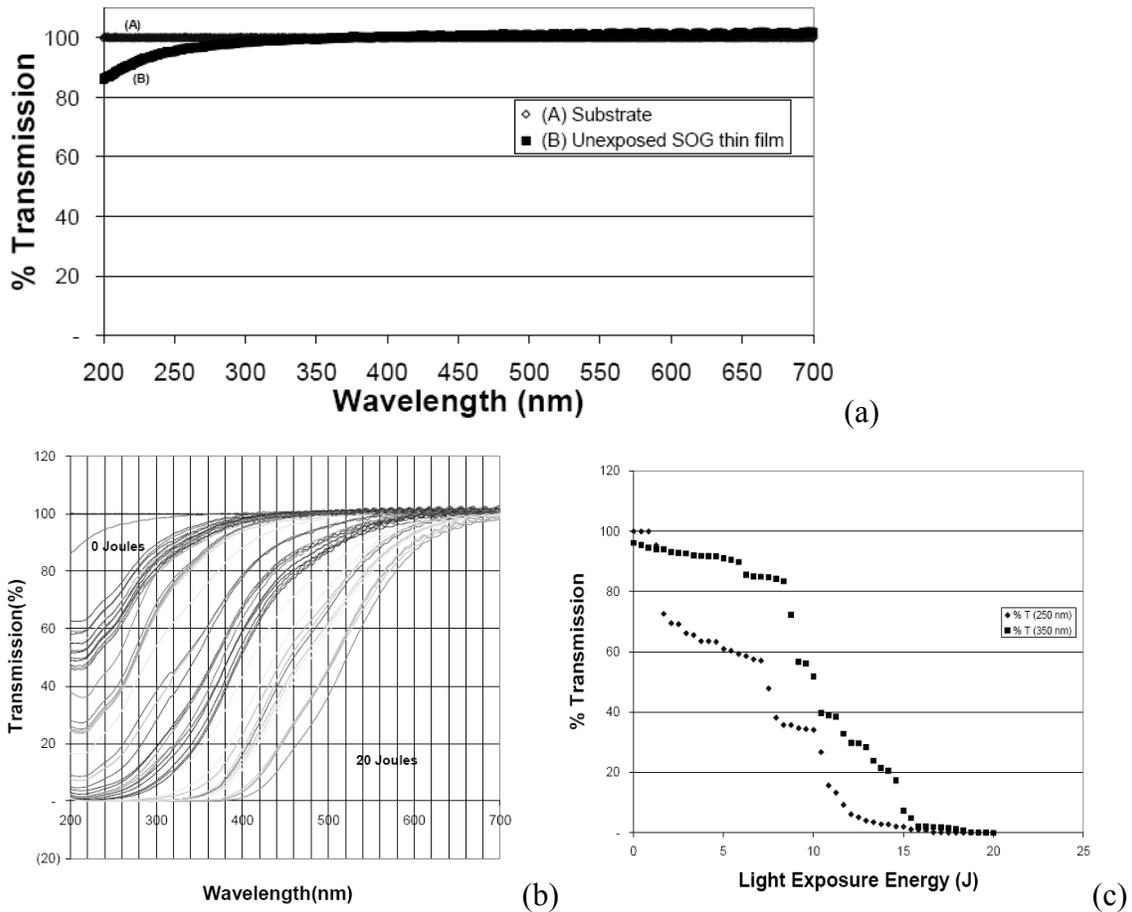


Figure 7: (a) the transmission curve of SOG vs fused silica, (b) the transmission curve of SOG at different UV exposure level, and (c) the transmission curve of SOG at λ_1 and λ_2 .⁵

The unexposed SOG is transparent between UV-Vis. However, the material changes its absorption coefficient between λ (200-600nm) after exposure to UV light. See Figure 7a. The steps in making such mask are 1). mix the chemical reagents to produce the SOG photo sensitive solution, 2). spin-coat 0.5-10 μ m thick film, 3). soft bake, 4). pattern the

SOG film with UV, and 5). hard bake the mask between 1000-1200°C. After hard bake, the SOG film has properties similar to those of fused silica and behaves very much like a bulk material. In addition, the hard bake has no adverse effect on the gray pattern.

Figure 7b shows the transmission of the SOG as a function of the wavelength for different UV exposure dosage. Figure 7c shows the transmission curve of the SOG as a function of the exposure dosage at two particular wavelengths. It is not clear from 7c that whether there is a region where the transmission and the exposure dosage is linear, is otherwise a piece-wise step-function, or that requires a look-up table. And in practice, one wants to select a wavelength from 7b where the transmission of the SOG spans from 0 to 1, which further limits the usable wavelengths.

An off shoot of this SOG technology is the possibility of making an amplitude + phase (perfect) CGH. I claim that it is possible to make an amplitude CGH using the above process first, then coat and soft bake a second layer of SOG, pattern and etch the second layer into a phase CGH without changing the global/local transmission property of the second layer, and thermally fix the second layer as the final step. Alternatively, a light sensitive polymer can be used to transmit and modulate the phase.

Yet another (very distant) off shoot of this SOG technology is the ability to improve the micro surface roughness of the glass and metal mirrors. Improving the micro surface roughness reduces the scattered light, but this is especially difficult to achieve for metal mirrors (to $R_a < 10 \text{ \AA}$ rms level) due to the inherently large metal grain size. We might choose to use SOG as a filler, and to smooth out the micro peaks and valleys (in the order of $R_a \sim 50 \text{ \AA}$ rms). By using the spin-on method and employing the surface tension property of the SOG solution, I believe it is possible to improve the micro surface roughness of a mirror without introducing low order aberration from the coating uniformity. In this proposed process, it is important to anneal the metal, so the mirror surface does not distort after the semi hard bake (apparently, one can not hard bake the metal at 1000°C, so $\sim 300^\circ\text{C}$ is chosen for the semi hard bake).

Calibration of mask alignment

Moire pattern can be employed to measure large scale misalignment⁶. Figure 8 shows two Moire patterns of two different underlying geometries. The underlying geometry in 8a is a straight grating with grating spacing of $3\mu\text{m}$, and 8b is a 10×10 array of different Fresnel zone plates. By themselves alone, it is difficult to map out the large scale distortions. However, when one superimposes two gratings on top of each other and rotate by a small angle with respect to each other (or two Fresnel zone plate arrays), a Moire pattern is formed. The ideal Moire pattern should produce straight fringes. From the deviation of the fringe pattern, one can infer the large scale distortion in the part. This distortion may be resulting from the alignment error between successive masks and patterning. One can then use this information to compensate the reproducible offsets.

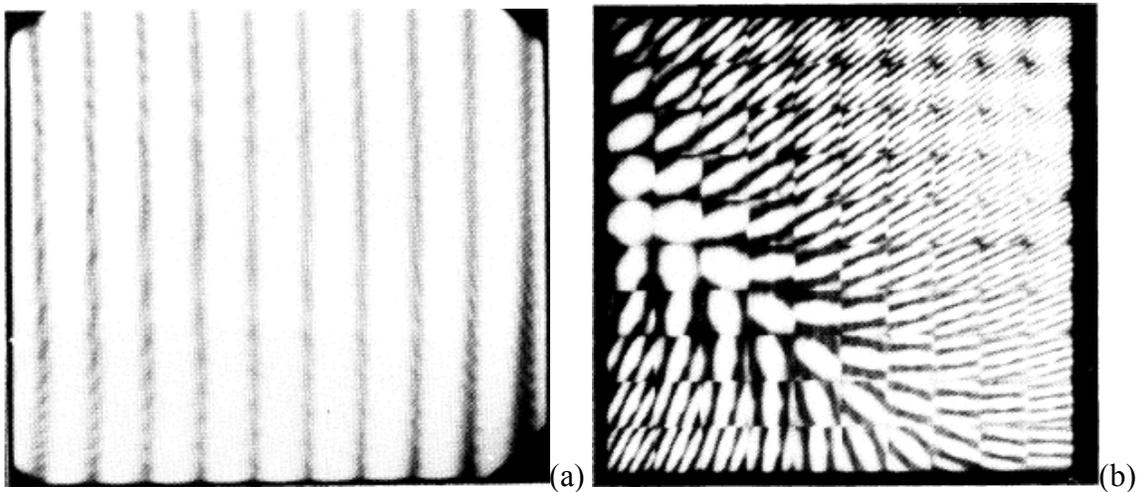


Figure 8: Moire pattern of two (a) gratings and (b) array of Fresnel zone plates⁶

A circular CGH writer

A Russian research group⁷ built a circular CGH writer specifically for making the CGH's for testing aspheric optics. These CGH's typically exhibits circular symmetry, but the laser intensity is modulated by an acousto optic modulator (AOM) to allow some deviation from printing only the circular patterns. A chromium film is used as the sacrificial material, and it is oxidized by heating from a focused Argon laser. However, the oxidation reaction depends strongly on the laser power and duration. Therefore, a power meter is inserted to close the loop with the AOM. See Figure 9. A chemical etchant then preferentially etches away the chromium and the chromium oxide patterns. The result is an amplitude CGH (chromium on glass). The researchers⁷ also pointed out the use of an amorphous silicon film, deposited by RF or magnetron sputtering methods,

as the sacrificial material. The underlying mechanism is the differential chemical solubility of the α -silicon film in the etchant after UV exposure. However, α -silicon is susceptible to oxidation even at the room temperature and pressure. Therefore, it has a life time or shelf life of only 3-6 months.

Two very interesting patterns were made with this circular CGH writer. See Figure 10. In fact, neither pattern looked circular, and the AO modulator clearly demonstrated the capability of the setup. Figure 10a is an amplitude CGH that produce an optical vortex, and 10b is a Fresnel zone plate for a focusing a collimated beam at 45° angle of incidence. These two patterns are best understood by thinking about modulations in both the radial r and rotational θ directions at the same time.

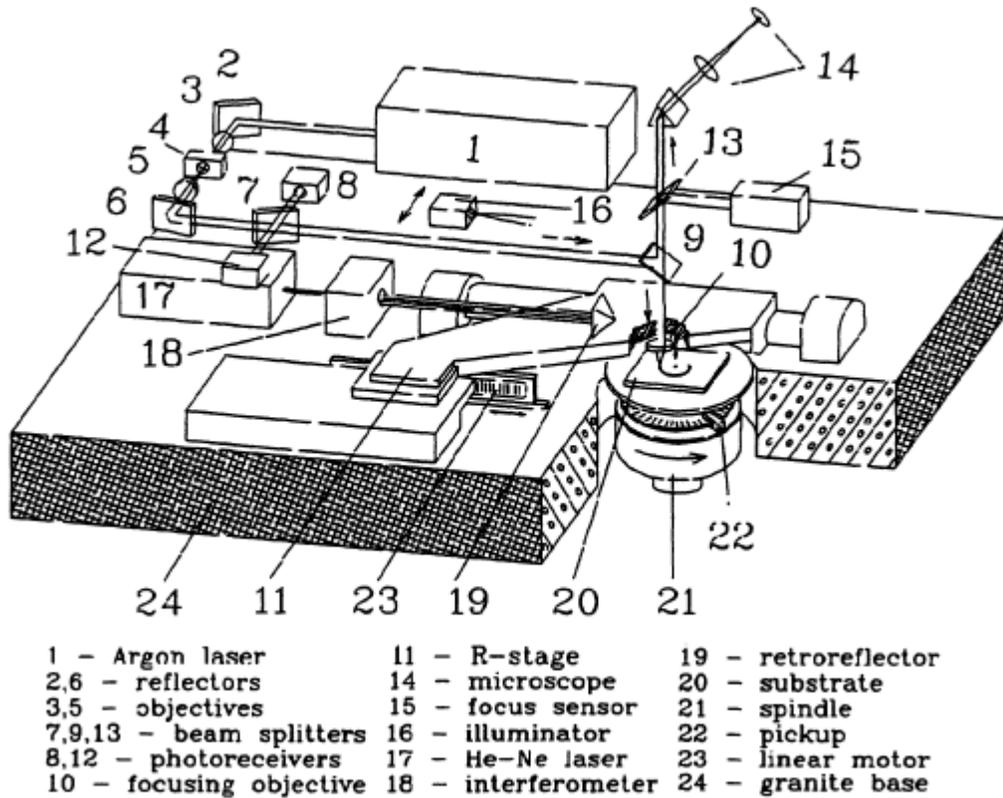


Figure 9: the circular CGH writer, with a power meter feedback loop to control the AOM and a distance measuring interferometer that controls the radial position of the beam. ⁷

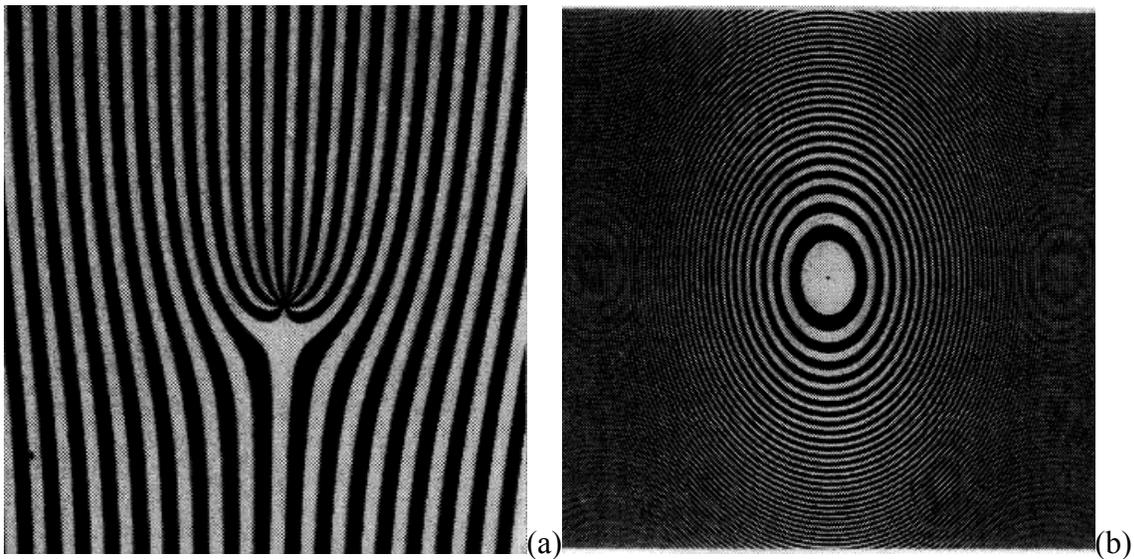


Figure 10: (a) an optical vortex mask and (b) an elliptical Fresnel zone plate printed on the circular CGH writer ⁷

Two-photon four-level material:

First, let's clarify the terminology. The four-levels in this context⁸ are the four electronic energy levels exhibited by the molecule, not a 4-level phase CGH. See Figure 11. This molecule, poly-(alkyl- α -cyanoacrylates), is a common material for adhesives (super glue). Because of the shift in the energy levels, the population of electrons in the C level is allowed only by the depopulation of electrons in the B level through *phonon* relaxation. Therefore, the direct electronic transition from A to C is forbidden. To use this as a holographic recording material, the molecules must first be "pumped" with a UV photon ($\lambda_1 \sim 300\text{nm}$), and then "written/read" with an IR photon ($\lambda_2 \sim 750\text{nm}$). The modulation is recorded in the local refractive index differential due to the concentrations of monomers and polymers, and a change of local density $\Delta\rho$. It should be pointed out that the cross linking of the monomers does not stop immediately after the removal of either photon; there is a self-development chemical process that takes place after the removal of photons. However, contrary to our instinct, this process actually improves density of the polymer matrix and the diffraction efficiency of the hologram.

