optica

3D computer-generated holography by non-convex optimization

JINGZHAO ZHANG,[1,†] NICOLAS PÉGARD,[1,2,†] JINGSHAN ZHONG,[1] HILLEL ADESNIK,[2,3] AND LAURA WALLER[1,*]

[1]Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, California 94720, USA
[2]Department of Molecular & Cell Biology, University of California, Berkeley, California 94720, USA
[3]Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA
*Corresponding author: waller@berkeley.edu

3D computer-generated holography uses a digital phase mask to shape the wavefront of a laser beam into a user-specified 3D intensity pattern. Algorithms take the target 3D intensity as input and compute the hologram that generates it. However, arbitrary patterns are generally infeasible, so solutions are approximate and often sub-optimal. Here, we propose a new non-convex optimization algorithm that computes holograms by minimizing a custom cost function that is tailored to particular applications (e.g., lithography, neural photostimulation) or leverages additional information like sample shape and nonlinearity. Our method is robust and accurate, and it out-performs existing algorithms.   © 2017 Optical Society of America

1. INTRODUCTION

Computer-generated holography (CGH) can create custom 3D intensity patterns by modulating the wavefront of a coherent light beam. This ability to control light volumetrically is critical to applications such as 3D displays [1], optical trapping [2], and neural photostimulation [3–6]. Typically, 3D CGH is implemented by encoding the phase of a coherent wavefront in the Fourier space (pupil plane) of an imaging system with a spatial light modulator (SLM). The SLM modulates the 2D Fourier phase of the wavefront, allowing one to control the 3D intensity distribution near the image plane.

Computing the 2D Fourier phase mask that generates a specific 3D intensity pattern is usually a non-trivial and ill-posed problem, for several reasons. First, most 3D intensity patterns are not physically feasible. The wave-field must propagate through free space while satisfying Maxwell's equations. Intensity must be conserved at each depth plane, and resolution is limited to the coherent diffraction limit set by the numerical aperture (NA). As a result, arbitrary 3D intensity patterns cannot be created exactly. Another limitation is the finite degrees-of-freedom (DOF) for patterning. Holograms must be phase-only, and the SLM dynamic range and pixel count impose further limits. Consider, for example, the DOF mismatch of our setup: we have $\sim 10^9$ resolvable voxels, but only $\sim 10^6$ SLM pixels. Given all these considerations, 3D CGH algorithms must search for the best *approximate* solution to the problem, within the constraints imposed by physics and hardware. The resulting optimization task is non-convex with non-unique solutions in a high-dimensional space (millions of pixels) that cannot be explored exhaustively.

Most existing algorithms can be classified into two categories. The first, *superposition algorithms* [7–9], decompose the desired 3D intensity into discrete 2D planes and digitally back-propagate each to the SLM plane. Each plane is considered independently and the solution is the average of the phases of the resulting complex fields. This class of algorithms is computationally efficient and always yields a solution. However, interference between planes is not accounted for, so the results are sub-optimal, and quality degrades quickly with more than 2–3 depth levels.

The second category of 3D CGH algorithms is *iterative projection algorithms* [10]. These methods iteratively update the solution by projecting it onto a subspace computed from the set of constraints. Gerchberg and Saxton showed that alternating projections onto the intensity constraints at the hologram plane and one focal plane will not diverge (though it is not guaranteed to converge) [11]. Similar concepts apply in 3D [12,13], and recent variations use multiple constraints in each iteration [2,14,15]. This class of algorithms is easy to implement and gives good results for single-plane intensity patterns. However, the strong enforcement of constraints may cause the algorithm to diverge when attempting to generate physically infeasible patterns. Introducing hyper-parameters can improve results [16], but it is difficult to tune without an explicitly defined cost function.

Here, we present a new algorithm, termed Non-Convex Optimization for VOlumetric CGH (NOVO-CGH) that combines a customizable cost function with an optimization strategy inspired by prior work in phase retrieval [17]. Phase retrieval from through-focus intensity measurements represents a similar mathematical problem to 3D CGH, since both take a 3D intensity

distribution as input and solve for the 2D phase that generates it. However, there is a major difference in that the 3D intensity patterns in the phase retrieval case are always physically feasible, whereas they are generally not in the 3D CGH case.

For 3D CGH, where there is no exact solution and many approximate solutions, results depend heavily on the cost function used. Our NOVO-CGH algorithm enables a search for optimal holograms with custom cost functions that can be chosen by the user to best respond to a specific application. For example, by specifying dark and bright regions with intensity thresholds, as well as regions where target intensity is unspecified, we give the algorithm flexibility to find a suitable solution. While our algorithm cannot expand the set of physically feasible intensity patterns, we will demonstrate that, within this same set, it can find better solutions than other algorithms.
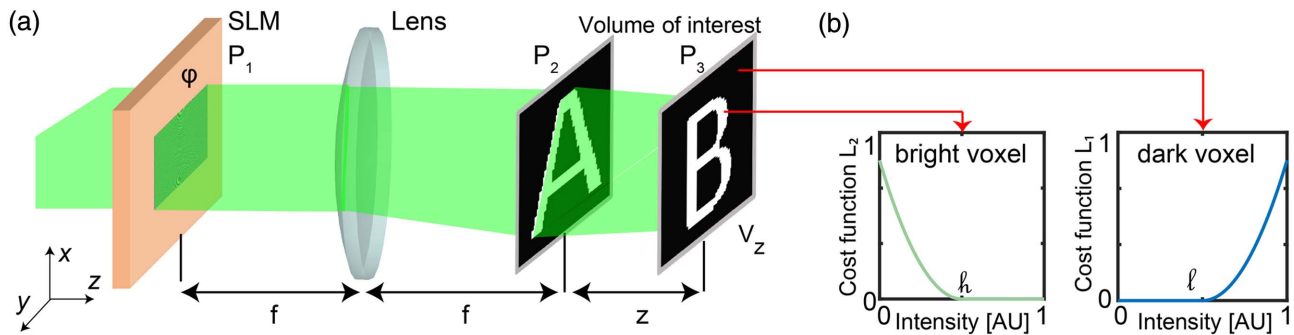
In Section 2, we mathematically describe the 3D CGH problem, review previous methods, and introduce our algorithm. In Section 3, we discuss the specific case of binary 3D intensity patterns for which we propose a customized multi-part cost function. In Section 4, we demonstrate improved performance over existing methods, both in simulation and experiment. Finally, we show a specific application in neural photostimulation, achieving enhanced spatial confinement of two-photon absorption when using NOVO-CGH. In Supplement 1, we provide additional details and implementations of NOVO-CGH for applications with non-binary target holograms.

