

Rapid hologram generation through backward ray tracing and adaptive-resolution wavefront recording plane

Jianying Zhu(朱建英), Yong Bi(毕勇), Minyuan Sun(孙敏远), and Weinan Gao(高伟男)

Citation: Chin. Phys. B, 2024, 33 (**11**): 114204. DOI: 10.1088/1674-1056/ad7c2e
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Accelerated generation of holograms with ultra-low memory symmetrically high-compressed look-up table

Yan Yang(杨燕), Jianying Zhu(朱建英), Minyuan Sun(孙敏远), and Yong Bi(毕勇)
Chin. Phys. B, 2024, 33 (**4**): 044201. DOI: 10.1088/1674-1056/ad1e67

A crossed focused vortex beam with application to cold molecules

Meng Xia(夏梦), Yaling Yin(尹亚玲), Chunying Pei(裴春莹), Yuer Ye(叶玉儿), Ruoxi Gu(顾若溪), Kang Yan(严康), Di Wu(吴迪), Yong Xia(夏勇), and Jianping Yin(印建平)
Chin. Phys. B, 2021, 30 (**11**): 114202. DOI: 10.1088/1674-1056/abf915

Hybrid vector beams with non-uniform orbital angular momentum density induced by designed azimuthal polarization gradient

Lei Han(韩磊), Shuxia Qi(齐淑霞), Sheng Liu(刘圣), Peng Li(李鹏), Huachao Cheng(程华超), Jianlin Zhao(赵建林)
Chin. Phys. B, 2020, 29 (**9**): 094203. DOI: 10.1088/1674-1056/aba09d

(3+1)-dimensional localized self-accelerating Airy elegant Ince-Gaussian wave packets and their radiation forces in free space

Dongdong Li(李冬冬), Xi Peng(彭喜), Yulian Peng(彭玉莲), Liping Zhang(张丽萍), Xingyu Chen(陈星宇), Jingli Zhuang(庄经立), Fang Zhao(赵方), Xiangbo Yang(杨湘波), Dongmei Deng(邓冬梅)
Chin. Phys. B, 2017, 26 (**12**): 124202. DOI: 10.1088/1674-1056/26/12/124202

Phase-only stereoscopic hologram calculation based on Gerchberg-Saxton iterative algorithm

Xinyi Xia(夏心怡), Jun Xia(夏军)
Chin. Phys. B, 2016, 25 (**9**): 094204. DOI: 10.1088/1674-1056/25/9/094204

Rapid hologram generation through backward ray tracing and adaptive-resolution wavefront recording plane

Jianying Zhu(朱建英)[1,2], Yong Bi(毕勇)[1], Minyuan Sun(孙敏远)[1], and Weinan Gao(高伟男)[1,†]

[1] *Applied Laser Research Center, Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *University of Chinese Academy of Sciences, Beijing* 100049, *China*

(Received 13 July 2024; revised manuscript received 4 September 2024; accepted manuscript online 18 September 2024)

An advanced method for rapidly computing holograms of large three-dimensional (3D) objects combines backward ray tracing with adaptive resolution wavefront recording plane (WRP) and adaptive angular spectrum propagation. In the initial phase, a WRP with adjustable resolution and sampling interval based on the object's size is defined to capture detailed information from large 3D objects. The second phase employs an adaptive angular spectrum method (ASM) to efficiently compute the propagation from the large-sized WRP to the small-sized computer-generated hologram (CGH). The computation process is accelerated using CUDA and OptiX. Optical experiments confirm that the algorithm can generate high-quality holograms with shadow and occlusion effects at a resolution of 1024×1024 in 29 ms.

Keywords: computer-generated hologram (CGH), wavefront recording plane (WRP), backward ray tracing

PACS: 42.40.Jv, 42.40.–i **DOI:** 10.1088/1674-1056/ad7c2e

1. Introduction

Holographic technology can fully reproduce the light wave information of an object to satisfy all aspects of human stereoscopic vision, thus being regarded as a true 3D display technology. Computer-generated holograms (CGHs) are interference patterns calculated by simulating the definition, propagation, and diffraction of light waves from a virtual object using a computer. Unlike conventional holograms, CGHs do not require a recording medium or a complex process for coherent generation. However, they require a significant amount of computational time, which limits their application in real-time 3D display systems.