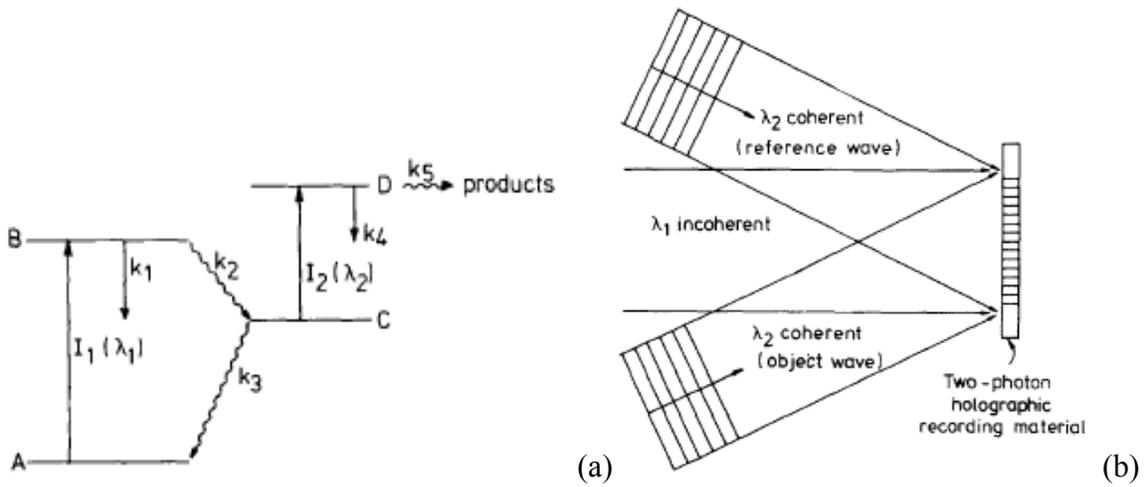


Figure 11: (a) a molecule having four electronic energy levels, λ_1 is the “pump”, λ_2 is the “write/read” photon, and the material is an “inactive” hologram with the absence of either photon, (b) the recording setup. ⁸

ALGORITHM DEVELOPEMENTS

Fourier Transform and the 4-F System

Readers are assumed to have some understandings of Fourier transform (FT) operator and the 4F optical system; the mid plane of the 4F system is the FT plane. In a CGH, the object to be encoded is transformed from the direct space $u(x,y)$ into the spatial frequency space $U(\xi,\eta)$ by the first ideal positive lens and the paraxial free space propagation. The complete or partial complex amplitude of $U(\xi,\eta)$ is then encoded on the CGH. The reconstruction of the object is performed by the second ideal positive lens and the paraxial free space propagation again. See Figure 12. These ideal lenses and the free space propagation behave like a FT operator on the object. And for the moment, let’s consider only coherent illumination.

Before starting to discuss the specific encoding methods, let’s classify the types of the objects to be FT’ed and the types of CGH one can choose to encode the transform. A general object may be described by $u(x,y) = a(x,y) \cdot e^{i\phi(x,y)}$. Most visual images can be described by its spatial amplitude $u(x,y) = a(x,y)$. However, some biological cells are phase objects and have very little or no amplitude modulation. They can be described by $u(x,y) = e^{i\phi(x,y)}$. In optical metrology, the aberrated wavefront can be considered as a

phase object too. Generally, the FT of the object $u(x,y)$ can be written as $U(\xi,\eta) = A(\xi,\eta) \cdot e^{i\Phi(\xi,\eta)}$, and $U(\xi,\eta)$ can be recorded onto a CGH. One may choose to implement only the $A(\xi,\eta)$ part of the complex amplitude $U(\xi,\eta)$ to form an amplitude CGH, or one may choose to implement only the $\Phi(\xi,\eta)$ part to form a phase CGH. A perfect CGH is a CGH that implements both the amplitude and phase.

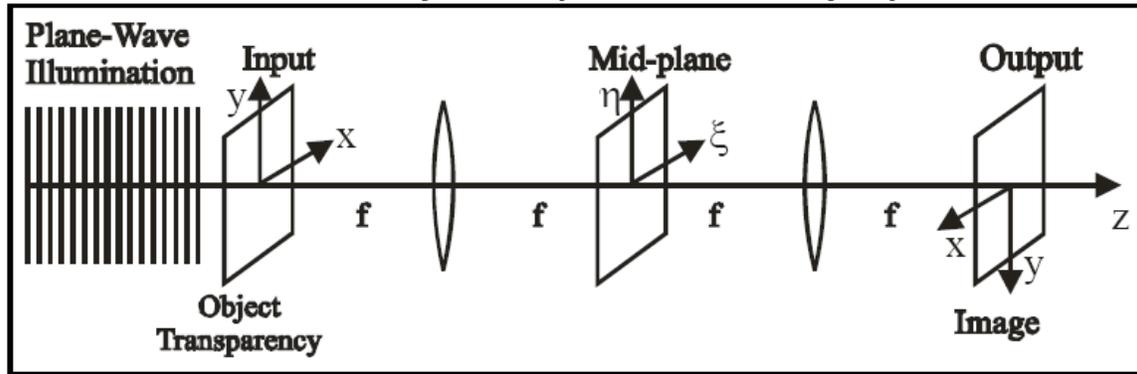


Figure 12: a 4F system (illustration taken from OPTI 512R class notes)

CGH images are often encoded on a carrier. Cosine and Square wave can be written as:

$$T_{\cos}(\xi,\eta) = \frac{1}{2} \cdot [1 + \cos(2\pi x_0 \xi)] \quad \text{Eq 4}$$

$$T_{\text{sq}}(a; x_0 \xi) = \sum_m a \cdot \text{sinc}(ma) \cdot e^{2\pi i m \xi_0 x} \quad \text{Eq 5}$$

a is the duty cycle of the square wave

To encode an image onto the cosine CGH, one would replace the amplitude and phase of Eq 4 with the computed $A(\xi,\eta)$ and $\Phi(\xi,\eta)$ of the object (Eq 6-7). Writing T_{\cos} in Euler formula, it is clear that there are three diffractive orders from this carrier (-1, 0, and +1 of the Fourier series of the object). Again, one might choose to make an amplitude-only CGH by dropping out the $\Phi(\xi,\eta)$ term, or one might choose to make a phase-only CGH by dropping out the $A(\xi,\eta)$ term. The author learned that about 25% of the information is encoded in the amplitude term $A(\xi,\eta)$, and 75% in the phase term $\Phi(\xi,\eta)$.

$$\begin{aligned} T_{\cos}(\xi,\eta) &= \frac{1}{2} \cdot \{1 + A(\xi,\eta) \cdot \cos[2\pi x_0 \xi + \Phi(\xi,\eta)]\} \\ &= \frac{1}{2} \cdot [1 + A(\xi,\eta) \cdot e^{2\pi i x_0 \xi} \cdot e^{i\Phi(\xi,\eta)} + A(\xi,\eta) \cdot e^{-2\pi i x_0 \xi} \cdot e^{-i\Phi(\xi,\eta)}] \quad \text{Eq 6} \end{aligned}$$

$$\begin{aligned} t(x,y) &= \frac{1}{2} \cdot [\delta + u(x+x_0,y) + u(-x+x_0,y)] \\ &= \text{DC spike} + \text{displaced true image} + \text{displaced twin image} \quad \text{Eq 7} \end{aligned}$$

There are infinite numbers of diffractive orders for a square wave grating. The strength at each order is determined by the Fourier series coefficient $a \cdot \text{sinc}(ma)$. The 0th order has the strongest diffraction amplitude. However, an image can not be encoded at the 0th order, because it is a constant and has no amplitude or phase modulation. The ± 1 order has the next highest strength. If the image were to be encoded on the +1 order, one may simply replace the phase $\Phi(\xi, \eta)$ of the object into the exponential term in Eq 5 again. However, the Fourier series coefficient $a \cdot \text{sinc}(ma)$ must be equated to $A(\xi, \eta)$; a becomes the local amplitude/duty cycle modulation. To solve for $a(\xi, \eta)$:

$$\begin{aligned}
 A(\xi, \eta) &= a(\xi, \eta) \cdot \text{sinc}[ma(\xi, \eta)], m = 1 \\
 &= \sin[\pi a(\xi, \eta)] / \pi \\
 a(\xi, \eta) &= \arcsin[\pi A(\xi, \eta)] / \pi, \text{ normalize the argument of arcsine to } (0, 1) \\
 &= \arcsine[A(\xi, \eta)] / \pi, \text{ range of } a \text{ is } (0, 1/2) \\
 &\text{substitute for } A(\xi, \eta) \text{ and } \Phi(\xi, \eta), \\
 T_{\text{sq}}(a; x_0 \xi) &= T_{\text{sq}}\{\arcsine[A(\xi, \eta)] / \pi ; x_0 \xi + \Phi(\xi, \eta)/2\pi\} \qquad \text{Eq 8}
 \end{aligned}$$

Again, Eq 8 will form a true image (at $m = +1$), a twin image (at $m = -1$), a DC spike (at $m = 0$), but also many other noisy diffractive orders. One may choose to drop out the amplitude term $A(\xi, \eta)$ or the phase term $\Phi(\xi, \eta)$ for ease of CGH implementation.

To extend this encoding methodology, one can choose any arbitrary periodic structure to encode the image, by the Fourier series decomposition of the periodic structure and choosing a particular order to encode the image ($m \neq 0$). The presence of the twin image is undesirable, because it reduces the field of view. There is a method for suppressing the twin image by using a particular form of quaternary CGH. This will be discussed later. In addition, one may choose to quantize the amplitude $A(\xi, \eta)$ or the phase $\Phi(\xi, \eta)$ into N discrete levels directly, after performing the operation $u(x, y) \rightarrow U(\xi, \eta)$, without encoding it on a carrier. Last, there are a few very interesting multiplexing techniques for encoding multiple images on the same CGH. However, one needs to incorporate an additional orthogonal degree of freedom for each additional image to be encoded. See OPTI 627 lecture notes.

Hough Transform

A Hough transform (HT) CGH is proposed as an image filter¹⁰. It is important to identify lines and edges in image processing and machine vision. HT is often used to extract this information, and it is very robust against missing pixels. HT can be extended to extract circular and elliptical edges, and it is similar to Radon transform. .

A line in 2D space can be written as $y = mx + b$. Alternatively, the line can be parameterized into an angle θ and a radial distance r from some reference (same as defining the slope and y intersection of the line). Therefore, if a line can be transformed from the (x,y) coordinate into this (r,θ) coordinate, the line appears as a point in this new coordinate. Figure 13 illustrates such transformation, and it can be written as:

$$H(\rho, \theta) = \iint f(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy \tag{Eq 9}$$

which looks like the projection slice theorem and the Radon transform.

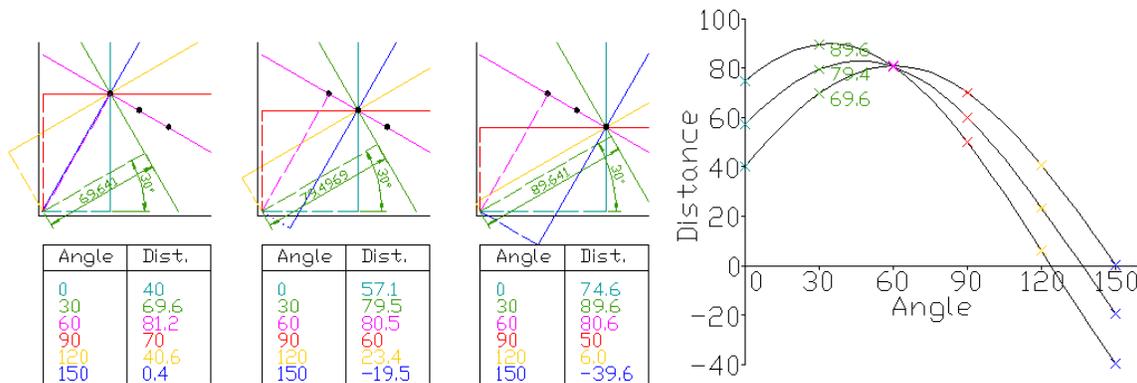


Figure 13: Hough transform of 3 collinear points that defines a line, and how the line is mapped onto the (r,θ) space as an unambiguous point (r_0,θ_0) (illustrations taken from Wikipedia.org)

To implement this transformation with an optical setup, two cylindrical lenses are used¹⁰. Figure 14 illustrates this optical transformation. L1 is a cylindrical lens with a focal length $FL_1 = d$. L1 performs the Fourier transformation of the line in 1D along the “x” direction, and a line transforms into a point in 1D FT. L2 is a cylindrical lens with a focal

length $fL_2 = d/2$, but placed at the orthogonal direction relative to L1. L2 performs the 1:1 imaging of the line onto the Hough transform space. By rotating the cylindrical lens pair, lines with different slopes can be transformed to the θ direction on the Hough transform plane. To do this HT transformation using CGH, one needs to encode N sets of cylindrical lens pair in order to detect lines at N discrete slopes. See Eq 10.

