

Comprehensive Holographic Parallel Beam Modulation inside Material based on Automatic Differentiation

Hengyang Li,¹ Huaizhi Zhang,¹ Jiaming Xu,¹ Shuo Li,¹ Xiao Li,¹ Wei Cheng,¹ Yu Xiao,¹ Gang Xu,¹ Xiahui Tang,¹ and Yingxiong Qin^{*1}

School of Optical and Electronic Information & Wuhan National Laboratory for Optoelectronics, Optics Valley Laboratory, National Engineering Research Center for Laser Processing, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, China

(*Electronic mail: qyx@hust.edu.cn)

(Dated: 13 April 2023)

Holographic beam modulation is widely applied in optical tweezers, hard-brittle material marking, high-density storage, etc. To generate high-fidelity arbitrary 3-dimensional (3D) parallel multifoci inside the material, the spherical aberration compensation automatic differentiation (SACAD) algorithm is presented. All polarization components are included and the spherical aberration compensation is embedded in the physical model. The technique of automatic differentiation is used in the error backpropagation procedure, ensuring efficient parallel computing of pixel-by-pixel gradients. In several simulation tests, the root mean square errors of the generated 3D multifoci distributions are all less than 0.01 and the diffraction efficiencies are all beyond 90%, outperforming the results of the established algorithms. In the experiments, we have verified the advantage of SACAD algorithm in complicated 3D internal marking with spherical aberration compensated. Since the SACAD algorithm can achieve high fidelity and efficiency phase retrieval with a straightforward procedure, it has the potential to become a well-received solution for internal parallel beam modulation.

Keywords: Phase retrieval, Spherical aberration, Automatic differentiation, Beam modulation

I. INTRODUCTION

Lasers have proven to be a mature and versatile tool with great flexibility and applicability for precision machining. The past two decades have witnessed the development of direct laser writing that serves various relevant fields^{1–4} especially the precise machining of hard-brittle material. The utilization of ultrafast laser internal focus realize internal machining without surface damage. However, the low efficiency of the point-by-point sequential workflow is a significant drawback in industrial scenarios.

To this end, parallel multifoci array^{5–7} is put forward, bringing a significant reduction in processing time by transferring point-by-point exposure to simultaneous exposure. Recent advancements towards parallel multifoci array problems mainly concentrate on phase-only modulation with computer-generated holographs (CGHs)^{8–12}. Researchers employed CGHs in 3D structure processing and fabrication, including in-bulk marking^{13,14}, high-density storage¹⁵, microfluidics^{16,17}, etc. Meanwhile, internal parallel multifoci requires a precise modulation of the light complex amplitude, while the largest obstacle exists in the spherical aberration (SA) caused by interface refractive index (RI) mismatch. Objective lens with glass thickness compensation is regarded as a solution of SA¹⁸.

However, the SA compensation based on special-designed objective lens is limited owing to the inflexibility in multi-depth processing and in diverse materials. Besides, the design of the above established CGH algorithms often pay primary attention to the beam modulation in the air. Therefore, they ignore the effect of SA on complex amplitude and different cross-interface behaviors of different polarization components, so that their evaluation functions cannot precisely estimate the real light intensity distribution inside the material.

For adopting internal beam modulation, there are methods introduce a SA compensation phase pattern and directly overlay it on the CGH^{19,20} as a quick fix. In experiments and engineering, these methods have obtained remarkable results, while in some cases only a compensation phase pattern might be insufficient.

In this article, we report the spherical aberration compensated automatic differentiation (SACAD) algorithm to generate CGHs for modulating 3-dimensional (3D) parallel multifoci inside the material. Unlike conventional methods that calculate the beam splitting phase and SA compensation phase separately, we discard this elaborate step-by-step design method. Instead, a comprehensive physical model is used, comprising the effect of interface RI mismatch and all polarization components. A cost function that directly evaluates the light intensity distribution inside the material is established. In the procedure of minimizing the cost function, we use efficient nonlinear optimization methods based on gradient back-propagation and automatic differentiation. The SACAD algorithm could realize higher light intensity fidelity and diffraction efficiencies compared with other mainstream methods in 10-second optimization. Experiment results have verified that the CGHs produced with SACAD algorithm can realize complex 3D structure marking with SA compensation inside the fused silica.

The article is structured as follows: in section 2, a comprehensive numerical beam-propagation model including the influence of interface RI mismatch will be described; in section 3, we will define the cost function and use automatic differentiation to realize error backpropagation; in section 4, the gradient-based optimization algorithm is presented; in section 5, the SACAD algorithm will be compared with other common-used 3D multifocal phase modulation algorithms; in section 6, we will give experimental results about the effect of

SA compensation and in-bulk 3D parallel marking. Conclusion is drawn in section 7.

II. FORWARD PHYSICAL MODEL

We follow a typical inverse design procedure by firstly determining a forward model, for the purpose of calculating the electric field in focal region with a given the electric field at the incident pupil. The effect of the interface RI mismatch needs to be included in the forward model. Our numerical expression is based on the well-known vectorial diffraction integration presented by Richard-wolf^{21,22} and is built with reference of the theory presented by Török²³.

Eq. 1 explicitly determines the electric field of the focal region inside the material in the high numerical aperture (NA) scenario:

$$\mathbf{E}_2 = -\frac{i}{2\pi n_1 k_0} \mathcal{F} \left\{ \frac{\mathbf{T}(\theta_1, \theta_2, \phi) \mathbf{E}_1 e^{ik_0 \psi(\theta_1, \theta_2, d_{nom})} e^{ik_2 z z}}{\cos(\theta_1)} \right\}, \quad (1)$$

where \mathbf{E}_2 and \mathbf{E}_1 are the electric field vectors in the focal region inside the material and at the incident pupil, including three orthogonal polarization components respectively. $\psi(\theta_1, \theta_2, d_{nom}) = d_{nom} (n_2 \cos \theta_2 - n_1 \cos \theta_1)$ in the first exponential term represents the SA phase introduced by the material, and $k_2 z$ in the second exponential term describes the phase accumulation when beam propagates along the z axis ($k_2 = n_2 k_0 \cos \theta_2$). θ_1 and θ_2 are the aperture angles outside and inside the material and ϕ is the polar angle. z is the axial coordinate with the zero point located at the actual focus inside the material. $k_0 = 2\pi/\lambda$ is the wave number in vacuo. n_1 and n_2 are the RIs outside and inside the material. $\mathcal{F}\{\cdot\}$ denotes the 2D Fourier transform. The matrix \mathbf{T} expresses the influences of the lens on the different polarized incident components. The incident beam we used in all the simulations and experiments is linearly polarized in x direction, so \mathbf{E}_1 only has one component E_1 and \mathbf{T} can be simplified as

$$\mathbf{T}(\theta_1, \theta_2, \phi) = A(\theta_1) \begin{bmatrix} \tau_p \cos^2 \phi \cos \theta_2 + \tau_s \sin^2 \phi \\ (\tau_p \cos \theta_2 - \tau_s) \cos \phi \sin \phi \\ -\tau_p \cos \phi \sin \theta_2 \end{bmatrix}, \quad (2)$$

where τ_p and τ_s are the Fresnel transmission factors for s-polarized and p-polarized light, respectively. $A(\theta_1)$ is the apodization function of the objective lens, and here we take $A(\theta_1) = f \sqrt{\cos \theta_1}$ omitting irrelevant constants. The corresponding schematic diagram is given in Fig. 1(a) and derivation process is detailed in supplementary material.

