Journal of
Applied Physics

Theoretical and practical aspects of the design and production of synthetic holograms for transmission electron microscopy F

Cite as: J. Appl. Phys. **131**, 031101 (2022); https://doi.org/10.1063/5.0067528
Submitted: 17 August 2021 • Accepted: 21 December 2021 • Published Online: 18 January 2022

Paolo Rosi, Federico Venturi, Giacomo Medici, et al.

COLLECTIONS

F This paper was selected as Featured

openaccess

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Surface phonon polaritons for infrared optoelectronics
Journal of Applied Physics **131**, 030901 (2022); https://doi.org/10.1063/5.0064234

Growth of bulk β-Ga_2O_3 single crystals by the Czochralski method
Journal of Applied Physics **131**, 031103 (2022); https://doi.org/10.1063/5.0076962

Microsphere-assisted microscopy
Journal of Applied Physics **131**, 031102 (2022); https://doi.org/10.1063/5.0068263

AIP
Publishing

Theoretical and practical aspects of the design and production of synthetic holograms for transmission electron microscopy

Cite as: J. Appl. Phys. **131**, 031101 (2022); doi: 10.1063/5.0067528
Submitted: 17 August 2021 · Accepted: 21 December 2021 ·
Published Online: 18 January 2022

View Online Export Citation CrossMark

Paolo Rosi,[1] Federico Venturi,[1,2] Giacomo Medici,[1] Claudia Menozzi,[1] Gian Carlo Gazzadi,[3] Enzo Rotunno,[3] Stefano Frabboni,[1,3] Roberto Balboni,[4] Mohammadreza Rezaee,[5] Amir H. Tavabi,[6] Rafal E. Dunin-Borkowski,[6] Ebrahim Karimi,[5] and Vincenzo Grillo[3,a)]

AFFILIATIONS

[1] FIM Department, University of Modena and Reggio Emilia, 41125 Modena, Italy
[2] Faculty of Engineering, University of Nottingham, Nottingham NG7 2RD, United Kingdom
[3] CNR-Nanoscience Institute, S3 Center, 41125 Modena, Italy
[4] CNR-Institute for Microelectronics and Microsystems, 40129 Bologna, Italy
[5] Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada
[6] Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons and Peter Grünberg Institute, Forschungszentrum Jülich, 52425 Jülich, Germany

[a)] **Author to whom correspondence should be addressed:** vincenzo.grillo@nano.cnr.it

ABSTRACT

Beam shaping—the ability to engineer the phase and the amplitude of massive and massless particles—has long interested scientists working on communication, imaging, and the foundations of quantum mechanics. In light optics, the shaping of electromagnetic waves (photons) can be achieved using techniques that include, but are not limited to, direct manipulation of the beam source (as in x-ray free electron lasers and synchrotrons), deformable mirrors, spatial light modulators, mode converters, and holograms. The recent introduction of holographic masks for electrons provides new possibilities for electron beam shaping. Their fabrication has been made possible by advances in micrometric and nanometric device production using lithography and focused on ion beam patterning. This article provides a tutorial on the generation, production, and analysis of synthetic holograms for transmission electron microscopy. It begins with an introduction to synthetic holograms, outlining why they are useful for beam shaping to study material properties. It then focuses on the fabrication of the required devices from theoretical and experimental perspectives, with examples taken from both simulations and experimental results. Applications of synthetic electron holograms as aberration correctors, electron vortex generators, and spatial mode sorters are then presented.

I. INTRODUCTION

The transmission electron microscope was developed primarily to study matter at the highest spatial resolution. However, over time, the quantum wave nature of electrons has attracted increasing interest for both fundamental reasons and applications. The wave nature of electrons provides analogies with light optics. For non-relativistic and monochromatic electrons, the Helmholtz equation can be used to describe both electrons and photons. Concepts such as the refractive index and lenses can also be considered in both contexts with similar results. For electrons, electrostatic and magnetostatic potentials result in the retardation (or anticipation) of an electron wave. Analogies between the two fields have been used widely in electron microscopy. However, light optics has provided a broad range of applications beyond imaging, with recent progress (e.g., superoscillation microscopy) triggered by the concept of structured light waves, whereby a wave front and its spatial intensity distribution can be controlled in a manner that goes beyond the use of conventional optical elements. Recently, the concept of structured waves has been extended to matter waves, primarily to electrons. Structured electron waves include electron beams with helical wave fronts (i.e., electron vortex beams), self-accelerating beams, and non-diffracting beams, as well as orbital angular momentum analyzers. One can also fabricate conventional electron optical devices such as lenses, diffractive elements, and aberration correctors using a holographic approach. The key technology for electrons is the use of synthetic holograms to modulate the phase and amplitude of the electron wave. The word "hologram" comes from the Greek term for "whole writing." The ability to write both the intensity and the phase of an electron wave is achieved by the creation of an interference pattern, which is related to the relative phases of two waves.

Even though Gabor's original concept of holography was intended for electron optics,[1] holograms have seen wider applications in light optics, becoming a ubiquitous concept (e.g., on Canadian dollar notes). In electron microscopy, holography normally refers to the recording of the interference of a wave perturbed by a semi-transparent specimen and a reference wave. In the present context, the recreation of a perturbed wave from a calculated interference pattern is of primary interest. For the sake of clarity, this approach is referred to as "*synthetic holography*," while a calculated pattern is referred to as a "*computer-generated hologram*." The two operations are inverse; when a fabricated interference pattern is illuminated, an electron beam that has the phase and amplitude of the original semi-transparent specimen is generated. A similar approach was historically implemented when a lack of computing resources meant that researchers could not apply numerical Fourier transforms and had to illuminate the recorded electron interference patterns with lasers to recreate images of objects.

In order to construct a synthetic hologram, one needs to scale down the equivalent of a transparent electron micrograph to the electron's scale. Unfortunately, there is no electron optical analog of a transparent object (i.e., an object that applies negligible intensity reduction to the wavefunction). The best approximation is given by the thin layer of a material with low electron absorption, such as carbon or silicon nitride (Si_3N_4). Si_3N_4 can be produced routinely in the form of membranes that can be inserted along the electron path. Modern days nanofabrication techniques, such as Focused Ion Beam (FIB) milling and Electron Beam Lithography (EBL), allow to imprint thickness modulations on a membrane with lateral and depth scales of tens of nm. Such scales are the common ones needed to fabricate a synthetic hologram whose typical total dimensions are in the range of a few micrometers with details down to tens of nm. Therefore, the tools that are needed to modulate both the phase and the amplitude of an electron wave are available. Historically, developments have proceeded from rough amplitude modulations of electron waves to today's fine and precise control over amplitude and phase modulations in the form of complex patterns.

This Tutorial provides an overview of the theoretical and numerical calculation, fabrication, and analysis of synthetic electron holograms. The first chapter will focus on giving to the reader the required theoretical knowledge starting from the concept of holography, focusing particularly on off-axis holography, passing from how computer-generated became an import tool for scientists, and at last, how different types of synthetic holograms can be designed. The second chapter focuses on the two mainly used fabrication techniques in the production of synthetic holograms, also providing details on the calibration process and possible optimizations schemes. Last, in the third chapter, we show a series of examples of the possible uses of synthetic holograms.

II. SYNTHETIC HOLOGRAM FORMATION: FROM CALCULATIONS TO COMPUTER-GENERATED HOLOGRAMS

A. Theory of hologram formation

We begin by describing interference between a generic perturbed wave and a "known" reference wave to form a hologram. This approach allows us to describe both "imaging" holography and synthetic holography as a general theoretical framework. From a physical point of view, a hologram is generated by interference between a reference wave function $\Psi_{ref}(\vec{r})$ and the wave function of interest

$$\Psi_I(\vec{r}) = A_I(\vec{r})e^{i\varphi_I(\vec{r})}, \tag{1}$$

where $A_I(\vec{r})$ and $\varphi_I(\vec{r})$ are the phase and amplitude of the wave function of interest, respectively. Holography involves writing, in two dimensions, of an interference pattern between waves propagating in three-dimensional space.

If we now consider a specific plane with coordinates $\vec{\rho} = (x, y)$ and an out-of-plane direction z, the wave function in three dimensions is

$$\Psi_{holo}(\vec{r}) = \Psi_I(\vec{r}) + \Psi_{ref}(\vec{r}), \tag{2}$$

whereas in a specific plane it is

$$\Psi_{holo}(\vec{\rho}) = \Psi_I(\vec{\rho}) + \Psi_{ref}(\vec{\rho}), \tag{3}$$

with corresponding intensity

$$\begin{aligned} I_{holo}(\vec{\rho}) &= |\Psi(\vec{\rho})|^2 \\ &= |\Psi_I(\vec{\rho})|^2 + |\Psi_{ref}(\vec{\rho})|^2 + 2Re[\Psi_I(\vec{\rho})\Psi^*_{ref}(\vec{\rho})]. \end{aligned} \tag{4}$$

Alternatively,

$$\begin{aligned} I_{holo}(\vec{\rho}) = |\Psi_I(\vec{\rho})|^2 + |\Psi_{ref}(\vec{\rho})|^2 \\ + 2|\Psi_I(\vec{\rho})||\Psi_{ref}(\vec{\rho})|\cos(\varphi_I(\vec{r}) - \varphi_{ref}(\vec{r})), \end{aligned} \tag{5}$$

where $\varphi_{ref}(\vec{r})$ is the phase of the reference beam. The use of a reference wave allows the phase $\varphi_I(\vec{r})$ to be made visible as an intensity modulation. The reference wave should have a known form, such as a plane wave or spherical wave (sometimes substituted by a parabolic approximation). The process is referred to as "inline" or "on-axis" holography if the waves propagate in the same direction and as "off-axis" holography if the waves propagate in different directions. An in-depth comparison between the two schemes, both theoretical and experimental, can be found in the papers by Koch and Lubk[2] and by Latychevskaia *et al.*[3] It is worth mentioning that the on-axis and off-axis schemes are also possible for synthetic holography. Here, for an *in-line* synthetic hologram, the beam of interest is generated on the optic axis at different z values (i.e., at different defocuses). On the contrary, *off-axis* synthetic holograms are realized by using an inclined plane wave as a reference wave (details on how to do this are provided in the following paragraphs), just as for an *off-axis* hologram, and the desired wave function is generated in the Fraunhofer plane of the hologram on one of the diffraction orders.

Throughout this paper, most of the discussion will refer to off-axis synthetic holograms and holography, if not specified otherwise.

B. "Image" holography for object phase reconstruction

Imaging holography is the basis of synthetic holography. If one considers $\Psi_I(\vec{r})$ as a wave function obtained after passing a partially electron-transparent sample with an unknown phase distribution, then holography can be used to extract this phase information.

Off-axis holography is performed by splitting a wavefront into two parts, typically using a biprism. In electron microscopy, a biprism normally takes the form of a metal or metal-coated wire that has a voltage applied to it, with one part of the beam traveling through a region of interest on a specimen. The relative tilt of the two parts of the electron wave introduced by the biprism allows them to interfere with one another. The object wave interacts with the sample and gains a phase that depends on the physical features of the sample. The intensity of the resulting hologram between the generic beam of interest in Eq. (1) and a tilted reference plane wave (with $\Psi_{ref}(\vec{\rho}) = e^{i\vec{g}\cdot\vec{\rho}}$) is described by the following expression:

$$\begin{aligned} I_{holo}(\vec{r}) &= \Psi^2_{holo}(\vec{\rho}) = |\Psi_I(\vec{\rho}) + \Psi_{ref}(\vec{\rho})|^2 \\ &= 1 + A_I^2(\vec{\rho}) + 2A_I(\vec{\rho})\cos(\varphi_I(\vec{\rho}) + \vec{g}\cdot\vec{\rho}), \end{aligned} \tag{6}$$

where $\vec{g}$ is the in-plane component of the wave vector of the plane wave and is determined by the tilt angle introduced by the biprism. Three contributions to the intensity can be distinguished: the reference image intensity, the specimen image intensity, and a set of cosinusoidal fringes, whose local phase shift and amplitude are given by the phase and amplitude of the electron wave function in the image plane. The phase and amplitude of the wave function of interest can be extracted from the hologram by applying a Fourier Transform (FT) and reconstructing the complex wave function by means of an inverse Fourier Transform (IFT). If required, 2π phase discontinuities can be removed. The FT of Eq. (3) can be written in the form

$$\begin{aligned} FT[I_{holo}(\vec{r})] = \delta(\vec{k}_\rho) + FT[A_I^2(\vec{r})] + \delta(\vec{k}_\rho + \vec{g}) \otimes FT\left[A_I(\vec{r})e^{i\varphi_I(\vec{r})}\right] \\ + \delta(\vec{k}_\rho - \vec{g}) \otimes FT\left[A_I(\vec{r})e^{-i\varphi_I(\vec{r})}\right], \end{aligned} \tag{7}$$

where $\delta(.)$ is the Dirac delta function and $f_1 \otimes f_2$ represents the convolution of f_1 and f_2, k_ρ is the in-plane component of the wave vector k, with $k^2 = k_\rho^2 + k_z^2 = \frac{2m\omega}{\hbar} = \left(\frac{2\pi}{\lambda_{dB}}\right)^2$, where m is the electron mass, $\hbar$ is the reduced Planck constant and λ_{dB} is the electron's de Broglie wavelength. In this expression, the first two terms are the FTs of the reference and sample wave function, respectively, located at $\vec{k}_\rho = 0$. The last two terms are peaked at $\vec{k}_\rho = \pm\vec{g}$, correspond to the FTs of the desired image wave function and its complex conjugate and are known as *sidebands,* while those centered on the origin are referred to as a *center band*. The larger the value of $\vec{k}_\rho$, i.e., the larger the tilt of the reference wave, the further from the origin are the sidebands. The sidebands contain both amplitude and phase information about the wave function of interest. In order to recover the complex wave function, one of the sidebands is selected and isolated by applying a mask, shifted to the origin of reciprocal space and inverse Fourier transformed. The most commonly used mask is a circular one (to have the same resolution in the reconstructed phase image along all directions) that has soft edges and a radius that is no larger than one-third of the distance between the sideband and the origin (as the radius of the center band is twice that of the sideband).[4,5] Nonetheless, the circular mask radius may be larger than 1/3 of $|g|$ and reconstructions with hard masks (that are not necessarily circular masks) are also used. The choice of the mask depends on the support/band limitations of the central/side bands.[6] The phase image may need to be "unwrapped" to remove 2π phase discontinuities, which appear at positions where the phase shift exceeds 2π, as IFT operations are calculated modulo 2π.[7] The number of phase wraps can sometimes be reduced by removing a constant phase gradient by repositioning the center of the sideband.

C. Synthetic hologram generation

Since the 1960s, as a result of advances in computational power, so-called *computer-generated holograms* (CGHs) have been introduced.[8,9] As Lesem *et al.*[10] stated in 1968, when referring to holograms for 3-D displays:

> "A properly illuminated hologram forms for the viewer a picture which is identical with that which he would observe if he were looking at the scene himself. A computer generated hologram yields such a 3-D picture, without the original scene ever having to exist."

Based on this simple explanation, it is possible to understand how CGHs allow desired patterns to be designed and tested without the need to create models for each iteration, reducing the time required to make a synthetic CGH (S-CGH). The term "synthetic" is used to underline the fact that the last step involves producing a hologram that will be inserted into a microscope or an optical bench. Figure 1 shows a representation of the two concepts (or modes) of holography: "conventional" image holography [Figs. 1(a) and 1(b)] and "synthetic" holography [Fig. 1(c)], where here by illuminating an off-axis S-CGH with a plane wave, it is

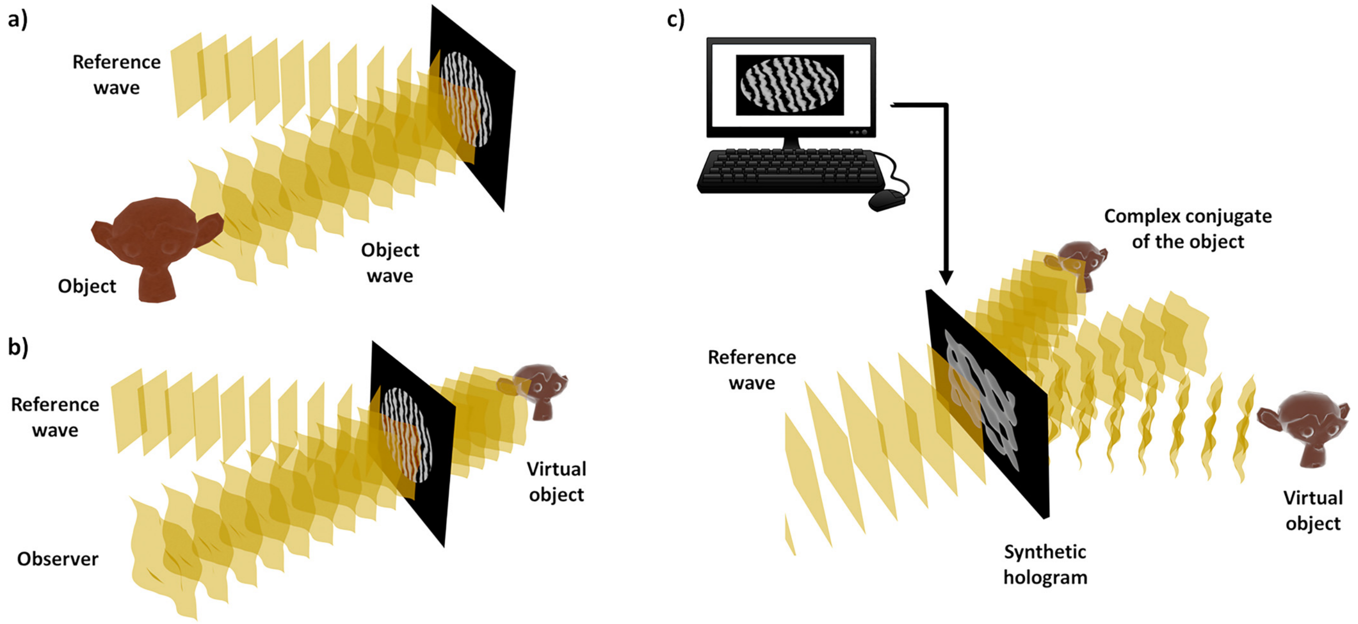

FIG. 1. Schematic diagram of (a) traditional “image” holography, (b) image reconstruction where an observer can reconstruct the image of an object by shining the hologram recorded in (a) with the same reference wave, while (c) schematically represents how “synthetic” off-axis holography is carried out: a hologram is generated using a computer and by illuminating it with a plane wave, it is possible to observe the object of interest (or desired wave function) in the Fraunhofer plane.

possible to observe the desired wave function in the Fraunhofer plane on one of the diffraction orders.