In order to address the significant computational cost, several acceleration algorithms have been proposed, mainly divided into point-source-based algorithms,[1–4] polygon-based algorithms,[5–8] and layer-based algorithms.[9–11] Additionally, deep learning is a major research direction in computational holography. Deep-learning-based CGH methods can offer high-speed, high-quality image reconstruction.[12–16] For instance, a recent study introduced an autoencoder-based neural network for phase-only hologram generation, achieving high-fidelity 4K resolution holograms in 0.15 seconds with fewer speckles compared to existing methods.[17] Point-source CGH is the simplest and most practical algorithm for computing holograms.[1–3] The current acceleration algorithms for point-source CGH mainly include the look-up table (LUT) method[18,19] and the WRP method.[4] The method that combines the LUT and the WRP achieves higher computational speed.[20] Moreover, the computational framework of point-source CGH is inherently compatible with parallel computing architectures, particularly those utilizing hardware acceleration techniques.[3,21] However, the standard 3D point cloud formulation does not support many basic effects such as occlusion and shading.[22] Moreover, ray tracing enables realistic reconstruction with shadows, highlights, and global illumination.[23–25] Consequently, several studies have explored point-source holography algorithms that integrate WRP with backward ray tracing.[26] To generate CGH for large objects, a recent study proposed a ray tracing CGH algorithm based on multiple off-axis WRPs.[27] This approach achieved a reconstruction speed of 9.8 FPS using 2×2 WRPs and shifting ASM propagation. However, its drawback is apparent: the computation time scales with the object's volume. Larger objects necessitate multiple WRPs, significantly increasing computational costs.

To address the aforementioned challenges, we propose a high-speed hologram computation method for large objects that combines backward ray tracing with adaptive resolution WRP and adaptive angular spectrum propagation. This method ensures computational speed is independent of the reconstructed image size while achieving high-quality holographic reconstructions with realistic lighting, shadows, and occlusion effects. By leveraging OptiX and CUDA for accelerated computation, we can generate CGHs with a resolution of 1024×1024 for large 3D objects in just 29 ms. A 25 FPS holographic reconstruction video provided in the appendix demonstrates the effectiveness of the algorithm.

†Corresponding author. E-mail: wngao@mail.ipc.ac.cn

© 2024 Chinese Physical Society and IOP Publishing Ltd

2. Proposed method

The algorithm operates in two steps: initially, backward ray tracing combined with an illumination model is used to obtain the bigger-sized WRP near the object. In the second step, a smaller-sized CGH is calculated using an adaptive angular spectrum diffraction method.

2.1. Wave-recording plane method for backward ray tracing

In ray tracing, each pixel on the CGH surface acts as a ray emitter that emits rays in fixed directions around it. When these rays intersect with the surface of an object, a new ray is generated at the intersection point based on the illumination model, reflection laws, and refraction laws. This process continues recursively until the ray either intersects with a light source or reaches the maximum recursion depth. The complex amplitude contribution of the ray to the emitting pixel is calculated based on the distance between the intersection point and the emitter.

$$U_{\mathrm{CGH}}(x,y)=\sum_{j}^{N}\frac{A_j}{r_j}\exp\left(\mathrm{i}\frac{2\pi}{\lambda}r_j\right), \tag{1}$$

$$r_j=\sqrt{(x-x_j)^2+(y-y_j)^2+z_j^2}. \tag{2}$$

Here , N represents the number of rays, j represents the sequence number of the ray, A_j represents the intersection amplitude calculated from the illumination model, and r_j represents the distance between the intersection point and the ray emitter. Figure 1 illustrates a comparison between conventional point-source ray tracing and backward ray tracing.

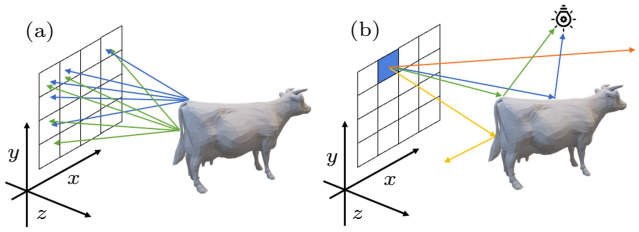


Fig. 1. Schematic diagram of (a) traditional point source method and (b) backward ray tracing method.