Numerical simulation results of single-focus intensity distribution inside the material using this physical model are also given in Fig. 1. Comparing Fig. 1(b) and (c), the focus deteriorates apparently because of the RI mismatch, resulting in lower peak intensity and longer depth of focus. The validity of the model's numerical results is strongly confirmed by recent work²⁴ in experiments.

On the basis of Eq. 1 and Eq. 2, we will further discuss the physical factors resulting in the focus deterioration illustrated in Fig. 1(b) and the existing ways to eliminate them.

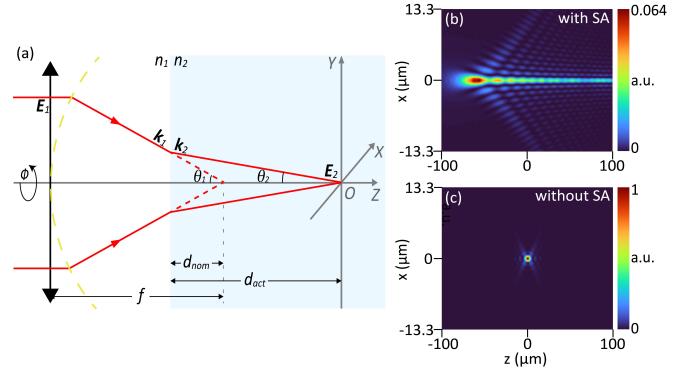


FIG. 1. Illustration of internal focus and SA. (a) is a schematic diagram of focusing inside the material. The zero point of the coordinate system in the diagram is set on the true focus position. f is the focal length of the objective lens in the air, and d_{nom} and d_{act} are the distance from nominal focus to the interface. \mathbf{E}_1 and \mathbf{E}_2 are the vectorial electric field at the incident pupil of the objective lens and near the focus. (b) and (c) show the shapes of the focus with and without SA. The color maps are normalized to the maximum of each image. The peak intensities are labeled.

Firstly, after passing through the air-material interface, the transmittance ratio differs with the polarizations and locations at the incident pupil, equivalent to an amplitude mask. The effect of the amplitude mask is not significant in multifocal modulation because it basically equally influences each focus. Hence, the nonuniformity of the amplitude mask is generally ignored. Secondly, according to Snell's law, the convergence angle is smaller inside the material and the internal focus will move from d_{nom} to d_{act} . For this shifting of the whole focus region, a common compensation measure is simply moving the material farther from the objective lens. Thirdly, the RI mismatch disturbs the wavefront of the converging spherical wave, causing spherical aberration (SA). Normally a pure SA phase mask determined by NA, d_{nom} , and λ could be extracted from $\psi(\theta_1, \theta_2, d_{nom})$ ^{20,25}. In this way, one can realize SA compensation by adding a phase pattern opposite to the SA phase using some adaptive optics means²⁶.

Briefly, it can be found that the established methods obtain the final CGH by combining two phase patterns, one for generating multifoci, and the other for compensating SA. Instead, we decided to directly employ the complicated but comprehensive physical model mentioned above to realize "one-stop" CGH design. In this way, factors such as polarization, defocusing and SA, etc. are not considered separately.

III. COST FUNCTION AND ERROR BACKPROPAGATION

The cost function is used to evaluate the deviation of the calculated intensity distribution inside the material from the target intensity distribution. Here, the cost function is defined in a "MSELoss" way and ensures all polarized components