Generation of the desired object and reference wave functions, as well as the interferometric process, can be carried out computationally. Most of the computational steps described in this article have been carried out using a modified version of *Stem_Cell* software.[11] In this software, the interferometric process, which consists of overlapping wave functions in a given plane, is carried out computationally. A CGH generated by such a set of operations (i.e., the interference intensity pattern) is exported to a file, which can be used to fabricate an S-CGH.

The fabrication process requires modern state-of-the-art machines and well-developed processes, which are described in Secs. II D and II E. This is because the typical dimension of an S-CGH ranges from a few μm to hundreds of μm, while the smallest features can be only a few tens of nm in size. An S-CGH shows the desired function when it is illuminated by an incident (reference) beam. One of the typical optical setups used to test an off-axis S-CGH is shown in Fig. 2(a), where the TEM is used in Low Mag mode. In fact, due to the dimensions of the S-CGH, the illumination needs to be widespread and as paraxial as possible. This is usually achieved by switching off the condenser-objective lens (the main imaging lens). Moreover, as a periodicity of 100 nm (typical for most S-CGHs) corresponds to a hscattering angle of only 20 μrad, the required very long focal length is usually not accessible when the main imaging lens (the objective lens) is switched on. In low-angle diffraction (LAD) mode, an off-axis S-CGH acts as a *transmission diffraction grating*, with each of the diffracted beams centered on a different position in the diffraction plane.

A manufactured off-axis S-CGH can be mounted in one of the condenser aperture planes of an electron microscope to generate the desired wave function at the sample plane [as shown in Fig. 2(b), in the case of an off-axis Fraunhofer S-CGH].

A simple recipe for producing an off-axis S-CGH is to take the formula for an “image hologram” and to invert it. By illuminating a material that introduces the same amplitude (or intensity) modulation as in Eq. (6), it is possible to obtain an object from which one of the diffracted beams corresponds to the wave that passed through the sample.

Depending on the type of interaction with the beam, synthetic holograms can be divided into three primary categories: (i) amplitude holograms; (ii) phase holograms; and (iii) mixed (amplitude-phase) holograms. In this way, they can encode (i) only the phase or (ii) both the phase and the amplitude of a wave function of interest. The character of the hologram (phase, amplitude, or both) and the type of encoding are independent. For example, a phase hologram can encode both the amplitude and the phase of a wave function, but with some restriction on efficiency. With respect to an incident plane wave, the transmittance function of an (amplitude-phase) S-CGH can be written in the form

$$T_H(\vec{\rho}) = A(\vec{\rho})e^{i\Delta\varphi(\vec{\rho})}, \tag{8}$$

where $\vec{\rho}$ is the transverse spatial coordinate with respect to the propagation direction of the beam, while $A(\vec{\rho})$ and $\Delta\varphi(\vec{\rho})$ are

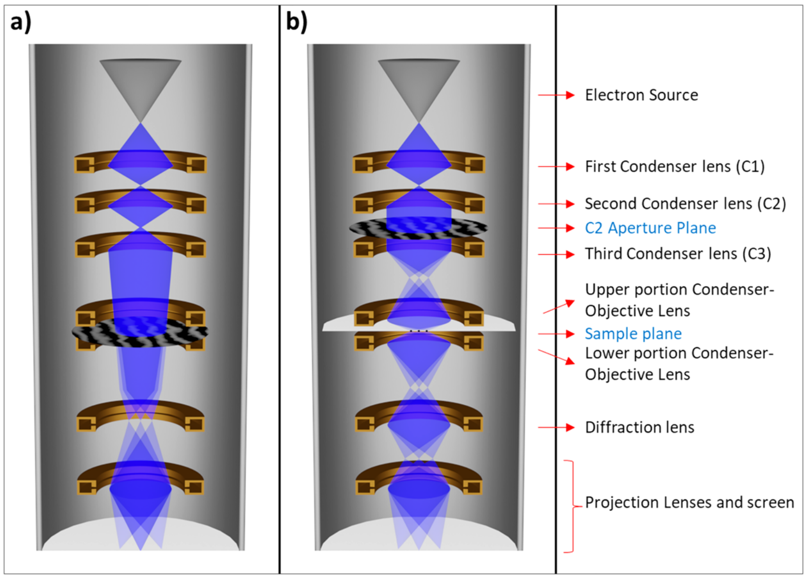
a)
b)
Electron Source
First Condenser lens (C1)
Second Condenser lens (C2)
C2 Aperture Plane
Third Condenser lens (C3)
Upper portion Condenser-Objective Lens
Sample plane
Lower portion Condenser-Objective Lens
Diffraction lens
Projection Lenses and screen

FIG. 2. Rendered images of (a) the Low-Mag TEM testing configuration where the S-CGH is inserted in the sample plane, while (b) shows a typical working condition where the S-CGH is inserted in the second condenser aperture plane. The condenser system here reported comprises three condenser lenses, as in most 300 kV TEMs, while the S-CGH in consideration is a Fraunhofer S-CGH.

amplitude and phase modulations. An amplitude hologram modifies the amplitude of the incident wave $A(\vec{\rho})$ and keeps the phase unchanged, such that $\Delta\varphi(\vec{\rho}) = \text{constant}$;[9] a phase hologram modifies the phase of the incident wave by modulating $\Delta\varphi(\vec{\rho})$, such that $A(\vec{\rho}) = \text{constant}$;[12] a mixed hologram modifies both $A(\vec{\rho})$ and $\Delta\varphi(\vec{\rho})$.[13,14] It should be noted that a "phase" S-CGH always has an additional absorption effect that depends on the thickness of the material and its chemical composition, while even a pure-phase S-CGH also has an effect on the amplitude of the wave function. An alternative way to achieve pure phase modification is to substitute a material-based hologram with a structured electric and/ or magnetic field, which introduces the desired phase modulation. It is then more challenging to design a complex and arbitrary phase shift.[15,16] This paper does not concentrate on such phase elements.

D. Different types of holograms

1. Amplitude holograms

In light optics, binary holograms (characterized by a local transmittance that is 0 or 1) are produced from partially transparent elements, such as gratings that are made from metals or substrates that can block a light beam in some regions in the transverse plane. They are considered to be the simplest types of S-CGHs that can be fabricated. Amplitude modulation in the transmission is usually achieved by covering parts of the beam with a material that can prevent light from passing through it (an opaque material that absorbs the beam), by deflecting the beam to a high angle or by reflecting part of the incident beam.

In electron optics, every scattering event that strongly modifies the electron beam beyond the simple phase effect can be considered as an amplitude effect. In particular, strong elastic changes of momentum due to atomic and thermal scattering contribute to a broadly diffuse intensity, while inelastic scattering largely alters the coherence of the electrons (the effect of the beam coherence in electron holography will be discussed in Sec. II E 2). Both effects contribute to remove intensity from the diffraction direction. These scattering processes are usually stronger for heavy materials and thicker substrates. The blocking of electrons can be achieved by using a thick sputtered layer of a high-atomic-number element such as Au or Pt. By doing so, the wave front amplitude is fully preserved or completely blocked, locally. Since a hologram absorbs or scatters electrons, its action is non-unitary and the overall intensity is reduced by a factor that is proportional to the blocked area in the incident beam cross section. By definition, amplitude holograms block part of the electron beam and have limited efficiency. Since absorption modulation is an amplitude-dominated effect, such holograms result in a diffraction pattern that is symmetrical between the positive and negative orders. Moreover, it is impossible to concentrate the intensity on a single diffraction spot and a large part of the intensity is directed to the 0th order transmitted beam.

In order to gauge the absorption of a material, a useful parameter is the mean free path for plasmon excitation (particularly for amorphous light material, since plasmon inelastic scattering can be considered the most important process).[17] The mean free paths of several materials are reported in Table I for 200 keV electrons.

TABLE I. Theoretical and experimental mean free paths for 200 keV electrons. The experimental values are total inelastic mean free paths, which include single electron excitations such as inner-shell ionization edges. The terms in parentheses are mean free paths for collective valence electron (i.e., plasmon) excitations.[17]

Material	Au	Ag	Pt	Si_3N_4	SiO_2	Al_2O_3	a-C
Theoretical (nm)[17]	76.1	88.3	76.4	135.3	133.6	135.7	106
Experimental (nm)[18]	84 (120)	100 (125)	82 (120)	…	155	140	160

It should also be noted that the construction of pure amplitude holograms, in which the absorptive material is alternated with vacuum, is complicated at small sizes—because of the probability that long and thin parts of the hologram may collapse or join together during fabrication or under electron beam illumination.

2. Phase holograms

In light optics, one way to implement a phase modulation is by etching grooves of a desired structure on a (transparent or reflective) surface, in order for the optical path inside (or upon the reflection from) the material to vary from one ray to another, thereby locally changing the phase of the outgoing wave function. In contrast, in electron optics, phase modulation is achieved by exploiting the relationship between scalar and vector potentials. The phase shift of an electron wave function[19] is given by the expression

$$\Delta\varphi(\vec{\rho}) = C_E \int_{-\infty}^{+\infty} V(\vec{\rho},\ z)dz - \frac{e}{\hbar}\int_{-\infty}^{+\infty} A_z(\vec{\rho},\ z)dz, \tag{9}$$

where

$$C_E = \frac{2\pi}{\lambda}\frac{e}{E}\frac{E_0 + E}{2E_0 + E}, \tag{10}$$

$V(\vec{\rho},\ z)$ and $A_z(\vec{\rho},\ z)$ are the scalar electrostatic potential and the z component of the magnetic vector potential, respectively, e is the absolute value of the electron charge, $\hbar = h/2\pi$ is the reduced Planck constant, λ is the relativistic electron wavelength, E_0 is the electron energy at rest, and E is the energy of the moving electrons. Typical electron energies in a TEM are 200 or 300 keV, resulting in corresponding values of $C_{E\_200\,\mathrm{keV}} = 7.3\ \times 10^{-3}\,\frac{\mathrm{rad}}{\mathrm{V\,nm}}$ and $C_{E\_300\,\mathrm{keV}} = 6.6\ \times 10^{-3}\,\frac{\mathrm{rad}}{\mathrm{V\,nm}}$. In a non-magnetic material, only the scalar electrostatic potential contributes to the phase shift. It can often be approximated by the mean inner potential V_{mip}, which provides a local acceleration to the electrons,[20] modifying the electron-optical path. The phase variation due to V_{mip} and the local thickness $t(\vec{\rho})$ is given by a simplified version of Eq. (9),

$$\Delta\varphi(\vec{\rho}) = C_E \int_0^{t(\vec{\rho})} V_{mip}dz = C_E V_{mip} t(\vec{\rho}). \tag{11}$$

When choosing a material for a synthetic hologram, one must take into account the robustness, electrical conductivity, and the value of V_{mip}. Most phase S-CGHs are currently made using Si_3N_4, which can be used to fabricate a nearly pure phase mask since it is almost transparent to an incoming electron beam. S-CGHs can be obtained by "carving" grooves in a free-standing Si_3N_4 membrane.

TABLE II. Si_3N_4 thickness required to introduce a 2π phase shift for different electron energies for $V_{mip} \approx 15$ V.

E (keV)	60	80	120	200	300
t(nm)	36.9	41.5	48.5	57.4	64.2

TABLE III. Theoretical and experimental mean inner potentials of the materials in Table I. Most values are taken from Refs. 30 and 31. For amorphous C, the mean inner potential depends on the density of the material.

Material	Au	Ag	Pt	Si_3N_4	SiO_2	Al_2O_3	a-C
Theoretical (V)	25 – 31	18.74 – 23	20 – 27	11.3 – 17.6	∼15.1	15.7 – 16.7 Ref. 26	10.1 – 11.3 Ref. 27
Experimental (V)	21 – 30	17 – 23	∼25	∼15 Ref. 21	∼17	16.9 ± 0.36 Ref. 26	9.09 – 10.7 Refs. 28, 29

The calculated value of V_{mip} for Si_3N_4 has been estimated to be ≈15 V.[21] The same value was reported by Harvey *et al.*[22] A slightly higher value was reported by Bhattacharyya *et al.*,[23] while a value of ∼10 V was found by Shiloh *et al.*[24] The specific Si_3N_4 preparation process and tension may affect the precise value. On the assumption that $V_{mip} \approx 15$ V, the required thickness to introduce a 2π phase shift for several electron energies is shown in Table II.

In recent years, new materials have been explored for the production of phase S-CGHs, with promising results shown for amorphous C.[25] The mean inner potentials of the materials from Table I are reported in Table III.

A schematic diagram of the operating principle of a phase mask is shown in Fig. 3 for an incoming plane wave, whose wave front is shown in red. A phase ramp, which results from the thickness profile of the phase mask, introduces a linear phase shift to a plane wave in a specific direction or azimuthal angle. The thinner region does not alter the electron beam wave front significantly, while the thicker region can be chosen so that it introduces a phase shift such as 2π to an incoming electron beam. The phase-shifted

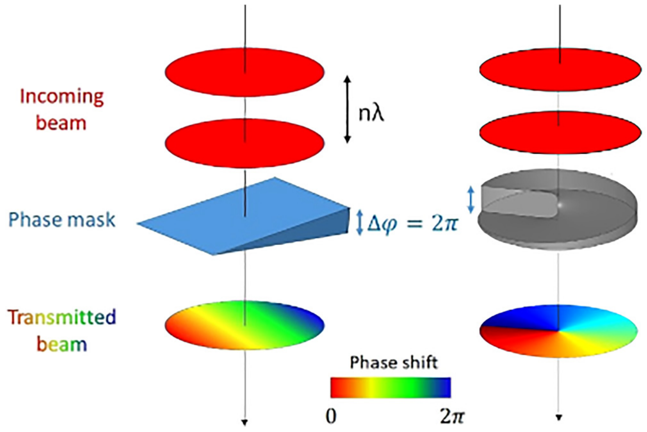

FIG. 3. Schematic diagram showing the phase-shifting effect on an incident plane electron wave of (left) a phase ramp and (right) a spiral phase mask. Readapted from "New approaches for phase manipulation and characterization in the transmission electron microscope," with the permission of Federico Venturi.[32]

wave front, which is shown using a color gradient, causes the electron beam to be deflected (left) or to carry OAM (right). The phase masks shown in Fig. 3 are *in-line* S-CGHs and are the simplest types that can be designed. Nevertheless, great manufacturing precision is required, as the phase shift is encoded in the pointwise thickness profile of the material. Such an *in-line* S-CGH does not require a reference beam.

On the other hand, for an *off-axis* S-CGH, the encoded beam (or desired beam) and hologram diffraction pattern are decoupled from one another. The off-axis S-CGHs are, therefore, less sensitive to small imperfections.

3. Amplitude-phase holograms

In the most general class of holograms, the material encodes a change in both the amplitude and the phase of the electron beam. In a strict sense, all phase holograms are amplitude-phase holograms, since any variation in thickness affects both the amplitude and the phase of the electron wave upon propagation.[33] Moreover, given the practical difficulties of fabricating complex amplitude gratings that are mechanically stable, amplitude gratings are often fabricated on continuous Si_3N_4 membranes.[34] Depending on how transparent the thick parts of the grating are, all levels between amplitude and phase gratings can be obtained. The smart use of two materials can be used to cancel amplitude effects in a phase hologram or to add an amplitude envelope to a phase grating. This approach could allow for the joint amplitude and phase encoding of wave functions but has not been explored in detail. Even the aperture effect that encloses a phase hologram (and truncates the beam) is still a type of amplitude filtering. In general, amplitude and phase modulations have slightly different effects, and the two contributions are superimposed. Therefore, it is difficult to control the amplitude and phase simultaneously using different holograms.