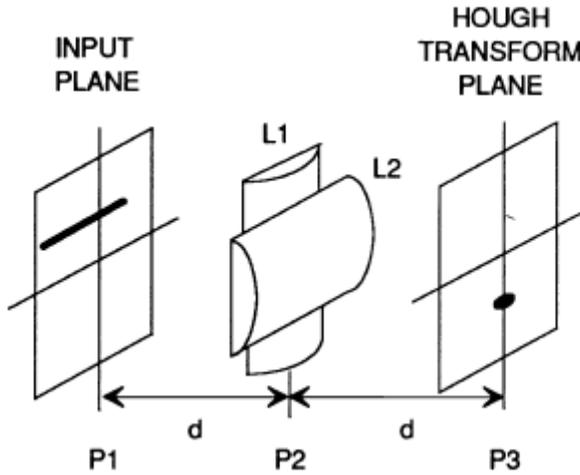


Figure 14: Hough transform optics, L1 has $fL_1 = d$ as a 1D FT optics and L2 is a 1:1 imaging lens in the orthogonal direction. ¹⁰

ideal lens: $t(x,y) = \exp[jk(x^2+y^2)/f_L]$, $k = 2\pi/\lambda$

ideal cylindrical lens: $t(x,y) = \exp[jkx^2/f_L]$

$$t(x, y) = \sum_{i=1}^N \exp\left\{\frac{jk}{2f_L} [(x \cos \theta_i + y \sin \theta_i)^2 + 2(y \cos \theta_i - x \sin \theta_i)^2]\right\} \quad \text{Eq 10}^{10}$$

From Eq 10, it is clear that there are N sets of cylindrical pair oriented at θ_{1-N} directions. The first term is a cylindrical lens with focal length f_L (the 1D Fourier transform lens), and the second term is a cylindrical lens with focal length $f_L/2$ (the 1:1 imaging lens). However, in order to reduce the speckle pattern, each cylindrical lens pair is added a phase tilt. In addition, a global phase tilt is added to the CGH to recenter the pattern. See Figure 15 and Eq 11. This HT CGH can be very useful for simultaneous tracking of multiple targets (view HT CGH as a match filter). And as mentioned earlier, the method can be modified to identify circular and elliptical curves in the (x,y) space.

$$\exp[j2\pi\alpha_c(x+y)] \sum_{i=1}^N \exp\left\{\frac{jk}{2f_L}[(x \cos \theta_i + y \sin \theta_i)^2 + 2(y \cos \theta_i - x \sin \theta_i)^2]\right\} \exp[j2\pi\alpha_s(y \cos \theta_i - x \sin \theta_i)]$$

Eq 11¹⁰

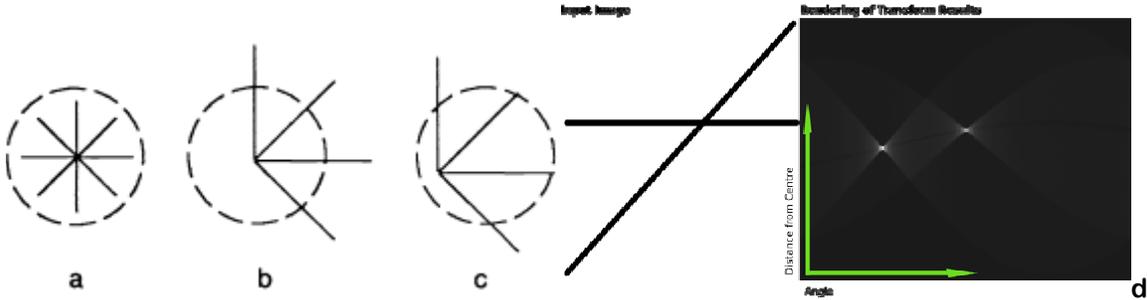


Figure 15: the above illustration is in the HT plane, and each line represents a set of parallel lines in the orthogonal (x,y) space; (a) Eq 10 with $N = 4$, (b) tilts are added to the 1:1 imaging lenses to reduce the speckle patterns from the interference of the 4 slices, (c) a global tilt is added to the CGH to recenter the field of view, and (d) a simulation ¹⁰

Errors and noise

Errors in the CGH can show up as noise in the reconstructed images. Quantization errors are a source of noise. Analytical treatments of noise arise from different quantization errors can be found in OPTI 627 class notes.

One method of minimizing the quantization error is called error diffusion (ED). Using binary CGH as an example, ED first assigns a binary value to a gray scale pattern $\sum_{ij} f(x_i, y_j)$, $f_{ij} = [0,1]$, into $f_{b,ij} = \{0,1\}$. The error at each pixel is $\epsilon_{ij} = f_{ij} - f_{b,ij}$. The error ϵ_{ij} is then distributed to the neighboring unprocessed pixels of (i,j). ED can be 1D or 2D, and it can be binary or multi levels. The Floyd and Steinber's error diffusion algorithm spread the error to four unprocessed neighboring pixels, $(i,j+1)$, $(i+1,j-1)$, $(i+1,j)$, and $(i+1,j+1)$. The ED processed images can have a few distinct artifacts, and they are called "worm", "tear", and "checker board".

Historically, optical data processing was a very active research topic. It is a very math- and statistics-heavy subject. An example of such research is Ref 11; it discusses different error diffusion methods and how to do signal processing using the quantized amplitude and phase CGH's. SNR is a merit function for characterizing different encoding methods.

APPLICATIONS

Digital image hiding

A scrambled secret image is imbedded into a host image in the FT domain¹², and the secret image always has an array size smaller than the host image.

A secret image is first Arnold transformed (AT) in the direct spatial domain. Figure 16 shows the process of Arnold transform. The resulting image becomes unrecognizable after a few AT iterations. However, AT eventually restores the original image after N iterations, and the process is periodic with a period N . Eq 12 describes a simple AT operation, and it has some unique Eigen function properties. See <http://mathworld.wolfram.com/ArnoldsCatMap.html>.

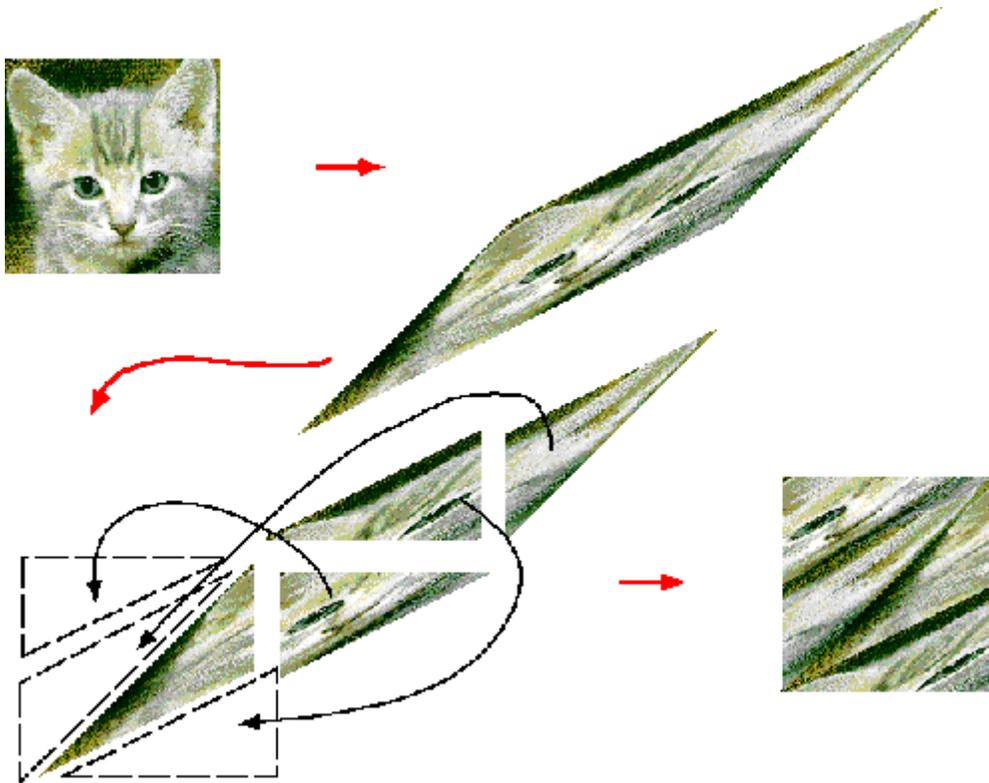


Figure 16: process of Arnold transform; after a few iterations, the cat is no longer recognizable and the image looks like noises; however, the original image is eventually restored. See <http://hypatia.math.uri.edu/~kulenm/diffeqaturi/victor442/index.html>.

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \bmod n \quad \text{Eq 12}$$

Discrete cosine transform (DCT) is applied to both the host image and the scrambled secret image. Eq 13 shows the 2D forward discrete Fourier analysis of an image $A(i,j)$ into the coefficients $B(k_1,k_2)$.

$$B(k_1, k_2) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} 4 \cdot A(i,j) \cdot \cos \left[\frac{\pi \cdot k_1}{2 \cdot N_1} \cdot (2 \cdot i + 1) \right] \cdot \cos \left[\frac{\pi \cdot k_2}{2 \cdot N_2} \cdot (2 \cdot j + 1) \right] \quad \text{Eq 13}$$

The coefficients of the scrambled secret image $B_{\text{secret}}(k_1, k_2)$ are scaled and added to the coefficients of the host image $B_{\text{host}}(k_1, k_2)$, or $B_{\text{CGH}}(k_1, k_2) = B_{\text{host}}(k_1, k_2) + c \cdot B_{\text{secret}}(k_1, k_2)$. The constant c should be small, so the reconstructed host image from CGH coefficients $B_{\text{CGH}}(k_1, k_2)$ can still closely resemble the original host image.

In order to extract the secret image, a few “keys” must be known; the host image, the scale constant c , and the parameters of the Arnold transform.

Another method of scrambling an image is by applying a coding and decoding mask in the FT domain¹³. Figure 17 shows the hardware implementation of such method, but the concept can be applied to the digital images as well. First, a conjugate pair of coding and decoding mask is generated. The hologram $h(x,y)$ contains the FT of the image $a(x,y)$ scrambled by the coding mask $c(x,y)$. The image reconstruction is only possible if the decoding mask and its “location” are given.

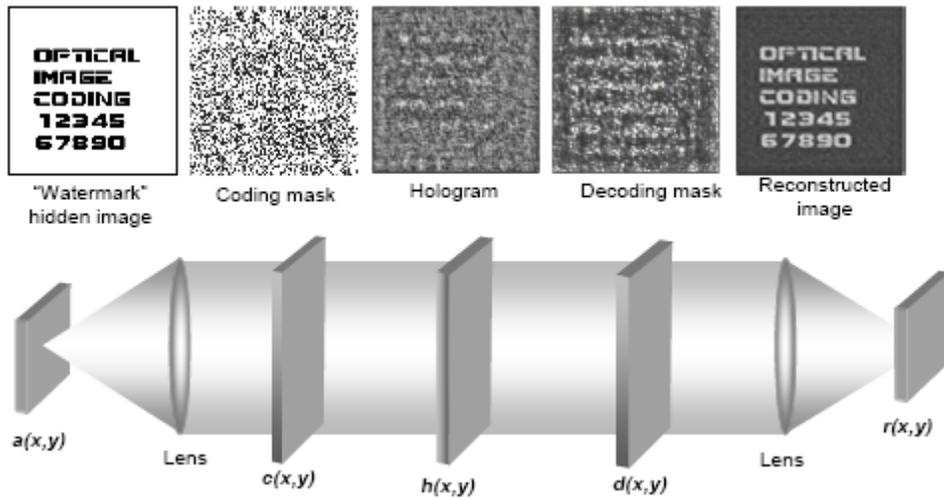


Figure 17: scrambling an image by coding and decoding mask.

Security devices

It is possible to use a CGH as an authentication device. This device does not need to form an image, but rather served like a seal or stamp¹⁴. Figure 18 shows such a device. There are three distinct surface characteristics on this device; a like structures on top of the grating, and a specific micro surface texture pattern. Such a “diagonal” grating, many point- device is manufactured in steps; CGH pattern is etched (such as the grating structure), interference holographic pattern is projected (such as the point-like structures on top of the grating), and the flat area is embossed with some specific micro surface texture. It is not clear from the article as to what kind of inspection machine is needed to authenticate this security device.

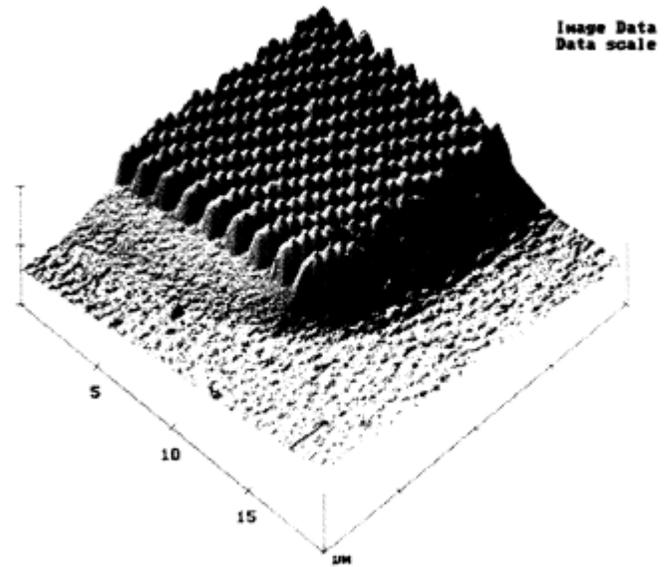


Figure 18: a multi level multi pattern security device.

Viewing angle enhancement with a superresolution phase mask

The bandwidth required to display an active CGH motion picture is very large. Consider the following problem: a display of 10mm x 10mm with a viewing angle of 14°, having

4096 x 4096 pixels of 2.5 μ m size. This amounts to more than 16Mbits/frame¹⁵. An RGB display updates at 33Hz will increase the bandwidth by another 100x. The computational demand is just too great. Therefore, the goal is to reduce the number of pixels (increase the pixel size) while maintaining the viewing angle (related to the large diffraction angle and small pixel size), but achieving a similar perceived image quality (not necessarily equal to the SNR of the displayed images) with the larger pixel size. One-step phase retrieval algorithm, FPGA, and binary CGH using a spatial light modulator (SLM) are designed to bring such display within reach.

