High-speed single-shot optical focusing through dynamic scattering media with full-phase wavefront shaping

Ashton S. Hemphill, Yuecheng Shen, Yan Liu, and Lihong V. Wang

Citation: Appl. Phys. Lett. 111, 221109 (2017);
View online: https://doi.org/10.1063/1.5009113
View Table of Contents: http://aip.scitation.org/toc/apl/111/22
Published by the American Institute of Physics
The depth of optical focusing inside biological tissue is limited by the diffusion of light, which prevents the formation of an optical focus at depths greater than ~1 mm (the optical diffusion limit).\(^1 \text{-} 3\) This limit greatly reduces the utility of optical imaging and manipulation techniques.\(^4 \text{-} 7\) Obviously, the ability to focus light within and through turbid media would be invaluable to biophotonics, permitting the use of optical methods such as optogenetics, microsurgery, optical tweezing, and phototherapy in deep tissue.

In order to focus light through or within scattering media, a variety of wavefront shaping methods have been developed, such as feedback based wavefront shaping, transmission matrix measurement, and optical time-reversal or optical phase conjugation (OPC).\(^8 \text{-} 20\) Of these, OPC holds the greatest promise for biological applications because it has the shortest average mode time (i.e., the runtime required per degree of freedom utilized) due to its global determination of the optimal wavefront.\(^21\) Both analog and digital OPC (DOPC) have been reported. Importantly, while analog OPC provides high speed and a large number of controls, DOPC has a much greater fluence reflectivity but currently lacks controls.\(^21 \text{-} 24, 26, 27\)

In addition, DOPC is typically limited by the speed of image acquisition (capture, transfer, and processing) and by the time it takes the spatial light modulator (SLM) to display the phase map. These low speeds have limited the use of DOPC \textit{in vivo}, where the motion of scatterers causes rapid decorrelation of the wavefront (on the order of milliseconds) and breaks the time reversal symmetry.\(^22, 29, 30\)

Recently, a fast DOPC system has been reported which controls $1.5 \times 10^5$ optical degrees of freedom with an effective latency of 5.3 ms and a system runtime of 7.1 ms. This was achieved by using a quasi-single-shot measurement method, in which a reference image is acquired before the beginning of the runtime, in conjunction with a digital micro-mirror device (DMD).\(^30\) Another recently developed system controls $2.6 \times 10^5$ optical degrees of freedom, focusing light through scattering media with an effective latency of 3.5 ms and a system runtime of 4.7 ms. This system also utilizes a quasi-single-shot measurement method and employs a high-speed ferroelectric SLM for phase modulation.\(^31\) However, because of their specific wavefront measurement methods and display devices, these methods do not yield full-phase wavefront compensation and are sufficient for only binary-amplitude and binary-phase correction.

In biomedical applications, full-phase compensation allows for greater accuracy in the displayed optimal phase map and thus significantly increases the focusing capability. This increased focusing ability, along with high speed, is paramount for applications in living tissue, where scattering is strong and the speckle correlation time is short.

Typically, full-phase wavefront shaping is performed using phase-shifting holography, requiring a minimum of three images in order to calculate the optimal phase map.\(^32 \text{-} 34\) However, by utilizing off-axis holography, a full-phase compensation map may be recovered with a single...
Here, we demonstrate such an off-axis DOPC system. In conjunction with a graphics processing unit (GPU) for parallel processing, this method allows for rapid single-shot phase recovery. Employing this fast phase measurement method along with a high-speed SLM, we focus light through scattering media with full-phase optimization, achieving an effective system latency of approximately 9 ms, on the order of the in vivo correlation time.

For this system, we utilized a high-speed SLM (HSP256-0532, Meadowlark Optics; 256 × 256 pixels) (Fig. 1). Illumination was supplied by a 5 W continuous wave laser (Verdi V-5, Coherent) at 532 nm. The sample arm and both reference arms were split by a pair of polarizing beam splitters, with half-wave plates used to adjust the ratios of illumination to each arm. In both reference arms, a pair of lenses (LA1540-A & LA1484-A, Thorlabs) expanded the beams to fully illuminate the SLM. The primary reference arm (R1) was used to acquire the optimal phase map, while the higher-intensity secondary reference arm (R2) was utilized in wavefront playback. The sample arm (S) was scattered using an opal diffuser (10DIFF-VIS, Newport), which provided 4π steradian scattering. The scattered light was then guided by a collection lens and a mirror to the SLM, where it interfered with the vertically polarized primary reference arm.

To facilitate off-axis holography, the angle between the sample and reference beams was chosen to satisfy the Nyquist criterion at the sCMOS camera (pco.edge 5.5, PCO). A camera lens imaged the surface of the SLM onto the sCMOS camera and provided 4:1 pixel matching of the SLM and sCMOS camera. The speckle size of the scattered light was set to approximately 4 × 4 camera pixels, corresponding to roughly 1 × 1 SLM pixels. With a region of interest (ROI) of 1036 × 1034 pixels at the sCMOS camera, our system provided ~65 000 controls for optimization.