E. Calculation of holograms

1. Encoding the phase in phase holograms and amplitude holograms

In this section, it is shown how to calculate a hologram for a target wave function for the off-axis case, i.e., when a target wave function $\Psi(k)$ is reproduced in the first diffraction order. We assume that the reference wave is a plane wave with an in-plane wave vector $\vec{g}$, as determined by the inclination of the reference wave.

In "imaging holography," the condition for good reconstruction is a large enough fringe spacing and an object with a narrow frequency band, in order to be able to isolate and demodulate the phase properly. Similarly, in synthetic holography, a desired function $\Psi(k)$ should have a compact support so that its extension in Fourier space is smaller than a reference frequency $\vec{g}$. The required phase modulation $\Delta\varphi(\vec{\rho})$ needs to be calculated based on the desired diffraction shape. The hypothesis is that one can control the phase $\alpha(\vec{\rho})$ of the desired diffracted beam at the plane of the hologram. This should simply be the phase of the inverse FT of the object $\Psi(k)$, with the addition of a phase gradient as a result of the off-axis tilt. One can numerically calculate $\alpha(\vec{\rho}) = arg\{FT^{-1}(\Psi(k))\} + \vec{g}\cdot\vec{\rho}$. For example, $\alpha(\vec{\rho}) = \ell\theta + \vec{g}\cdot\vec{\rho}$ for a vortex beam, where ℓ is the desired winding number of the vortex and θ is the azimuth of the $\vec{\rho}$ coordinate in the hologram plane.

If $FT^{-1}(\Psi(k))$ had a constant amplitude over the hologram area (normally a circular top hat function), then $T_g = \exp(i\alpha(\vec{\rho}))$ would be the transmission function T_g of the desired phase plate and its Fourier transform would be $\Psi(k)$, apart from a tilt of $\vec{g}$. However, only the phase at the exit plane of the hologram is encoded. Methods to generalize this approach are discussed below. In the phase hologram case, the phase can be any function $\Delta\varphi = f(\alpha)$, with the periodicity condition $f(\alpha(\vec{\rho})) = f(\alpha(\vec{\rho}) + 2n\pi)$. For example, a sinusoidal grating that is used to generate vortex beams would be $\Delta\varphi = \varphi_0 \sin(\ell\theta + \vec{g}\cdot\vec{\rho})$. Since the fundamental requirement for holography is that the function $\alpha(\vec{\rho})$ is mainly bandwidth-limited (i.e., its FT is mainly contained within a frequency range $\sigma_k \ll g$), the transmission function of the full hologram is $T = \exp(i\, f(\,\alpha(\vec{\rho})))$, with approximate periodicity $\vec{g}$. It has a diffraction pattern that is given by many well-separated beams centered at $n\vec{g}$, where $n \in \mathbb{Z}$ is the diffraction order, and each diffracted beam can be spatially separated. For the first order beam, the hologram acts as the desired transmission function $T_g = \exp(i\alpha(\vec{\rho}))$. As f changes, so does the distribution between diffraction orders. For any form of f, the first diffraction order is only affected by a phase effect $T_g = \exp(i\alpha(\vec{\rho}))$.

For an amplitude hologram, it is possible to assume a simplified form of interference, where one retains only the cross term in Eq. (5). The simplest form of interference is just a cross term $T = \cos(\alpha(\vec{\rho}))$, which is clearly an amplitude modulation. However, as in the case of a phase hologram, one can use any function of the form $T = f(\alpha(\vec{\rho}))$. Analogously to a phase hologram, a sinusoidal amplitude grating that generates a vortex takes the form $T = \frac{A_0}{2}(1 + \sin(\ell\theta + \vec{g}\cdot\vec{\rho}))$, where positivity of the amplitude hologram is enforced. Even in this case, for any form of f, the first order beam is only affected by a phase effect $T_g = \exp(i\alpha(\vec{\rho}))$. Therefore, amplitude and phase holograms are the same at the level of individual diffracted beams. However, the phasing and amplitude ratio between the diffraction orders are different. For example, the first diffraction order is in phase with the zero order for an amplitude hologram, whereas there is generally a dephasing close to $\pi/2$ for a phase hologram. Amplitude effects are discussed in Secs. II E 4 and II E 5.

It is also important to mention Fresnel holograms. In this case, the desired intensity is not obtained in the Fraunhofer plane, but in an intermediate (Fresnel) plane. Although the concept is identical, the Fourier transform is then substituted by the Fresnel integral and

$$\alpha(\vec{\rho}) = arg\{\Psi(\mathrm{k}) \otimes \mathrm{P}(-\Delta \mathrm{z}))\} + \vec{g}\cdot\vec{\rho}\ , \tag{12}$$

where $\otimes$ is a convolution integral and $P(\Delta z) = \frac{1}{i\lambda\Delta z}\exp\left(i\frac{\pi}{\lambda\Delta z}(x^2 + y^2)\right)$.

Δz here is the free propagation distance at which the hologram's diffraction is observed. If instead of the free space propagation a lens is used, then the effective value of Δz must be scaled to account for the effective optical distance.[19]

The Fresnel transform is sometimes defined as fractional Fourier transform[35] so we are tempted to just assume it is perfectly analogous to the FT. In reality, there is an important practical difference between the two: the Fresnel propagation depends, in general, on both the position and momentum coordinates. This means that, for example, shifting the hologram changes both the position and the shape of the diffracted beam. Moreover, in a TEM microscope, the Fresnel distances are typically not calibrated and particular care should be considered in the match between simulations and experiments.[36,37]

For more details about a specific case of Fresnel hologram, the reader should refer to Sec. IV C.

2. Coherence

Another aspect that so far was neglected is the coherence of the beam. While the calculation is typically carried out with a simple Fourier transform, more in general, this approach is correct only for a perfectly coherent beam. Coherence is defined in different contexts. In quantum mechanical term, a coherent state is a pure quantum mechanical state defined by a single wave function. A pure state always produce interference so we can describe coherence as the ability to produce significant interference.

In the imaging holography with a biprism, the most generic form of the interference figure is

$$I = 1 + A_I^2(\vec{\rho}) + 2\boldsymbol{\mu} A_I(\vec{\rho})\cos(\varphi_I(\vec{\rho}) + \vec{g}\cdot\vec{\rho}), \tag{13}$$

The additional μ factor describes the coherence as, indeed, the ability to produce diffraction fringes.[38,39]

The electron beam in a microscope is always partially coherent.[20] The most consistent description of the state can be done by the formalism of density matrix or equivalently by the Wigner function.[40,41] In practical term, it can be demonstrated that this is equivalent to saying that a single wavefunction is substituted by a set of wave functions having different energy and momentum and that do not interfere with each other. The energy distribution determines the so-called longitudinal coherence (or temporal coherence), while the extension of the electron source assumed as a collection of independent emitters is the main responsible for the transverse coherence (or spatial coherence).

For holography, the spatial coherence is mainly responsible for the loss of interference (contrast?) and it is clear that the same arguments hold for synthetic holography.

A look at the optical scheme in Fig. 2 convinces us that a nominal plane wave impinging on the hologram is actually an incoherent sum of plane waves with slightly different momentum. This spread in the Fraunhofer diffraction determines the size of the diffraction spot. This is, therefore, the point spread function of the intensity of the synthetically generated electron beam. A simple convolution can be added in simulation to account for this effect, but it is clear that the effect must be reduced for many practical applications of synthetic holography.

A practical approach in holography is the use of limiting apertures to limit the effective part of the source contributing to the imaging and wave formation. This approach can be extended to synthetic holography to improve the final result.

It is worth finally mentioning that for Fresnel holograms, the convolution is not a valid simulation approach. High resolution microscopy has invented a series of effective approaches based on "damping factors,"[42] while a more general approach is to consider the effective incoherent sum of diffraction from component plane waves.

3. Diffraction efficiency and groove profile

A key parameter that defines the performance of an off-axis S-CGH is its diffraction efficiency, which can be defined as the ratio between the intensity measured in a specific diffraction order and the intensity of the incoming beam or the total transmitted intensity.[22] According to the first definition, efficiency is

$$\eta_n^{(inc)} = \frac{I_n}{I_{inc}}, \tag{14}$$

where I_n is the intensity of order n and I_{inc} is the intensity of the incident beam. $\eta_n^{(inc)}$ is then known as the absolute diffraction efficiency. The total transmitted efficiency is

$$I_{trans} = \sum_n I_n, \tag{15}$$

and the second definition of diffraction efficiency is

$$\eta_n^{(trans)} = \frac{I_n}{I_{trans}}. \tag{16}$$

In amplitude and phase S-CGHs, the beam has to pass through a material and the total transmitted intensity is reduced, typically by 50% and up to 30%–40%, respectively. The reduction results from absorption and other inelastic processes, even in high-transmittance materials such as Si_3N_4. Low intensities might have some implications in applications due to signal/noise ratios, both positive and negative, but in most cases, a high brightness source is ideal to improve the transmitted intensity by the synthetic hologram. Henceforth, we use η to refer to $\eta_n^{(trans)}$, the so-called transmitted efficiency. When a distinction is necessary, the appropriate symbol is used. As mentioned above, the efficiency of an S-CGH depends on whether it is phase- or amplitude-modulated. However, the efficiency also depends on the groove profile/ thickness pattern of the hologram. In order to establish the relationship between groove profile and efficiency, we begin by explaining how an incoming wave function is transformed after its interaction with an S-CGH. This interaction depends on the groove pattern. According to Eq. (8), the transfer function $T_H(\vec{\rho})$ describes the amplitude and the phase of a beam that has passed through a diffraction grating. An alternative representation of the transfer function, specifically for a phase S-CGH, is given by the expression

$$T_H(\vec{\rho}) = e^{i\tilde{V}t(\vec{\rho})}, \tag{17}$$

where $\tilde{V} = C_E V_{mip} + i\gamma$ is the complex index of refraction, $\gamma = \frac{1}{\lambda_{mfp}}$ is the absorption coefficient, λ_{mfp} is the mean free path of an electron, and $t(\vec{\rho})$ is the thickness profile. Since the transfer function is independent of the incident wave function $\Psi_{inc}(\vec{\rho})$, the transmitted wave function can be written in the form

$$\Psi_t(\vec{\rho}) = \Psi_{inc}(\vec{\rho})T(\vec{\rho}) = \Psi_{inc}(\vec{\rho})e^{i\tilde{V}t(\vec{\rho})}. \tag{18}$$

In most case studies, the incoming wave is assumed to be a plane wave and can be ignored, as it has a flat-phase wave front and its FT depends mainly on the transfer function. A generic diffraction grating is characterized by a periodic wave function $f(\alpha)$, which describes its groove pattern. It is usually dimensionless and normalized from zero to unity with a period 2π, such that $f(\alpha + 2\pi) = f(\alpha)$. The function can be expanded as a Fourier series in the form

$$f(\alpha) = \sum_{n=-\infty}^{\infty} c_n e^{in\alpha}, \tag{19}$$

where $n \in \mathbb{Z}$ and the nth Fourier coefficient

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} f(\alpha)e^{-in\alpha}d\alpha. \tag{20}$$

Each value of n represents one diffraction order. If $f(\alpha)$ is real-valued, then $c_n = c^*_{-n}$, where the asterisk denotes a complex conjugate and c_0 is real. The Fourier power spectrum of $f(\alpha)$ is given by the expression

$$S = \sum_n |c_n|^2. \tag{21}$$

For a bi-dimensional grating with a single type of groove profile, the periodic function $f(\alpha)$ has

$$\alpha = \alpha(\vec{\rho}), \tag{22}$$

where $|\vec{\rho}|$ and θ, the azimuthal angle, are polar coordinates that define the grating and $|\vec{\rho}|$ is measured in units of the grating spatial period Λ in the $\theta = 0$ direction.

For an amplitude S-CGH, the Fresnel transmission function $T(x, y)$ is proportional to the grating function

$$T(\vec{\rho}) = bf(\alpha(\vec{\rho})), \tag{23}$$

where b is a constant and $0 \leq b \leq 1$.

The wave function impinging on the S-CGH is denoted $\Psi_{in}(\vec{\rho})$. The output wave function can then be determined as follows:

$$\Psi_{out}(\vec{\rho}) = T(\vec{\rho})\Psi_{in}(\vec{\rho}) = bf(\alpha(\vec{\rho}))\Psi_{in}(\vec{\rho}). \tag{24}$$

For a phase S-CGH, the transmission function is given by the expression

$$T(\vec{\rho}) = e^{i\tilde{a}f(\alpha(\vec{\rho}))}, \tag{25}$$

where $\tilde{a} = a_1 + ia_2$ is a complex number, $a_1 = C_E V_{mip} t_M$, $a_2 = \gamma t_M$ and t_M is the maximum thickness difference between a peak and a valley. It can be shown that, depending on the microscope accelerating voltage, $a_2 \sim 7\% - 8\%\ a_1$ for Si_3N_4. Furthermore, the product of t_M and $f(\alpha(\vec{\rho}))$ yields the local thickness of the grating $t(\vec{\rho})$. Therefore, for this type of hologram, the output wave function is given by the expression

$$\Psi_{out}(\vec{\rho}) = \Psi_{in}(\vec{\rho})e^{i\tilde{a}f(\alpha(\vec{\rho}))}. \tag{26}$$

The efficiency of the S-CGH can be estimated/ calculated from the power transmission spectrum, given by the sum of the Fourier coefficients of the transmission function Fourier series expansion

$$\mathcal{T}(\vec{\rho}) = \sum_n |\tau_n|^2, \tag{27}$$

where the intensity of the nth diffraction order is modulated by the transmission coefficient $|\tau_n|^2$. For amplitude S-CGHs, $\theta_n = \frac{n\lambda}{\Lambda}$ is the diffraction angle for the nth order diffracted beam. In an ideal phase hologram, the transmission should be unitary. Therefore, the ideal phase plates are the optimal choice to convey the largest intensity on the diffracted-shaped beams. In practice, the absorption for real phase plates is often relatively high and the overall efficiency on the generation of intense beams could, in some cases, favor amplitude holograms in terms of intensity: a precise balance depends on the details for the synthetic hologram design and its material.

4. Comparison between grating profiles

In this section, a series of grating profiles are presented for both amplitude and phase S-CGHs. In each case, the grating profile function and the Fourier coefficients of the transmission function τ_n are given. Calculations describing how the equations were obtained are given in Appendix B.

a. Sinusoidal/cosinusoidal profile. The simplest profile is sinusoidal/cosinusoidal. As they have the same characteristics, only one is considered. For an amplitude S-CGH with a cosinusoidal profile, $f(\alpha) = \frac{1}{2}(1 + \cos(\alpha(\vec{\rho})))$. The transmission function is

$$T(\vec{\rho}) = \frac{b}{2}(1 + \cos(\alpha(\vec{\rho}))), \tag{28}$$

where $0 \leq b \leq 1$ is a constant. For a phase S-CGH, the transmission function can be written in the form

$$\begin{aligned} T(\vec{\rho}) &= e^{i\frac{\tilde{a}}{2}1 + \cos\alpha\vec{\rho}} = e^{i\frac{a_1}{2}\cos\alpha\vec{\rho}}e^{-\frac{a_2}{2}\cos\alpha\vec{\rho}}e^{i\frac{\tilde{a}}{2}} \\ &= e^{ia_1'\cos\alpha\vec{\rho}}e^{-a_2'\cos\alpha\vec{\rho}}e^{i\tilde{a}'}. \end{aligned} \tag{29}$$

For both types of holograms, $f(\alpha)$ is normalized between zero and unity, so that for an amplitude S-CGH, the power transmittance changes locally between 0 for full absorption and 1 for no absorption. For an ideal phase S-CGH, the power transmittance is always 1 and it is possible to estimate the optimal phase shift introduced by the local thickness profile to maximize the intensity of one of the diffraction orders (usually $n = \pm 1$). For an amplitude S-CGH, the squared modulus of the Fourier coefficient of the transmission function is

$$|\tau_n(\vec{\rho})|^2 = \begin{cases} \frac{1}{4}b^2 \; for\ n = 0, \\ \frac{1}{16}b^2 \; for\ n = \pm 1, \\ 0 \text{ for the other orders,} \end{cases} \tag{30}$$

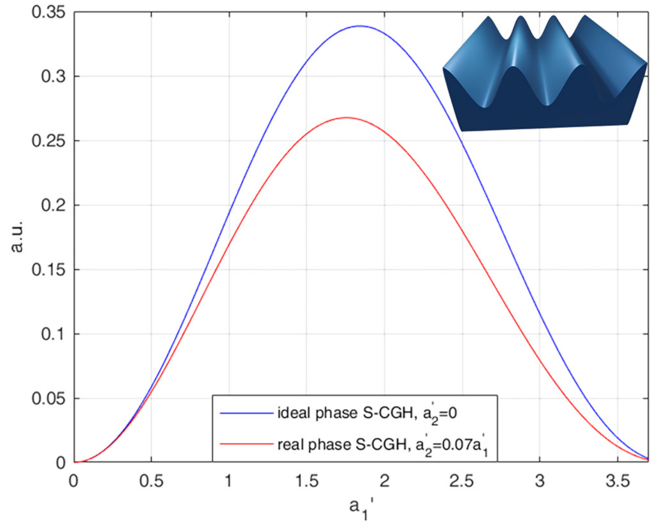

FIG. 4. $|\tau_1|^2$ plotted for a phase S-CGH with a cosinusoidal profile as a function of a_1'. As an inset is shown the 3-D rendering of a cosinusoidal profile. Here, "ideal" means no absorption ($a'_2 = 0$), while for the "real" case we assumed $a_2' = 0.07a_1'$. The choice of this specific relation of a_2' has been done considering as support material Si_3N_4 and 200 keV electrons.