2. THEORY

3D CGH algorithms take the target 3D intensity pattern, $V(x, y, z)$, as an input and return a 2D phase, ϕ, to be displayed on the SLM. Given the general setup in Fig. 1(a), a coherent plane wave, A_0, is incident on a phase-only SLM, which patterns the wavefront to generate the complex field $P_1(x, y) = A_0 \exp(i\phi(x, y))$. $\mathcal{T}_f$ denotes the transfer function from Fourier space at P_1 to image space at P_2, and $\mathcal{T}_z$ denotes propagation between P_2 and P_3 for the simple case of a two-plane target intensity pattern. Extension to more depth planes is straightforward.

Under the paraxial approximation, the complex fields can be written as $P_2 = \mathcal{T}_f P_1$ and $P_3 = \mathcal{T}_z P_2$:

$$P_2(x, y) = \frac{1}{i\lambda f} \iint P_1(x', y') \exp\left[\frac{2i\pi(xx' + yy')}{\lambda f}\right] dx' dy', \quad (1)$$

$$P_3(x, y) = \iint \frac{P_2(x', y')}{\sqrt{i\lambda z}} \exp\left[\frac{i\pi((x - x')^2 + (y - y')^2)}{\lambda z}\right] dx' dy',$$
(2)

where λ is wavelength, f is the lens' focal length, and z is the signed distance from P_2 to P_3. Adopting the notation in [18], Eqs. (1) and (2) become $\mathcal{T}_f = \mathcal{V}\left[\frac{1}{\lambda f}\right]\mathcal{F}$ and $\mathcal{T}_z = \mathcal{F}^{-1}\mathcal{Q}[-\lambda^2 z]\mathcal{F}$, where $\mathcal{F}$ denotes Fourier transform, $\mathcal{V}$ is a scaling constant, and $\mathcal{Q}$ is multiplication by a quadratic-phase exponential. Thus, $\mathcal{T}_f^{-1} = \mathcal{F}^{-1}\mathcal{V}[\lambda f]$, and $\mathcal{T}_z^{-1} = \mathcal{F}^{-1}\mathcal{Q}[\lambda^2 z]\mathcal{F}$. All operations are linear in the complex field, except for taking intensity, which is quadratic.

A. Existing Algorithms

1. Superposition Algorithm

Superposition algorithms aim to simplify the 3D CGH problem by splitting the desired intensity into segments and solving each independently. Our implementation uses the random superposition (RS) algorithm [2], except that instead of decomposing the intensity pattern into points, we decompose it into planes. This accelerates the computation and reduces errors due to intra-plane interference. As with other algorithms, the number of planes along the z-axis can be adjusted, and uniform sampling is not required. The phase hologram is computed as

$$\phi = \arg\left[\sum_{z \in Z} \mathcal{T}_f^{-1} \mathcal{T}_z^{-1}\left[\sqrt{V_z} \exp(j\phi_z)\right]\right], \quad (3)$$

where V_z is the 2D plane, $V(\cdot, \cdot, z)$, at depth z of V, and ϕ_z is a 2D random phase taking values uniformly in $[0, 2\pi]$. The arg function takes in a complex number and returns its phase. Since no iteration is required, the algorithm is fast, with complexity being proportional to the number of planes.

The simplicity of the superposition algorithm is also what limits its accuracy. Since amplitude information is discarded in the last step, it may not provide the optimal phase-only pattern. More importantly, the algorithm achieves 3D multiplexing by averaging the holograms corresponding to each plane. This assumes that each plane is independent, ignoring wave interference effects between planes. This approximation rapidly becomes invalid for dense 3D patterns with multiple depth planes.

Fig. 1. (a) Experimental setup for 3D CGH. A SLM imposes a phase pattern (hologram) on an incoming collimated plane wave. A Lens (focal length, f) in $2f$ configuration transforms the resulting wavefront into a 3D intensity distribution in the volume-of-interest. 3D CGH algorithms aim to generate a target 3D intensity distribution, $V(x, y, z) = V_z(x, y)$, within the volume-of-interest by designing, ϕ, the phase pattern on the SLM. (b) For our binary NOVO-CGH algorithm, we introduce a custom cost function with two parts, L_1 and L_2, which are truncated square functions defined in Eq. (9). L_1 penalizes intensities above a certain threshold, l, for dark voxels of the target pattern, and L_2 is the opposite (penalizes intensity below a threshold, h, for bright voxels). Total cost is the sum of all pixels' cost values.

2. Iterative Projection Algorithms

We also compare our algorithm with two popular iterative projection algorithms. Both are based on the Gerchberg–Saxton (GS) algorithm: *sequential* GS [12] and *global* GS [14] (see Supplement 1), and both use the superposition algorithm for initialization.

The sequential GS algorithm (also known as the *ping-pong* algorithm) digitally propagates the estimated complex field through each depth plane, one-by-one, while replacing the amplitude of the estimate with the desired amplitude at that plane. Then, the complex field is propagated to the SLM, where only phase is retained. This hologram becomes the estimate for starting the next iteration. This procedure is equivalent to stochastic gradient descent (SGD) with a fixed step size [19]. It diverges easily for infeasible target intensities, so we make a small modification that relaxes constraints slightly to enable convergence in most cases [16]; instead of forcing the dark areas to be zero, we let $\mathcal{P}_{V,z}E(x, y) = \mu \exp(j \arg[E(x, y)])$, where μ is a small number.

The global GS algorithm corresponds to batched SGD. Instead of sequentially propagating through all depth planes, the current phase hologram estimate is propagated to all depth planes at once, and the holograms from each are averaged at each iteration.

Though sequential and global GS algorithms are similar, their performance is very different. Sequential GS tends to diverge when the target 3D intensity is not physically feasible or becomes dense, while global GS is more stable because it simultaneously averages the contributions of all planes at each iteration.

B. Non-Convex Optimization for Volumetric CGH (NOVO-CGH)

Here, we propose a new algorithm, NOVO-CGH, that gives improved performance when it is not possible to exactly satisfy all the intensity constraints (the target pattern is physically infeasible). In this case, the choice of cost function and optimization procedure become critical. Previous algorithms used either fixed or implicit cost functions. Here, to enable more flexible addition of priors and customized cost functions, we start by formulating the 3D CGH problem as an optimization problem:

$$\phi^* = \arg \min_{\phi} L(I(\phi), V), \tag{4}$$

where $I(\phi)$ is the 3D intensity distribution generated by the 2D phase, ϕ, on the SLM, and L is a user-defined cost function.