The complex amplitude of the WRP can be calculated using Eq. (1), replacing $U_{\mathrm{CGH}}(x,y)$ by $U_{\mathrm{WRP}}(x,y)$. A schematic diagram of conventional WRP and backward ray tracing based WRP is given in Fig. 2.

In the conventional WRP approach, the maximum diffraction angle θ for reconstructing a 3D object from the CGH can be expressed as

$$\sin\theta=\frac{\lambda}{2p}. \tag{3}$$

The radius W_j on the WRP is the following equation:

$$W_j=|z_j|\tan\left(\sin^{-1}\left(\frac{\lambda}{2p}\right)\right)\approx|z_j|\frac{\lambda}{2p}. \tag{4}$$

As a practical matter, θ can be smaller. As shown in Fig. 2,

$$\theta_o=\tan^{-1}\left(\frac{W}{d+z_j}\right)\approx\frac{W}{d+z_j}, \tag{5}$$

where W represents half of the CGH width. Meanwhile, equation (4) can be rewritten as

$$W_j=z_j\frac{W}{d+z_j}=\frac{W}{1+d/z_j}. \tag{6}$$

Since z_j is small compared to d, W_j is also small compared to W. Therefore, a lot of computation is reduced.

The number of rays emitted per pixel is determined by the angular range and angular interval, where the angle range is $[-\theta_o,\theta_o]$. In the horizontal direction,

$$N_{\mathrm{Horizontal}}=\frac{2\cdot\theta_o}{\Delta\theta}, \tag{7}$$

where $\Delta\theta=p/z_j$. p represents the object sampling interval. Based on the actual application scenario and the angular resolution of the human eye, we set p to 0.1 mm. The distribution of rays in the vertical direction is the same as in the horizontal.

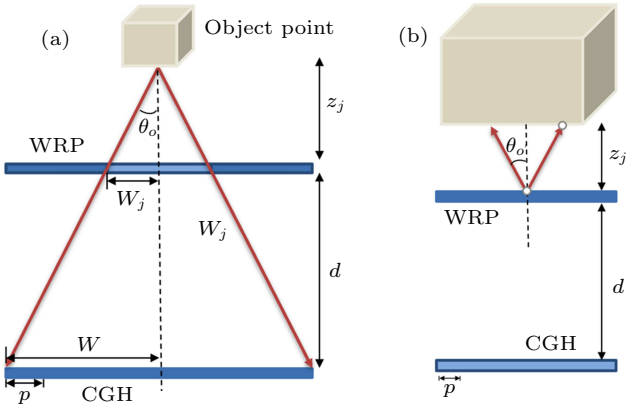


Fig. 2. Schematic comparisons of (a) the traditional point source method and (b) the backward ray tracing method.

2.2. Diffraction calculation from the large-sized WRP to the small-sized CGH.

The second step involves diffraction calculation from the large-sized WRP to the small-sized CGH. There are various methods for diffraction calculation between parallel planes with different sampling intervals. As early as 2012, Weng *et al.* utilized shifted Fresnel diffraction[28] to perform holographic calculations of the WRP method for objects exceeding the size of the CGH.[29] In addition, there are improved algorithms based on the angular spectrum propagation method, as described in scaled angular spectrum method[30] and adaptive-sampling angular spectrum method.[31] Compared to the Fresnel diffraction method, the improved angular spectrum method does not rely on the paraxial approximation and overcomes

the limitations of the angular spectrum method in far-field applications. It also offers more flexible sampling intervals and ranges. In subsequent calculations, we will build upon this approach.

In general ASM, using FFT results in circular convolution instead of linear convolution, which introduces errors. To mitigate these errors, zero-padding of the input optical field is typically required. From a physical perspective, as shown in Fig. 3, FFT extends the input optical field periodically, and this extended field can affect the output field. Therefore, zero-padding the input field is necessary to reduce interference. The required zero-padding range t is given by the following formula:

$$t = \frac{1}{2}S_2 - \frac{1}{2}S_1 + z \cdot \tan\theta, \tag{8}$$

where S_1 and S_2 represent the ranges of the source plane and the destination plane, respectively. Δ_1 and Δ_2 are the sampling intervals of the source plane and the destination plane, respectively. z is the diffraction distance. θ is the maximum diffraction angle, which can be obtained from Eq. (3).