The first challenge is to increase the replay field by eliminating the twin image from the CGH. A particular type of quaternary phase CGH will do just that. Since about 75% of the information is carried in the phase term $\Phi(\xi,\eta)$ of the complex amplitude, the amplitude term is simply “thrown away”. Experiences show that the reconstructed image from such phase CGH is almost always of reasonable quality. To eliminate the twin image from such CGH, the phase term $\Phi(\xi,\eta)$ will be encoded twice; once in a cosine CGH and a second time in a sine CGH. The cosine and the sine CGH are then shifted by a phase $\pi/2$ (or i) apart. The twin image from the cosine CGH then cancels out the twin image from the sine CGH exactly. See Eq 14.

$$\begin{aligned}
 T_{\cos} &= \cos[2\pi x_0 \xi + \Phi(\xi,\eta)] = \frac{1}{2} \cdot (e^{2\pi i x_0 \xi} \cdot e^{i\Phi} + e^{-2\pi i x_0 \xi} \cdot e^{-i\Phi}) \\
 T_{\sin} &= \sin[2\pi x_0 \xi + \Phi(\xi,\eta)] = \frac{1}{2} i \cdot (e^{2\pi i x_0 \xi} \cdot e^{i\Phi} - e^{-2\pi i x_0 \xi} \cdot e^{-i\Phi}) \\
 T_{\text{quad}} &= T_{\cos} + i \cdot T_{\sin} \\
 &= e^{2\pi i x_0 \xi} \cdot e^{i\Phi} \\
 &= e^{2\pi i x_0 \xi} \cdot \text{phase}[U(\xi,\eta)]
 \end{aligned}
 \tag{Eq 14}$$

recall that $u(x,y) \rightarrow U(\xi,\eta) = A(\xi,\eta)e^{i\Phi(\xi,\eta)}$. And with this quaternary CGH, Eq 14 shows that only the true image is reconstructed. However, the original image $u(x,y)$ is now shifted by an amount $u(x_0+x,y)$. There is no DC spike in a sine/cosine quaternary phase CGH. This method can also be applied to a square wave quaternary phase CGH. The twin images will be cancelled out in the same way. A simulation of such binary vs quaternary CGH is shown in Figure 19. The removal of the twin image effectively increases the replay field.

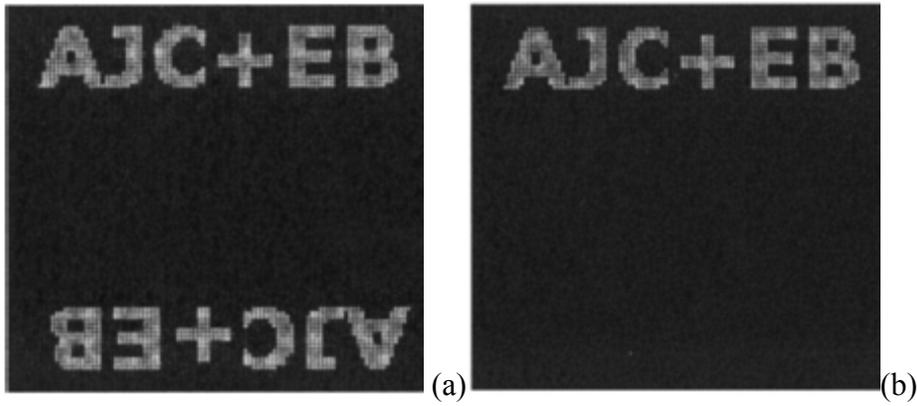


Figure 19: simulated reconstruction of (a) binary CGH with true and twin images and (b) quaternary CGH with a displaced true image and no twin image.

To implement this quaternary CGH, one's natural instinct is to make the SLM a four phase modulation instead of the original binary modulation. However, this would require smaller pixel sizes in order to maintain the same spatial sampling rate, which is just the opposite of the initial intent of resolving the bandwidth issue. A very clever technique of achieving a quaternary phase is as followed: let the SLM remains a binary phase of $\{0, \pi\}$ for its $P \times P$ pixels, etch a binary phase mask having randomly distributed $\{0, \pi/2\}$ $M \times M$ pixels. Let M be an integer multiple of P , and place the binary phase mask in front of the SLM. The result is a quaternary CGH having $\{0, \pi/2, \pi, 3\pi/2\}$ phases. See Figure 20.

There is a penalty associated with this method – noise. The ratio M/P introduces additional noises in the reconstructed image. For $M/P = 2$, the SNR is decreased by 3dB. However, this degradation asymptotically slows down. In fact, for $M/P = 12$, the SNR only decreases by a total of 4.6dB. This is a wonderful realization, because one may use a larger SLM pixel dimension to reduce the P^2 numbers of DFT calculations, but at the same time maintain the viewing angle using a randomly distributed binary phase mask having M^2 pixels (again, M is an integer multiple of P).

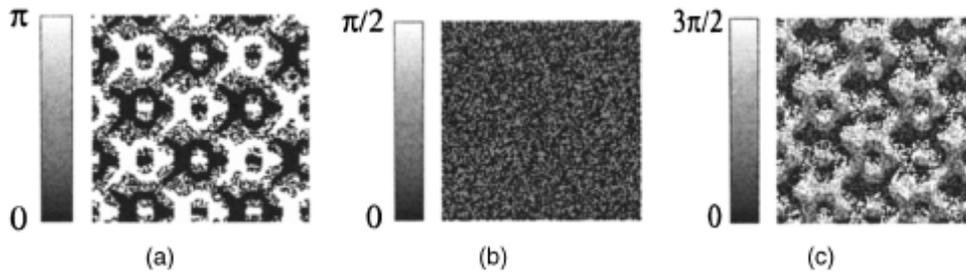


Figure 20: (a) binary phase CGH $\{0,\pi\}$ from SLM, (b) randomly distributed binary phase CGH $\{0,\pi/2\}$, and (c) combination of (a) and (b) forming a quaternary CGH.

The last piece of this technology utilizes of the integration process performed by the human eyes. It has been shown how the noise in each frame of the CGH image is independent and identically distributed (i.d.d.). Even though each image frame has the same SNR, but the human eyes can perceive and integrate over multiple frames at a frame rate $>30\text{Hz}$. Therefore, the noise in each frame averages out. It has been shown that the average of 4 subframes is perceived much less noisy than the four single frames along. See Figure 21.

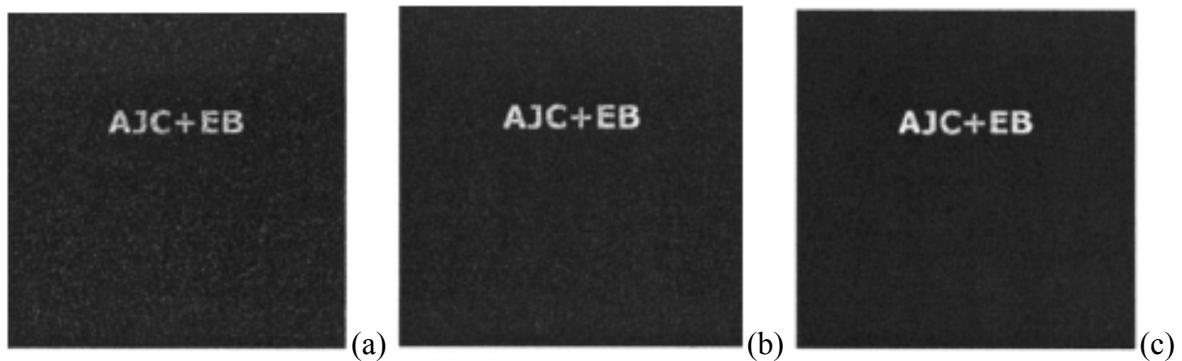


Figure 21: (a) $N = 1$, (b) $N = 4$, and (c) $N = 16$

Parallax 2D encoding CGH

This research¹⁶ argues that forming a 3D motion picture through 2D parallax CGH projection is much more efficient than the direct 3D CGH reconstruction. DFT is again implemented in hardware by the basic multiply-and-accumulate operations. However, the math it uses to describe the parallax view rendering, how to calculate its CGH pattern, and some primitive shape used to reconstruct the image are not very easily understood.

A light-light switch

Nonlinearity and frequency doubling of a BBO I (β - B_2O_4 , beta Barium Borate) crystal is used as the “recording” media¹⁷. However, it should be clarified right away that no signal can be stored in this material, and the “hologram” is only “instantaneous”. This “hologram” behaves more like a switch for turning on/off the interconnections, or for encoding/ decoding signals. There is no “twin” image, and it has the advantage of zero response time. Figure 22 shows such frequency doubling “hologram”.

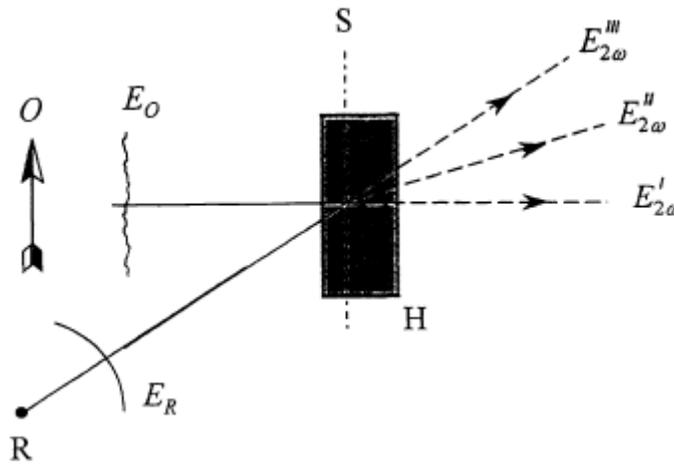


Figure 22: a second-harmonic generation crystal BBO is used as a hologram; E_O and E_R have frequencies ω , and the “reconstructed image” E^m has a frequency 2ω .

In order to understand the “reconstruction”, the interference effect must be modified to account for the up-conversion process. Following the steps in Eq 15, we see that the “reconstructed image” (the second term in Eq 15, denoted as $E^m_{2\omega}$ in Figure 22) is not the same as the original object E_O . It is the average of the original reference + object waves, and at frequencies 2ω and k_2 . A pictorial representation of Eq 15 is shown in Figure 23. In addition, the “reconstructed image” is modified in both the axial position and the transverse size. Consider a typical one-photon process reconstructed image $O_{1,\omega}$, an intermediate image $O_{2,2\omega}$, and the two-photon reconstructed image O_3 . The transverse size of the images $O_{1,\omega}$ and $O_{2,2\omega}$ are identical. However, because of the up-conversion process, the longitudinal size is doubled. The size and location of the intermediate image $O_{2,2\omega}$ is further modified by the spherical reference wave to produce the final image O_3 ; this can be understood as a focusing effect. One interesting by-product of this two-photon process is the auto-correlation of the image from $E^l_{2\omega}$ or the first term of Eq 15.

of the null corrector is to form an ideal reference wavefront, so the deviation from the test optics can be compared and measured by their interferogram. Eq16-20 shows how the two beams interfere with each other and their resulting (intensity) interferogram. Using the LCD to as an active programmable hologram, it becomes unnecessary to manufacture individual CGH's to match particular test surfaces.

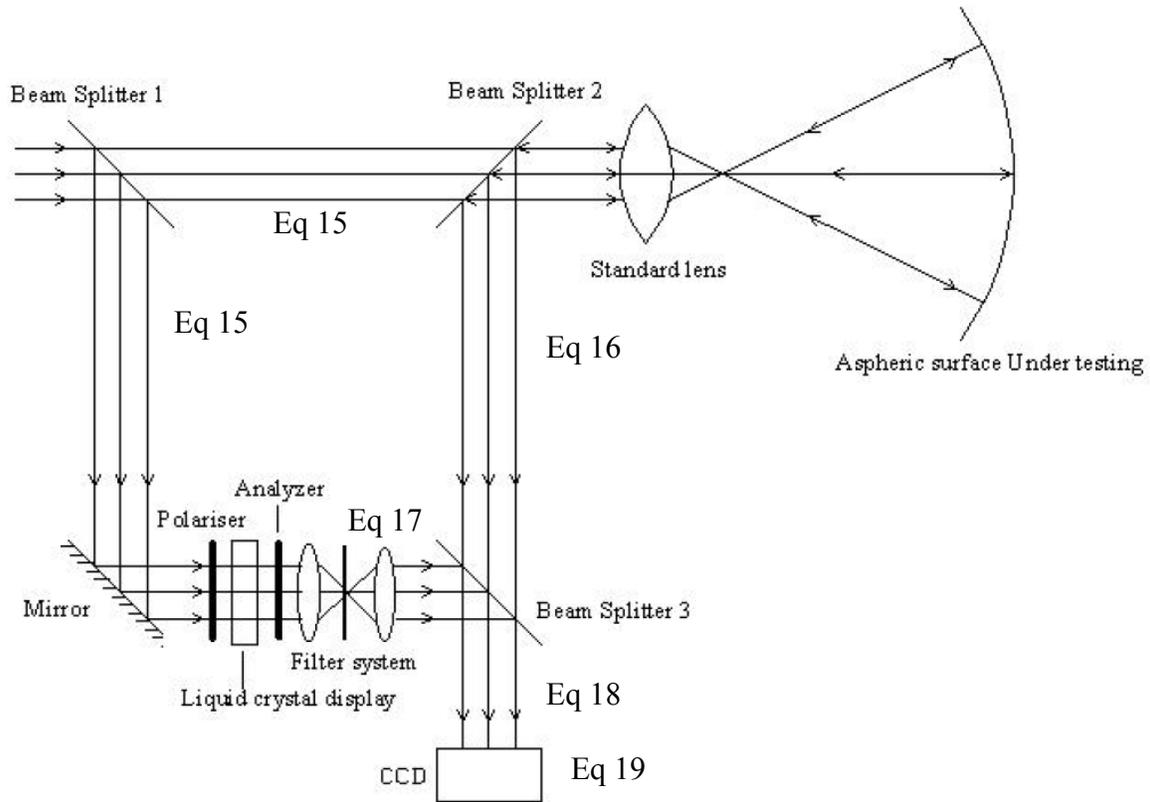


Figure 24: an interferometer setup utilizing an LCD to modulate the phase of the reference beam.

$$U(x, y) = u_0 \exp(ikz) \quad \text{Eq 16}$$

$$U_o(x, y) = u_0 \exp[ik(z_0 + o(x, y))] \quad \text{Eq 17}$$

$$U_r(x, y) = u_0 \exp[ik(z_0 + r(x, y))] \quad \text{Eq 18}$$

$$U'(x, y) = U_o(x, y) + U_r(x, y) \quad \text{Eq 19}$$

$$\begin{aligned} I(x, y) &= U'(x, y) \cdot U'^*(x, y) = (U_o + U_r)(U_o^* + U_r^*) \\ &= |U_o|^2 + |U_r|^2 + U_o U_r^* + U_r U_o^* \\ &= 2u_0^2 + 2u_0^2 \cos\{k[(z_0 - z_r) + (o(x, y) - r(x, y))]\} \end{aligned} \quad \text{Eq 20}$$

It must be pointed out that the intensity and phase of a LCD can not be controlled independently. However, with the help of the polarizer and analyzer oriented at some particular angle relative to the nematic angle of the LCD, there can be a range of phase modulation where the intensity variation is negligible. See Figure 25.