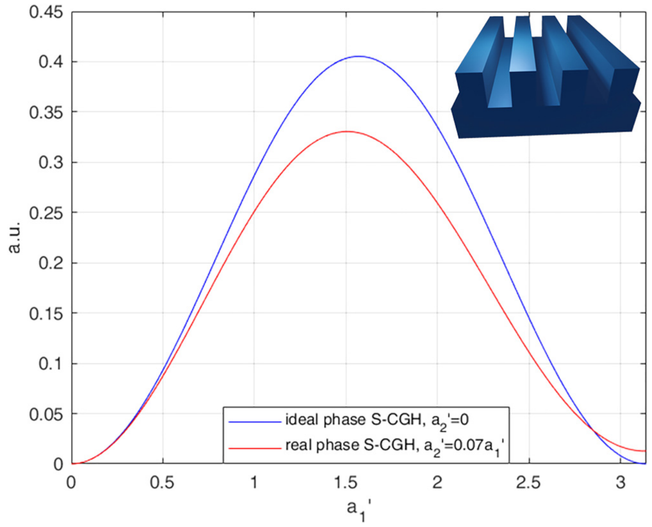

FIG. 5. $|\tau_1|^2$ for a squared profile as a function a_1'. As an inset is shown the three-dimensional profile. Here, "Ideal" means no absorption ($\boldsymbol{a'}_2 = 0$), while for the "real" case we assumed $a_2' = 0.07a_1'$.

for the nth order, whereas for a phase S-CGH, the corresponding expression is

$$|\tau_n(\vec{\rho})|^2 = J_n^2(\tilde{a}')\, e^{-2a'_2} \approx \left[J_n^2(a'_1) - \frac{a'^2_2}{4}(J_{n-1}(a'_1) - J_{n+1}(a'_1))^2\right] e^{-2a'_2}. \tag{31}$$

Figure 4 shows the characteristic efficiency of the first diffracted order for a phase S-CGH, realized on a Si_3N_4 membrane, plotted as a function of the parameters a_1' and a_2'. The efficiency reaches a maximum when $a_1' \approx 1.84$ rad. In the "real phase S-CGH" profile, the contribution of absorption is appreciable, with the main contribution originating from the exponential term $e^{-2a'_2}$. The ideal value of a_1', which maximizes $|\tau_1(\vec{\rho})|^2$ of an ideal phase S-CGH, corresponds to a peak-to-valley phase difference of ~3.68 radians (1.17π), which corresponds to $t_M \approx 33.6$ nm for 200 keV electrons.

b. Squared profile. The second profile considered here is a squared profile, for which $f(\alpha) = \frac{1}{2}(1 + \mathrm{Sign}(\sin(\alpha(\vec{\rho}))))$. For an amplitude S-CGH, the transmission function takes the form

$$T(\vec{\rho}) = \frac{b}{2}(1 + \mathrm{Sign}(\sin(\alpha(\vec{\rho})))), \tag{32}$$

while for a phase S-CGH the transmission function can be written

$$\begin{aligned} T(\vec{\rho}) =& e^{i\frac{\tilde{a}}{2}(1+\mathrm{Sign}(\sin(\alpha(\vec{\rho}))))} = e^{i\tilde{a}'\mathrm{Sign}(\sin(\alpha(\vec{\rho})))} e^{i\tilde{a}'} \\ =& e^{ia'_1\mathrm{Sign}(\sin(\alpha(\vec{\rho})))}\ e^{-a'_2\mathrm{Sign}(\sin(\alpha(\vec{\rho})))} e^{i\tilde{a}'}. \end{aligned} \tag{33}$$

For an amplitude grating, the square modulus of the Fourier coefficients of the transmission function is

$$|\tau_n(\vec{\rho})|^2 = \begin{cases} \frac{1}{4}b^2 \text{ for } n = 0, \\ \frac{1}{n^2\pi^2}b^2 \text{ for } n = \text{odd}, \\ 0 \text{ for } \ n = \text{even}, \end{cases} \tag{34}$$

while for a phase grating, it is

$$|\tau_n(\vec{\rho})|^2 = \begin{cases} [\cos^2(a'_1)\cosh^2(a'_2) + sin^2(a'_1)\sinh^2(a'_2)]e^{-2a'_2} \text{ for } n = 0, \\ \frac{4}{n^2\pi^2}[\sin^2(a'_1)\cosh^2(a'_2) + cos^2(a'_1)\sinh^2(a'_2)]e^{-2a'_2} \text{ for } n = \text{odd}, \\ 0\ for\ n = even. \end{cases} \tag{35}$$

Figure 5 shows the efficiency of the first diffraction order for a phase S-CGH. The maximum is reached when $a_1' \approx 1.57$ rad, so the optimal peak-to-valley phase difference corresponds to $\Delta\varphi \approx \pi$ for an ideal phase S-CGH.

c. Triangular profile. The third case is a triangular profile, which can be described (for an isosceles triangle) by the function $f(\alpha) = \frac{1}{\pi}(\mathrm{Sign}(\sin(\alpha(\vec{\rho}))))(\pi - \mathrm{Mod}(\alpha(\vec{\rho}),\ 2\pi))$. For an amplitude S-CGH, the transmission function is

$$T(\vec{\rho}) = b\frac{1}{\pi}(\mathrm{Sign}(\sin(\alpha(\vec{\rho}))))(\pi - \mathrm{Mod}(\alpha(\vec{\rho}),\ 2\pi)), \tag{36}$$

while for a phase S-CGH, it is

$$T(\vec{\rho}) = e^{i\tilde{a}\frac{1}{\pi}(\mathrm{Sign}(\sin(\alpha(\vec{\rho}))))(\pi - \mathrm{Mod}(\alpha(\vec{\rho}),2\pi))}, \tag{37}$$

where *Mod*(p, q) is the remainder after dividing p by q. For an amplitude S-CGH with a triangular modulation, the efficiency of the nth diffracted order is proportional to

$$|\tau_n|^2 = \begin{cases} \frac{1}{4}b^2 \text{ for } n = 0, \\ \frac{4}{n^4\pi^4}b^2 \text{ for } n = \text{odd}, \\ 0 \text{ for } n = \text{even}, \end{cases} \tag{38}$$

while for a phase S-CGH with a triangular modulation, it is

$$|\tau_n|^2 = \frac{(a_1^2 + a_2^2)[1 + 2(-1)^{n+1}e^{-a_2}(\cos(a_1)) + e^{-2a_2}]}{[a_1^4 + 2a_1^2a_2^2 + a_2^4 + (n\pi)^4 - 2n^2a_1^2\pi^2 + 2n^2a_2^2\pi^2]}. \tag{39}$$

Figure 6 shows how the efficiency changes as different parameters are varied. If a_2 is non-zero, i.e., if absorption is considered, then the efficiency is reduced and the peak moves to lower values of a_1. $|\tau_1|^2$ has a maximum at $a_1 \approx 4.31\,rad$ for an ideal phase S-CGH, whereas it is at $a_1 \approx 4$ rad for a real S-CGH.

d. Blazed profile. An interesting triangular profile is a blazed profile, which is similar to a sawtooth blade and can be described by the function $f(\alpha) = \frac{1}{2\pi}(\mathrm{Mod}(\alpha(\vec{\rho}), 2\pi))$. For an amplitude S-CGH, the transmittance function is

$$T(\vec{\rho}) = b\frac{1}{2\pi}(\mathrm{Mod}(\alpha(\vec{\rho}), 2\pi)), \tag{40}$$

while for a phase S-CGH, it is

$$T(\vec{\rho}) = e^{i\tilde{a}\frac{1}{2\pi}(\mathrm{Mod}(\alpha(\vec{\rho}),2\pi))}. \tag{41}$$

For an amplitude S-CGH, the efficiency of the nth order of diffraction is

$$|\tau_n|^2 = \begin{cases} \frac{1}{4} \text{ for } n = 0 \\ \frac{1}{4\pi^2n^2} \text{ for } n \neq 0 \end{cases}, \tag{42}$$

while for a phase S-CGH, it is

$$|\tau_n|^2 = \frac{(1 + e^{-2a_2} - 2\cos(a_1)e^{-a_2})}{[(a_1 + 2\pi n)^2 + a_2^2]}. \tag{43}$$

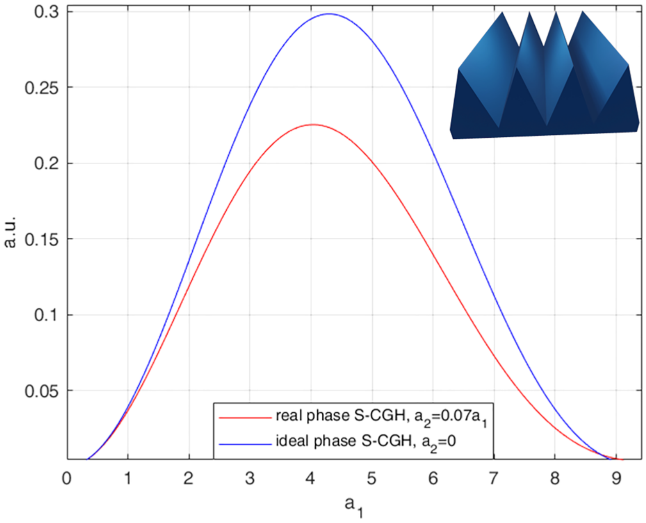

FIG. 6. $|\tau_1|^2$ for a triangular profile plotted for a phase S-CGH as a function of a_1. As an inset is shown the three-dimensional profile. Here, "Ideal" means no absorption ($a_2' = 0$), while for the "real" case we assumed $a_2' = 0.07a_1'$.

Figure 7 shows two features of a blazed profile for a phase S-CGH. First, in the ideal case, maximum efficiency is reached when the peak-to-valley distance is equivalent to a phase difference of 2π. Second, by tuning the shape, it is possible to reach an efficiency of almost 100% in one of the first diffraction orders (Fig. 8).

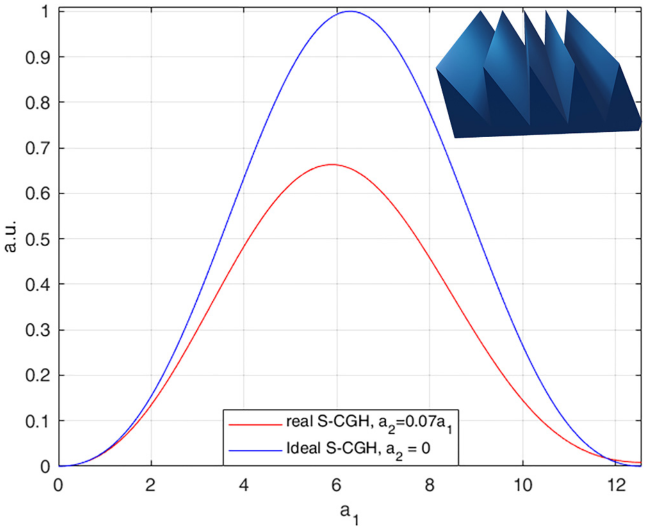

FIG. 7. Relative transmitted efficiency in the −1 diffracted order for a blazed profile with ($a_2' = 0.07a_1'$) and without absorption ($a_2' = 0$). The maximum relative efficiency is reached for $a_1 \sim 2\pi$, which means that when no absorption is considered, ideal efficiency is obtained when the peak-to-valley phase difference due to thickness is 2π. As an inset is shown in the three-dimensional profile.

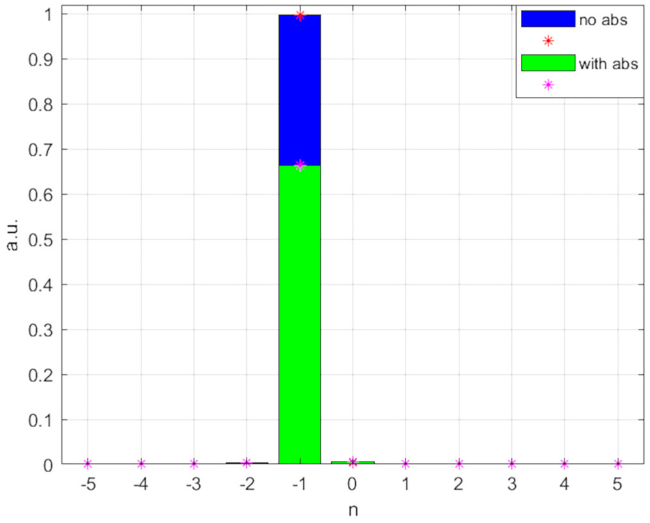

FIG. 8. Comparison between $|\tau_n|^2$ for different diffraction orders for the ideal case of a blazed phase S-CGH and a "real case" with absorption. In the ideal case, only the minus one order survives and all others ideally have zero intensity. If absorption is considered, $\mathcal{T}$ is no longer unitary.

This profile is the only one that theoretically allows for the whole transmitted wave to be directed to one of the first diffraction orders.

If a blazed profile is not perfect and the grooves are more similar to scalene triangles, then the transmitted intensity is no longer concentrated in one of the first diffracted orders, but spreads to others. This is usually visible when looking at a diffraction pattern of a real blazed hologram affected by fabrication limitations. A more in-depth analysis regarding the optimization of a real blazed phase S-CGH is described in Sec. IV A 3

5. *Efficiencies of the profiles*

The efficiencies of the profiles that have been described are now compared, distinguishing between the amplitude and phase S-CGHs. Efficiency is one of the critical parameters to consider during the design of a synthetic hologram. Here, one histogram is shown for each groove pattern, with the diffraction order on the horizontal axis and the transmitted efficiency $\eta_n^{(t)}$ on the vertical axis. For ease of visualization, only orders between -5 and $+5$ are shown.

a. Amplitude S-CGH. Figure 9 shows the intensity distribution between diffraction orders for different profile shapes (sinusoidal, squared, triangular, and blazed) for an amplitude S-CGH. The central or zeroth order peak always has the highest efficiency. The total transmitted intensity is never 100%, since the hologram absorbs some incoming electrons. To a first approximation, if only absorption from the opaque part is considered and that from the supporting layer is neglected, the best performing shape is the squared profile, for which 50% of the intensity is transmitted. The worst performing shape is the blazed profile, for which only 33% is transmitted.

b. Ideal phase S-CGH. Figure 10 shows the intensity distribution between diffraction orders for different profile shapes for a phase S-CGH. The calculations have been carried out such that the phase difference maximizes the intensity in one of the first two diffraction orders. The zeroth order peak is always less intense than the first orders. While the phase difference between peak and valley can be tuned in a phase S-CGH, this is not possible for an amplitude S-CGH, for which the zeroth diffraction order is always the most intense. For an ideal phase S-CGH, in which absorption is omitted, the total transmitted intensity is almost 100% for all of the shapes considered here. Key values of efficiencies are reported in Table IV.

F. Encoding both amplitude and phase in a synthetic hologram

1. *Encoding amplitude and phase in a phase hologram*

Unlike the other S-CGHs presented so far, which were aimed at generating a desired wave function in all non-zero diffraction orders (apart from a multiplicative factor for the angular momentum), mixed holograms generate a desired wave function in only one specific diffraction order. The method is based on tuning the peak-to-valley phase difference in each region of the S-CGH profile. This yields a local change in efficiency, which changes the wave front phase at the exit of the hologram, resulting in a change in the intensity of the beam. If $A(\vec{\rho})$ and $\varphi(\vec{\rho})$ are the amplitude and phase of a desired wave function, $B(\vec{\rho})$ is a normalized bounded positive function of amplitude, $C(\vec{\rho})$ is an analytical function of the amplitude and phase profiles of the desired field, and Λ is the period of the diffraction grating, then the profile to be fabricated takes the form[43]

$$T_{Mix}(\vec{\rho}) = \exp\left[iB(\vec{\rho})\mathrm{Mod}\left(C(\vec{\rho}) + \frac{2\pi\rho_{\theta=0}}{\Lambda},\ 2\pi\right)\right], \qquad (44)$$

where

$$B(\vec{\rho}) = 1 + \pi^{-1}\mathrm{sinc}^{-1}(A(\vec{\rho})), \qquad (45)$$

$$C(\vec{\rho}) = \varphi(\vec{\rho}) - \pi B(\vec{\rho}). \qquad (46)$$

And $\mathrm{sinc}^{-1}()$ is the inverse of sinc function in the interval of $[-\pi, 0]$. Areas characterized by a full 2π phase shift contribute to the amplitude of the +1st diffraction order, whereas other areas spread intensity over other orders, limiting the intensity of the 1st order. The beam of interest is generated with the correct phase and amplitude information only in the 1st diffraction order.