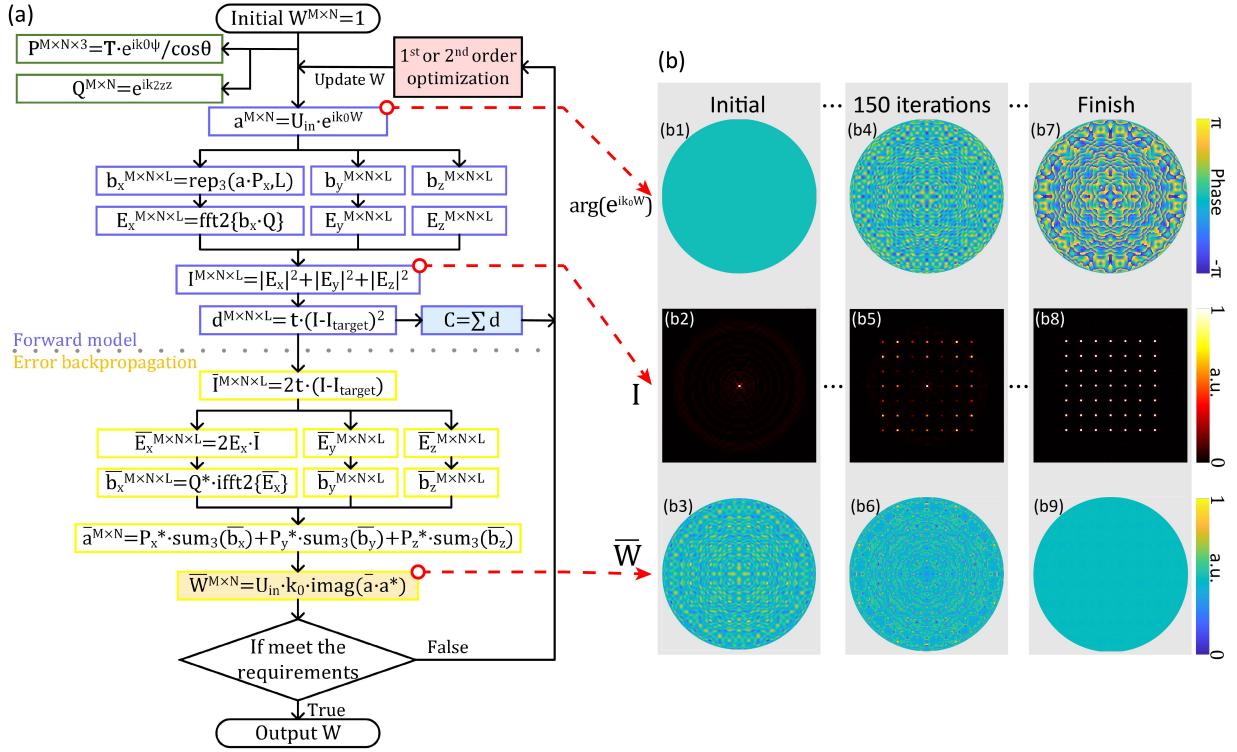


FIG. 2. Iteration procedure flowchart. In (a), P and T contain 3 elements corresponding to the 3 polarization components. The superscripts of variables denote their matrix sizes, M and N are the transversal resolution, and L is the number of layers. Function $rep_3(a \cdot P_x, L)$ means repeating the 2-dimension matrix $a \cdot P_x$ for L times along the third dimension. $\sum d$ is the sum of all elements in d . $sum_3(\bar{b}_x)$ means sum the elements along the third dimension. Function $fft2$ and $ifft2$ means carrying out 2D fft or ifft. $imag(..)$ denotes taking the imaginary part. Asterisk denotes taking the complex conjugate. Variables with a $\bar{}$ overline denotes the partial differentiation of the cost function on every element in this variable. (b) shows the transforms of $arg(e^{ik_0 W})$, I , \bar{W} with iteration. The color maps in (b1), (b4), and (b7) stick to $-\pi$ and π . The color maps in (b2), (b3), (b5), (b6), (b8) and (b9) are normalized to the maximum and minimum of each image.

included as

$$C = \sum_{n=1}^N t_n \left| \sum_{p=x,y,z} |E_{n,p}|^2 - I_{target,n} \right|^2, \quad (3)$$

where p denotes each of the polarization components in focal region. $E_{n,p}$ denotes one of the electric field components at the n th pixel, and $I_{target,n}$ is the target light intensity at the n th pixel. N is the number of all the simulated pixels, which means that the cost function does not only calculate the deviation of a few specific focal positions (like the GSW algorithm⁹ does), but also takes into account other positions where the target value is 0. t_n is the weight of each pixel according to demand, and generally, acceptable results can be obtained by choosing the t_n in the focal positions higher than that in the dark area where target value is 0.

It is noteworthy that unlike the GS algorithm²⁷ where the target distribution is usually in a normalized form, normalization is not recommended here. The reason is that the bulk target light field consists of multiple transversal planes but does not necessarily contain the plane where the original focus is located in. Normalization at this point will possibly result in undesired light intensity at the original focus, i.e., 0th-order beam (under the premise of the initial phase is planar phase).

The recommended approach is calculating the total light intensity first and assigning it to each focus proportionally.

According to this cost function, the partial derivatives of a certain polarization component at a certain pixel in the focal region can be obtained as

$$\overline{E_{n,p}} = 4t_n E_{n,p} \left(\sum_{p=x,y,z} |E_{n,p}|^2 - I_{target,n} \right), \quad (4)$$

and $\overline{E_{n,p}}$ will be further backpropagated to every pixel at the incident pupil. Here, the term “backpropagation” refers to the traditional definition found in machine learning²⁸. In the procedure of backpropagation, the technique of reverse-mode automatic differentiation is used. The optimization variables of this model are pixel-by-pixel 2D phase distribution (10^5 to 10^6 dimensions), so calculating the finite-difference gradients would be extremely time-consuming. However, the automatic differentiation provides a “recipe” to convert the forward model into the error backpropagation procedure in a more systematic and straightforward way²⁹. Automatic differentiation has been utilized in solving a range of nonlinear optimization problems in optics, especially in ptychography^{30–32}. A detailed explanation is given in supplementary material.

To give a clear picture of the algorithm, the procedure of the

forward model and the error backpropagation is shown in Fig. 2(a). As the variable to be optimized, W contains the optical path difference in each pixel at the incident pupil. The forward model procedure is in blue, whose output variable is the aforementioned cost function C . The error backpropagation procedure is in yellow, whose output variable is the gradient \bar{W} (of the same size as W). C and \bar{W} are together used as resources for the first- or second-order optimization algorithms. Fig. 2 also indicates that three orthodox polarization components are computed independently, ensuring the intensity field is constructed with all polarized electric field components.

The information on three key variables of the optimization procedure is featured in Fig. 2(b). Figures are picked from a simple computing instance that the target intensity distribution is 7×7 foci matrix on the focus plane. Initially, W is 1 as the planar phase pattern shown in Fig. 2(b1), intensity in focal region is a focus peak in Fig. 2(b2), and gradient \bar{W} is obvious in Fig. 2(b3). After 150 iterations, the phase pattern vaguely appears in Fig. 2(b4), and part of the focus matrix can be seen in Fig. 2(b5) though the center focus is much stronger. After the computing completes, the phase pattern is regular and sharp in Fig. 2(b7), the same intensity distribution as target is obtained as shown in Fig. 2(b8), and the gradient \bar{W} almost vanishes in Fig. 2(b9). Fig. 2(b) indicates that not only the multifocal modulation phase but also the SA compensation phase are generated from optimization. Two kinds of phase patterns are treated as a fusion and the SACAD algorithm makes no distinction between them.

In Fig. 2(a), it can be found that 2-D FFT and iFFT are used in the algorithm. In the framework of laser micro-nano processing and optical micromanipulation, high precision of foci position is enticing. Unfortunately, due to the property of the discrete Fourier transform, the transversal sampling interval in the image plane is as large as $\lambda f/L_0$ (L_0 is the diameter of the incident pupil), and the image plane must maintain the same resolution with the incident pupil. Hence, the resolution and sample interval may limit the accuracy and arbitrariness of beam modulation. To overcome this obstacle, inspired by Rabiner³³ and Jurling et.al^{34,35}, we utilize the chirp Z-transform (CZT), which allows breaking the fixed pixel pitch relationship between incident and image plane in FFT. Using the CZT in the forward model is straightforward, but its inverse transform in error backpropagation procedure is not trivial, and some discussion on it has been made^{36,37}. In this work, we provide a practical reverse-as-inverse approach when the chirp contour is evenly spaced and on the unit circle. Although not universally applicable, this approach is very enough in our problem and efficient in coding and computing. Further description can be found in the supplementary material.