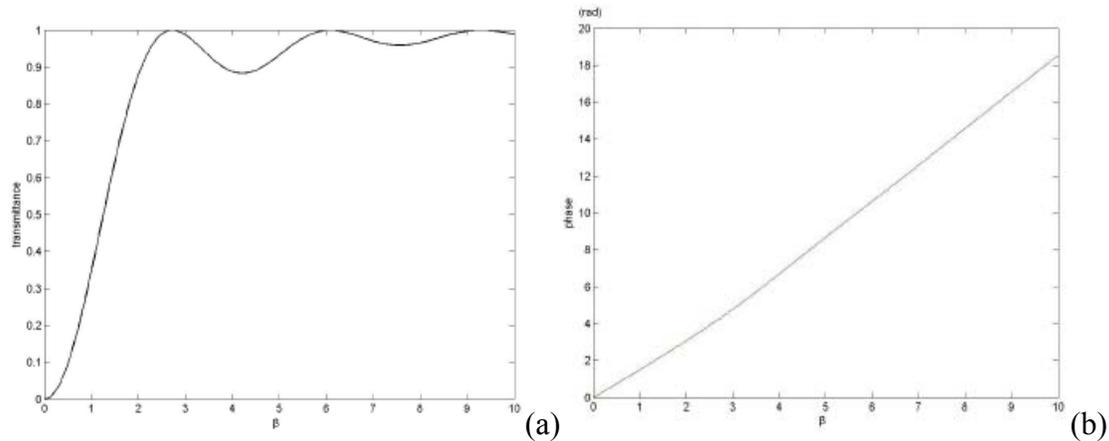


Figure 25: (a) intensity and (b) phase as a function of β (β is proportional to the difference between the ordinary and extraordinary refractive indices of the liquid crystal $\Delta n = n_o - n_e$, and it is controlled by the capacitance across the individual LCD pixel)

However, the LCD has the added functionality of performing phase-shifting. The reference wave is phase shifted by $0, \pi/2, \pi$, and $3\pi/2$, and the resulting four interferograms are to reconstruct the surface deviation of test optics. The LCD used is a Sony LCX023 1.3” display. It has 1024x768 pixels, with a pixel spacing of $26\mu\text{m}$. The thickness of the liquid crystal layer is $d = 5\mu\text{m}$, $\beta = \pi d(n_o - n_e)/\lambda$, and $\lambda = 0.6328\mu\text{m}$.

ENCODING IMAGES ON THE 0th ORDER OF A QUATERNRY CGH

There are two main goals in this section: to develop a mathematical description of the Quaternary Phase-only CGH, and to simulate/encode/decode digital image hiding using CGH in Matlab. The algorithms are shown, and two variations of the CGH are derived. Figure 26 and 27 are the flow chart and the images applied.

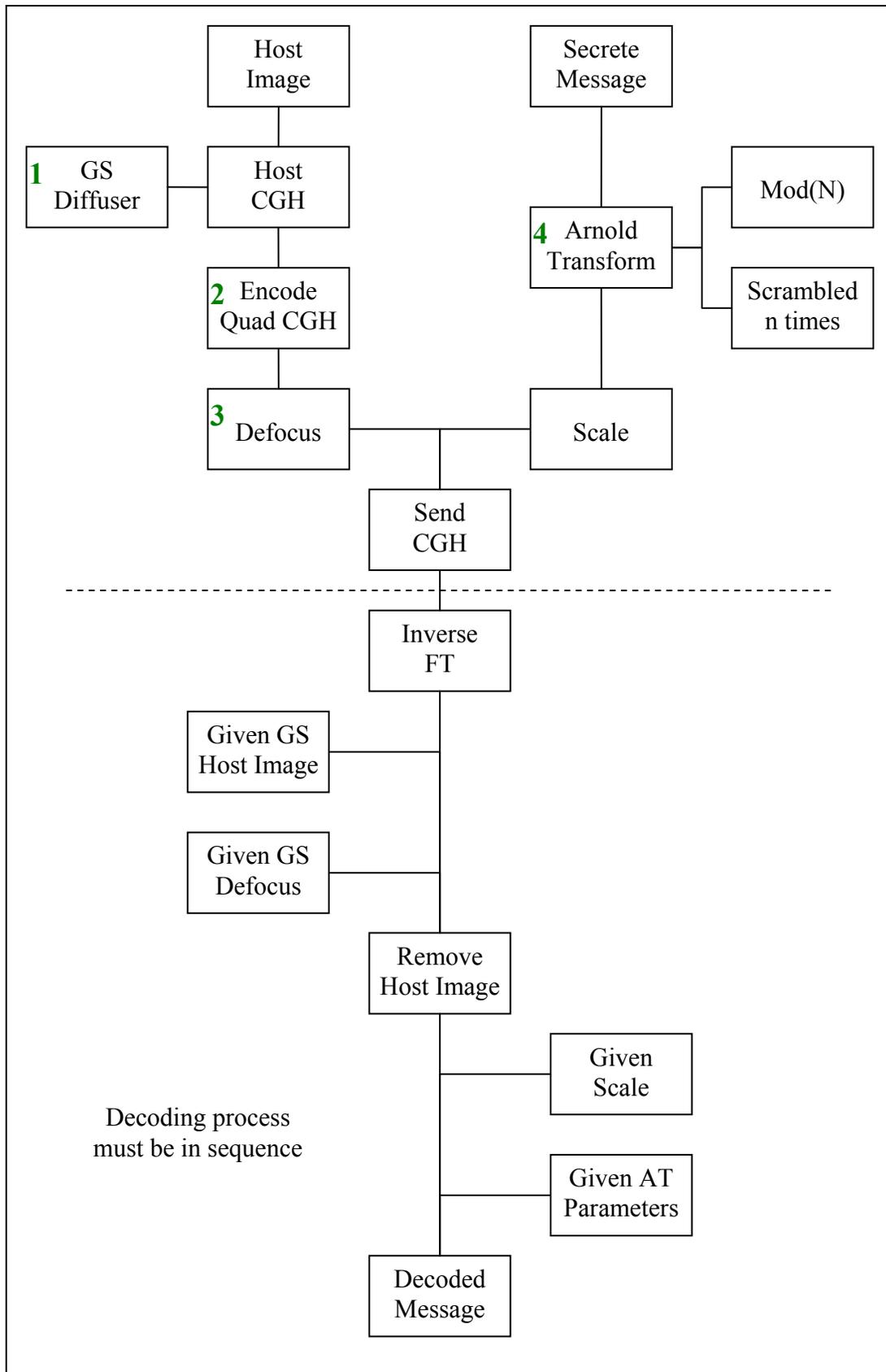


Figure 26: a flow chart of the coding and decoding a secret message inside a host image

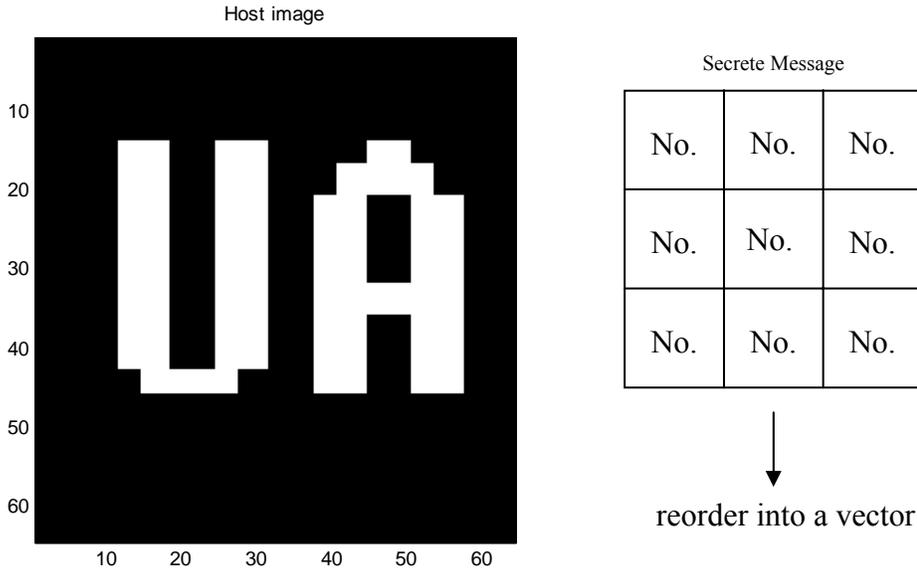


Figure 27: a host image of 64x64 pixel, and a secrete message of equal or smaller in size.

The following sections discuss the algorithms developed for the **labeled** boxes and their simulation results. Subroutines are put into functions.

BOX 1: Gerchberg-Saxton (GS) Diffuser

GS diffuser is used to more-uniformly redistribute the object in the Fourier domain, by attaching an “optimal random” phase to the object in the direct space. The iterative algorithm is shown below; the algorithm goes back-and-forth between the direct and the FT space, and each time keeping the phase and destroying the amplitude information. Typically, the “optimal random” phase is achieved after 5-10 iterations. Figure 28 compares two images with and without the GS diffuser. It is clear that the GS diffuser effectively spreads out the information in the FT domain, and “structures” are eliminated.

```
function [phase] = GSDiffuser(object)
phase = rand(size(object));
for n = 1:10;
    diffuser = exp(2*pi*i*phase);
    u = object .* diffuser;
    U = fftshift(fft2(fftshift(u)));
    U_PhaseOnly = U./abs(U);
    u = fftshift(ifft2(fftshift(U_PhaseOnly)));
    phase = angle(u);
end;
```

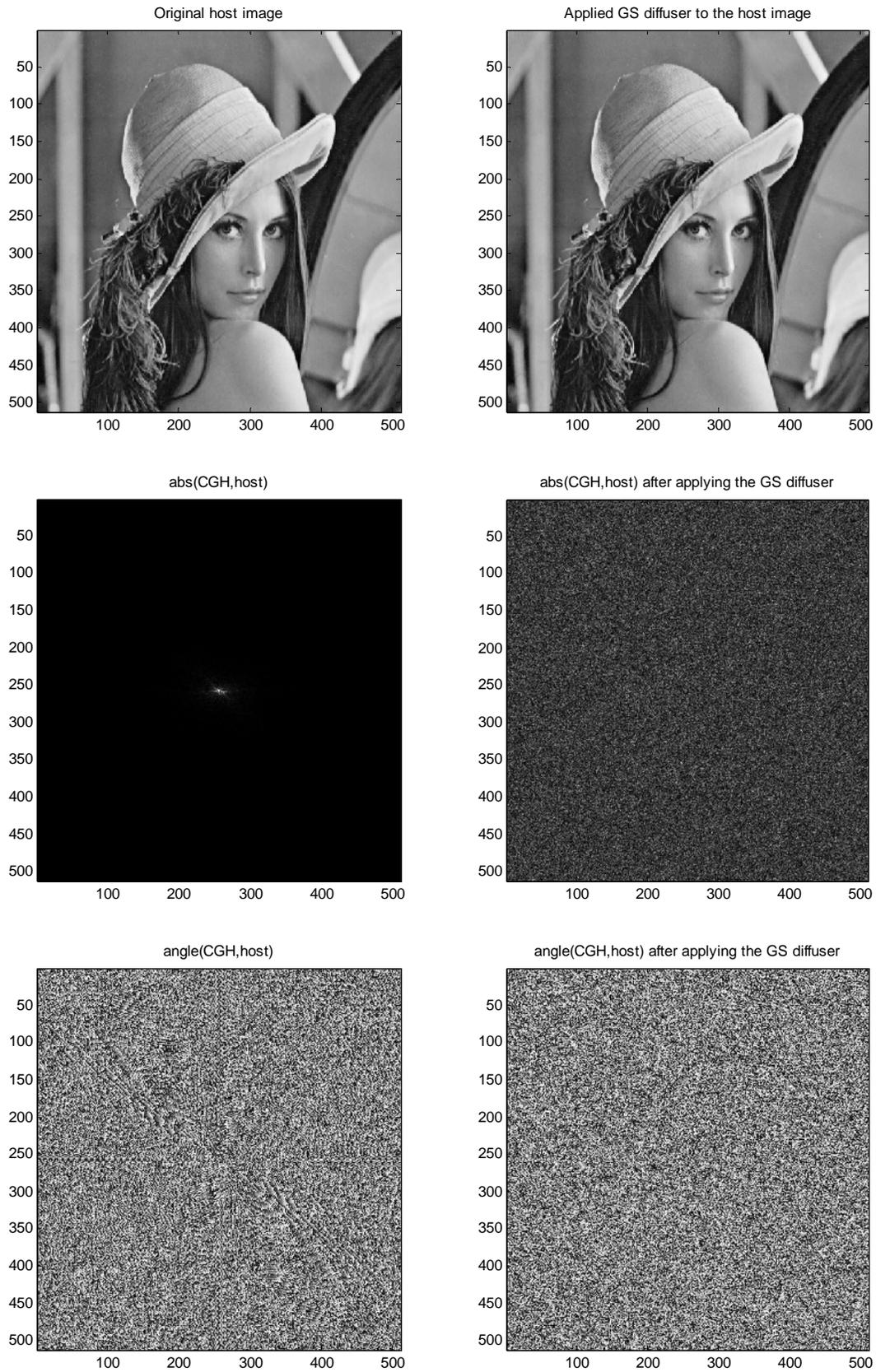


Figure 28: left column shows the FT(original image), and right column shows the FT(after applying a GS diffuser); “structures” are obvious on the two bottom left plots.