This problem is non-convex (the Hessian matrix of the cost function is not positive semi-definite), but when the forward model, $I(\phi)$, and the cost function have a well-defined derivative, we can still solve the problem using first or second-order optimization methods. Here, we use L-BFGS [20], a quasi-Newton gradient descent procedure that always converges to a stationary point [21] and only requires limited amounts of computer memory. Our NOVO-CGH algorithm is summarized in Algorithm 1.

Algorithm 1. *NOVO-CGH Algorithm*

1:　**Procedure** NOVO-CGH (V)
2:　　$\phi_0 \leftarrow$ superposition(V)　　▷ Initialized by superposition
3:　　$n \leftarrow 0$
4:　　**while** $n <$ maxiter **do**
5:　　　$n \leftarrow n + 1$
6:　　　$\Delta\phi \leftarrow L - BFGS[\nabla_{\phi_{n-k}} L, \nabla_{\phi_{n-k+1}} L, ..., \nabla_{\phi_n} L]$
7:　　　$\phi_n \leftarrow \phi_{n-1} - \Delta\phi$
8:　　**return** ϕ_n

To compute the gradient, $\frac{\partial L}{\partial \phi}$, directly, we consider the 2D complex field at P_1 given by $E = \exp(i\phi)$ and derive

$$\frac{\partial L}{\partial \phi} = \frac{\partial L}{\partial I} \frac{\partial I}{\partial E} \frac{\partial E}{\partial \phi}, \tag{5}$$

$$\frac{\partial E}{\partial \phi} = \mathrm{diag}[-\sin(\phi) + i \cos(\phi)]. \tag{6}$$

In the following notation, we convert I and V into 1D vectors of size $N \times 1$, where N is the total number of voxels in I. $\frac{\partial L}{\partial \phi}$ has size $1 \times M$, and $\frac{\partial I}{\partial E}$ has size $N \times M$, where M is the number of pixels on the SLM. For these vectors, we let $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote the transpose, conjugate, and Hermitian operators, respectively. To find $\frac{\partial I}{\partial E}$, we first define $I_z(x, y) = I(x, y, z)$, then compute the gradient for each value of z.

In discrete space, we can represent linear operations as matrix multiplication, giving $\mathcal{T}_f = \gamma F^H$ and $\mathcal{T}_z = F^H K_z F$, where F is the discrete Fourier transform (DFT) matrix, γ is a scaling coefficient determined by the number of pixels, and K_z is a diagonal matrix representing Fresnel propagation. Hence, $\mathcal{T}_z \mathcal{T}_f = \gamma' F^H K_z$. Since intensity is point-wise amplitude squared, that is, $I_z = \|\mathcal{T}_z \mathcal{T}_f E\|^2 = \mathrm{diag}[\gamma' F^H K_z E]^H \gamma' F^H K_z E$, we can write

$$\frac{\partial I}{\partial E} = \sum_{z \in Z} \frac{\partial I_z}{\partial E} = \sum_{z \in Z} 2\gamma'^2 K_z^* F \, \mathrm{diag}[F^H K_z E]. \tag{7}$$

The only remaining requirement is that the cost function should be defined with respect to the intensity in the volume-of-interest and differentiable. NOVO-CGH can therefore be implemented with any cost function for which both quantities, $L(I(\phi), V)$ and $\frac{\partial L}{\partial I}$, can be expressed within the discretized volume. By combining $\frac{\partial L}{\partial I}$ with Eqs. (5)–(3), we deduce the gradient which is required for *L*-BFGS optimization.

In Sections 3 and 4, we propose cost functions for several specific applications. In Supplement 1, we consider alternate implementations of NOVO-CGH for 3D holography with variable-intensity patterns and multi-photon holography. In each case, we derive the cost function and corresponding gradient.

3. BINARY 3D HOLOGRAPHY

We first consider the case of *binary* intensity targets, which is relevant to applications in optical trapping and neural photostimulation [22,23], where precise control over light-sculpting capabilities is critical. Neurons can be modified optogenetically to be made photosensitive, which in theory enables the triggering [24] or inhibition [25] of neural activity on demand with light. In practice, a neuron's response to light is highly nonlinear, and activating custom neuron ensembles requires the ability to concentrate light on targeted neurons while keeping the intensity received by non-targeted neurons below some threshold value to prevent unwanted activation. With NOVO-CGH, custom cost functions can be tailored to match the nonlinear physiological response of the neurons.

Although the choice of cost function is flexible, we propose the following cost function:

$$L(I, V) = \sum_{\substack{x,y,z \\ \text{for } V(x,y,z)=0}} L_1(I(x, y, z)) + \sum_{\substack{x,y,z \\ \text{for } V(x,y,z)=1}} L_2(I(x, y, z)), \tag{8}$$

where the specific L_1 and L_2 functions [shown in Fig. 1(b)] are

$$L_1(I) = \begin{cases} (I - l)^2, & \text{if } I > l \\ 0, & \text{otherwise} \end{cases}. \tag{9}$$

L_2 is similar, but it penalizes low rather than high intensity. This cost function does not penalize intensity in voxels that should be dark, as long as it is below a certain threshold value, and then it quadratically increases the cost for intensity values above this threshold. The second part of the cost function, L_2, penalizes intensities below a threshold value for a desired bright voxel.