IV. OPTIMIZATION ALGORITHM

The optimization algorithm relies on the gradient and the value of cost function. As shown in Fig. 2, it can be summarized in the following steps:

1. Initialize W . Incidentally computing static variables \mathbf{P}

and Q .

2. Obtain the cost function C using the forward model.
3. Obtain the gradient vector \bar{W} using error backpropagation
4. Perform first- or second-order optimization with C and \bar{W} and update W along the gradient direction.
5. If the iteration stop condition is satisfied, output W . Otherwise, go back to step 2.

Concerning step 1, the initial value of W is chosen to be ones. It is encouraged to choose a better initial value such as the direct superposition of the defocusing phases of all foci. A clever initial value will reduce the time consumption of the optimization procedure, but here initial values remain 1 with the consideration of ensuring the same starting points in following comparisons.

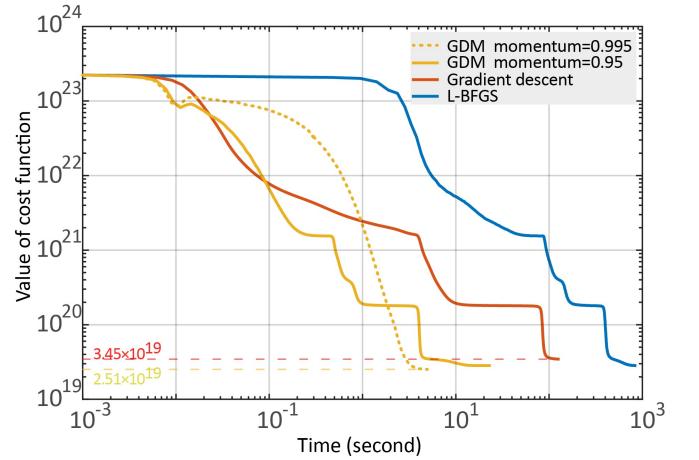


FIG. 3. The reduction of cost function with time. The single-loop time consumptions in different optimization algorithms are different and total numbers of loops is also unequal. GDM with 0.995 momentum: 3100 loops. GDM with 0.95 momentum: 14000 loops. Gradient descent: 48000 loops. L-BFGS: 1920 loops.

Concerning step 4, automatic differentiation provides analytic gradient values, so some gradient-required nonlinear optimization methods are applicable here. We have tested the gradient descent with fixed step size, as well as the gradient descent with momentum (GDM). Besides, the BFGS algorithm is considered as one of the best quasi-Newton methods and the L-BFGS algorithm^{38,39} further reduces the consumption of memory. Fig. 3 illustrates the convergence speeds and results of the above 3 algorithms in calculating the CGHs for two layers of 3×3 foci matrix. GDM algorithm shows significant advantage in both computing speed and optimizing result. The L-BFGS algorithm is the slowest because of the complexity of the variable updating rule, but its convenience lies in that a specified step size is not required. Based on the above analysis, in all the following comparisons, simulations and experiments, the GDM algorithm is employed with proper step size and momentum.

All the calculations were performed on a personal desktop with an AMD 5950X processor and Nvidia RTX3080Ti GPU. The program was written, compiled, and run in MATLAB R2022a. A demonstration code is available in GitHub.⁴⁰

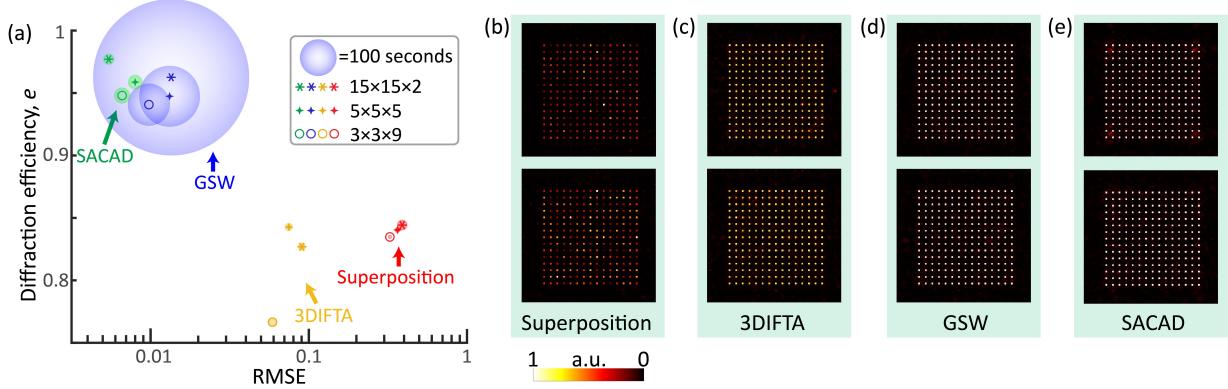


FIG. 4. Comparison in diffraction efficiency, uniformity, and time consumption of 4 algorithms. (a) 3 kinds of 3D target distribution are $15 \times 15 \times 2$ (labeled with hexagon), $5 \times 5 \times 5$ (labeled with four-pointed star), and $3 \times 3 \times 9$ (labeled with ring) foci matrix. Foci are evenly distributed inside a region with $42 \mu\text{m}$ length, $42 \mu\text{m}$ width, and $160 \mu\text{m}$ height, and the target intensities of foci remain the same. Different algorithms are in different colors. The area of the disk denotes the time consumption, and the disk in legend corresponds to 100 seconds. In (b)-(e), the simulated intensities of the $15 \times 15 \times 2$ foci matrix generated with 4 algorithms are shown. The color map was normalized to the maximum of each image.

V. COMPARISON WITH EXISTING ALGORITHMS

Several methods have been presented for the phase-only modulation of 3D multifoci in high NA scenario. We will give evaluation criteria, briefly introduce three kinds of widely used algorithm, and compare their numerical simulation results with SACAD's.

Firstly, diffraction efficiency e and the normalized root mean square error ($RMSE$) are defined as

$$e = \frac{\sum_m I_m}{I_0}, \quad RMSE = \sqrt{\frac{\sum_m (I_m - \langle I_m \rangle)^2}{\sum_m \langle I_m \rangle^2}} \quad (5)$$

where I_m and I_0 are the m th focal intensity and original single focal intensity (only SA compensation phase on CGH). $\langle I_m \rangle$ is the average of I_m . Pseudocodes for all three algorithms are given in supplementary material.

A. Existing algorithms

1. Superposition algorithm

The superposition algorithm solves the phase retrieval problem by modulating all target foci separately⁴¹. The phase masks for all separated foci in different positions are obtained initially, and summed together according to their intensity weights. Our implementation introduced a random phase offset for every phase mask according to the reference¹¹. The simplicity of the superposition algorithm is also what limits its accuracy. Except in the case of very few foci, with a high probability this algorithm provides not an optimal solution but just a feasible solution because it considers all focal points independently and ignores lateral and interlayer crosstalk.