After the sCMOS camera captured the interferogram, data processing was performed using a GPU (GTX 1070, Nvidia) and the compute unified device architecture (CUDA).

The angle, \( \theta \), between the sample and reference beams produced periodic fringes within the interferogram which served as a carrier frequency in k-space. The angle \( \theta < \theta_{\text{max}} = \arcsin(\lambda/2d) \), where \( \theta_{\text{max}} \) is the angle at which the carrier frequency is equal to the Nyquist frequency, \( \lambda \) is the wavelength of the laser, and \( dx \) is the pixel size of the sCMOS camera.

This frequency modulation allows the phase map to be isolated through filtering in k-space. The intensity of the hologram is given by \( IH = |R + S|^2 = |R|^2 + |S|^2 + RS^* + R*S, \) where \( R \) is the complex amplitude of the reference beam, \( S \) is the complex amplitude of the sample beam, and \( * \) denotes complex conjugation. In an on-axis configuration, the zeroth order, \( |R|^2 + |S|^2, \) overlaps with the real image, \( R*S, \) and the virtual image, \( S*R^*, \) in k-space.

However, when an angle is introduced, the reference beam becomes \( R(x) = \sqrt{\theta} \exp(i\sin \theta x) \) and the intensity of the hologram becomes \( IH = I_R + I_S + \sqrt{\theta} \exp(-i\sin \theta x)S + \sqrt{\theta} \exp(i\sin \theta x)S^*. \) The real and virtual images are therefore located at \(( k \sin \theta/2\pi, 0)\) and \((-k \sin \theta/2\pi, 0)\), respectively, about the center frequency in k-space and can be separated by spatial filtering. In our system, this filtering was achieved digitally, using GPU based processing (please see supplementary materials for pseudocode) for fast Fourier transformation (FFT), multiplication of the terms in k-space by a pre-determined mask, and inverse FFT for recovery of the phase map.

The system runtime, from the time that the sCMOS camera began acquisition to the display of the compensation wavefront, was 11.48 ± 0.02 ms (see Fig. 2). However,
because a rolling shutter was used in the acquisition of the interferogram by the sCMOS camera to achieve a shorter run-time, the effective system latency was shorter than the system runtime.30,31 This shutter utilized dual outside-in readout: Starting from the top and bottom rows of the ROI, pairs of rows are exposed for 500 μs and then acquired. Acquisition next moves toward the center, with a 9.17 μs delay in the start of the exposure of each subsequent pair of rows. In our system, the central 1034 rows of the sCMOS camera were used. The effective system latency, from the average exposure start time to the playback of the optimal wavefront, was therefore calculated to be 9.13 ms. The system latency, defined as the time constant in the exponential fit of the peak-to-background ratio (PBR) versus the speckle correlation time, was also determined experimentally to be 9.11 ms.30,31

The performance of our system was quantified by calculating the ratio between the experimental and theoretical PBRs of the focus achieved through full-phase DOPC. Specifically, the PBR was defined as the ratio between the average intensity of the focal peak, the area where the intensity was greater than half the maximum intensity, and the average mean intensity when five random wavefronts were displayed by the SLM.30,31 The focus achieved by our system when focusing through the opal diffuser can be seen in Fig. 3(a), while Fig. 3(b) demonstrates the focal intensity distribution in the vertical dimension. The experimental PBR of our system was 1.2 × 10^{-3}, with the background intensity coming from an area of 1.04 × 1.04 mm² and the focal size being 2.9 × 10^{-3} μm², as determined by the full width half maximum. The theoretical PBR is given by \( \frac{\pi N}{4M} \), where \( N \) is the optical degrees of freedom optimized by the SLM and \( M \) is the number of optical modes within the focus. The speckle size at the sCMOS camera was 6.8 × 10^{-2} μm², calculated by the FWHM of the autocovariance function of the speckle field captured by the sCMOS camera, which had a pixel size of 6.5 μm. Because the sCMOS camera utilized 4:1 pixel matching with the SLM, which has a pixel size of 24 μm, the size of the speckles at the SLM was 5.8 × 10^{-2} μm². \( M \) was calculated by comparing the area of the focus (2.9 × 10^{-3} μm²) and the area of a speckle grain at the CCD camera (4.4 × 10^{-2} μm²). Based on this calculation, \( M = 6.6 \), and the theoretical PBR of our system was 7.9 × 10^{-3}. The experimental PBR was thus calculated to be 15% of the theoretical PBR, with the discrepancy likely attributable to imperfect alignment, aberration of the reference beam by optics, and SLM substrate curvature.

To measure the system latency, we produced an optical focus through dynamic scattering media with a controllable speckle correlation time.21,22,30,39–43 To control the speckle correlation time, we mounted the opal diffuser on a linear translation stage with a motorized actuator (LTA-HS, Newport) and quantified the relationship between the speckle correlation time and the movement speed.