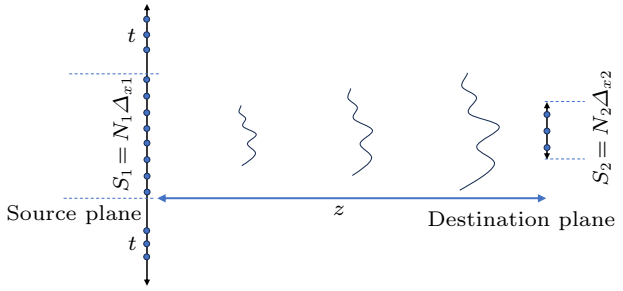


Fig. 3. Schematic diagram of diffraction propagation.

In the frequency domain, zero-padding effectively reduces the sampling interval. Thus, non-uniform fast fourier transform (NUFFT) can be used to directly achieve the corresponding smaller sampling interval without the need for actual zero-padding, thereby reducing the number of samples required for the transformation and increasing computational speed.

Based on the required zero-padding range t, the sampling interval of the transfer function can be obtained

$$\Delta_\mathrm{f} = \frac{1}{S_1 + t}. \tag{9}$$

Additionally, as shown in Fig. 3, the maximum sampling frequency is given by

$$f_\mathrm{max} = \frac{L}{\lambda\sqrt{L^2 + z^2}}, \tag{10}$$

where $L = (S_1 + S_2)/2$. In addition, according to the Nyquist sampling theorem, $f_\mathrm{max} < 1/(2\Delta_1)$. Therefore, f_max should be taken as the smaller value of the two. The number of samples M is given by $M = \lceil 2f_\mathrm{max}/\Delta_\mathrm{f} \rceil$. Here, $\lceil\cdot\rceil$ represents the operation of rounding up to the nearest integer.

After obtaining the required number of samples, the diffraction optical field can be calculated using NUFFT and INUFFT transformations. The third type of NUFFT, accelerated by the fast Gaussian gridding method proposed by Greengard and Lee,[32] was used,

$$U_1(m\Delta_\mathrm{f}) = \mathrm{NUFFT}\{u_1(n_1\Delta_1)\}, \tag{11}$$

$$u_2(n_2\Delta_2) = \mathrm{INUFFT}\{U_1(m\Delta_\mathrm{f})H(m\Delta_\mathrm{f})\}. \tag{12}$$

Using the above formulas, the diffraction calculation from the large-sized WRP to the small-sized CGH can be achieved. Compared to the traditional FFT-based diffraction methods, the use of NUFFT inevitably increases the computational complexity. However, NUFFT can adaptively adjust sampling parameters, thereby avoiding zero-padding, which reduces computational complexity in another aspect. Overall, for diffraction calculations involving objects of a size close to that of the CGH, the computation times of both methods are comparable, as evidenced in several previous NUFFT-based diffraction algorithms.[31,33] When dealing with larger objects, NUFFT-based algorithms allow for flexible adjustment of sampling parameters, resulting in a computational time reduction proportional to the ratio of the sampling intervals between the object plane and the target plane.

3. Experiment

To demonstrate the efficiency of the proposed algorithm, we compared its computational efficiency with the MO-WRP method proposed in Ref. [27], which also utilizes backward ray tracing and is suitable for generating CGHs of large objects. The computational platform for both algorithms consisted of a computer running Microsoft Windows 10, equipped with an AMD Ryzen 5 3600 6-core processor (3.60 GHz), 32 GB of memory, and Microsoft Visual C++ 2017 (Intel C++ Compiler version 12.1). Additionally, we used an Nvidia GeForce RTX 2080 Ti GPU board, OptiX 6.5, and CUDA 10.1 for GPU programming. Table 1 shows the relevant CGH computation parameters.

Table 1. CGH computation parameters.

Parameters	Values
CGH and WRP resolution	1024×1024
CGH pixel size	3.74 μm
WRP pixel size	$N \cdot 3.74$ μm
CGH size	3.83 mm×3.83 mm
WRP size	$(3.83 \cdot N$ mm)×$(3.83 \cdot N$ mm)
Wavelength	532 nm
Distance between WRP and model	0.05 m
Distance between WRP and CGH	0.45 m

In the experiments, we set the area of the reconstructed image to be 1, 4, and 9 times the area of CGH (corresponding to $N = 1, 2, 3$, respectively). We selected a teapot model (41472 points, 16188 polygons) as the object to compare the computation times of the two methods. The model is shown in Fig. 4. The model size was adaptively adjusted to match the dimensions of the selected WRP. In the MO-WRP algorithm, the corresponding number of WRPs was 1×1, 2×2, and 3×3, respectively.