2. Weighted Gerchberg-Saxton (GSW) algorithm

The GSW algorithm is the most widely-used and effective multifocal modulation algorithm⁹. Similarly, this algorithm initially requires phase masks for all separated foci in different positions. Then it introduces the adaptive weights and updates them with iteration according to the results of each iteration and the target intensities. Essentially, the GSW algorithm similarly adds a phase offset for each phase mask and sums them up, but the phase offsets are determined by iteration rather than by random selection. The GSW algorithm can achieve good uniformity and high diffraction efficiency.

3. 3D Iterative Fourier transform algorithm (3DIFTA)

With the rapid growth of hardware's computing speed and memory capacity, direct iterative operation between three-dimensional light fields becomes realistic. 3DIFTA uses Ewald cap and 3D Fourier transform to establish the relationship between the incident beam and the volumetric light field in the focal region⁴². Recent study presented a modification solution for adopting it into internal beam modulation¹⁴. It is claimed that this algorithm can deal with the inter-layer crosstalk problem better¹³.

B. Comparisons on diffraction efficiency, RMSE and time consumption

We have carried out a comparison between the 4 algorithms in simulation, where 3 kinds of target intensity distributions were chosen as test instances. Fig. 4 shows that the superposition algorithm performs worst in terms of diffraction efficiency and uniformity. Results of 3DIFTA are acceptable in uniformity but weak in diffraction efficiency, additionally requiring huge memory space. Both the SACAD and GSW al-

gorithms significantly outperform them in all 3 test instances, but the difference in computing time is obvious. The iteration process of the GSW algorithm includes the integral of the light intensity for each focus, so the time consumption is relatively high and proportional to the number of target foci. The SACAD algorithm uses less than 10 seconds in all test instance. Overall, taking only several multiples of the time of the explicit algorithm (superposition algorithm), the SACAD algorithm surpasses the leading GSW algorithm in terms of diffraction efficiency and fidelity.

Not only leads in the evaluation criteria, SACAD algorithm is also straightforward and friendly in use. the superposition algorithm, GSW algorithm and 3DIFTA toward beam modulation inside the material requires some modifications including changing Z-direction defocusing phase and additionally overlaying an SA compensation phase. However, these modifications are built-in mechanisms of SACAD algorithm. Figuring out the parameters of the objective lens and the material is all one needs.

VI. EXPERIMENTS AND RESULTS

We generated several CGHs with the SACAD algorithm, adopting them into experiments to verify the performance of SA compensation and the capability of internal 3D parallel multifocal modulation.

A. Experiment setup

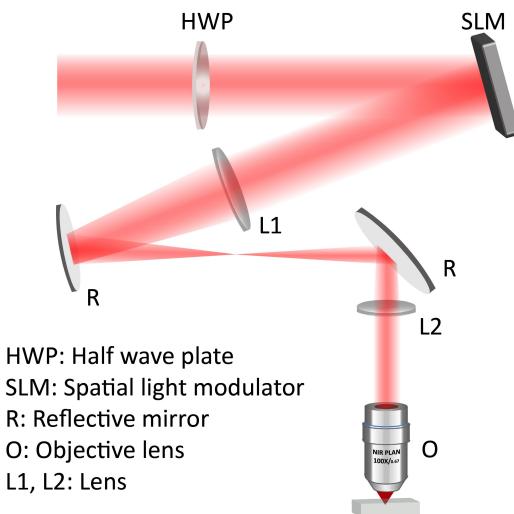


FIG. 5. Experiment setup. L1 and L2 are lens with $f_1 = 400\text{mm}$ and $f_2 = 200\text{mm}$.

To verify the validity of the algorithm, we built the experiment platform as shown in Fig. 5. A picosecond laser (edgewave INNOSLAB PX100-1-GF, Nd:YVO₄, 1064nm) was used as the light source. The half wave plate was used

to adjust the polarization direction of incidence to fit the response direction of the LCOS spatial light modulator (SLM) (Hamamatsu X15223). The CGHs were downloaded to the SLM. A 4f system was used to relay the optical path and match beam diameter with the incident pupil size of the objective lens (OptoSigma PAL-50-NIR-HR 0.67NA). The fused silica sample is fixed on an x-y-z translation stage.

After the internal marking is completed, the fused silica sample is polished on the lateral sides for observation. We used a back-illuminated microscope to observe the internal marking results.

B. Experiment results

We have carried out experiments to verify the multifocal SA compensation. The primary SA phenomena inside the material are the elongation of the depth of focus and the weakening of the energy at the focal position, which can be observed by the size and shape of the material modified region inside the fused silica. Fig. 6(a) gives a phase pattern with SA compensation for three foci inside the material at a depth of 2000 μm with a horizontal interval of 2.4 μm and a longitudinal separation of 70 μm . The phase pattern in Fig. 6(b) does not take into account the effect of SA. Simulations with these two phase patterns show that in Fig. 6(c) that the light intensity distribution of the three foci with SA compensation is almost identical to the focal shape in vacuo, while the light intensity distribution in Fig. 6(d) is severely disturbed. Fig. 6(e) and Fig. 6(f) give a comparison of the material modified regions observed from the side. In Fig. 6(e), the modified regions are shorter and consistent in shape, and all three foci can be clearly distinguished. The darker colors also reflect the greater light intensity in modified regions around foci. While in Fig. 6(f), the modified regions are so long that they become indistinguishable due to SA. Fig. 6(e) and Fig. 6(f) were obtained by maintaining the laser pulse output while moving the sample. The same phenomenon can be observed in Fig. 6(j) and Fig. 6(k) with a spiral structure as in Fig. 6(i), realized using the phase patterns in Fig. 6(g) and Fig. 6(h). In addition, SA can also cause a notable loss of fidelity when attempting to achieve a complicated light intensity distribution inside the material. We have marked an array of 14 foci with an energy ratio of 2:3:4:5:6:7:8 [Fig. 6(m)] inside the material at 1000 μm depth, with two foci of the same energy in a row. In Fig. 6(n), the modified regions produced by the SA-compensated CGH clearly reflect the same variation trend as the increasing trend of the light intensity proportions, from the weakest point above to the strongest point below in order. Moreover, due to the degradation of the focal intensity, the two focal positions with the lowest energy proportion in Fig. 6(p) did not reach the modification threshold of fused silica, causing the marking failure. In another two-layer structure, Fig. 6(q) and Fig. 6(r) illustrate that the defects due to SA are more non-negligible, exacerbating interlayer crosstalk in the bulk light field and obstructing the phase modulation of the complicated light intensity distribution. These comparisons indicate that SACAD algorithm is urgently needed in large-depth internal