Calculating the derivative of this cost function gives

$$\frac{\partial L}{\partial I} = 2(I - l)^T \operatorname{diag}[I > l]\operatorname{diag}[\mathbf{1} - V]$$
$$+ 2(I - h)^T \operatorname{diag}[I < h]\operatorname{diag}[V], \qquad (10)$$

where $\mathbf{1}$ is an $N \times 1$ vector with all elements equal to 1 and

$$[I > l](x, y, z) = \begin{cases} 1, & \text{if } I(x, y, z) > l \\ 0, & \text{otherwise} \end{cases}. \qquad (11)$$

Since all matrix products in the derivative are either Fourier transforms, inverse Fourier transforms, or element-wise multiplications, the memory requirement is only $O(N)$, and computing the gradient remains spatially and temporally efficient.
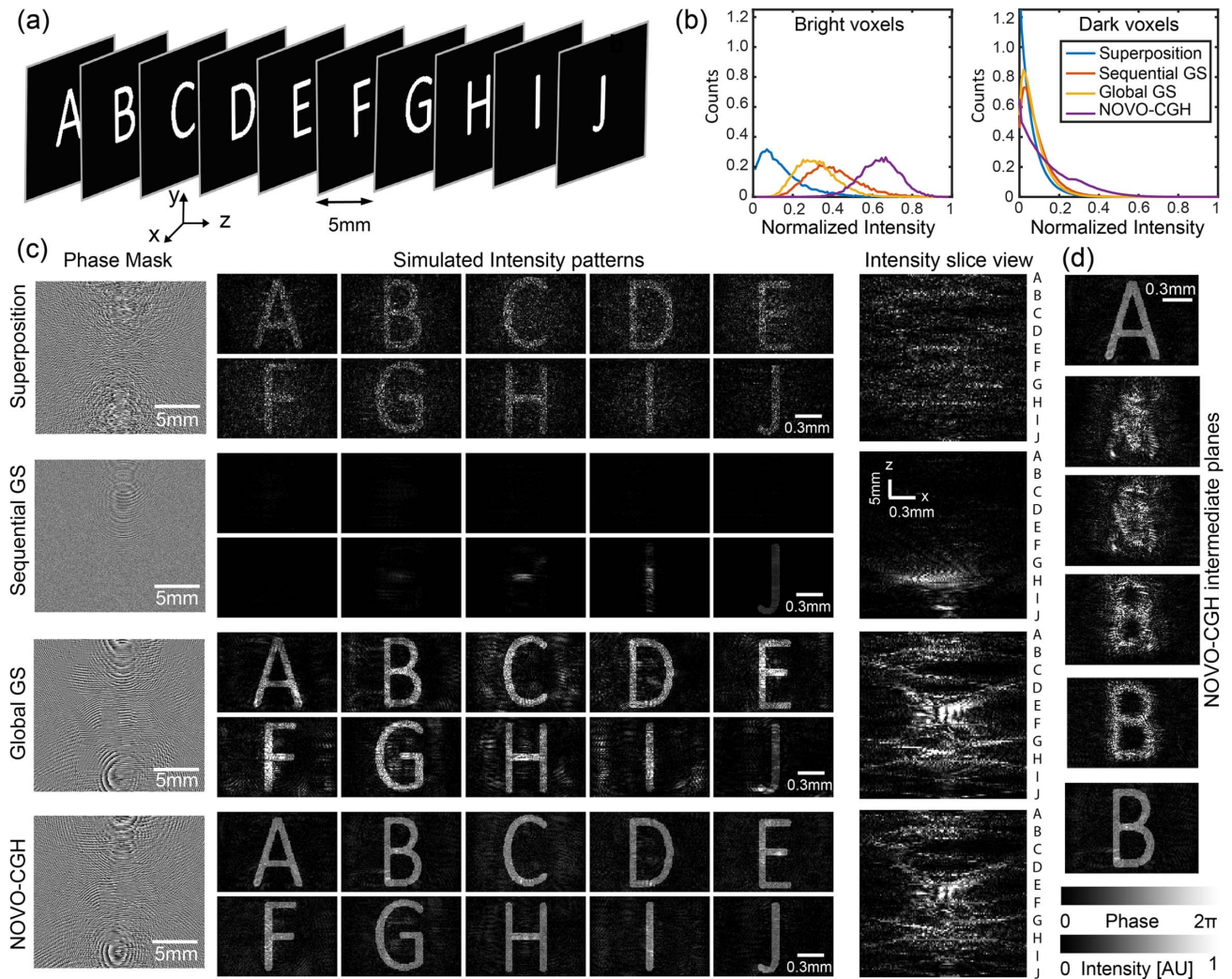
4. RESULTS

We demonstrate our algorithm's improved performance over previous techniques, in both simulation and experiment, for two applications: *multi-plane displays* and *neural photostimulation*.

In simulations, we have ground truth information and use two metrics to numerically evaluate performance—*accuracy*, α, and *efficiency*, η. Accuracy measures the resemblance of the generated intensity pattern to the target pattern, defined as

$$\alpha(I, V) = \frac{\sum_{x,y,z} I(x, y, z) V(x, y, z)}{\sqrt{\left[\sum_{x,y,z} V(x, y, z)^2\right]\left[\sum_{x,y,z} I(x, y, z)^2\right]}}, \qquad (12)$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product, and $\| \cdot \|_2$ denotes the L_2 norm. Efficiency quantifies how much of the energy goes into the bright areas of the target pattern, defined as

$$\eta(I, V) = \frac{\text{Effective energy}}{\text{Total energy}} = \frac{\sum_{V(x,y,z)=1}^{x,y,z} I(x, y, z)}{\sum_{x,y,z} I(x, y, z)} = \frac{\langle I, V \rangle}{\|I\|_1}, \qquad (13)$$

Fig. 2. Simulation results comparing NOVO-CGH to existing algorithms. (a) Target 3D intensity pattern containing 10 letters at evenly spaced depth planes (5 mm separation). (b) Quantitative comparison of simulation results between superposition [2], sequential Gerchberg–Saxton (GS) [12], global GS [14], and our binary NOVO-CGH algorithm. Normalized intensity histograms compare intensity values within bright and dark voxels separately, showing that NOVO-CGH yields more uniform intensity. (c) (Left) Computed 2D phase mask to be displayed on the SLM. (Center) Resulting intensity distributions at each patterned depth plane. (Right) Intensity slice in the *x-z* plane. (d) Intermediate intensity at planes between the "A" and "B" planes for NOVO-CGH. All intensity images have the same colorbar. Propagation images through intermediate planes are shown in Visualization 1.
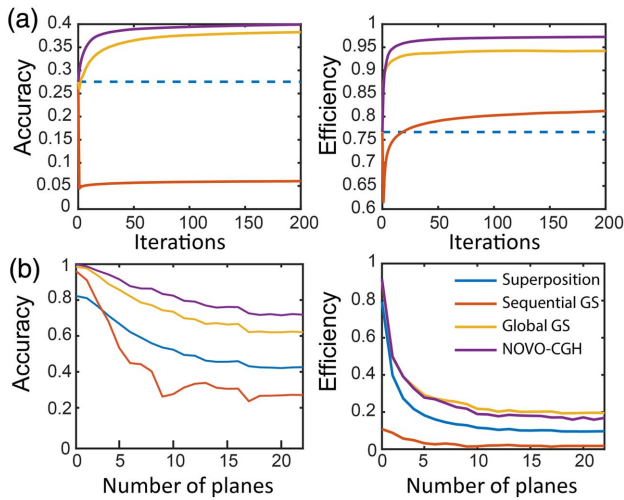
Fig. 3. Numerical comparison of algorithm performance in simulation. (a) Efficiency and accuracy versus number of iterations for a target intensity with 3 depth planes. Superposition algorithm is not iterative, so is drawn as a dotted line. (b) Efficiency and accuracy degrade as the target intensity pattern becomes more complicated (contains more depth planes).

where $\| \cdot \|_1$ denotes the L_1 norm. Note that efficiency and accuracy are not necessarily related—for example, a bright spot within the letter can have high efficiency but poor accuracy.