BOX 2: Quaternary Phase-Only CGH

Early in the Algorithm Developments section, the general approach to encoding an image on a carrier frequency was discussed. Brief discussion shows how to encode the FT of an object on a cosine wave, a square wave, and a particular type of Quaternary CGH (that eliminates the twin image). Here, a new type of Quaternary Phase-Only CGH is developed. Images can be encoded on the 0th order of this “four-step staircase” carrier. First, the staircase is decomposed into their Fourier Series coefficients, consisting of four square waves at four different heights. Two variations of the staircase geometry are explored, and their diffraction orders/strengths are compared.

Recall a simple square wave having a duty cycle “a”, in the coordinate “p”, and has a displacement measured from the origin “p₀”. See Eq 21. Quad_Type1 can be written as a summation of four square waves at different displacements “p₀’s” and heights “A’s”. See Eq 22 and Figure 29.

$$T_{sq}(a;x_0p;p_0;A) = A \cdot \sum_m a \cdot \text{sinc}(ma) \cdot e^{2\pi i m \xi_0 (p-p_0)} \quad \text{Eq 21}$$

for the moment, let

$$\begin{aligned}
 T_{\text{Quad_Type1}} = & T_{sq_1}(a_1;x_0p;p_1;A_1) & a_1 = 1/4, p_1 = 1/8, A_1 = 1/4, \\
 & + T_{sq_2}(a_2;x_0p;p_2;A_2) & a_2 = 1/4, p_2 = 3/8, A_2 = 2/4, \\
 & + T_{sq_3}(a_3;x_0p;p_3;A_3) & a_3 = 1/4, p_3 = 5/8, A_3 = 3/4, \\
 & + T_{sq_4}(a_4;x_0p;p_4;A_4) & a_4 = 1/4, p_4 = 7/8, A_4 = 4/4, \quad \text{Eq 22}
 \end{aligned}$$

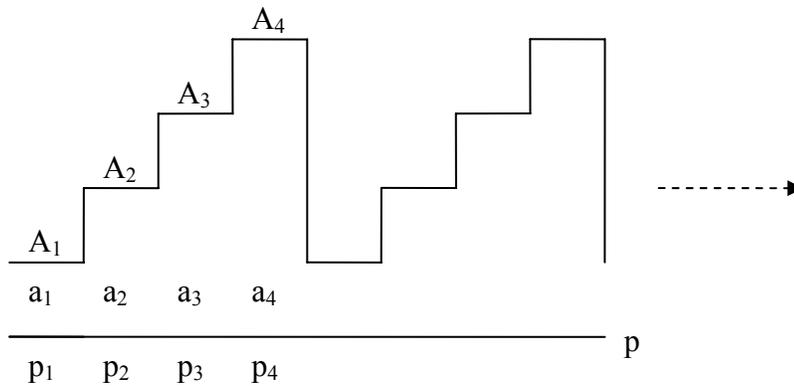


Figure 29: Quad_Type1 carrier

With a slight variation, Quad_Type2 can be written as a summation of four square waves with different duty cycles “a’s”. See Eq 23 and Figure 30. Since all the parameters of the square waves are defined, these staircases can be reconstructed using “for-loops” to first construct the individual square waves from their Fourier Series coefficients, and then adding up four different square waves to yield the final shapes.

$$\begin{aligned}
 T_{\text{Quad_Type1}} = & T_{\text{sq_1}}(a_1; x_0 p; p_1; A_1) && \text{for the moment, let} \\
 & + T_{\text{sq_2}}(a_2; x_0 p; p_2; A_2) && a_1 = 1/4, p_1 = 1/2, A_1 = 1/4, \\
 & + T_{\text{sq_3}}(a_3; x_0 p; p_3; A_3) && a_2 = 2/4, p_2 = 1/2, A_2 = 1/4, \\
 & + T_{\text{sq_4}}(a_4; x_0 p; p_4; A_4) && a_3 = 3/4, p_3 = 1/2, A_3 = 1/4, \\
 & && a_4 = 4/4, p_4 = 1/2, A_4 = 1/4, \quad \text{Eq 23}
 \end{aligned}$$

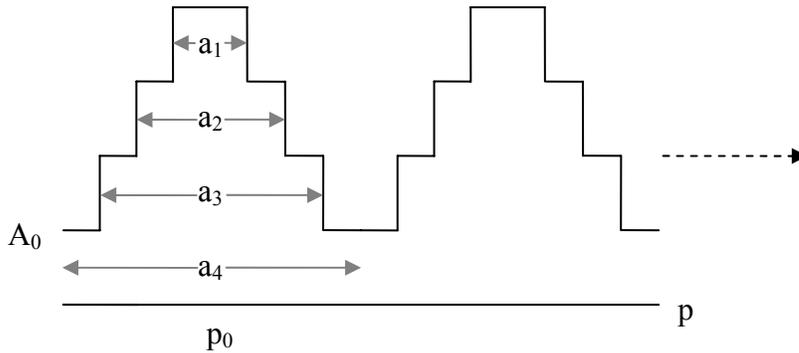


Figure 30: Quad_Type2 carrier

Figure 31a shows the synthesis of these two types of the quaternary CGH from their Fourier Series coefficients. 256 terms were used (see the function sq_wave below), and the Gibbs phenomenon is not too significant. Figure 31b shows the phase shifts, and 31c and 31d are plots of the Fourier Series coefficients for the Quad_Type1 and Quad_Type2. It can be seen that the Quad_Type1 has every 4mth order equal to zero, and Quad_Type2 had every 2mth order equal to zero. These two plots are not properly scaled relative to each other (because the strength of the 0th order for Quad_Type2 is expected to be higher than Quad_Type1). Symmetry can be used to explain the “missing orders” in the Quad_Type2 in comparison to the Quad_Type1.

```

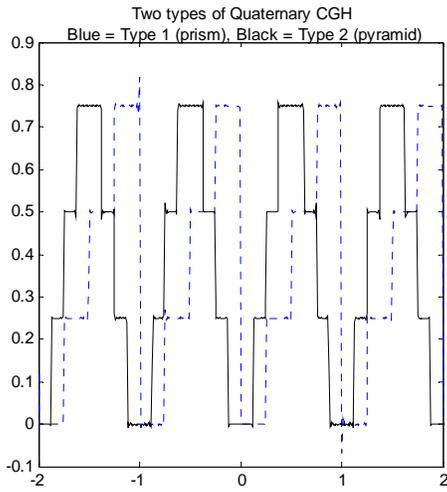
function [y] = sq_wave(a, del_p, m, p)
y = zeros(size(p));
for n = 1 : 2*m + 1;

```

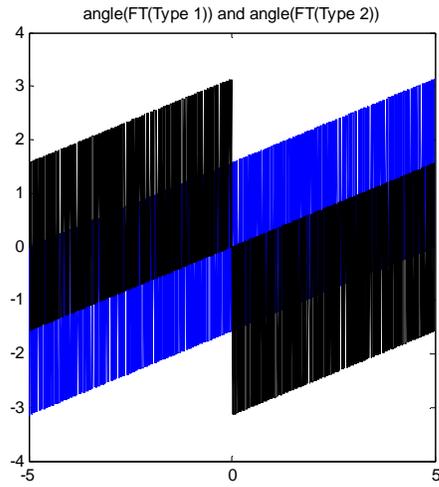
```

y_n = a * sinc((n-m-1)*a) * exp(2*pi*i*(n-m-1)*(p-del_p));
y = y + y_n;
end;

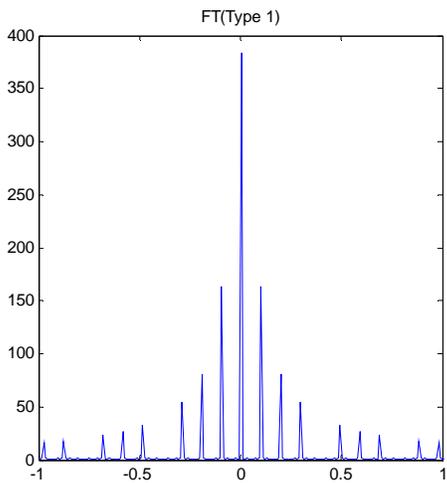
```



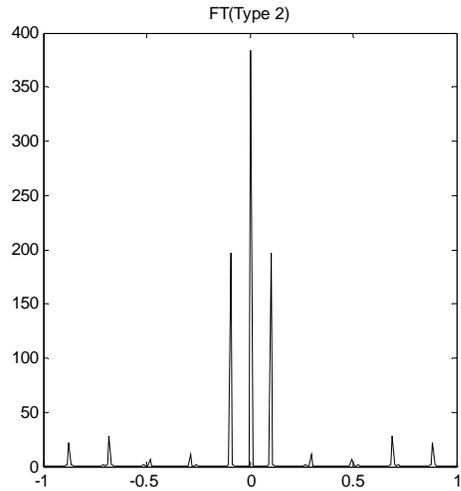
(a)



(b)



(c)



(d)

Figure 31: Matlab simulation of the “staircase waves”, constructed from the Fourier Series coefficients

Now that the type of the carrier wave is chosen, the encoding method can be determined. Upon examining the function “sq_wave”, it is realized that no modulation can be encoded in the 0th order of the square wave; both the sinc and the exponential functions become ones for $m = 0$. However, if A_{1-4} represents the steps in an exponential function $e^{2\pi i A}$, then one may encode an image in the 0th order of the “staircase wave” by varying the combined width/strength of the $e^{2\pi i A}$ in a period. Specifically, reassign:

$$A_1 = e^{2\pi i(0/4)} = +1$$

$$A_2 = e^{2\pi i(1/4)} = +i$$

$$A_3 = e^{2\pi i(2/4)} = -1$$

$$A_4 = e^{2\pi i(3/4)} = -i$$

Therefore, at the 0th order, the “staircase waves” become:

$$T_{\text{Quad_Type1}} = [+1 \cdot \text{rect}(a_1, p-p_1) \quad a_1 = a_2 = a_3 = a_4 = 1/4$$

$$+ i \cdot \text{rect}(a_2, p-p_2) \quad p_{1-4} \text{ varying}$$

$$- 1 \cdot \text{rect}(a_3, p-p_3)$$

$$- i \cdot \text{rect}(a_4, p-p_4)] \text{ convolve with comb}(p) \quad \text{Eq 24}$$

$$T_{\text{Quad_Type2}} = [+1 \cdot \text{rect}(a_1, p-p_1) \quad p_1 = p_2 = p_3 = p_4 = 1/2$$

$$+ i \cdot \text{rect}(a_2, p-p_2) \quad a_{1-4} \text{ varying}$$

$$- 1 \cdot \text{rect}(a_3, p-p_3)$$

$$- i \cdot \text{rect}(a_4, p-p_4)] \text{ convolve with comb}(p) \quad \text{Eq 25}$$

In order to assign a complex value to each period, one may locally vary both a_{1-4} and p_{1-4} in Eq 24, or just a_{1-4} in Eq 25. A complex value may be represented by its real and imaginary parts (see Figure 32 for the physical interpretation of duty cycle “a”):

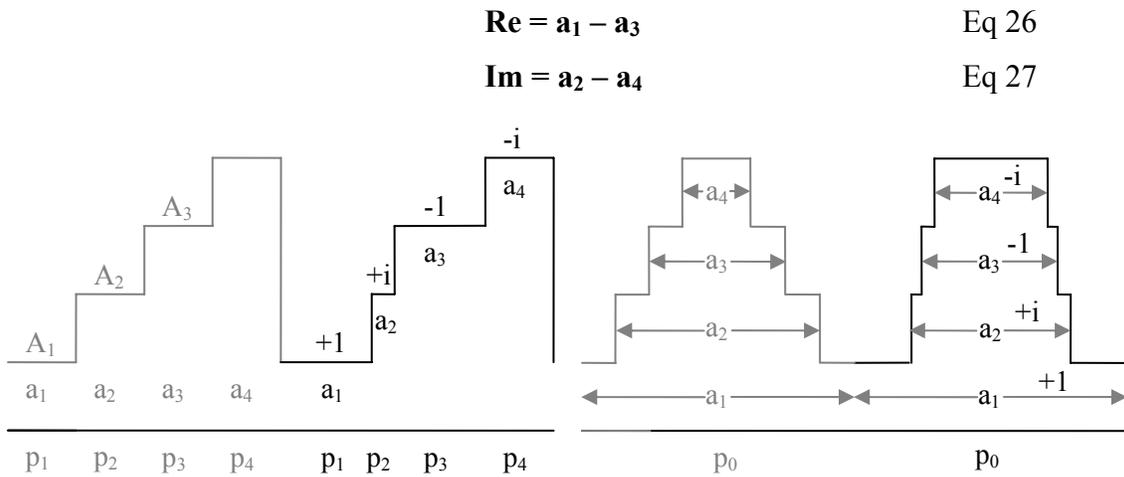


Figure 32: four square rect’s inside a unit cell, each discrete step represents a phase $\{0, \pi/2, \pi, 3\pi/2\}$, each duty cycle “a” assigns a strength to the phase, and Eq 26 and 27 represent the complex amplitude of the unit cell.

One of the conditions placed on a_{1-4} is that the duty cycle is always non negative:

$$\mathbf{a_{1-4} \geq 0} \quad \text{Eq 28}$$

Last, we place a constraint on the length of the period:

$$\mathbf{\Sigma a_i = 1} \quad \text{Eq 29}$$

One must realize that Eq 29 is slightly over constraining a particular degree of freedom; we may allow $\Sigma a_i = 1 \pm \epsilon$, providing that there is no gap between any two adjacent periods anywhere on the CGH plane. Since there are four unknowns and four statements, there is at least one possible solution; Eq 28 is an inequality, so it gives rise to more than one possible set of solutions. The complex amplitude can also be expressed as:

$$\mathbf{Amp = (Re^2 + Im^2)^{1/2}} \quad \text{Eq 30}$$

$$\mathbf{Phase = \tan^{-1}(Im/Re)} \quad \text{Eq 31}$$

All of the derivations and modeling so far is leading up to answering this one question: how can one encode both amplitude and phase information on the $m = 0$ order of this Quaternary Phase-Only CGH. The recipe is:

1. Take an object u ;
2. $U = FT(u)$;
3. Assign the real part of U , $\text{real}(U)$, to $\text{Re} = a_1 - a_3$;
4. Assign the imaginary part of U , $\text{imag}(U)$, to $\text{Im} = a_2 - a_4$;
5. Adjust the length a_{1-4} to fill each period.