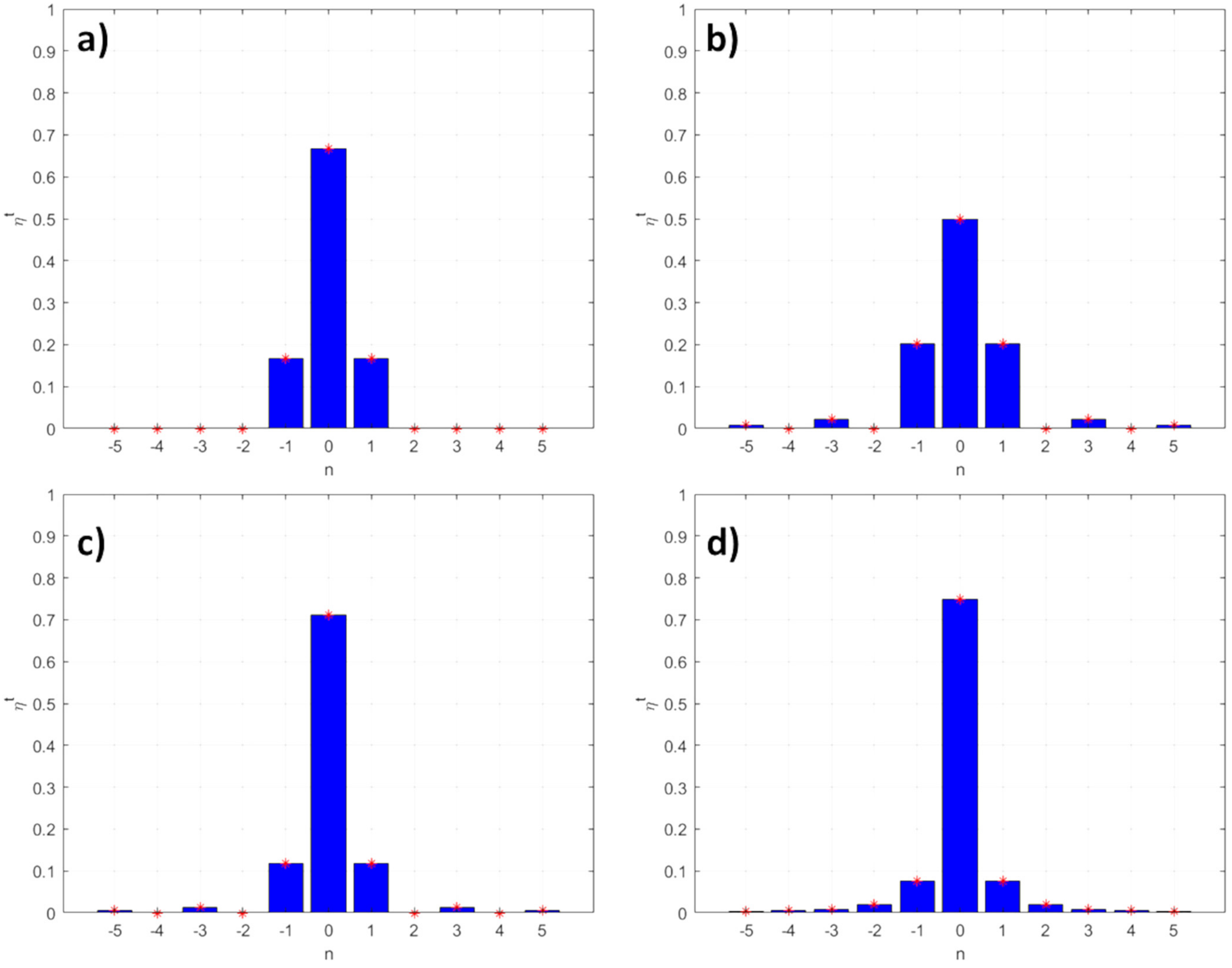

FIG. 9. Transmitted efficiency of diffraction orders $n \in [-5, 5]$ for an amplitude S-CGH: (a) sinusoidal profile, (b) squared profile, (c) triangular profile, and (d) blazed profile.

2. Encoding amplitude and phase in an amplitude hologram

The approach described in Sec. II E 1 for phase-only holograms is based on the modulation of the peak-valley value. This modulation locally varies the efficiency of the grating, and therefore, the amplitude encoding. The same method can be used to achieve amplitude and phase encoding using an amplitude hologram.

For the sake of simplicity, we start by considering a binary mask (i.e., a rectangular profile) and first encode the phase, before adding modulation to the width of the groove that is related to the local efficiency of the hologram, as outlined in Eq. (33).

In simple terms, the center of each groove is related to the phase modulation, while the width is related to the amplitude of the wave (as it can be appreciated in Fig. 11). Instead of a rectangular groove, one can choose any profile.

If a phase-only modulation is chosen such that $f(\alpha) \propto \cos(kx + \alpha(\vec{\rho}))$, then the center of the fringes corresponds to the condition $\cos(kx + \alpha(\vec{\rho})) \;= 1$. An amplitude modulation can be achieved by substituting the 1 with a "bias" function of the form $\cos(q(\vec{\rho}))$, where $q(r)$ is a function of the local desired efficiency. The relation $\cos(kx + \alpha(\vec{\rho})) = \cos(q(\vec{\rho}))$ can then be used to find the clipping points at the sides of the groove. Further mathematical approaches based on this principle are possible (e.g., Ref. 44).

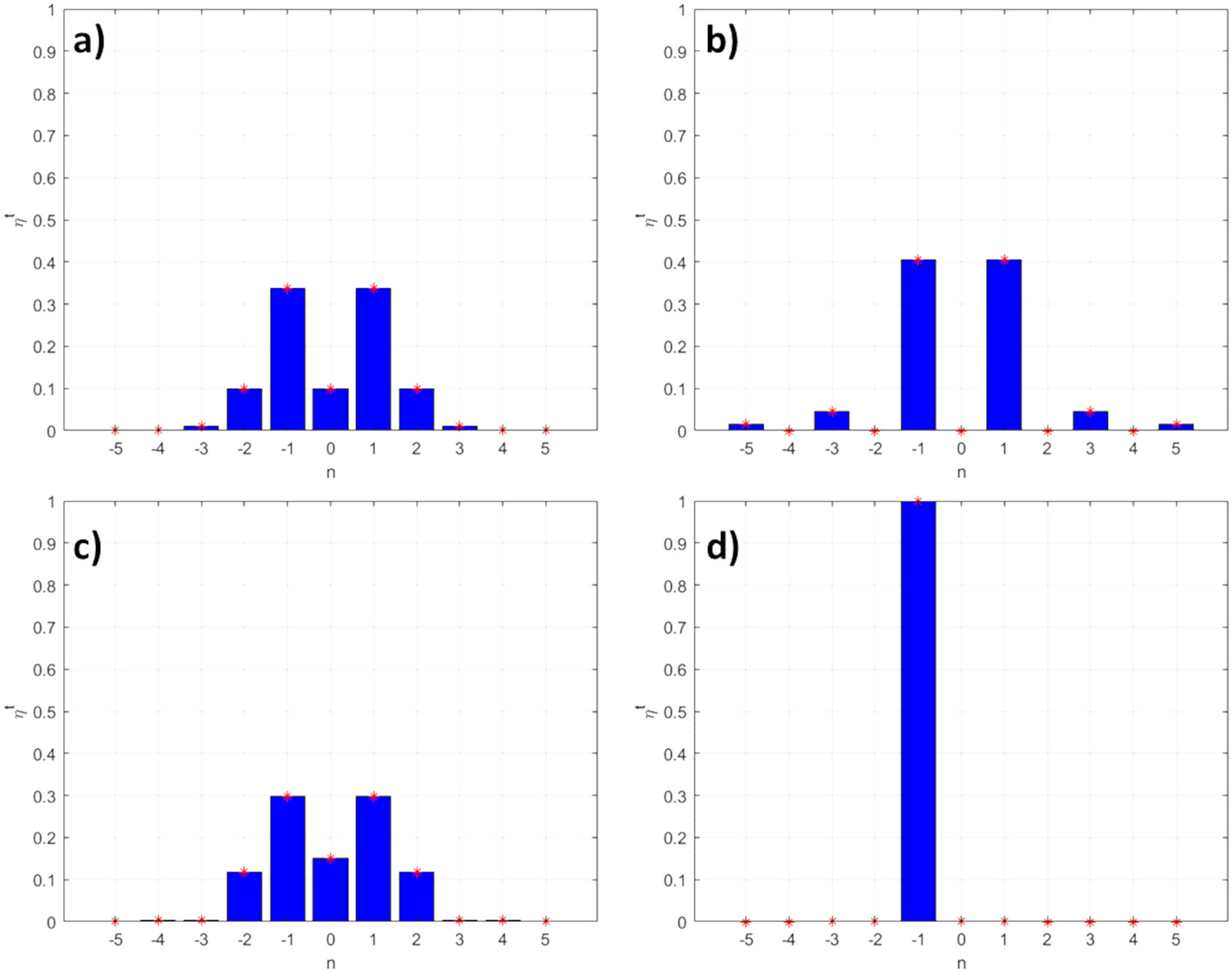

FIG. 10. Transmitted efficiency of diffraction orders $n \in [-5, 5]$ for an ideal phase S-CGH: (a) sinusoidal profile; (b) squared profile; (c) triangular profile; (d) blazed profile.

This conceptual scheme can be extended to include ideas for the phase hologram encoding of amplitude and phase, as seen above. Furthermore, the above approach is more exact, as it accounts for the amplitude modulation effect on the phase shift and the phase effect on the amplitude.

G. Sampling effect and choice of groove shape

When deciding on the design of a hologram and particularly on the groove shape, the practical problem of the limited number of addressable or calculated pixels should be considered. Typically,

TABLE IV. Transmission power function, n = ±1 square modulus of Fourier coefficients and transmitted efficiency for two kinds of S-CGH and all of the groove profiles considered here.

Profile shape	$\mathcal{T}_{\text{amp}}$ (%)	$\lvert\tau_{\pm 1}\rvert^2$ (%)	$\frac{\lvert\tau_{\pm 1}\rvert^2}{\mathcal{T}_{\text{amp}}}$ (%)	$\mathcal{T}_{\text{phase}}$ (%)	$\lvert\tau_{\pm 1}\rvert^2$ (%)	$\frac{\lvert\tau_{\pm 1}\rvert^2}{\mathcal{T}_{\text{phase}}}$ (%)
Cosinusoidal	37.5	6.25	16.67	100	33.86	33.86
Squared	50	10.13	20.26	100	40.53	40.53
Triangular	35.13	4.11	11.69	100	29.82	29.82
Blazed	33.3	2.53	7.60	100	100	100

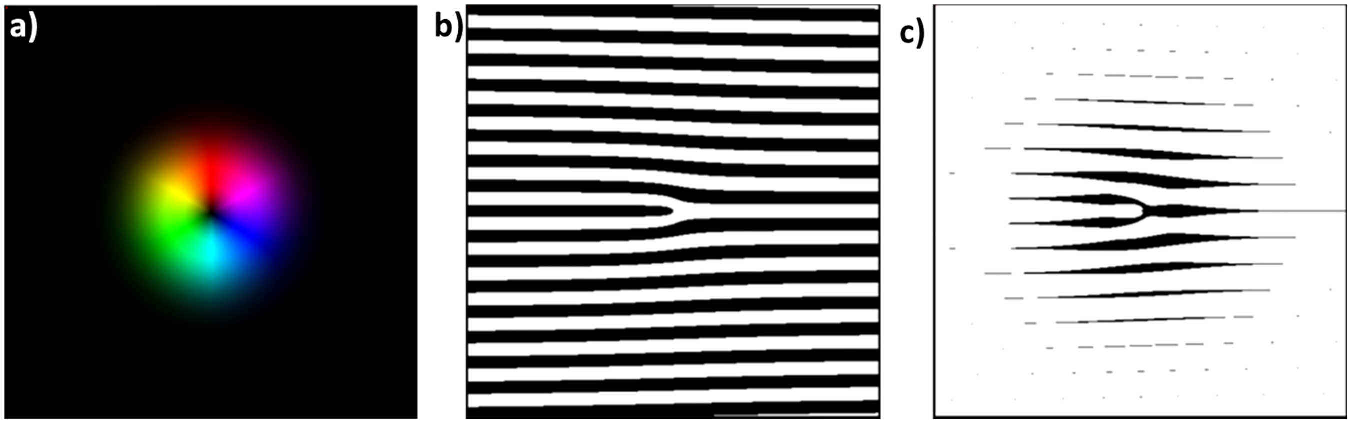

FIG. 11. Example of encoding the amplitude and phase of a Laguerre Gauss beam (ℓ = 1, p = 0) in an amplitude hologram: (a) false color representation of the LG beam (saturation represent the amplitude and hue represent the phase); (b) characteristic pitchfork amplitude hologram encoding the phase, and (c) the modulation of the groove width giving the amplitude envelope.

a hologram uses a square with between 1000 and 4000 pixels on each side. Beyond 8000 pixels, it is computationally and experimentally demanding to build an S-CGH. A typical groove is sampled with $n_{pg} = 5$ to 20 pixels. If the resolution in the groove positioning is given approximately by $1/n_{pg}$, then the phase is defined to be within $2\pi/n_{pg}$. The resulting problem in phase shaping is greatest for the rectangular groove, and to some extent, for the blazed groove, as any discontinuity is defined by the size of the pixel. In contrast, a sinusoidal groove has the advantage that each pixel intensity defines the phase with no discontinuity. In other words, even if the center of a groove is not defined by a single pixel, it can be calculated with sub-pixel precision as a weighted position average, whereas for a rectangular groove the phase is defined only on a discrete grid. A falsely encoded phase profile can result in an additional intensity between the diffraction orders in the hologram's Fourier transform. Under specific conditions, it is possible to recognize many (n_{pg}) copies of the same beam. This effect is related to the Talbot effect.[45] A second point is the bandwidth of the function to be encoded. The carrier frequency $|\vec{g}|$ must be larger than the bandwidth of the signal. For a sinusoidal pattern, at least 4 pixels are needed per period. If the bandwidth is B, then $|\vec{g}| \gg 2B$ and

$$n_p = K_{max} \gg 8B. \tag{47}$$

For the case of a vortex beam with a top hat amplitude cutoff, $B \approx a\,\ell$ with $a \approx 1/\pi$, so for a beam with $\ell = 1000$, approximately 4000 pixels are required. A different groove shape could result in a different maximum winding number for the vortex that can be generated. One should also consider the fact that a groove shape depends on the fabrication approach. When using EBL, it is more difficult to fabricate a groove that is not rectangular. Further details about vortex beam generation and fabrication techniques are given below.

III. PRODUCTION OF HOLOGRAMS: ELECTRON BEAM LITHOGRAPHY AND FOCUSED ION BEAM MILLING

The final step of S-CGH production is the fabrication of the designed pattern on a chosen substrate. The most common substrate of choice is currently silicon nitride (Si_3N_4), while the two fabrication techniques that are typically used to make S-CGHs are focused ion beam (FIB) milling and electron beam lithography (EBL).

A. Focused ion beam milling

FIB milling is a powerful tool for the fabrication of designed patterns. A FIB instrument is used to generate a focused high-energy beam of accelerated ions, which are then directed toward a sample surface to remove material by sputtering. Although Ga ions are the most widely used ions for this purpose, Au, Ir, Ar, He, Xe, O, N, and Si ions are also available. A higher-atomic-number element provides a higher milling yield, whereas a lower-atomic-number element offers greater accuracy in reproducing the desired pattern. FIB milling exploits so-called *knock-on sputtering*. For this to happen, the ion needs to be accelerated by a potential in the 1–50 kV range.[46] During FIB milling, an incoming ion hits a surface atom and transfers part of its kinetic energy to it, such that the atom is displaced from its equilibrium lattice position and collides with neighboring atoms, which can result in their release from the substrate. The incoming ion after several impacts loses almost entirely its primary energy and can be trapped in the target substrate, leading to ion implantation and a change in the properties of the target substrate. Although FIB ions can themselves also be exploited for imaging, so-called *dual beam* FIB instruments include an SEM column, which can be used for non-destructive electron imaging. Depending on the manufacturer of the FIB machine, there are differences in the procedure for the fabrication of an S-CGH. These differences are associated primarily with the electronics and software that manage beam scanning and patterning. Most of the following discussion is based on the authors' experience with FEI (now Thermo Fisher Scientific) instruments.

After an S-CGH is designed with the aid of a computer and dedicated software, the resulting image can be fed directly to the patterning software that comes with a dual beam machine, or it needs to be converted in a file format that can be read by the software. In the first case, the most common image file formats are.bmp or.png; in the second case, vectorial (.dxf or.gdsII) or stream files (.str) are used.