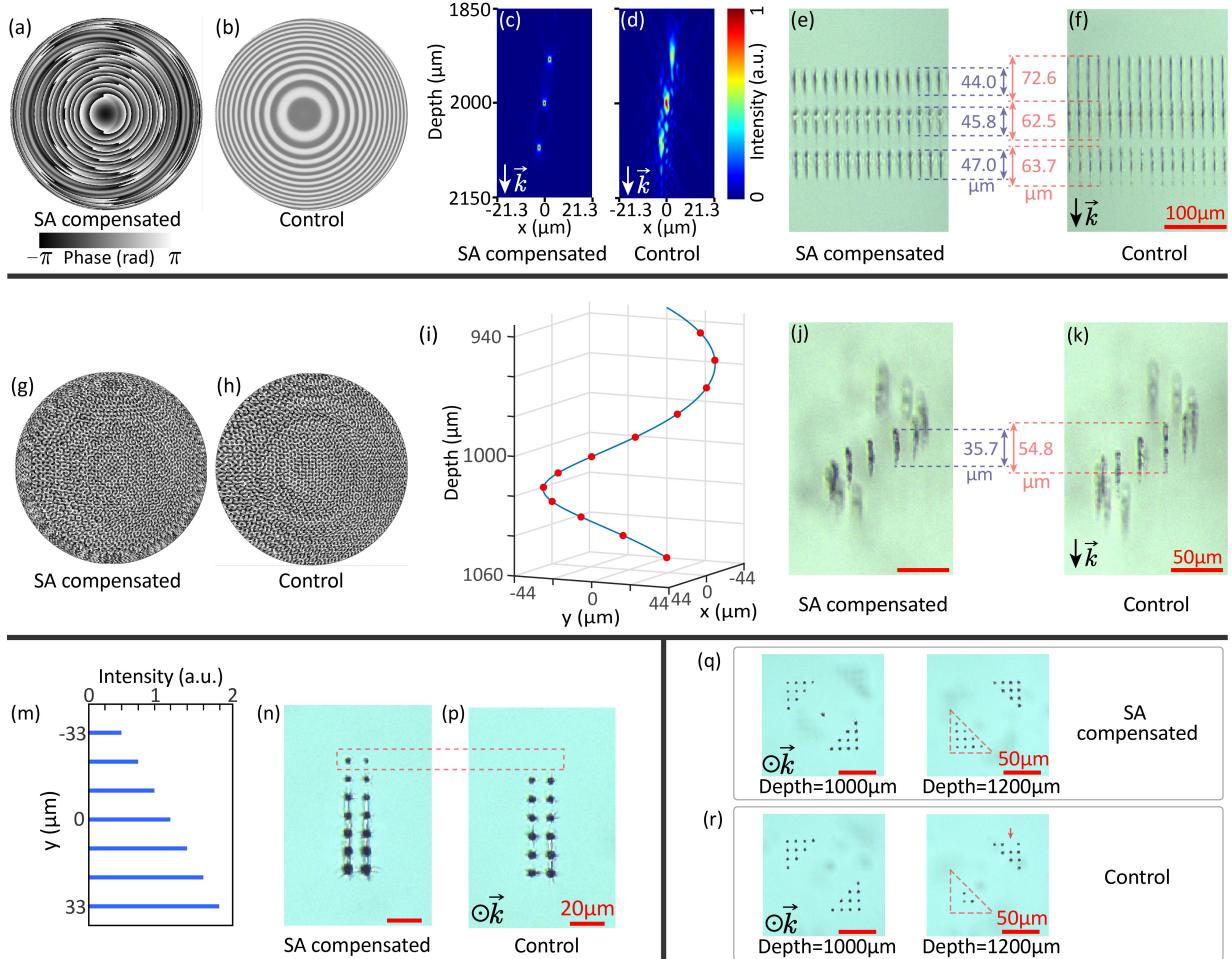


FIG. 6. Experiment results of comparisons between the SA compensation and control (control groups ignore the influence of the material). Both SA compensated images and control images in the same groups use the same microscopic imaging parameters and the same laser pulse energies. For the 3-foci experiment, (a) and (b) are the phase patterns with 2-mm-depth SA compensation and ignoring material influence. (c) and (d) are the simulated intensity, and the color map was normalized to the maximum of each image. (e) and (f) are the material modified regions observed from the side. Blue dash lines measure the axial lengths of material modified regions in the image with SA compensated, and orange dash lines measure the counterparts in control group. Axial length is labeled in the figure. For the spiral-structure experiment, (g) and (h) are the phase patterns, (i) is a sketch map of the spiral structure with the 3D scale labeled, and (j) and (k) are observed images with the axial length of material modified regions labeled. For the unequal-energy foci array experiment, (m) indicates the energy proportion of the foci corresponding to the observed images (n) and (p). For the two-layer structure experiment, observed images of SA compensated group (q) and control group (r) in two depth is shown. Red dash boxes and arrow in (n)-(r) indicate some marking failures caused by SA. The direction of light propagation is indicated in the figure.

bulk processing.

Finally, two 3D structures were produced in experiments to verify the complicated 3D foci matrix modulation capability. First structure using the CGH in Fig. 7(a), with a depth of 500 μm inside the material and a layer interval of 100 μm . Fig. 7(c1)-(c4). The four layers are marked with the letters “H”, “U”, “S”, and “T”, which are abbreviations for the affiliation of all authors. All foci are designed to be of equal energy. The other structure is a 9 \times 9 focal matrix tilting along the diagonal generated using the CGH in Fig. 7(b). The lateral spacing of all focal points is 9.5 μm , and the diagonal angle with the xoy plane is 66°. The slice images in Fig. 7(d1)-(d4) show that the modified zones are clear and have generally the same shape at different depths. Similar structures can be employed

in laser direct writing of binary optical elements and in high-dimensional information storage.