Line 5 is still ambiguous at this point; how are the length a_{1-4} adjusted to fill exactly one period, and what is the maximum Re/Im or Amp/Phase “encode-able”? To answer this question, we need to start with a unit circle on the complex plane. See Figure 33. If we assign the entire period to a_1 , then $a_2 = a_3 = a_4 = 0$, the maximum amplitude can be encoded is 1 (on the unit circle). However, everywhere other than $\{0, \pi/2, \pi, 3\pi/2\}$, the $\text{length}(\text{Re}) + \text{length}(\text{Im}) > 1$; sum of two sides of a triangle is always greater than the

third side. This violates Eq 29, because each period has exactly a length of 1. It can be shown that a reduced circle having a radius equal to $1/\sqrt{2}$ never violates Eq 29 (at 45° , $\text{length}(\text{Re}) + \text{length}(\text{Im}) = 1$; everywhere else, $\text{length}(\text{Re}) + \text{length}(\text{Im}) < 1$). Therefore, in order to satisfy Line 5, $\text{Amp}(U)$ must first be scaled to $(0, 1/\sqrt{2})$.

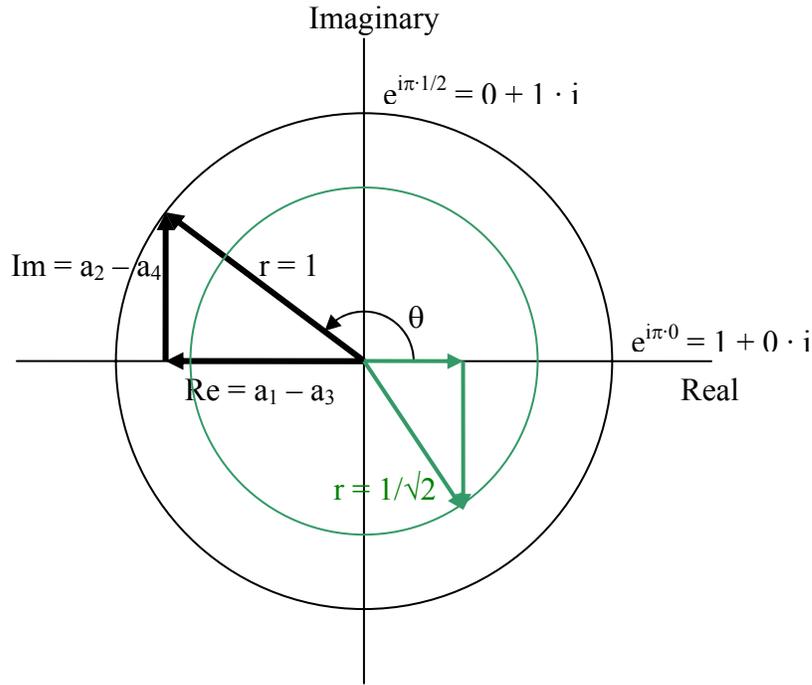


Figure 33: the maximum amplitude that can be encoded in a Quad_Type1 or Quad_Type2 CGH having a period length of 1 is $1/\sqrt{2}$.

What happens if a vector is inside the $r = 1/\sqrt{2}$ circle (having a radius $r < 1/\sqrt{2}$), or on the $r = 1/\sqrt{2}$ circle but is not at $\{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$? At these points, $\text{length}(\text{Re}) + \text{length}(\text{Im}) < 1$ and it means the period of the “staircase” is only partially filled. As mentioned before, we require continuity between any two adjacent periods (no gap allowed). Therefore, at these points, we must add a filler to fill in the “gap”. For simplicity, we define a $\text{Filler} = [1 - \text{length}(\text{Re}) - \text{length}(\text{Im})]/4$, and add the filler to each of the a_{1-4} for a given period.

The Matlab code for assembling such Quaternary CGH is shown below in the next two pages. For each pixel of the original U , it is further divided into 32×32 resolution cells. On average, each a_{1-4} occupies 32×8 resolution cells.

```

function [CGH_quad1 Check Error]=Quad1(U);

%scale U to the max amplitude of 0.707
Max = max(max(abs(U)));
Scale = 1/sqrt(2);
U = U / Max; %normalize U to (0,1)
U = U * Scale; %scale U to (0,0.707)

%find the "empty space" in each pixel
Re = real(U);
Im = imag(U);
Empty = 1 - (abs(Re) + abs(Im));
Filler = Empty/4; %always >= 0
Check = Filler < 0; %pixels having round off errors
Error = sum(sum(Check.*Filler)); %always a negative number, ideally = 0

%define some CGH parameters
[m n] = size(U);
ResCell = 32; %divide each U pixel into 32x32 resolution cells -> quantization error
p = linspace(0,1,ResCell);
CGH = ones(ResCell*m,ResCell*n);

a1 = Filler + (Re .* (Re>0)) + 1/ResCell; %add 1/Rescell,
a2 = Filler + (Im .* (Im>0)) + 1/ResCell; %so rect is 1 at both +1/2 and at -1/2
a3 = Filler - (Re .* (Re<0)) + 1/ResCell; %Re .* (Re<0) always -, but a3 always +
a4 = Filler - (Im .* (Im<0)) + 1/ResCell; %Im .* (Im<0) always -, but a4 always +

del_a1 = a1/2;
del_a2 = a1 + a2/2;
del_a3 = a1 + a2 + a3/2;
del_a4 = a1 + a2 + a3 + a4/2;

for i = 1 : m; %row index
    for j = 1 : n; %column index
        rect_a1 = rect(p,del_a1(i,j),a1(i,j),1/4);
        rect_a2 = rect(p,del_a2(i,j),a2(i,j),2/4);
        rect_a3 = rect(p,del_a3(i,j),a3(i,j),3/4);
        rect_a4 = rect(p,del_a4(i,j),a4(i,j),4/4);
        %everthing that is not a1, a2, or a3 is a4

        u = ResCell * (i-1) + 1;
        v = ResCell * i;
        x = ResCell * (j-1) + 1;
        y = ResCell * j;
        CGH(u:v,x:y) = ones(ResCell,1) * (rect_a1 + rect_a2 + rect_a3 +
rect_a4 - 1/4);
    end;
end;

CGH_quad1 = CGH;

```

```

function [CGH_quad2 Check Error]=Quad2(U);

%scale U to the max amplitude of 0.707
Max = max(max(abs(U)));
Scale = 1/sqrt(2);
U = U / Max; %normalize U to (0,1)
U = U * Scale; %scale U to (0,0.707)

%find the "empty space" in each pixel
Re = real(U);
Im = imag(U);
Empty = 1 - (abs(Re) + abs(Im));
Filler = Empty/4; %always >= 0
Check = Filler < 0; %pixels having round off errors
Error = sum(sum(Check.*Filler)); %always a negative number, ideally = 0

%define some CGH parameters
[m n] = size(U);
ResCell = 32; %divide each U pixel into 32x32 resolution cells -> quantization error
p = linspace(0,1,ResCell);
CGH = ones(ResCell*m,ResCell*n);

a1 = Filler + (Re .* (Re>0)) + 1/ResCell; %add 1/Rescell,
a2 = Filler + (Im .* (Im>0)) + a1; %so rect is 1 at both +1/2 and at -1/2
a3 = Filler - (Re .* (Re<0)) + a2; %Re .* (Re<0) is always -, but a3 is always +
a4 = Filler - (Im .* (Im<0)) + a3; %Im .* (Im<0) is always -, but a4 is always +

for i = 1 : m; %row index
    for j = 1 : n; %column index
        rect_a1 = rect(p,1/2,a1(i,j),1/4);
        rect_a2 = rect(p,1/2,a2(i,j),1/4);
        rect_a3 = rect(p,1/2,a3(i,j),1/4);
        rect_a4 = rect(p,1/2,a4(i,j),1/4);
        %everthing that is not a1, a2, or a3 is a4

        u = ResCell * (i-1) + 1;
        v = ResCell * i;
        x = ResCell * (j-1) + 1;
        y = ResCell * j;
        CGH(u:v,x:y) = ones(ResCell,1) * (rect_a1 + rect_a2 + rect_a3 +
rect_a4 - 1/4);
    end;
end;

CGH_quad2 = CGH;

```

Figure 34 shows the results and comparisons of the above two CGH coding methods. In Quad_Type1 CGH, the boundaries between the periods are very easily identified (by the black and white strips). The images from the non-zero orders seem to be “accumulating” only on one side, but this is not the case with the Quad_Type2 CGH. In addition, the reconstructed image from the Quad_Type 2 CGH was originally 180° out of rotation; the reconstruction was reoriented for the purpose of comparisons.

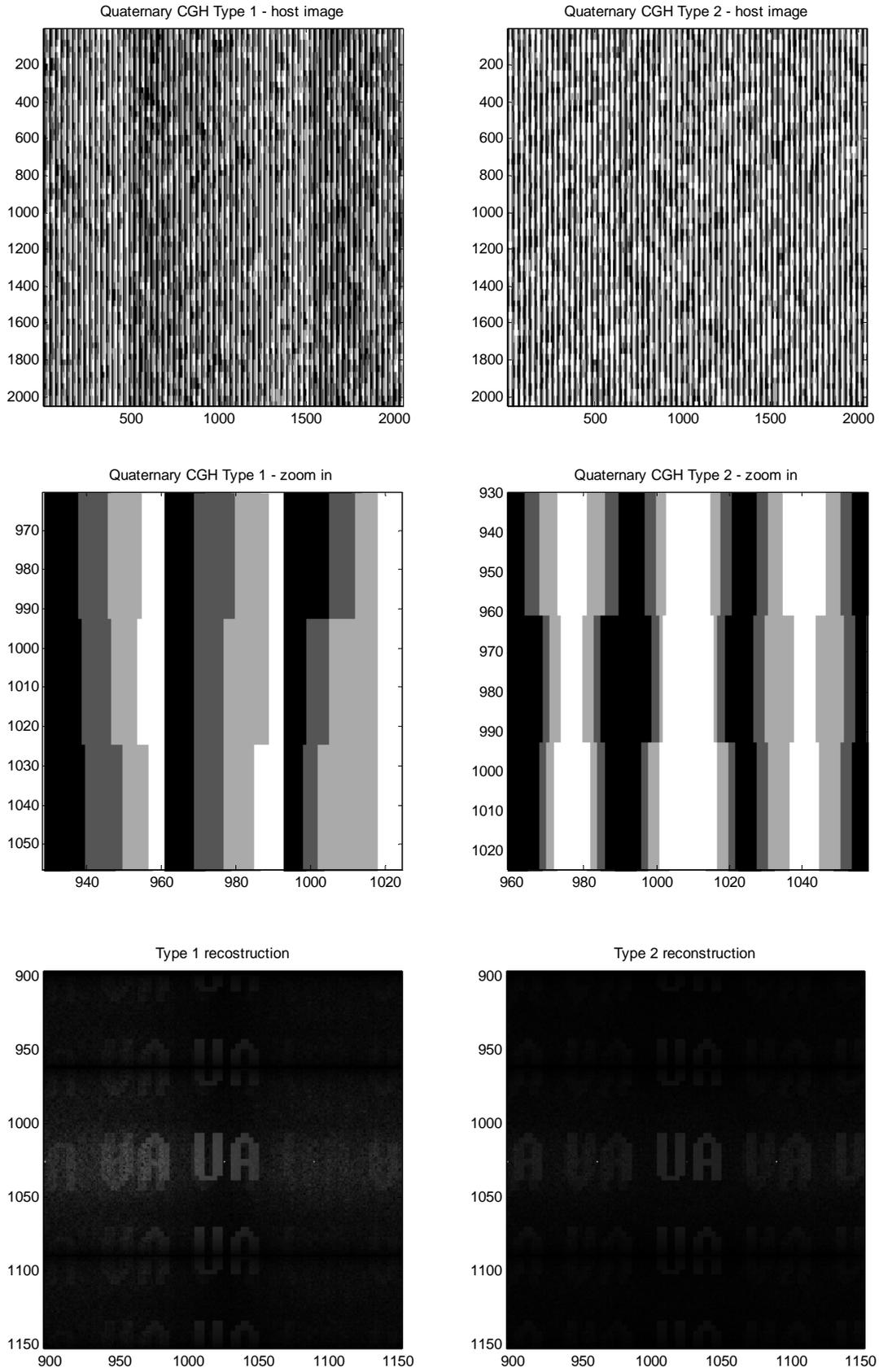


Figure 34: simulation of the Quad_Type1 and Quad_Type2 CGH and their reconstructed images; a few “sinc” diffractive orders can be seen in the background.

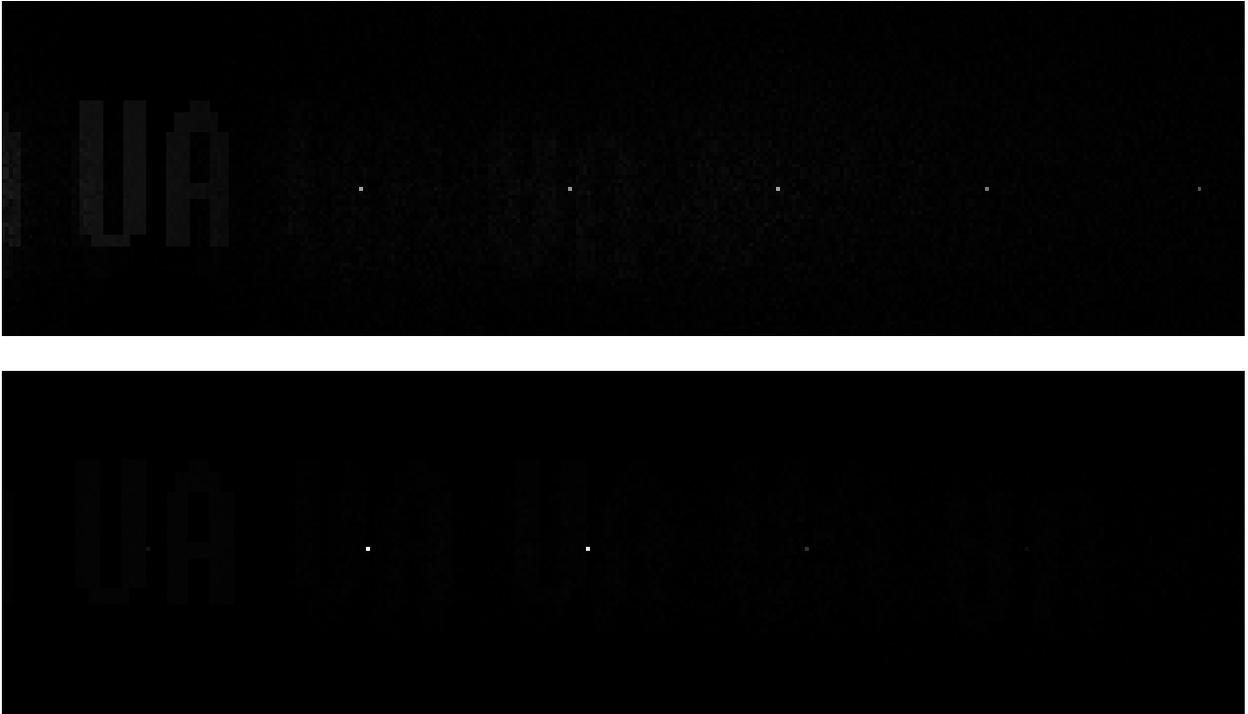


Figure 35: diffractive order from the “sinc” function can be seen in the background of the reconstructed image; bottom picture is from the Quad_Type2 CGH, which exhibits fewer visible orders than the Quad_Type1 CGH.