Fig. 4. The teapot model used in the calculations.

The calculation time of these algorithms running on a GPU is shown in Table 2 and Fig. 5.

Table 2. Computation time of the two algorithms (in units of ms).

Size of the reconstructed image	MO-WRP	Proposed method
3.83 mm×3.83 mm	21	22
7.66 mm×7.66 mm	74	26
11.49 mm×11.49 mm	206	29

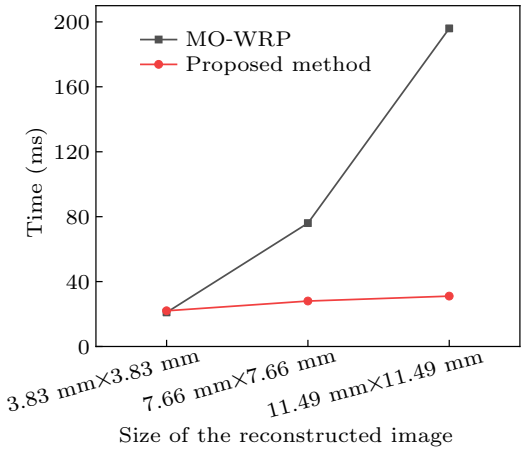


Fig. 5. Comparisons of reconstruction time between the two methods.

As can be seen, the computation time of the MO_WRP algorithm increases with the size of the reconstructed image, while the proposed algorithm is almost unaffected by this aspect.

Figure 6 illustrates the optical setup for holographic reconstruction. A 532-nm green laser is collimated by a beam expander (Thorlabs GBE20-A) and directed onto the SLM (Holoeye GAEA-2) loaded with the CGH. After passing through a 4f system to filter out the zero-order light, a lensless camera (Nikon D90) is placed on the reconstruction plane to record the reconstructed images. Figure 7 shows optical reconstruction images from different angles. The highlights, shadows, and occlusion effects on the teapot are clearly visible, demonstrating the effectiveness of the algorithm. Additionally, a dynamic reconstruction video of the teapot model at 25 FPS is provided in supporting information (see visualization 1), further confirming the high-speed performance of the algorithm.

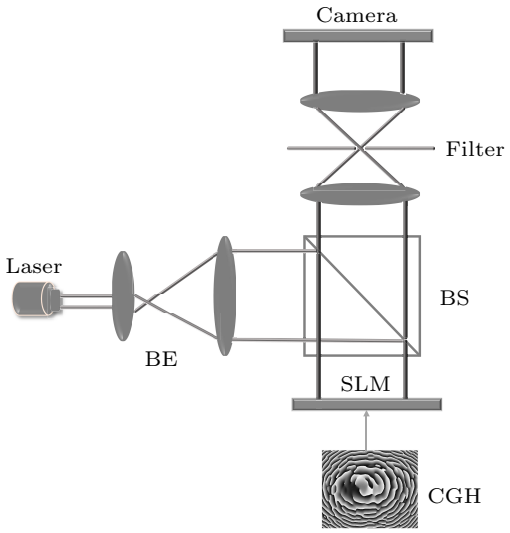


Fig. 6. Schematic diagram of the optical experimental setup.

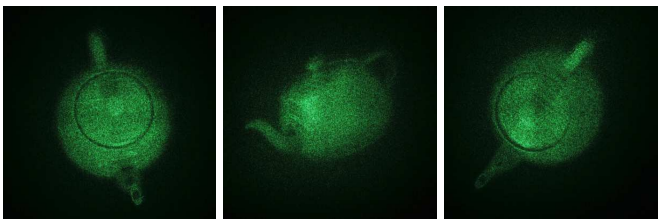


Fig. 7. The optical reconstruction images from different angles.