A. Multi-Plane Display

To compare our method to the superposition and iterative projection algorithms, we use each method to generate phase holograms of the same 3D target intensity patterns. We first simulate a relatively dense pattern that will not have a feasible solution [Fig. 2(a)]. The target intensity pattern contains 10 depth planes with different letters at evenly spaced z distances. Each letter extends an area of 1 mm × 1 mm, and the axial range spans 45 mm. The system

properties for simulations match our experimental setup (Fig. S1a of Supplement 1). The system NA is limited by the size of the SLM (NA ≈ 12.5 μm ∗ 1280/(2 × 200 mm) = 0.04).

As expected, the superposition algorithm generates a noisy result, because it ignores coherent interference from wave propagation between planes (see Fig. 2). Sequential GS also performs poorly, since it heavily favors the last plane used as a constraint. Global GS produces a better result, but it still has unwanted stray light in each plane. Our NOVO-CGH algorithm produces the most accurate result, particularly in the dark areas. Figure 2(b) plots the intensity distribution within bright and dark voxels, showing that our algorithm yields less variation. Note that all simulated intensity patterns have the same total energy, but some images appear darker because energy is spread or cropped by the colorbar. Additional slice views at the depth planes where the letters are [Fig. 2(c)] and at intermediate depths [Fig. 2(d)] show enhanced light confinement capabilities with NOVO-CGH. Media 1 shows the intensity at all planes through the 3D volume.

To quantify performance, Fig. 3 shows accuracy and efficiency plots for the simulated results. The three iterative algorithms generally improve with each iteration, eventually converging. Sequential GS exhibits a drop in performance during the first iteration because the algorithm is initialized with the superposition algorithm result, which is reasonably good but does not satisfy any single constraint. Therefore, projecting it onto any one of the constraints drags the solution further away from other constraints. This would not happen if the pattern was physically feasible. In Fig. 3(b), we show the accuracy and efficiency of each algorithm's converged solution as a function of the number of intensity planes in V (increasing complexity of the 3D pattern). As expected, when more planes are added, performance degrades, since the solution becomes less feasible. NOVO-CGH always produces the most accurate result, though the global GS algorithm has slightly higher efficiency.

To experimentally verify our algorithm, we conducted experiments (using the same parameters as simulations) over a volume
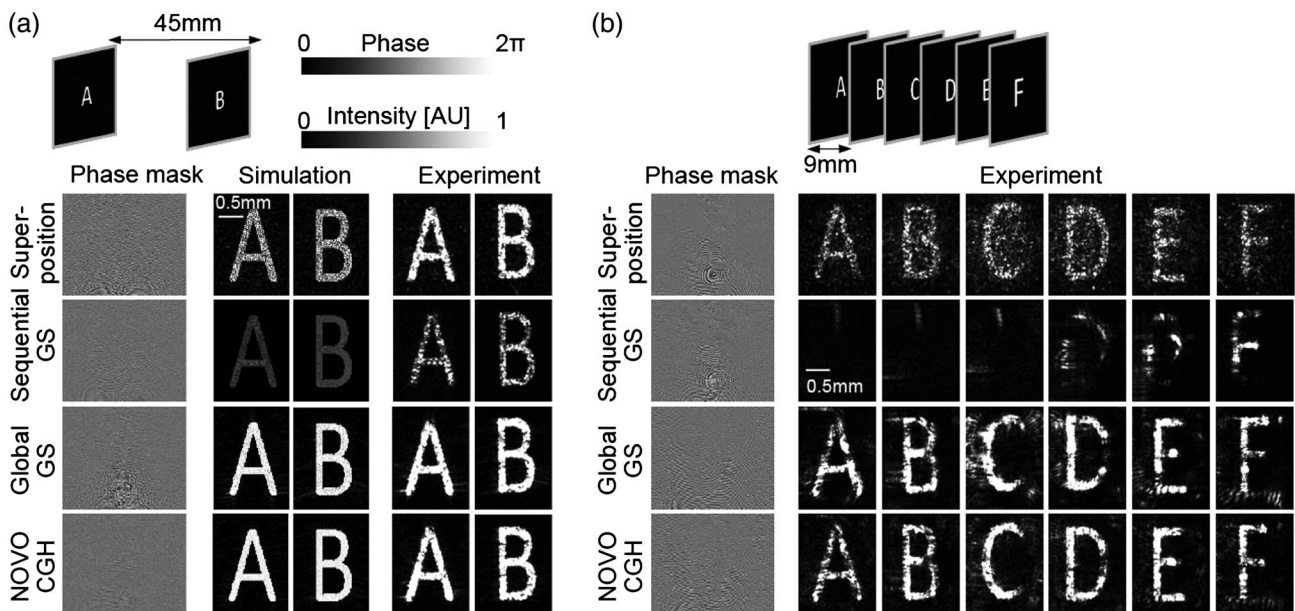
Fig. 4. Experimental results demonstrating improved 3D intensity patterning with NOVO-CGH, as compared to previous algorithms. (a) Comparison of simulation and experimental results for a two-plane intensity target. Calculated phase holograms are shown at left. (b) Experimental measurements for a six-plane intensity pattern. All images use the same camera settings.

of 1.5 mm × 1.5 mm × 35 mm. First, we show a simple 3D tar-get intensity pattern consisting of only two planes [Fig. 4(a)]. In this case, all algorithms perform well. However, when the target intensity contains more planes with different letters, the solution becomes less feasible and performance varies. In the case of six letter planes [Fig. 4(b)], NOVO-CGH achieves higher efficiency, less speckle, and better fidelity than the other algorithms.