Generally speaking, any file format fed to the software will be used to tell the FIB controller where to position the beam and how long to stay at a certain position. A pixel position in the image is converted to a position in the coordinate system of the beam controller, while the pixel intensity is proportional to the time for which the milling beam spends at that position, i.e., the *dwell time*. This last parameter is what one can use to select between these formats. Most of the aforementioned image files are 8-bit ones, which means that the vertical milling resolution in the milling is limited to 256 intensity levels. If higher fidelity in the profile shape is needed, then a different file format is required. This usually translates into the need to use vectorial file formats (.dxf or.gdsII) or a direct coordinate and milling time file format such as stream files (.str), where the resolution in z dimension is no longer a limiting factor.

An additional distinction between image, vectorial, and stream files is the order in which the points in the pattern are scanned. For a picture or vectorial format, the FIB pattern handling software allows a choice of scanning direction (e.g., line by line or column by column, in different directions, or spiraling). In addition, all software packages typically allow to choose the number of passes across the sample. The total milling time can, therefore, be subdivided into longer dwell time for fewer passes or shorter dwell time for more passes. These aspects will be covered in detail later in this chapter.

1. Optional procedure: Au coating

In general, when performing an observation of a synthetic hologram using low-angle diffraction, a central spot and additional lateral spots can usually be identified. As outlined above, the central spot in the Fraunhofer plane is referred to as the 0th diffraction order, whereas the lateral spots are non-zero diffraction orders that arise from periodicities in the sample. The part of the electron wave function that impinges on the S-CGH, which contains the patterning periodicities, is diffracted and contributes to the intensities of the diffraction spots, i.e., electron holograms encoding the wave function of interest. All parts of the wave function that impinge on unpatterned areas in the surroundings of the S-CGH, along with unscattered electrons and a contribution from non-ideal S-CGH fabrication, will contribute to the intensity of the central spot. In order to minimize the intensity contribution from surrounding unpatterned areas which can suppress the contribution from the S-CGH, it is possible to first deposit a thick (∼150 nm) layer of Au by sputtering or evaporation, followed by FIB removal of this Au layer only in the area where the S-CGH will be fabricated. Although this procedure is rapid and straightforward, the downside is an increase in the SiN surface roughness due to the roughness of the Au coverage, which is subsequently projected onto the SiN surface after Au removal by FIB milling. An alternative Au coating procedure using EBL, which is less straightforward and more time-consuming, preserves the initial SiN surface roughness but may leave residues.

B. FIB milling calibration

After the file format is chosen, the milling process requires calibration to be able to mill reproducible patterns with a well-defined groove depth, and in the case of phase S-CGHs, to obtain a desired phase change. Several factors play a role in determining the milling yield, including the beam current, dwell time, and the number of passes. The calibration process requires to produce a series of simple patterns, controlled by one of the source files described above, each of which has the same size (i.e., number of pixels) but different milling times and different real pixel sizes. For example, it is possible to use a square-shaped pattern, in which half of the side is milled twice as much as the other half, such as shown in Fig. 12(a).

The next step involves reproducing the pattern on a membrane of known thickness. The milling time defined by the user should be quite long since the process has to be manually stopped once one side of the pattern (the white one in this case) completely breaks [as in Fig. 12(e)]. A clear sign that the membrane is about to break is the appearance of holes as in Fig. 12(d). By knowing the total pattern size, the beam current, the milling time, and the physical size of each pixel, it is possible to determine the dose, and therefore, the milling rate for that specific pixel size and current. Figure 12 shows an example of such a procedure, in which the total milling time is increased by changing the number of repetitions while keeping the pixel dwell time at 10^{-4} s, illustrating (c) slight bending, (d) local milling through, and (e) severe milling through the membrane. In this way, the milling depth can be measured and an estimate of the milling rate can be obtained.

Another critical parameter is the pixel size in the image or stream file. The pixel size is the area every pixel from the image will occupy on the substrate, and it is equivalent to the square of the distance between neighboring pixels. The pixel area can be varied in many ways. For instance, instruments controlled by a Raith scan and control unit typically allow the user to choose the pixel size once the pattern image is loaded, to modify the pattern and to impose custom sampling conditions and milling mode on the designed pattern. Instruments from Thermo Fisher Scientific, instead, define the pixel position in the imaging reference frame. Here, the pixel size is defined as the ratio between the desired S-CGH diameter (in nm) and the pixel-to-pixel distance (or step-size) times the CGH lateral solution. This means that, for example, the pixel size of a S-CGH of 50 μm fabricated starting from a CGH of $1024 \times 1024 px$ and step-size of 2 is

$$\begin{aligned}\text{pixel size} &= \frac{\text{S} - \text{CGH diameter}}{(\text{step size})(\text{CGH lateral resolution})}\\ &= \frac{50\,000\ \text{nm}}{2 \times 1024} \approx 24.41\ \text{nm}. \qquad (48)\end{aligned}$$

Moreover, since the pattern to be reproduced is shown in the FIB imaging reference frame, it does not scale by changing the

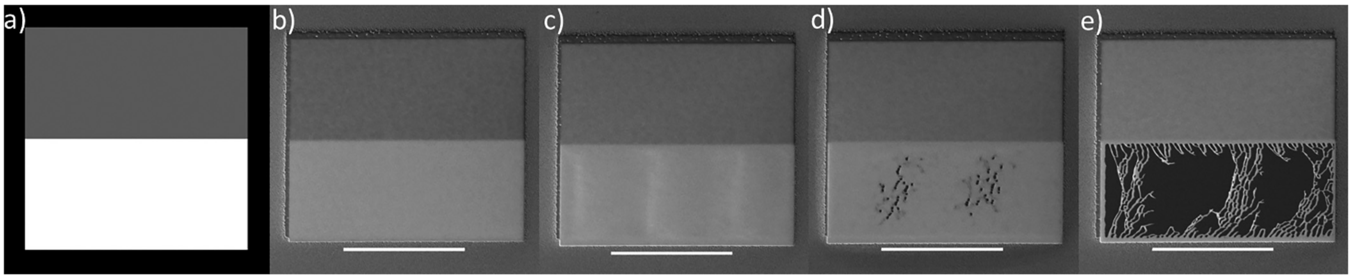

FIG. 12. (a) Representative calibration pattern image where white corresponds to a pixel of maximum intensity, i.e., to the longest dwell time, while black corresponds to "zero" intensity, i.e., to no milling (gray has half the value of white in this case). (b), (c), (d), and (e) are SEM images illustrating the fabrication of the pattern shown in (a) on a Si_3N_4 membrane. The white scale bar is 10 μm long. From (b) to (e), the total milling time and the number of repetitions were increased linearly. The pixel size was ~30 nm, the ion current was ~104 pA, and the accelerating voltage was 30 kV. Low electron energy was used to enhance surface sensitivity during imaging.

magnification (differently with respect to the software-defined patterns), then for the same pattern, the choice of magnification leads to different pixel sizes.

This information can be used as a starting point for pattern milling aimed at achieving the desired phase shift. Greater accuracy in the calibration may be achieved by repeating the procedure using different pixel sizes and ion beam currents. The calibration should, in principle, be valid while the ion beam aperture, which defines the beam current and spot size, remains unchanged.

This method is effective for determining the milling rate. By increasing the number of tests, it is possible to decrease the error statistically. As a rule of thumb, we repeat the procedure four to six times for each ion current that will be used for patterning.

Apart from ordinary surface profilometry using methods such as atomic force microscopy, complementary TEM measurements can help to improve the fabrication depth accuracy and to examine if a fabricated S-CGH works appropriately. These methods include low-angle diffraction (LAD), energy-filtered TEM (EFTEM), and low magnification off-axis electron holography; Whereas, LAD is available on most modern TEMs and can easily be used to achieve camera lengths of 1.4 km; EFTEM and low magnification off-axis electron holography are less commonly used. The first method requires an energy filter. The second method requires a biprism and the use of free lens control, which can damage the biprism if it is performed carelessly.

By fabricating a series of diffraction gratings that are identical to each other apart from the overall milling time, it is possible to estimate the correct milling time by comparing the diffraction intensity using LAD. For example, for an S-CGH with a sinusoidal modulation, the intensities of the central spot and the first order diffraction peak will depend on whether it has been properly milled. It is good practice to start with larger variations in milling time to be able to assess a wide range of parameters. Subsequently, the process should be refined using a smaller range of parameters. As a rule of thumb, such a process needs two to three iterations to find the best milling time and is therefore time-consuming. Before a good calibration is achieved, at least seven to ten patterns need to be optimized by changing the pixel size or milling current from one run to another. Furthermore, care should be taken to avoid a 2π ambiguity in the fabrication of a phase S-CGH when a large range of thickness values is explored.

EFTEM mapping is a complementary technique, which can be used to provide real space thickness information about the pattern. This technique exploits inelastic interactions between incident electrons and the sample, with scattered electrons losing a small amount of energy that can be measured using an energy filter. The proportion of electrons that have undergone inelastic scattering compared to electrons that have undergone elastic scattering or any scattering at all, can yield a value proportional to the local thickness by using the log-ratio method.[10] This value, multiplied by the electron mean free path, provides the local thickness, which can be compared to the intended thickness. In this way, it is possible to reconstruct an x-y map with additional thickness information. As the thickness determines the phase shift, it is possible to use the resulting thickness map in computer simulations of electron beam propagation to understand how the hologram's phase and amplitude information influence the details of LAD patterns.

The use of low magnification off-axis electron holography as an alternative method to validate the quality of a S-CGH and to calibrate the FIB machine requires setting up the TEM in a non-standard configuration. This technique allows the phase and amplitude of a large region of interest of a sample to be measured directly. The region of interest is usually limited to $30 \times 30\,\mu m^2$ and the approach requires the milling of a large window near the S-CGH for the reference wave. A linear gradient may need to be removed from the recorded phase image during post-processing.

It is necessary to point out that the calibration process needs to be repeated every time the substrate material for the S-CGH is changed. If more complex patterns are required, then the methods can provide valuable information for their fabrication. It is advisable to carry out a new calibration for every new pattern or experimental condition if high accuracy is required.

Figure 13 summarizes the fabrication process of an S-CGH using FIB milling. It also provides an intuitive recipe for calibration of the FIB machine. For simplicity and clarity, only the main steps are shown.

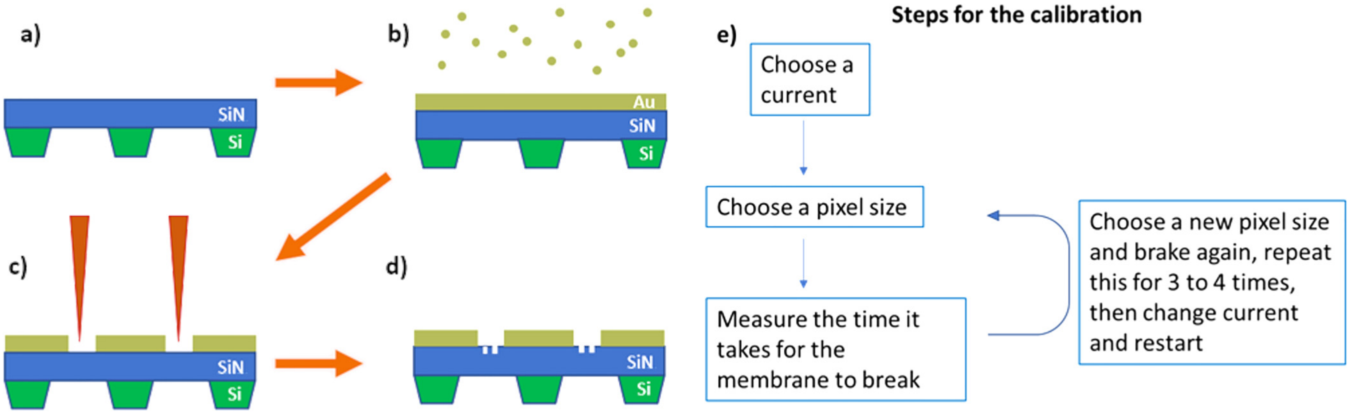
a)
SiN
Si
b)
Au
SiN
Si
c)
SiN
Si
d)
SiN
Si
e)
Steps for the calibration
Choose a current
Choose a pixel size
Measure the time it takes for the membrane to break
Choose a new pixel size and brake again, repeat this for 3 to 4 times, then change current and restart

FIG. 13. Schematic diagrams showing the typical steps in the fabrication of an S-CGH using FIB milling: (a) Fresh device; (b) Au evaporation; (c) Au removal and FIB patterning; (d) Grooves in the membranes; (e) Simple algorithm for the calibration process.

C. Optimization of FIB milling pattern reproducibility

Once the milling process has been calibrated, it is possible to start S-CGH fabrication. The calibration process focuses on estimating the milling rate of the FIB instrument, while the optimization process is used to fine tune the parameters to achieve an optimal result. Parameters that can be optimized include beam current, pixel size, distance and dwell time, the number of passes or repetitions of the pattern, and the scanning strategy. Even the membrane thickness before S-CGH milling will influence the result. This section contains some tips and tricks.

1. Optimization of ion current

The choice of the ion current is related to the choice of ion probe size, which ultimately defines the hologram resolution. The primary parameters that should be considered are the total milling time and the pixel size. The pixel size is related to the intrinsic resolution of the S-CGH, with finer details in the profile requiring a smaller pixel size. In general, a higher pattern resolution is desirable. However, there is a limit to how small the pixel size can be, since a resolution that is too high or hologram area that is too large can result in a file whose size cannot be handled by the patterning software, while a pattern resolution that is higher than the milling resolution will not be reproduced properly in the S-CGH.

A Ga ion source on a high-end instrument, at the lowest current, can have a spot size of approximately 5 nm or less. Although the size scales as the square root of the current, the patterning resolution also depends on other factors, such as the local milling time or instabilities, resulting in a larger effective spot size. A lower current is needed for higher resolution, at the cost of a longer patterning time as the sputtering rate depends on the current. However, long continuous patterning times have a higher probability that a drift of the stage or a beam defocus may occur. Although these effects can be reduced by using machines with interferometric stages and higher beam stability, normally a trade-off between ion current and total patterning time must be found. As a rule of thumb, patterning times longer than two hours are not recommended. For these reasons, the current should be chosen carefully to achieve the best resolution for a reasonable patterning time.

2. Optimization of dwell time, repetition number, pixel distance, and scan direction

The local milling time, or dwell time, is one of the parameters that can be optimized alongside the pixel distance (if available), the number of repetitions and patterning strategy, or scan direction. The dwell time influences the final shape of a milled pattern. Figure 14 shows an example of a box pattern, which illustrates the difference between using short dwell times with many repetitions and long dwell times with few repetitions, for the same total dose. The former approach [Fig. 14(a)] results in a rectangular box profile with mild redeposition on the sidewalls, while the latter approach [Fig. 14(b)] results in a sloped profile with redeposition effects along the horizontal direction of the serpentine scan.[46]

The use of too many repetitions can also be detrimental. A drift of a few nm can occur during the "homing" phase at the end of a repetition, though rarely, a small drift of a few nanometers can occur, leading to smearing of the end result. A trade-off between the number of repetitions and the dwell time is required, while avoiding the use of long dwell times and large numbers of repetitions.

The pixel-to-pixel distance can determine the amount by which adjacent pixels overlap. Clearly, the use of a very large pixel-to-pixel distance (i.e., a highly negative overlap) is detrimental, as the end result is a dotted pattern. Conversely, the use of a very short pixel-to-pixel distance increases the patterning time and file size. A −50 to 50% pixel overlap is ideal in the production of S-CGHs; however, pixel-to-pixel distance does not affect as much the final resolution of the S-CGH as other factors (primarily the ion current, i.e., the probe size).

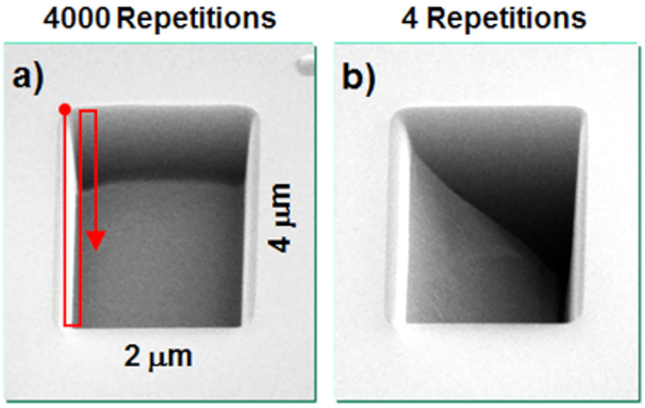

FIG. 14. SEM image (tilted view) of a box pattern milled using: (a) A short dwell time and many repetitions; (b) A long dwell time and few repetitions. The serpentine beam scan is shown using a red line.