4. Conclusion

This study proposes an advanced method combining backward ray tracing and adaptive resolution WRP for rapidly computing large-sized objects. Optical experiments confirm that the algorithm can generate high-quality holograms with shadows and occlusion effects at a resolution of 1024×1024 within 29 ms. This algorithm holds promising potential for future applications in dynamic holography.

Acknowledgement

Project supported by the Special Project of Central Government Guiding Local Science and Technology Development in Beijing 2020 (Grant No. Z201100004320006).

References

[1] Lucente M E 1993 *J. Electron. Imaging* **2** 28

[2] Kim S C and Kim E S 2008 *Appl. Opt.* **47** D55
[3] Pan Y, Xu X, Solanki S, Liang X, Tanjung R B A, Tan C and Chong T C 2009 *Opt. Express* **17** 18543
[4] Shimobaba T, Masuda N and Ito T 2009 *Opt. Lett.* **34** 3133
[5] Li X, Liu J, Pan Y and Wang Y 2017 *Opt. Commun.* **390** 22
[6] Im D, Cho J, Hahn J, Lee B and Kim H 2015 *Opt. Express* **23** 2863
[7] Matsushima K and Nakahara S 2009 *Appl. Opt.* **48** H54
[8] Wang F, Shimobaba T, Zhang Y, Kakue T and Ito T 2021 *Opt. Express* **29** 35442
[9] Chen J S and Chu D 2015 *Opt. Express* **23** 18143
[10] Zhao Y, Cao L, Zhang H, Kong D and Jin G 2015 *Opt. Express* **23** 25440
[11] Zhang H, Cao L and Jin G 2017 *Appl. Opt.* **56** F138
[12] Horisaki R, Takagi R and Tanida J 2018 *Appl. Opt.* **57** 3859
[13] Shi L, Li B, Kim C, Kellnhofer P and Matusik W 2021 *Nature* **591** 234
[14] Liu S C and Chu D 2021 *Opt. Express* **29** 27373
[15] Shi L, Li B and Matusik W 2022 *Light: Sci. Appl.* **11** 247
[16] Liu K, Wu J, He Z and Cao L 2023 *Opto-Electron. Adv.* **6** 220135
[17] Wu J, Liu K, Sui X and Cao L 2021 *Opt. Lett.* **46** 2908
[18] Jia J, Wang Y, Liu J, Li X, Pan Y, Sun Z, Zhang B, Zhao Q and Jiang W 2013 *Appl. Opt.* **52** 1404
[19] Yang Y, Zhu J, Sun M and Bi Y 2024 *Chin. Phys. B* **33** 044201
[20] Pi D, Liu J, Han Y, Yu S and Xiang N 2020 *Opt. Express* **28** 9833
[21] Ikawa S, Takada N, Araki H, Niwase H, Sannomiya H, Nakayama H, Oikawa M, Mori Y, Kakue T, Shimobaba T and Ito T 2020 *Chin. Opt. Lett.* **18** 010901
[22] Blinder D, Chlipala M, Kozacki T and Schelkens P 2021 *Opt. Lett.* **46** 2188
[23] Wang Y, Chen Z, Sang X, Li H and Zhao L 2018 *Opt. Commun.* **410** 768
[24] Liu S, Wei H, Li N, Liu Z and Zhang J 2019 *Opt. Commun.* **443** 76
[25] Ichikawa T, Yamaguchi K and Sakamoto Y 2013 *Appl. Opt.* **52** A201
[26] Wang Y, Sang X, Chen Z, Li H and Zhao L 2018 *Opt. Commun.* **429** 12
[27] Sun M Y, Yuan Y, Bi Y, Zhang S, Zhu J Y and Zhang W P 2020 *Opt. Express* **28** 34994
[28] Muffoletto R P, Tyler J M and Tohline J E 2007 *Opt. Express* **15** 5631
[29] Weng J, Shimobaba T, Okada N, Nakayama H, Oikawa M, Masuda N and Ito T 2012 *Opt. Express* **20** 4018
[30] Shimobaba T, Matsushima K, Kakue T, Masuda N and Ito T 2012 *Opt. Lett.* **37** 4128
[31] Zhang W, Zhang H and Jin G 2020 *Opt. Lett.* **45** 4416
[32] Lee J Y and Greengard L 2005 *J. Comput. Phys.* **206** 1
[33] Zhang W, Zhang H and Jin G 2020 *Opt. Lett.* **45** 1543