B. Two-Photon NOVO-CGH for Neural Photostimulation

1. Enhanced Confinement Around a Single Target

Holographic two-photon photostimulation [26,27] is an imme-diate application that directly benefits from custom cost functions. To illustrate the benefits of NOVO-CGH, we first consider the simple case where we wish to illuminate the entire soma of a specific neuron, while avoiding surrounding neurons. The target intensity pattern is shown in Fig. 5(a). One plane with a disc pattern (approximately the size and shape of a neuronal soma in the mouse cerebral cortex) is surrounded by dark planes. By specifying a protected volume around the target, our algorithm seeks a solution with good axial confinement for use with dense neural networks. For this task, the sequential

GS algorithm fails to converge with the added dark planes, so we use only the center plane with the disc as input. The global GS and NOVO-CGH algorithms will have their computation scale linearly with number of dark planes added in the target in-tensity pattern.

Figure 5 shows that NOVO-CGH provides improved axial confinement in two-photon experiments. We use a femtosecond infrared laser source (Fidelity 18 W, $\lambda = 1040$ nm) and a micro-scope objective (20 × , 1.0 NA water immersion) to de-magnify the volume of operation. The target pattern [Fig. 5(a)] is a disc of radius 10 µm at a single depth plane (dark everywhere else). For characterization, we measure the two-photon absorption in 3D by observing fluorescence induced in a thin film at various depths (see Supplement 1).

Figures 5(b) and 5(c) shows cross-sections of the measured 3D two-photon absorption. All algorithms successfully generate a lateral disc pattern, but NOVO-CGH provides the best axial confinement. To explore the implications of our algorithm's im-provement on the relative threshold values for two-photon photo-stimulation, Fig. 5(d) shows normalized two-photon absorption as a function of an artificially increasing volume threshold. In any given volume size, the two-photon absorption threshold is consistently lower than NOVO-CGH compared to other algorithms, indicating better volumetric confinement of light.
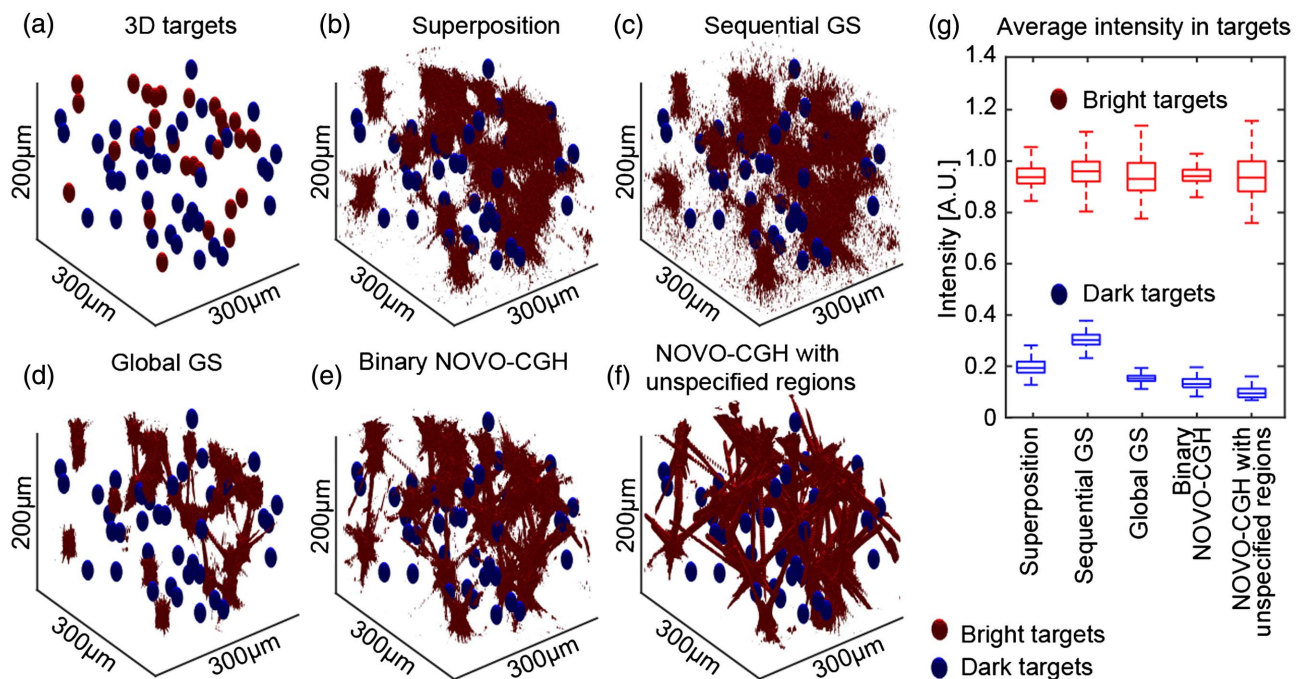
2. Multiple Targets with Unspecified Regions

A major advantage of NOVO-CGH for 3D light sculpting in optogenetic photostimulation and photolithography is the ability to define spatially dependent cost functions that enforce both bright and dark voxels, while leaving other areas unspecified. Dark voxels enforce a penalty in locations where illumination should be avoided, instead of enforcing it uniformly in all non-bright voxels (see Section 3 of Supplement 1). This strategy relaxes constraints and searches for holograms for which there ex-ists a pathway for photons to reach the bright targets while avoid-ing the dark targets, better protecting sensitive parts of the sample from receiving inadvertent illumination.