Several questions remain to be answered:

1. Are there logical statements one can write to replace this look-up table method, like the square wave type statement, and make the code more efficient?
2. How does the relaxation of Eq 29 help with 1).
3. Does the relaxation of Eq 29 help with any new type of encoding schemes?
4. Can we write the CGH as the summation of a cosine transform + a sine transform instead of the Fourier transform? Does this help with canceling out the twin image as in the $U = \cos(\Phi) + i \cdot \sin(\Phi)$ type CGH?
5. Can we place some symmetry condition on the filler to reduce/eliminate the strength of the images at the non-zero orders?

BOX 3: Defocus

Defocus is applied to the CGH to make the reconstruction unrecognizable, if one assumes a 4F system. In order to avoid aliasing, a “slow” quadratic phase factor is used. See Figure 37. Its effect on the reconstructed image is animated at different depth of focus.

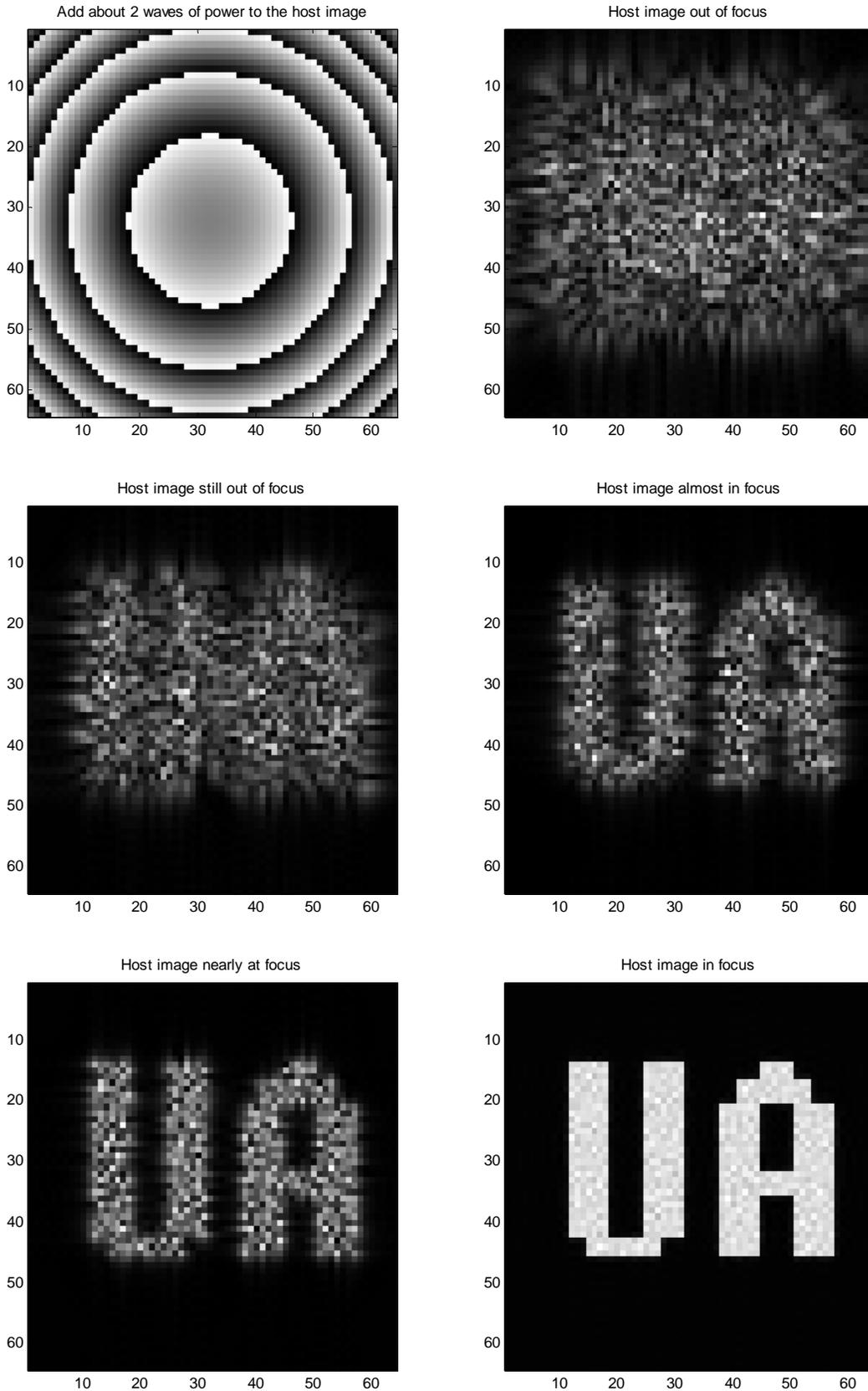


Figure 37: adding power = $\exp[i\pi(x^2+y^2)/(FL*\lambda)]$ to the CGH, and the reconstructed images at different depth of focus away from the original 4F system

BOX 4: Arnold Transform

The basic concept of Arnold Transform is mentioned previously. Here, the implementation is shown in more details. There are two important constructs in this image processing tool - the Arnold Transform matrix and the operation mod(N). The Arnold Transform matrix T is defined as:

```
for i = 2:t;
    for j = 2:t;
        AT(:,1) = 1;
        AT(1,:) = 1;
        AT(i,j) = AT(i-1,j-1) + AT(i-1,j);
    end;
end;
```

or

$$\mathbf{T}_{n \times n} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 2 & 2 & \cdots & 2 & 2 & 2 \\ 1 & 3 & 4 & \cdots & 4 & 4 & 4 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 1 & a_1 & a_2 & \cdots & a_{n-3} & a_{n-2} & a_{n-1} \\ 1 & 1+a_1 & a_1+a_2 & \cdots & \cdots & a_{n-3}+a_{n-2} & a_{n-2}+a_{n-1} \end{bmatrix}$$

Given a column vector $[x_i]$, Arnold Transform performs this operation indefinitely,

$$[x_{i,m+1}] = \{T_{n \times n} \cdot [x_{i,m}]\} \bmod(N)$$

After a certain number of operations, there will be a period P where $[x_p] = [x_1]$. P depends on the initial values of $[x_i]$ as well as N. This period is usually very big, hence it is impractical to try all possible P to decode a message. The inverse process does not work on decoding the message either, because there is no such thing as “inverse mod”. The “keys” for restoring the column vector $[x_{i,m}]$ to the initial values $[x_{i,1}]$ is if both P and N are given.

DECODING THE MESSAGE

Now we have all of the necessary tools to perform Figure 26 (GS diffuser, Quaternary Phase-Only CGH, Defocus, and Arnold Transform), we may begin encoding and decoding the secrete image.

A host image “UA” having 64x64 pixels is used. See Figure 27(left). GS diffuser is applied to the host image, and this host image attached with an “optimized” random phase is Fourier transformed. The amplitude and phase of the image in the FT domain are encoded using the Quaternary Phase-Only CGH technique. A small amount of defocus is added. This finishes the preparation of the host image.

A 3x3 secrete message such as one’s social security number is used as an example here. See Figure 27(right). This 3x3 matrix is Arnold Transformed and re-ordered to fill in the first column of an initially Zero(64,64) matrix. The scale factor is chosen and applied such that the “strength” of the secrete message is much smaller than the host message. This finishes the preparation of the secrete image; note that this secrete message is not FT and remains in the direct space.

The two matrixes are added together. The new product can be sent electronically or made into hardware and delivered.

In order to extract the secrete message, the end user must have five “keys”:

- 1). phase of the host message,
- 2). amount of the defocus,
in order to remove the host image,
- 3). scale of the secrete image,
- 4-5). Mod(N) and n scrambles,
in order ot unscramble the secrete image.

To decode the message, follow bottom half of Figure 26. Here is an example if any one of the “keys” is missing. See Table 1. In this example, the defocus is vs is-not removed from the CGH, and the perceived message (before decoding by Arnold Transform) look completely uncorrelated.

Defocus removed from the host image			Defocus not removed from the host image		
0.16	0	0	0.8331	2.9209	6.9048
0.23	0	0	0.8201	4.3899	8.3462
0.08	0	0	2.2457	6.3837	11.0386
0.17	0	0	5.5014	9.9813	14.7945
0.20	0	0	10.4138	15.1453	20.366
0.24	0	0	18.5343	22.3049	27.2953
0.16	0	0	27.8424	30.3077	34.1499
0.09	0	0	35.7816	35.5247	38.8826
0.04	0	0	34.4049	33.7792	39.7025
0	0	0	17.1673	24.9891	37.861
0	0	0	11.1289	16.9649	24.5205
0	0	0	20.0804	10.7161	25.6233
0	0	0	36.7678	69.2997	102.1498
0	0	0	91.4672	110.7673	131.025

Table 1: The first 3 columns and 14 rows of the 64x64 matrix. Both sides show the sought-after-secrete-message matrix, with and without a particular “key” missing, after an attempt to remove the host image (but prior to the de-scaling and completing one period of Arnold Transform). Right side shows a message with a “key” missing, and the message looks random. Left side shows a message with all the right “keys” so far, but it still needs to be further decoded by de-scaling and completing the Arnold Transform.

CONCLUSION

CGH is a subject that reaches different application areas. 3D optical display, volumetric data storage, optical metrology, and optical information/signal processing are just a few. A brief survey here examined the fundamentals of different “odd” CGH hardware, software, and applications.

The last part of the report proposed a method of encoding a message using a Quaternary Phase-Only CGH and Fourier Transform. Because of the vast number of degrees of freedom in this encoding/decoding process, this message can be safe-guard during transmission; and without the “keys”, the message appears to be random.

REFERENCES:

1. "Creation of diffractive optical elements by one step E-beam lithography for optoelectronics and x-ray lithography", Tallinn, Estonia, A A Aristov, S V Dubonos, R Ya Dyachenko, B N Gaifullin, V N Matveev, A A Svintsov, S I Zaitsev, Baltic Electronics Conference, Oct 7-11 1996, p483-486.
2. "Three dimensional design in electron beam lithography", V V Aristov, S V Dubonos, R Ya Dyachenko, B N Gaifullin, V N Matveev, H Raith, A A Svintsov, S I Zaitsev, J. Vac. Sci. Technol. B, 13(6), Nov/Dec 1995, p2526-2528.
3. "Proximity corrected 3D structure", S V Dubonos, B N Gaifullin, H F Raith, A A Svintsov, S I Zaitsev, Microelectronics Engineering, 27(1995), p195-198.
4. "Multilevel diffractive optical element manufactured by excimer laser ablation and half tone masks", F Quentel, J Fieret, A S Holmes, S Paineau, Proc. SPIE, 4274, p420 (2001).
5. "Continuous tone gray scale photomasks based on photosensitive spin-on-glass technology for deep UV lithography applications", E A Mendoza, F A Sigoli, H Paulus, L Q Giang, M Seifouri, E Lam, L Kempen, Proc. SPIE, 5256, p889 (2003).
6. "Synthetic phase holograms written by laser lithography", S Haselbeck, M Heissmeier, W P Hofmann, U W Krackhardt, B Manzke, P Savander, M Schrader, J Schwider, M Sperl, N Streibl, Proc. SPIE, 1718, p117 (1993).
7. "Fabrication of diffractive optical elements by direct laser writing with circular scanning", V P Koronkevich, V P Kiryanov, V P Korol'kov, A G Poleshchuk, V V Cherkashin, E G Churin, A A Kharissov, Proc. SPIE, 2363, p290 (1995).
8. "Two-photon four-level hologram Recording in poly-(alkyl- α -cyanoacrylates)", J Pinsl, F W Deeg, C Brauchle, Appl Phys B, 40, 77-84 (1986).
9. "Computer generated binary Fourier holograms", P Pombo, A Arrifano, J L Pinto, Proc. SPIE, 6252, 62521N (2006).
10. "Hough transform computer generated holograms: new output format", N Carender, D Casasent, F Coetzee, D Yu, Proc. SPIE, 1555, p182 (1991).
11. "Accuracy of CGH encoding schemes for optical data processing", D Casasent, F Coetzee, S Natarajan, T Xu, D Yu, H K Liu, Proc. SPIE, 1555, p23 (1991).
12. "All-digital holography and application in digital image hiding", Y Zhao, J Zhong, Proc. SPIE, 5636, p459 (2005).

13. "Method of digital hologram synthesis for hidden image optical coding in holographic security devices", R Lymarenko, O Budnyk, Proc. SPIE, 6341, 63411Y (2006).
14. "Combined optical/digital security devices", V I Girnyk, I V Tverdokhlebov, A A Ivanovsky, Proc. SPIE, 3973, p322 (2000).
15. "Viewing angle enhancement for two- and three-dimensional holographic displays with random superresolution phase masks" E Buckley, A Cable, N Lawrence, T Wilkinson, Applied Optics, 45(28), p7334 (2006).
16. "Reconfigurable image projection holograms", W Plesniak, M Halle, V M Bove, J Barabas, R Pappu, Optical Engineering, 45(11), 115801 (2006)
17. "Second harmonic generated hologram", Y N Denisyuk, A Andreoni, M Potenza, Proc SPIE, 3956, p119 (2000)
18. "Research on encode technology for aspherical surface measurement based on real time hologram", H Wang, Z Wang, H Zhao, A Tian, Proc. SPIE, 6279, 62797J (2007).