VII. CONCLUSION

In this article, we presented a versatile SACAD algorithm to modulate the light intensity distribution inside the material with built-in SA compensation. Simulation result shows that the SACAD algorithm surpasses the leading GSW algorithm in terms of diffraction efficiency ($e > 90\%$) and fidelity ($RMSE < 0.01$), and only needs a tenth of optimization time compared with GSW. In the experiment, complicated 3D structures inside the material with comprehensive SA com-

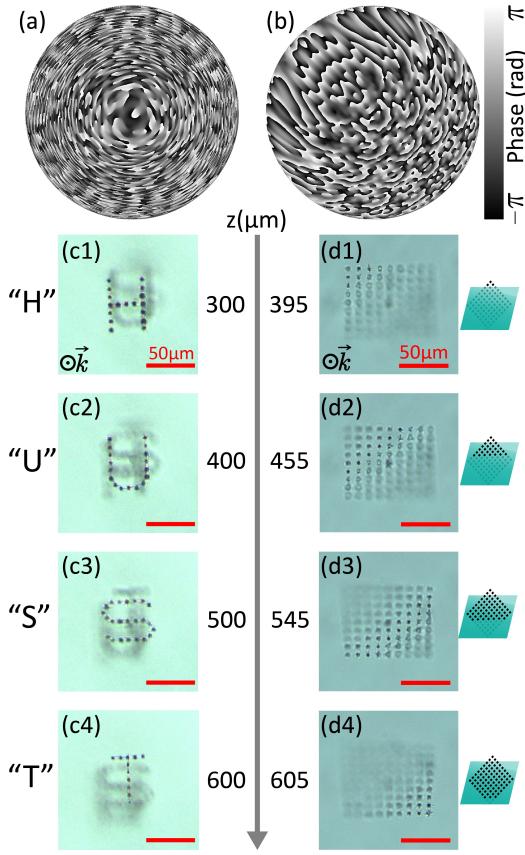


FIG. 7. Experiment results of 3D structures internal marking. (a) and (b) are the phase pattern generated with the SACAD algorithm and the color map was normalized to $-\pi$ and π . (c1)-(c4) are the four slice images with marking information “H”, “U”, “S” and “T”. (d1)-(d4) are four slice images of a tilting 9x9 focal matrix. The depths corresponding to all the images are labeled beside them. The direction of light propagation is indicated in the figure.

pensation are marked with high quality. Meanwhile, higher fidelity and higher power focus enable relative low energy laser to realize defect-free marking.

The SACAD algorithm differs from all established internal beam modulation algorithms as it makes no distinction between the multifocal modulation phase and the SA compensation phase. Both of them, as a fusion, are generated in a holistic optimization process. We hope this algorithm can provide a different prospective for CGH design. Instead of paying attention to the CGH solver and the physical mechanism of SA, the user can now focus on designing the intensity distributions inside the material that truly meet their original needs. Given the performances of the algorithm, we can envision extending this work to more complex situations such as multiple polarization components incidence, multiple wavelengths incidence, multiple phase modulators and even volumetric DOEs⁴³.

SUPPLEMENTARY MATERIAL

See supplementary material for the derivation of Eq. 1 and Eq. 2, the brief introduction of automatic differentiation, the usage and pseudocode of inverse CZT and the psuedocodes for 3 existing algorithms mentioned in the article.

ACKNOWLEDGMENTS

H.L. acknowledges the support from Key Research and Development Project of Hubei Province, China (2020AAA003). H.L. also wishes to acknowledge the support of Zengzhong Li, Yinghong Cheng and HGtech CO. LTD. for kindly providing convenience in the experiment.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available within the article and its supplementary material, and part of the codes are openly available in GitHub⁴⁰.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCE

- ¹Ke Sun, Dezhi Tan, Xinyuan Fang, Xintao Xia, Dajun Lin, Juan Song, Yonghong Lin, Zhaojun Liu, Min Gu, Yuanzheng Yue, and Jianrong Qiu. Three-dimensional direct lithography of stable perovskite nanocrystals in glass. *Science*, 375(6578):307–310, January 2022.
- ²Jingyu Zhang, Mindaugas Gecevičius, Martynas Beresna, and Peter G. Kazansky. Seemingly Unlimited Lifetime Data Storage in Nanostructured Glass. *Physical Review Letters*, 112(3):033901, January 2014.
- ³Masaaki Sakakura, Tsutomu Sawano, Yasuhiko Shimotsuma, Kiyotaka Miura, and Kazuyuki Hirao. Improved phase hologram design for generating symmetric light spots and its application for laser writing of waveguides. *Optics Letters*, 36(7):1065, April 2011.
- ⁴Matthias Häfner, Christof Pruss, and Wolfgang Osten. Laser direct writing: Recent developments for the making of diffractive optics. *Optik & Photonik*, 6(4):40–43, December 2011.
- ⁵Satoshi Hasegawa, Haruyasu Ito, Haruyoshi Toyoda, and Yoshio Hayasaki. Massively parallel femtosecond laser processing. *Optics Express*, 24(16):18513, August 2016.
- ⁶Patrick S. Salter and Martin J. Booth. Adaptive optics in laser processing. *Light: Science & Applications*, 8(1):110, November 2019.
- ⁷Masahiro Yamaji, Hayato Kawashima, Jun’ichi Suzuki, and Shuhei Tanaka. Three dimensional micromachining inside a transparent material by single pulse femtosecond laser through a hologram. *Applied Physics Letters*, 93(4):041116, July 2008.
- ⁸Roberto Di Leonardo, Francesca Ianni, and Giancarlo Ruocco. Computer generation of optimal holograms for optical trap arrays. *Optics Express*, 15(4):1913, February 2007.
- ⁹Jiaming Xu, Hengyang Li, Huaiyi Zhang, Shuo Li, Wei Cheng, Xilin Yao, Gang Xu, Yu Xiao, Xiahui Tang, and Yingxiong Qin. A general phase-modulating solution for generating optical array with arbitrary dimensions