To demonstrate in simulation, we consider a 3D distribution of sphere-like targets within the accessible volume [Fig. 6(a)]. There are two categories of targets: red regions denote bright tar-gets, where we wish to deposit energy, and blue regions denote dark targets, which should be protected from unwanted illumi-nation. The remainder of the volume is unspecified (may have any intensity value). With other algorithms [Figs. 6(b)–6(d)] and binary NOVO-CGH [Fig. 6(e)], the user would only specify $V(x, y, z) = 1$ within the red targets, then attempt to enforce darkness everywhere else within the volume. Here, we selectively relax the darkness constraint by applying a strong penalty on dark target voxels (see Supplement 1). We specify $V^+(x, y, z) = 1$ within the bright targets, $V^+ = 0$ elsewhere, and $V^-(x, y, z) = 10$ within the dark targets (see Fig. S2) and 0 elsewhere. Outside the bright and dark targets, $V^+ = V^- = 0$. The resulting 3D intensity generated by this variant of NOVO-CGH [Fig. 6(f)] finds a suitable optical pathway to illuminate the bright targets while avoiding the dark targets. For all targets within the volume, we compute the average received power and repeat the simulation 10 times with randomized target distributions (no overlap). Statistical data is shown in Fig. 6(g), confirming that dark targets are better protected from unwanted illumination with NOVO-CGH using unspecified regions.
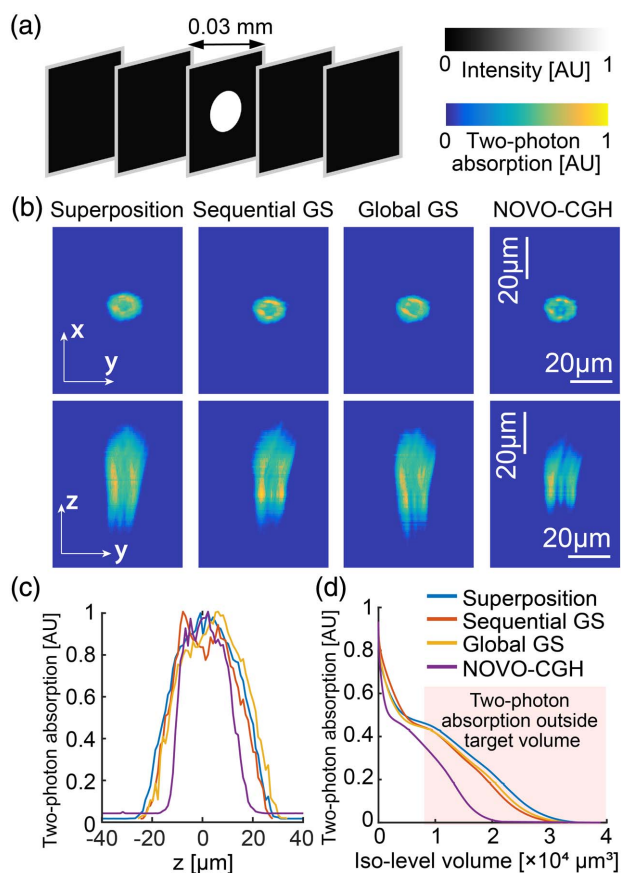

Fig. 5. Experimental results for neural photostimulation. (a) Target 3D intensity is a bright disc (10 µm radius) at one depth, surrounded by darkness. Planes are evenly spaced by 0.03 mm. (b) x-y and y-z cross-sections of two-photon absorption (fluorescence) measured along the center of the volume. (c) Two-photon absorption axial cross-cuts. (d) Sorting all voxels in descending order, we plot two-photon absorption versus volume. Given a neuron of volume 8×10^3 µm³, values in the pink region represent intensity in undesired regions.

Fig. 6. NOVO-CGH with unspecified regions. (a) Simulated 3D target intensity with 70 randomly distributed spheres. Red targets are desired bright regions, and blue targets are desired dark regions, with the remainder of the volume unspecified. (b)–(f) Resulting 3D intensity distributions from (b) superposition algorithm, (c) sequential GS, (d) global GS, (e) binary NOVO-CGH, and (f) NOVO-CGH with unspecified regions. We display in each case threshold envelopes of the volume intensity distribution (3% of maximal intensity, in red) alongside the blue regions. (g) For an equal amount of illumination within the red targets, we compute the amount of unwanted illumination in the blue targets, repeating the simulation 10 times with randomized distributions. NOVO-CGH out-performs other algorithms and further enhances confinement when unspecified regions are allowed.

C. Two-Photon and Variable-Intensity NOVO-CGH

In Supplement 1, we provide additional variations for NOVO-CGH. One enables the rendering of grayscale (variable intensity) 3D intensity patterns. For image rendering applications, we propose a threshold-based and Euclidean cost function, then derive their respective gradients. Simulation results (see Fig. S3 of Supplement 1 and Visualization 2) show that our method enhances accuracy. Finally, we look at the case of two-photon imaging and propose a cost function (and its derivative) that directly optimizes for I^2 instead of I, tailoring holograms to nonlinear light-matter interactions (see Fig. S4 of Supplement 1).
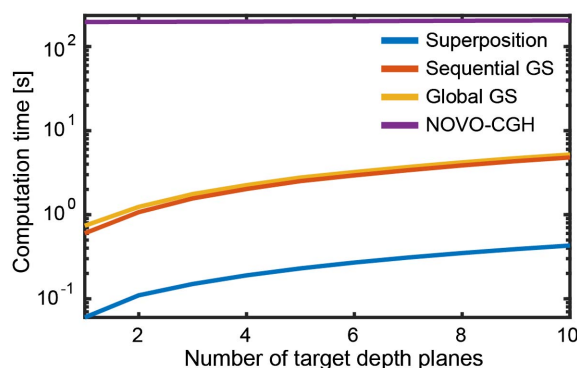
D. Computation

The NOVO-CGH algorithm is implemented in Matlab with the L-BFGS method [20]. All algorithms used a graphics processing unit (GPU) for accelerated computation. Runtimes for generating holograms of size 1280×1024 on our GPU (NVIDIA Tesla K20c) are shown in Fig. 7 as a function of the number of planes in the target volume intensity (which scales linearly with the number of voxels). The runtimes for superposition, GS, and NOVO-CGH algorithms all scale linearly with the number of planes. NOVO-CGH has the largest computational burden from computing the cost and the derivative, as well as a significant offset due to large overhead. Hence, the improvement in hologram quality comes at a cost of more computation time.

5. CONCLUSION

We have demonstrated a new algorithm for improved 3D CGH by optimization with customizable cost functions that can be explicitly defined as a function of light intensity or space. The method can be tailored to particular applications or samples. For the example case of binary intensity targets, we demonstrated improved performance for multi-plane displays and enhanced confinement for light sculpting. We further included results for different cost functions to show that our algorithm can be easily adapted to different applications.

These results show that the customizable cost function of NOVO-CGH can be adapted not only to a specific type of holography but also to sample-specific features. The result is an algorithm that is better suited to accomplish specific tasks. For instance, in neuroscience applications, the cost function

Fig. 7. Computation times on log scale for varying numbers of target depth planes. We use GPU computing to generate holograms of size 1272×1024, with 50 iterations for the GS algorithms and the NOVO-CGH algorithm.

can be tuned to match the response of a specific opsin to light and can also take into account the spatial distribution of neurons within the 3D volume-of-interest to enable photostimulation of specific sets of neurons on demand. The same principle may also be applied to holographic photolithography [28] to match the optical response of a specific photoresist and to prevent a material from inadvertently curing in locations that would affect the shape of the structures being created.