For this quantification, we used the sCMOS camera to record the speckle field at the SLM for a given movement speed. After recording the images corresponding to several translation speeds, we calculated the correlation coefficients between the first frame and each subsequent frame for each movement speed. The specific speckle correlation time was then obtained by fitting the correlation coefficient \( R \) versus time using the equation \( R = A \exp \left( -\frac{2r^2}{\tau_c^2} \right) + B \), where \( \tau_c \) is the time at which \( R = \frac{1}{e^2} \). Figure 4(a) shows this fit for a movement speed of 0.024 mm/s, which yields a speckle correlation time of 100 ms. The relationship between the...
speckle correlation time and movement speed was then fitted by the equation $t_c = d_s/v$, as shown in Fig. 4(b), where $d_s$ is the expected speckle size seen by the detection optical system at the rear surface of the scattering medium. Based on this fit, we found that $d_s = 2.42 \mu m$ for our system, and we were able to control the speckle correlation time by using the appropriate movement speed.

In this way, a high-contrast focus was achieved for correlation times greater than 5 ms. As shown in Fig. 5(a), the respective PBRs for each set speckle correlation time are given in Table I.

Table I. PBRs of foci with varying speckle correlation times. The number of realizations $n = 10$.

<table>
<thead>
<tr>
<th>Correlation Time (ms)</th>
<th>&gt;3 s</th>
<th>250</th>
<th>100</th>
<th>50</th>
<th>25</th>
<th>15</th>
<th>10</th>
<th>5</th>
<th>2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBR</td>
<td>1186 ± 27</td>
<td>1167 ± 31</td>
<td>1141 ± 23</td>
<td>1065 ± 32</td>
<td>891 ± 45</td>
<td>484 ± 28</td>
<td>293 ± 42</td>
<td>25 ± 7</td>
<td>2 ± 0.7</td>
</tr>
</tbody>
</table>

Figure 5(b) shows the foci formed through scattering media with correlation times varying from 2.5 ms to >1 s (a stationary diffuser). The control, a random phase map displayed by the SLM, is also shown in Fig. 5(b). Because the PBR is proportional to the speckle correlation coefficient, the PBR at each speckle correlation time can be fitted by an exponential function, $PBR = A \exp \left(-\frac{2B^2}{t_c^2}\right) + C$. Using this fit, we then found the effective system latency where $B = t_c$, and the PBR became $1/e^2$ of the maximum PBR with a non-moving diffuser.

While our system provides full-phase single-shot DOPC focusing, it still suffers from hardware-limited bottlenecks in image capture and phase display speeds. Currently, we use a pco.edge 5.5, which has a minimum exposure time of 0.5 ms. By using a faster camera such as a pco.edge 4.2, the exposure time can be reduced to 0.1 ms, thus reducing the system runtime by ~0.4 ms.

Other devices, such as DMDs and ferroelectric SLMs, may be used for faster wavefront modulation. These devices allow for phase display within 1 ms, reducing the system runtime by ~3 ms. However, the devices provide only binary-amplitude or binary-phase wavefront shaping. Because of this, the maximum theoretical PBRs of the system would be reduced to $N/(2\pi M)$ and $N/(\pi M)$, respectively, yielding lower efficiency than full-phase wavefront shaping.

Finally, it is possible to reduce the runtime of our system by reducing the ratio of the scCMOS camera to SLM pixels so that a smaller ROI may be imaged while still providing 4 × 4 camera pixels per speckle. By decreasing the ROI of the scCMOS camera, the runtime can be shortened by an equivalent amount of time. Therefore, while the pixel ratio of 4:1 scCMOS to SLM pixels as used requires 11.48 ms, a ratio of 2:1 would require 8.85 ms, while a 1:1 ratio would require only 7.13 ms. This reduction can be done by grouping multiple SLM pixels into superpixels while adjusting the camera lens so that each superpixel is matched to 4 × 4 scCMOS pixels (in order to satisfy the Nyquist criterion). By doing so, the runtime can be decreased by reducing the number of controls and thus the maximum theoretical PBR.

Off-axis holography dramatically reduces the time required for image acquisition and data transfer by minimizing the number of images needed. Compared to phase-shifting holography, there is an increased computational load from the use of fast Fourier transforms, but this is offset by the use of GPU parallel processing. As a result, our system provides full-phase DOPC focusing with a system runtime of 11.48 ms and an effective system latency of 9.11 ms. This runtime is on the order of other aforementioned high-speed systems using devices such as DMDs and ferroelectric SLMs for light modulation and the in vivo speckle correlation time while still providing full-phase optimization.

See supplementary material for further information regarding the algorithms and control code utilized in the measurement, computation, and display of the optimal phase map from an off-axis hologram.

We thank Jim Ballard for help in the revision and editing of the manuscript. This work was sponsored by NIH Grants No. DP1 EB016986 (NIH Director’s Pioneer Award) and R01 CA186567 (NIH Director’s Transformative Research Award).

1. V. Ntziachristos, Nat. Methods 7(8), 603–614 (2010).