- via mapping space. *Applied Physics Letters*, 122(9):091102, February 2023.
- ¹⁰Bangshan Sun, Patrick S Salter, Clemens Roider, Alexander Jesacher, Johannes Strauss, Johannes Heberle, Michael Schmidt, and Martin J Booth. Four-dimensional light shaping: Manipulating ultrafast spatiotemporal foci in space and time. *Light: Science & Applications*, 7(1):17117–17117, August 2017.
- ¹¹Jingzhao Zhang, Nicolas Pégard, Jingshan Zhong, Hillel Adesnik, and Laura Waller. 3D computer-generated holography by non-convex optimization. *Optica*, 4(10):1306, October 2017.
- ¹²Hengyang Li, Jiaming Xu, Huaizhi Zhang, Wei Cheng, Gang Xu, Xiaohui Tang, and Yingxiong Qin. Coordinate-transformation iteration algorithm for phase modulation of arbitrary axial light distribution. *Applied Optics*, 61(25):7424, September 2022.
- ¹³Haoran Ren, Han Lin, Xiangping Li, and Min Gu. Three-dimensional parallel recording with a Debye diffraction-limited and aberration-free volumetric multifocal array. *Optics Letters*, 39(6):1621, March 2014.
- ¹⁴Huaizhi Zhang, Jiaming Xu, Hengyang Li, Gang Xu, Yu Xiao, Wei Cheng, Xiaohui Tang, and Yingxiong Qin. Modulation of high-quality internal multifoci based on modified three-dimensional Fourier transform. *Optics Letters*, 48(4):900, February 2023.
- ¹⁵Min Gu, Xiangping Li, and Yaoyu Cao. Optical storage arrays: A perspective for future big data storage. *Light: Science & Applications*, 3(5):e177–e177, May 2014.
- ¹⁶Chenchu Zhang, Yanlei Hu, Wenqiang Du, Peichao Wu, Shenglong Rao, Ze Cai, Zhaoxin Lao, Bing Xu, Jincheng Ni, Jiawen Li, Gang Zhao, Dong Wu, Jiaru Chu, and Koji Sugioka. Optimized holographic femtosecond laser patterning method towards rapid integration of high-quality functional devices in microchannels. *Scientific Reports*, 6(1):33281, September 2016.
- ¹⁷Bing Xu, Wen-Qiang Du, Jia-Wen Li, Yan-Lei Hu, Liang Yang, Chen-Chu Zhang, Guo-Qiang Li, Zhao-Xin Lao, Jin-Cheng Ni, Jia-Ru Chu, Dong Wu, Su-Ling Liu, and Koji Sugioka. High efficiency integration of three-dimensional functional microdevices inside a microfluidic chip by using femtosecond laser multifoci parallel microfabrication. *Scientific Reports*, 6(1):1989, January 2016.
- ¹⁸C. Hnatovsky, R. S. Taylor, E. Simova, V. R. Bhardwaj, D. M. Rayner, and P. B. Corkum. High-resolution study of photoinduced modification in fused silica produced by a tightly focused femtosecond laser beam in the presence of aberrations. *Journal of Applied Physics*, 98(1):013517, July 2005.
- ¹⁹N. Huot, R. Stoian, A. Mermilliod-Blondin, C. Mauclair, and E. Audouard. Analysis of the effects of spherical aberration on ultrafast laser-induced refractive index variation in glass. *Optics Express*, 15(19):12395, 2007.
- ²⁰P. S. Salter, M. Baum, I. Alexeev, M. Schmidt, and M. J. Booth. Exploring the depth range for three-dimensional laser machining with aberration correction. *Optics Express*, 22(15):17644, July 2014.
- ²¹Emil Wolf. Electromagnetic diffraction in optical systems - i. an integral representation of the image field. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 253:349 – 357, 1959.
- ²²Richards B. and Wolf E. Electromagnetic diffraction in optical systems, II. Structure of the image field in an aplanatic system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 253(1274):358–379, December 1959.
- ²³P. Török, P. Varga, Z. Laczik, and G. R. Booker. Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive indices: An integral representation. *Journal of the Optical Society of America A*, 12(2):325, February 1995.
- ²⁴Qingfeng Li, Maxime Chambonneau, Markus Blothe, Herbert Gross, and Stefan Nolte. Flexible, fast, and benchmarked vectorial model for focused laser beams. *Applied Optics*, 60(13):3954, May 2021.
- ²⁵Alexander Jesacher and Martin J. Booth. Parallel direct laser writing in three dimensions with spatially dependent aberration correction. *Optics Express*, 18(20):21090, September 2010.
- ²⁶M. A. A. Neil, R. Juskaitis, M. J. Booth, T. Wilson, T. Tanaka, and S. Kawata. Adaptive aberration correction in a two-photon microscope. *Journal of Microscopy*, 200(2):105–108, November 2000.
- ²⁷R. W. Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.
- ²⁸David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.
- ²⁹PyTorch Contributors. Autograd mechanics. <https://pytorch.org/docs/stable/notes/autograd.html>.
- ³⁰Saugat Kandel, S. Maddali, Marc Allain, Stephan O. Hruszkewycz, Chris Jacobsen, and Youssef S. G. Nashed. Using automatic differentiation as a general framework for ptychographic reconstruction. *Optics Express*, 27(13):18653, June 2019.
- ³¹Delong Yang, Shaohui Zhang, Chuanjian Zheng, Guocheng Zhou, Lei Cao, Yao Hu, and Qun Hao. Fourier ptychography multi-parameter neural network with composite physical priori optimization. *Biomedical Optics Express*, 13(5):2739, May 2022.
- ³²Dustin B. Moore and James R. Fienup. Ptychography for optical metrology with limited translation knowledge. *Applied Optics*, 55(17):4596, June 2016.
- ³³Lawrence R. Rabiner, Ronald W. Schafer, and Charles M. Rader. The chirp z-transform algorithm and its application. *Bell System Technical Journal*, 48(5):1249–1292, 1969.
- ³⁴Alden S. Jurling and James R. Fienup. Phase retrieval with unknown sampling factors via the two-dimensional chirp z-transform. *Journal of the Optical Society of America A*, 31(9):1904, September 2014.
- ³⁵Alden S. Jurling, Matthew D. Bergkötter, and James R. Fienup. Techniques for arbitrary sampling in two-dimensional Fourier transforms. *Journal of the Optical Society of America A*, 35(11):1784, November 2018.
- ³⁶Vladimir Sukhoy and Alexander Stoytchev. Numerical error analysis of the ICZT algorithm for chirp contours on the unit circle. *Scientific Reports*, 10(1):4852, March 2020.
- ³⁷Vladimir Sukhoy and Alexander Stoytchev. Generalizing the inverse FFT off the unit circle. *Scientific Reports*, 9(1):14443, October 2019.
- ³⁸Jorge Nocedal. Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, 35(151):773–782, 1980.
- ³⁹C. G. Broyden. A new double-rank minimization algorithm. *Not. Am. Math. Soc.* 16:670, 1969.
- ⁴⁰Hengyang Li. Example code for SACAD. <https://github.com/LiHengyang1/SACAD>.
- ⁴¹M. Reicherter, T. Haist, E. U. Wagemann, and H. J. Tiziani. Optical particle trapping with computer-generated holograms written on a liquid-crystal display. *Optics Letters*, 24(9):608, May 1999.
- ⁴²G Shabtay. Three-dimensional beam forming and Ewald's surfaces. *Optics Communications*, 226(1-6):33–37, October 2003.
- ⁴³Nicolas Barré and Alexander Jesacher. Inverse design of gradient-index volume multimode converters. *Optics Express*, 30(7):10573, March 2022.