The "scanning strategy" determines the path that the beam follows. The most common approach involves the use of zig-zag scanning, as shown in Figs. 15(a) and 15(b). An alternative approach involves spiral patterning, as shown in Fig. 15(c).[47] It is important to use the best possible scanning strategy because the scanning direction and path contribute to determining where the material is redeposited. For zig-zag scanning, redeposition is mainly found on the opposite side to the scanning direction, as shown in Fig. 14(b). If long rows are being patterned, it is then suggested to scan the beam along the rows instead of perpendicular to them. For spiral scanning, the continuous "back and forth" motion should allow for a "cleaner" result. However, only few examples have been presented in the literature.[47]

D. EBL for S-CGH

EBL usually requires a series of steps and controlled processes to achieve a final result, but can be used to produce features as small as a few nm and to mass-produce S-CGHs. The typical workflow for the production of an S-CGH is shown in Fig. 16. In this case, a negative resist is used. It is also possible to use a positive resist together with reactive ion etching to transfer the pattern, however, it usually leads to poorer results.

Calibration procedures are also needed for EBL. These are less time-consuming when using a negative resist such as hydrogen silsesquioxane (HSQ), which polymerizes into SiO_x when illuminated by an electron beam and has a mean inner potential similar to that of Si_3N_4. A standard procedure for selecting the dose involves creating a dose matrix of small features of the pattern that one wants to reproduce. The milling rate does not need to be considered in this case, since the resist thickness dictates the peak-to-valley height.

The steps required for preparing an S-CGHs using EBL can be summarized as follows. First, a layer of negative resist is spin-coated on the TEM membrane, patterned into the desired S-CGH enclosure shape and developed. A layer of Au or any other metal (with a high atomic number) is evaporated onto the device, with the metallic layer used to block a portion of the incoming beam. The use of an adhesion promoter of Cr or Ti is encouraged before depositing the metal of choice. The device is then immersed in a resist remover to achieve lift-off of the metallic layers that were on the previously developed resist, in order to prepare the canvas for the S-CGH. HSQ or another resist of choice can now be spin-coated to

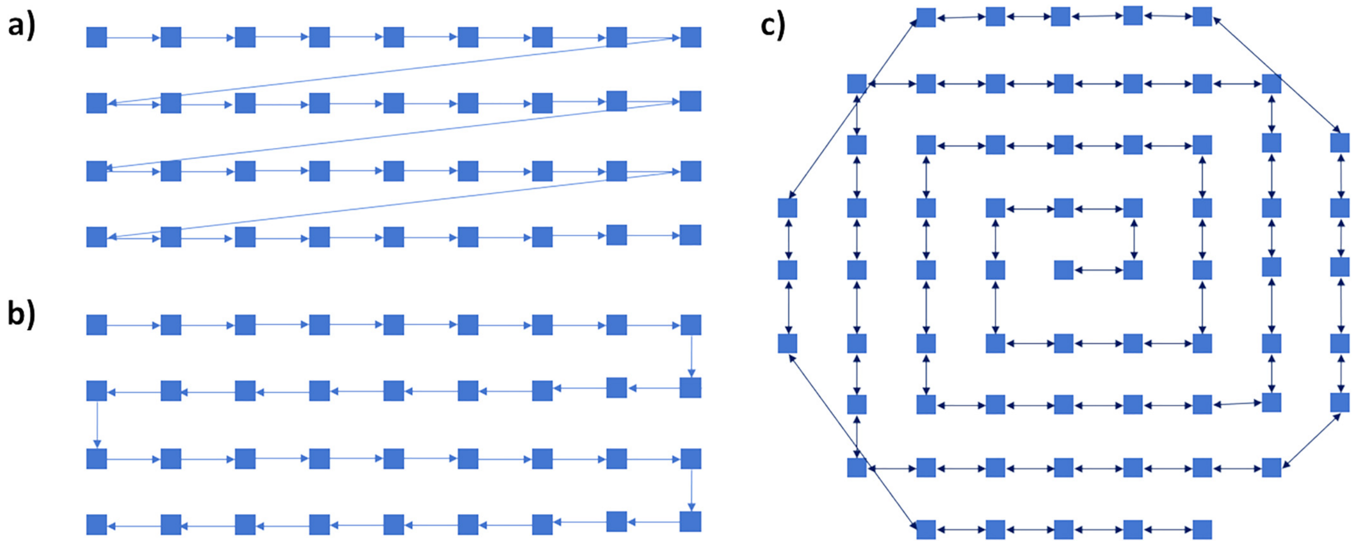

FIG. 15. Examples of patterning strategies: (a) and (b) show two different strategies for the zig-zag scanning, while (c) shows the less commonly used spiral scanning strategy. The distance between symbolic pixels has been increased for the ease of visualisation.

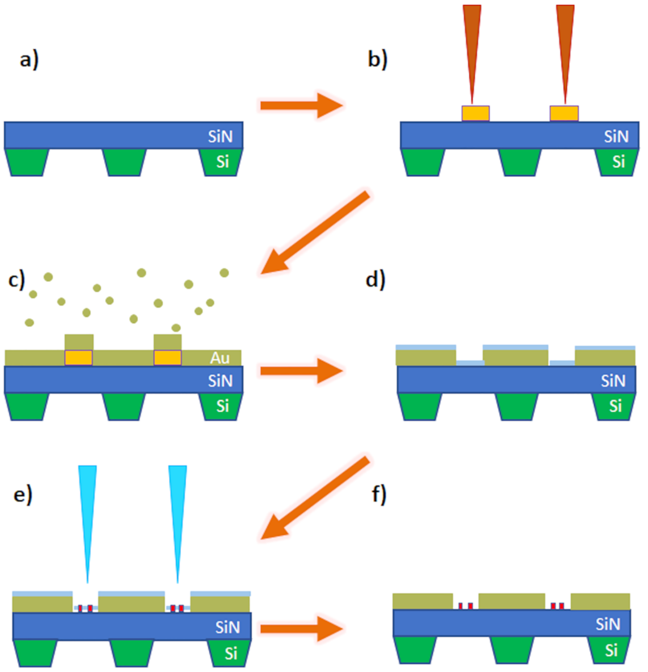

FIG. 16. Schematic diagrams showing: (a) A fresh device; (b) EBL and a developed (negative) resist, (c) Au evaporation, (d) lift-off and HSQ spin coating, (e) EBL, and (f) developing the HSQ.

a desired thickness, patterned, and developed. At this point, a few nm of metal or amorphous C can be flash-evaporated onto the developed pattern to balance the generation of secondary electrons in the TEM. More details about the fabrication process and the steps and exact parameters can be found in the paper by Mafakheri *et al.*[48] and many others related to the EBL technique.

Limitations of the EBL technique include the fact that the thickness is fixed, so one needs to fine-tune the spin coating process to achieve the required thickness for the phase shift. In addition, the pattern profile is either squared or sinusoidal and it is difficult to achieve a blazed profile. Most importantly, multiple steps are required to complete the process and the final devices are small and fragile, meaning that they have to be handled carefully during processing. However, the advantages of EBL are manifold. The Si_3N_4 membrane thickness can be reduced to only 15 nm as it is only a supporting layer, whereas for FIB milling, it is normally at least 75–100 nm before patterning. The use of a thinner membrane reduces inelastic scattering, background noise, and absorption. It also allows the use of a lower electron dose during patterning and results in the generation of fewer secondary electrons in the resist-supporting substrate, opening up the possibility to achieve sub-10-nm-sized features, if the process is well optimized.

Even for EBL-fabricated S-CGHs, it is possible to adjust the fabrication procedure to obtain finer details. As a result of the large number of steps, a tedious process of trial and error may be required. Examples of possible improvements include

- changing the pre-patterning baking temperature or adding a post-pattern baking step;
- searching for the proper dose and using proximity correction;
- adjusting the development temperature and time, as some resists provide higher contrast when they are developed at a lower temperature for longer than at room temperature,[49] while others behave in a similar manner when developed at higher temperature;[50]
- developing an understanding of the chemistry of the resist to find an optimal developer; and
- test the different thicknesses of silicon nitride or other supporting layers.

Some of the inherent limitations of EBL and FIB milling have recently been overcome by using a thermal scanning probe instead of an electron probe for patterning, resulting in higher accuracy and greater control in patterning depth and morphology.[51]

E. Experimental limitations of the use of synthetic holograms in microscopy

The use of S-CGHs can be effective for the realization of complicated phase patterns for wave front control. However, their primary drawback is that they are static. Their exchange with a different one in the aperture plane of the microscope usually requires the breaking of the vacuum of the microscope column. Alternative approaches, such as the use of multipoles of spherical aberrations correctors,[52] electrostatic fields,[16] or programmable phase plates[53,54] are still far from reaching the same level of arbitrary wave shaping with a similar number of pixels. Thin synthetic holograms are therefore still preferred for many experiments where a well-known effect is sought for, despite the fact that they require the insertion of additional material in the electron beam path, which can result in (1) inelastic scattering and decoherence; (2) a reduction in beam intensity; (3) contamination, damage, and aging of the device as a result of electron beam exposure; and (4) charging of the device during operation.

It should also be noted that the use of thin membranes as patterning media for S-CGHs typically suffers from local thickness variations on a scale of a few nm, resulting in a “frosted glass” effect that is similar to the effect on light crossing a turbulent or inhomogeneous medium. Even the elastically scattered part of the electron beam will therefore have a lateral spread in momentum due to the membrane. Furthermore, different forms of inelastic scattering will reduce the beam current and increase the lateral distribution.

Over time, the beam alters the groove profile from the desired phase profile. This effect is more significant if the synthetic hologram is in the condenser plane, where the electron beam current is higher. Experimentally, the quality of a synthetic hologram is found to deteriorate quickly due to contamination (local C deposition can form in only a couple of days). In contrast, damage (e.g., from knock-on effects and irradiation) tends to be slower, with minor profile alterations becoming apparent after one week of intensive use. It is, therefore, important to take care of vacuum quality in the TEM column and to be careful during operations such as sample exchange to decrease the probability of contamination. It is also important to avoid concentrating the electron beam to a spot on the S-CGH during any phase of operation. The most serious problem is potentially charging, in particular, because SiN is an insulating material, from which it can be difficult to dissipate the charge generated by the electron beam. As mentioned above, most of the membrane onto which the S-CGH is patterned is covered by a relatively thick Au layer that allows to dissipate the charge and the electrons only pass through the transparent area of the S-CGH. Whereas, the Au layer is efficient in removing charge and partially blocking the beam; the problem can persist in the uncovered area. Experimentally, in the steady state, synthetic holograms are often found to develop a charge density distribution that results in an approximately parabolic projected potential profile, which in turn adds a focusing effect to the hologram phase. It is possible to compensate for such an effect by using microscope lenses. However, the required compensation can depend on the electron dose, i.e., the higher the dose, the greater the effect. Furthermore, when using large synthetic holograms and unfavorable materials such as HSQ, a steady state is sometimes never reached and the additional phase contribution may vary over time. Possible solutions to this problem include the use of more conductive materials such as C, or coating both surfaces of the S-CGH with a thin layer of metal or C. The use of thinner synthetic holograms is also helpful. An alternative approach involves using amplitude S-CGHs, which are virtually all conductive. An example of this approach can be found in the paper by McMorran *et al.*[55] and Fig. 17(a) shows a CGH of a similar design to the one that they used to realize their amplitude synthetic holograms. As the presence of thin material bridges exposed to vacuum makes such structures mechanically unstable and difficult to fabricate, the structure can be strengthened by substituting the separate lines with a cross-grating. The diffraction orders are then dispersed in two directions, with an overall reduction in the efficiency of the order of interest and greater difficulty in isolating the desired beam, as shown in Fig. 17(b).

IV. EXAMPLES

A. Phase S-CGH design for the generation of electron beam vortices carrying orbital angular momentum

The generation of electron vortex beams (EVBs) was first demonstrated in 2009 and 2010 by three groups. The approaches involved using "spiral phase plates" constructed from thin films of graphite[56] and S-CGHs with pitchfork designs.[34,57] In some of the first experiments, EVBs were generated using amplitude S-CGHs or similar structures. Since then, most research groups have used phase or mixed amplitude-phase S-CGHs, which have higher efficiencies. New methods for the generation of EVBs have been presented[15,58,59] and the topic has matured sufficiently that most efforts are directed toward the measurement of OAM values and increasing applications in the fields of plasmonics, studies of magnetic materials, and chiral structures such as proteins. In a circular symmetrical reference system, an EVB has an angular-dependent helical phase term, which can be described by the expression

$$\varphi(l, \theta) = \ell\theta, \tag{49}$$

where ℓ is the OAM eigenvalue of the Schrödinger equation solved in cylindrical coordinate (also known as the topological charge or OAM quantum number) and θ is the angular coordinate (corresponding to the azimuthal angle). The wave function of a generic EVB is then given by the expression

$$\Psi_{helical} = A_0 e^{i\ell\theta}. \tag{50}$$

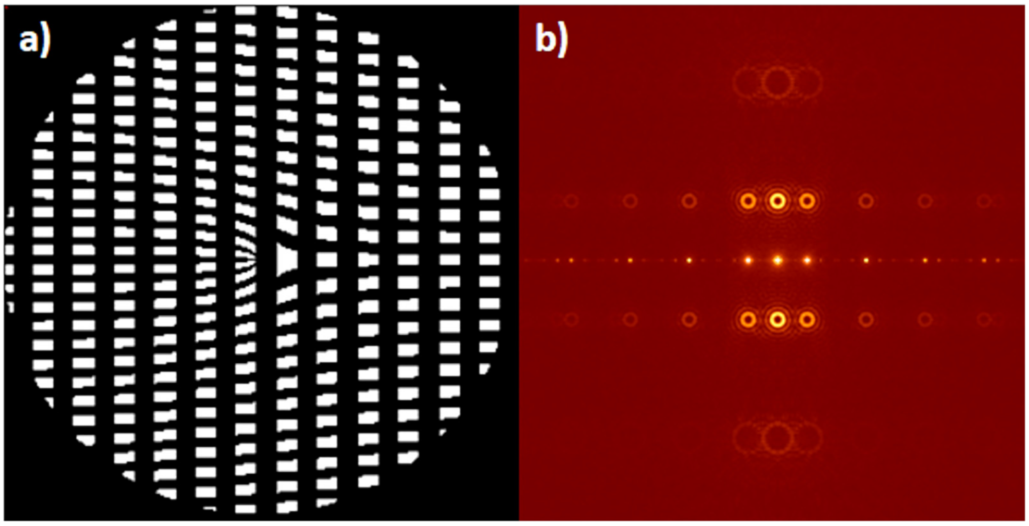

FIG. 17. (a) Example of an amplitude CGH with a grid-like structure for improved mechanical stability and (b) The resulting diffraction pattern, which forms a two-dimensional array of beams.

Some of the most prominent strategies for creating EVBs are described below. Further details about EVBs and vortex beams, in general, can be found elsewhere.[40,55,60–62]

1. Spiral design

The simplest way to generate an EVB using a S-CGH is to design an in-line[24,63] phase S-CGH that has a spiral/ helical form, similar to that shown in Fig. 3, in which a smoothly varying thickness profile is used to tune the phase shift imprinted on the wave front of the outgoing beam.

This design, in its simplest form, is an inline S-CGH and its realization requires good control of the fabrication process for the reasons outlined in Secs. III and IV. However, a well-calibrated machine makes fabrication straightforward. An EVB with topological charge ℓ can be generated using a spiral phase plate in which the total phase shift over a complete revolution is $\Delta\varphi = \ell \cdot 2\pi$.

A typical design of an EVB with a spiral phase is shown in Fig. 18(a), with phase ramps in six angular sections, in each of which the phase shift goes from 0 to 2π. The outgoing EVB, therefore, carries an OAM value corresponding to $\ell = 6$.

This design allows a superposition of EVBs to be generated. A beam that is generated from two superimposed and has no azimuthal current is referred to as a "petal beam." For example, a phase S-CGH can be used to generate an electron beam corresponding to a coherent superposition of $\ell = -5$ and $\ell = +5$ by summing the wave functions for EVBs with $\ell = 5$ and $\ell = -5$, and calculating the phase of the resulting wave function. Mathematically, the phase is $\Delta\varphi = arg(\sin(l\,\theta))$, corresponding to alternating values of 0 and π. Figure 18(d) shows the phase of a beam that carries a superposition of two beams with $\ell = \pm 5$, with white corresponding to a phase shift of π with respect to black areas. For a generic EVB generator with a spiral design, the enclosure is a circle just as for a conventional aperture and the physical dimension is typically $10 - 50\,\mu$m.