We hope that NOVO-CGH can provide a different prospective for CGH algorithms since it is different from existing algorithms. Instead of paying attention to the solver, the user can now focus on designing the cost function to define what a good solution means. Such an approach addresses problems more directly, giving improved performance at the expense of higher computational cost. We want to emphasize that our algorithm does not expand the set of possible solutions, as that is constrained by the physics and hardware. Instead, NOVO-CGH finds a better solution within the set, based on particular tasks. Open-source code for NOVO-CGH can be found in our GitHub repository [29].

†These authors contributed equally to this work.

See Supplement 1 for supporting content.

REFERENCES

1. C. Slinger, C. Cameron, and M. Stanley, "Computer-generated holography as a generic display technology," Computer **38**, 46–53 (2005).
2. R. D. Leonardo, F. Ianni, and G. Ruocco, "Computer generation of optimal holograms for optical trap arrays," Opt. Express **15**, 1913–1922 (2007).
3. S. Yang, E. Papagiakoumou, M. Guillon, V. de Sars, C.-M. Tang, and V. Emiliani, "Three-dimensional holographic photostimulation of the dendritic arbor," J. Neural Eng. **8**, 046002 (2011).
4. N. C. Pégard, A. R. Mardinly, I. A. Oldenburg, S. Sridharan, L. Waller, and H. Adesnik, "3D scanless holographic optogenetics with temporal focusing," Nat. Commun. In press (2017).
5. C. Lutz, T. S. Otis, V. DeSars, S. Charpak, D. A. DiGregorio, and V. Emiliani, "Holographic photolysis of caged neurotransmitters," Nat. Methods **5**, 821–827 (2008).
6. F. Anselmi, C. Ventalon, A. Bègue, D. Ogden, and V. Emiliani, "Three-dimensional imaging and photostimulation by remote-focusing and holographic light patterning," Proc. Nat. Acad. Sci. **108**, 19504–19509 (2011).
7. L. B. Lesem, P. M. Hirsch, and J. A. Jordan, "The kinoform: a new wavefront reconstruction device," IBM J. Res. Dev. **13**, 150–155 (1969).
8. D. Leseberg, "Computer-generated three-dimensional image holograms," Appl. Opt. **31**, 223–229 (1992).
9. T. Shimobaba, T. Ito, N. Masuda, Y. Ichihashi, and N. Takada, "Fast calculation of computer-generated-hologram on AMD hd5000 series GPU and OpenCL," Opt. Express **18**, 9955–9960 (2010).
10. O. Ripoll, V. Kettunen, and H. P. Herzig, "Review of iterative Fourier-transform algorithms for beam shaping applications," Opt. Eng. **43**, 2549–2556 (2004).
11. R. Gerchberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35**, 237 (1972).
12. M. Makowski, M. Sypek, A. Kolodziejczyk, G. Mikuła, and J. Suszek, "Iterative design of multiplane holograms: experiments and applications," Opt. Eng. **46**, 045802 (2007).
13. R. Piestun, J. Shamir, B. Weßkamp, and O. Bryngdahl, "On-axis computer-generated holograms for three-dimensional display," Opt. Lett. **22**, 922–924 (1997).
14. R. Piestun, B. Spektor, and J. Shamir, "Wave fields in three dimensions: analysis and synthesis," J. Opt. Soc. Am. A **13**, 1837–1848 (1996).
15. M. Pasienski and B. DeMarco, "A high-accuracy algorithm for designing arbitrary holographic atom traps," Opt. Express **16**, 2176–2190 (2008).
16. M. Makowski, M. Sypek, A. Kolodziejczyk, and G. Mikuła, "Three-plane phase-only computer hologram generated with iterative Fresnel algorithm," Opt. Eng. **44**, 125805 (2005).
17. Z. Jingshan, L. Tian, J. Dauwels, and L. Waller, "Partially coherent phase imaging with simultaneous source recovery," Biomed. Opt. Express **6**, 257–265 (2015).
18. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company, 2005).
19. L.-H. Yeh, "Analysis and comparison of Fourier ptychographic phase retrieval algorithms," Technical Report No. UCB/EECS-2016-86 (University of California, 2016).
20. D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," Math. Program. **45**, 503–528 (1989).
21. F. E. Curtis and X. Que, "A quasi-Newton algorithm for nonconvex, nonsmooth optimization with global convergence guarantees," Math. Program. Comput. **7**, 399–428 (2015).
22. L. Fenno, O. Yizhar, and K. Deisseroth, "The development and application of optogenetics," Annu. Rev. Neurosci. **34**, 389–412 (2011).
23. V. Nikolenko, K. E. Poskanzer, and R. Yuste, "Two-photon photostimulation and imaging of neural circuits," Nat. Methods **4**, 943–950 (2007).
24. E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth, "Millisecond-timescale, genetically targeted optical control of neural activity," Nat. Neurosci. **8**, 1263–1268 (2005).
25. B. Y. Chow, X. Han, A. S. Dobry, X. Qian, A. S. Chuong, M. Li, M. A. Henninger, G. M. Belfort, Y. Lin, P. E. Monahan, and E. S. Boyden, "High-performance genetically targetable optical neural silencing by light-driven proton pumps," Nature **463**, 98–102 (2010).
26. V. Nikolenko, B. O. Watson, R. Araya, A. Woodruff, D. S. Peterka, and R. Yuste, "SLM microscopy: scanless two-photon imaging and photostimulation using spatial light modulators," Front. Neural Circuits **2**, 5 (2008).
27. G. S. He, P. P. Markowicz, T.-C. Lin, and P. N. Prasad, "Observation of stimulated emission by direct three-photon excitation," Nature **415**, 767–770 (2002).
28. P. Wang and R. Menon, "Optical microlithography on oblique and multi-plane surfaces using diffractive phase masks," J. Micro/Nanolithogr., MEMS, MOEMS **14**, 023507 (2015).
29. N. Pégard and J. Zhang, "MATLAB code for NOVO-CGH," https://github.com/Waller-Lab/NOVOCGH (2017). GitHub repository.