2. Pitchfork design

A pitchfork design can be used for both amplitude[56] and phase[48] off-axis S-CGHs, with EVBs generated in the n^{th} diffraction order, where n can vary from 1 to infinity. The design is based on an interference pattern between a plane wave $\Psi = A_0 e^{i(k_x x + k_z z)}$ and a helical wave in the $z = 0$ plane, in the form

$$I = 2|A_0|^2(1 + \cos(k_x x - \ell\theta)), \qquad (51)$$

from which it is possible to find the phase term of the interference wave to design the pitchfork S-CGH. As described in Sec. II E 1, in order to generate a pitchfork S-CGH, the argument of the profile

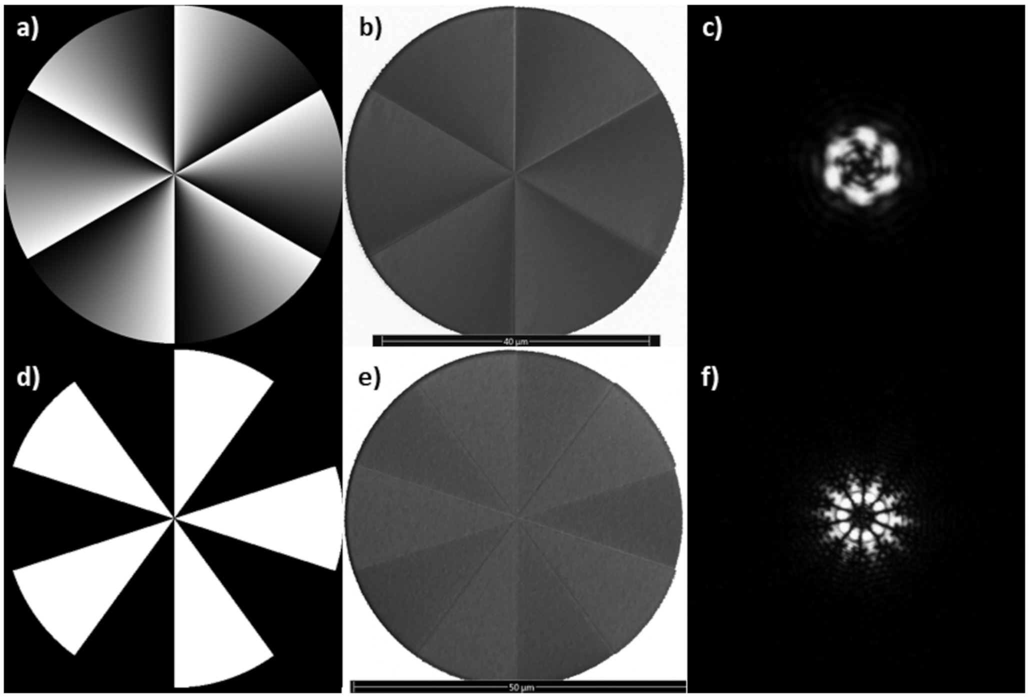

FIG. 18. (a) The phase of an EVB used to fabricate a phase-S-CGH with a spiral/ helical design for EVB generation for l = 6ℏ. The phase varies from 0 (black) to 2π (white) and goes from 0 to 12π over a complete revolution. (b) SEM image of a phase S-CGH corresponding to (a). (c) Experimental EVB in the Fraunhofer plane. (d) Phase and (e) phase S-CGH for a petal beam obtained from the coherent superposition of two EVBs with $\ell = 5$ and $\ell = -5$. (f) Experimental Petal beam in the Fraunhofer plane.

function is

$$\alpha(x, y) = \ell\theta + 2\pi x, \tag{52}$$

where x is one of the two in-plane coordinates, $\theta = ArcTan\left(\frac{y}{x}\right)$ and ℓ is the topological charge. The planar Cartesian coordinates x and y are expressed in units of the grating spatial period Λ.

Figure 19 shows the bi-dimensional profile functions $f(\alpha)$ for a pitchfork design with $\ell = 2$. Each pattern is obtained by combining the generic profile functions described in Sec. II E 4 and Eq. (52), such that

- $f_{sqrd}(\alpha) = \frac{1}{2}(1 + Sign(\sin(\ell\theta + 2\pi x)))$ [Fig. 19(a)],
- $f_{\cos}(\alpha) = \frac{1}{2}(1 + \cos(\ell\theta + 2\pi x))$ [Fig. 19(b)],
- $f_{trian}(\alpha) = \frac{1}{\pi}(Sign(\sin(\ell\theta + 2\pi x)))(\pi - Mod(\ell\theta + 2\pi x, \ 2\pi))$ [Fig. 19(c)],
- $f_{blzd}(\alpha) = \frac{1}{2\pi}(Mod(\ell\theta + 2\pi x, 2\pi))$ [Fig. 19(d)].

This design is versatile, as it can be used to generate both low-OAM and high-OAM EVBs. However, in the latter case, the features in the central part may be so small (in some cases even smaller than a pixel) that they are almost impossible to reproduce using either of the fabrication techniques discussed above. A common strategy involves masking out the central part up to a chosen radius. Although such a mask reduces the transmitted efficiency, an EVB with the correct OAM value is generated.[48]

3. Case study: Optimization and understanding of a blazed phase S-CGH with a pitchfork design

In recent years, we have worked on optimizing the fabrication process of a blazed phase S-CGH using FIB milling, in particular for a pitchfork with $\ell = 1$.[32] We have aimed at reaching the highest

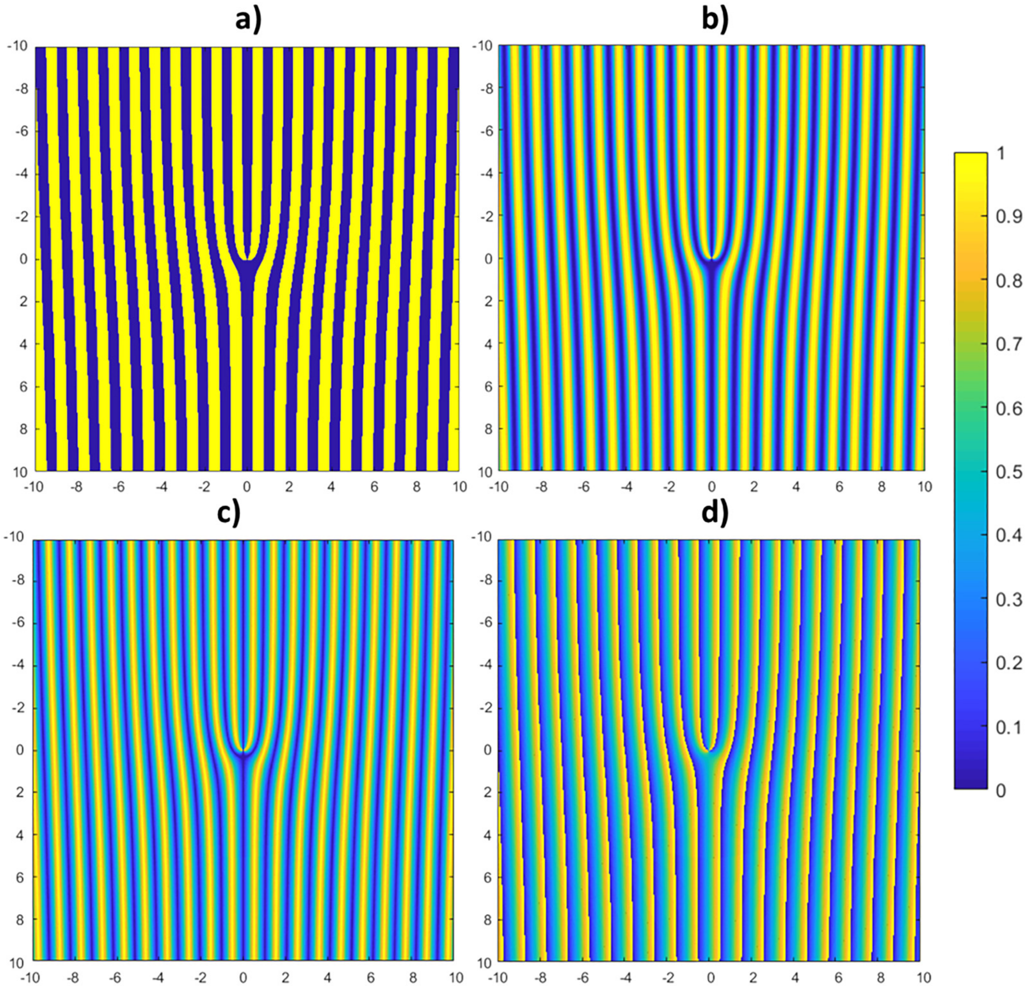

FIG. 19. Designs of a pitchfork S-CGH with $\ell = 2$ for (a) squared, (b) cosinusoidal, (c) triangular, and (d) blazed designs. The color bar represents the value of $f(\alpha)$ at the position of each pixel.

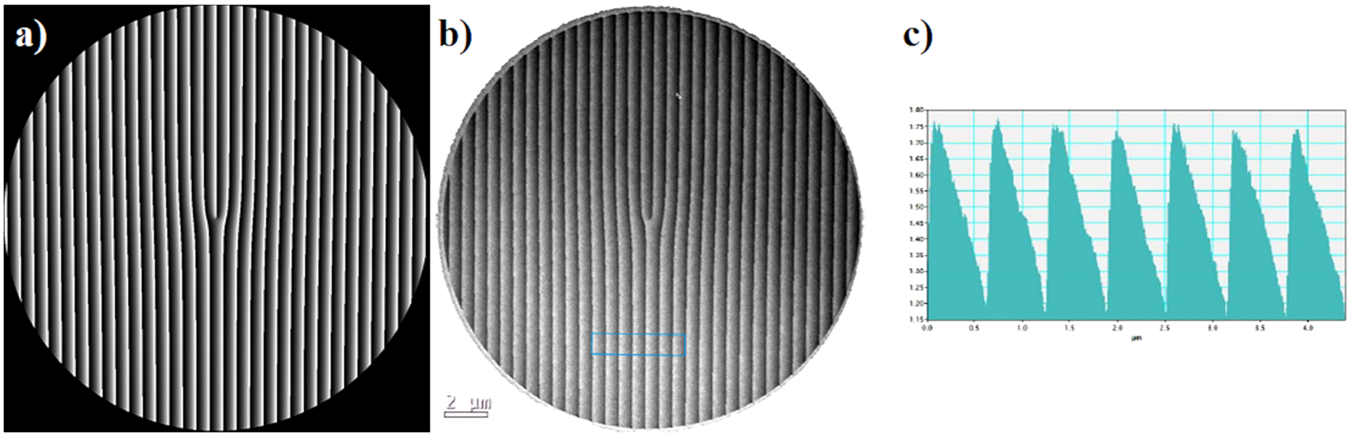

FIG. 20. (a) CGH for a blazed $\ell = 1$ pitchfork, (b) EFTEM thickness map of the fabricated S-CGH, and (c) profile of the region marked by a blue rectangle in (b).

diffraction efficiency for one of the two first diffraction orders (100%; see Table IV) by converging most of the intensity in the beam carrying the desired amount of OAM.

In order to reduce the number of variables, most parameters were kept constant, with only the number of passes and the maximum dwell time changed to tailor the phase shift and approach 2π. First, the number of passes was varied for rough optimization, then the maximum dwell time was optimized for finer optimization. The parameters that were kept constant and their values are given in Table V.

Figure 20 shows the CGH and the best-performing fabricated S-CGH. The patterning parameters for best performance, other than those reported in Table V, are

- Number of repetitions: 8 passes.
- Maximum dwell time: 91.6 μs.

These numbers can vary between both FIB machines and fabrication sessions, as factors such as laboratory environment, vacuum quality, and machine characteristics can influence the fabrication process.

The EFTEM image and line profile in Figs. 20(b) and 20(c) show that the in-plane periodicity of the pattern is ∼600 nm and the distance between peak and valley is ∼70 nm. This is slightly larger than the required value, which is ∼64 nm for 300 keV electrons, as reported in Table II. The shapes of the peaks approximate the ideal shape of a blazed profile, but differ slightly from one another, with sharp troughs but blunter peaks. These effects show some of the limitations of using FIB milling and contribute to the measured reduction in diffraction efficiency. Figure 21 shows that the best-performing sample was able to achieve 66.22% of the transmitted intensity in the +1st diffraction order, with the experimental diffraction intensity distributed between the orders in a different manner from that observed in Figs. 8 or 10(d).

TABLE V. Patterning parameters that were kept constant during the optimization of the fabrication of the $\ell = 1$ blazed phase S-CGH. The fabrication process was carried out on a FEI strata DB235M FIB-SEM equipped with a Ga+ source.

S-CGH diameter	Ion beam current	Ion beam accelerating voltage	CGH resolution	Step size	Effective pixel size of the S-CHG
20 μm	~260 pA	30 kV	1024 × 1024 px	2	9.8 nm

We used simulations to assess the origin of this behavior. First, we examined the effect of a non-ideal peak-to-valley phase difference by recalculating the intensity distribution for a ±10% phase mismatch from an ideal phase S-CGH. Figure 22 shows that even a 10% mismatch has almost a negligible influence on the diffraction intensity distribution, suggesting that the intensity distribution measured experimentally has a different origin. Although absorption affects the diffraction intensity, as shown in Sec. II E 4 and Fig. 8, it mainly decreases the total transmitted intensity, redistributing it almost evenly between the orders.

Even by considering the effect of both the absorption and the phase mismatch, it is still impossible to reproduce the same intensity distribution. We then focused on the profile shape of the S-CGH. Figure 24(c) shows that the actual shape is closer to a scalene triangle than to a blazed one. The scalene triangular profile function is

$$g(\alpha) = \begin{cases} Mod\left(\frac{1}{s}\alpha(\vec{\rho}), 2\pi\right) & for \;\; \alpha(\vec{\rho}) < s, \\ 1 + \frac{s}{(2\pi - s)} - Mod\left(\frac{\alpha(\vec{\rho})}{2\pi - s}, \; 2\pi\right) & for \; s \le \alpha(\vec{\rho}) < 2\pi. \end{cases} \tag{53}$$

This profile function is normalized between 0 and 1 and has its maximum for $\alpha(\vec{\rho}) = s$. The further s is from 0, the more it differs from an ideal blazed profile. Figure 23 shows the intensity distribution for $s = 1.1$. Although the shape difference was accentuated by choosing a high value of s, it is likely to be imperfect in the profile shape, including small differences between the shapes of adjacent "teeth,"

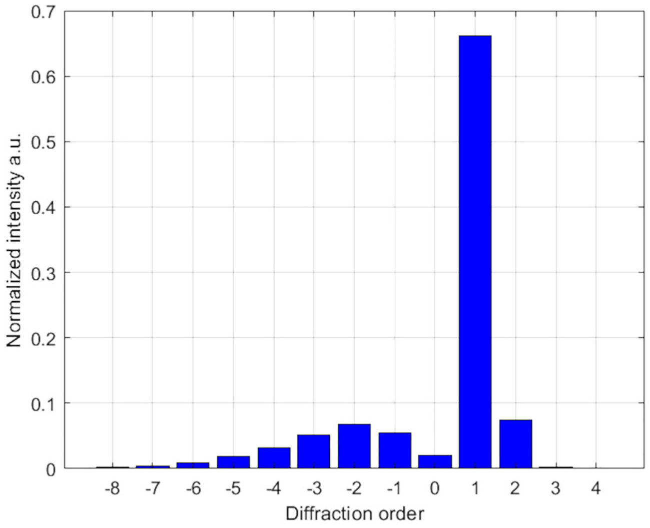

FIG. 21. Experimental distribution of diffraction intensities in the best-performing sample, with the total intensity normalized to unity.

which lead to the spreading of the diffraction intensity between the orders.

In summary, we have been able to model and discover the main factors that limit the operation of a real blazed phase S-CGH. Most of them are related to the limitations of the fabrication process. Imperfections of the shape and the phase mismatch usually result from the instabilities of the FIB machine. Optimization of the lateral resolution of the FIB machine is likely to bring the greatest improvement. Even a stage shift of only a few nm (for thermal or mechanical reasons) during the fabrication procedure can compromise the result. It may be possible to reduce some of the limitations by using a FIB machine that is designed for S-CGH production or by rethinking the fabrication steps. For example, it has been shown that gas-assisted FIB milling can improve the reproduction fidelity and patterning speed of blazed profiles.[64] However, some effects that arise from inelastic and diffuse scattering, including absorption and background noise, will always be present.

4. Generation of EVBs using Gaussian beams

The vortex beam generators that were described above are characterized by a hard aperture in the hologram plane. The resulting beams are sometimes referred to as "hypergeometric beams."[65] In light optics, a more suitable class of vortex beams has been derived based on a member of the Gaussian beam family: so-called Laguerre–Gaussian (LG) beams. Exact Gaussian beams are characterized by flat phase wave fronts at $z = 0$ and well-defined amplitude structures, with planes perpendicular to the optical axis that are equiphase surfaces. An in-depth mathematical description can be found in the book by Guenther.[66] An important parameter is the Gouy phase term, which is related to the transverse confinement of the beam and introduces anomalous behavior in the phase of the beam when it passes through the focus.[67–70] In a TEM, an exact Gaussian beam or a coherent Gaussian beam cannot be obtained easily. In fact, while in a TEM, the source emission intensity shape at the early crossover is typically Gaussian, this shape is an effect of partially coherent superposition of beams. Moreover, if the beam extent is limited by apertures, these generate diffraction effects that ruin the Gaussian intensity profile. However, it is still possible to generate a Gaussian-like beam that reproduces the intensity of an exact beam by converging the beam.

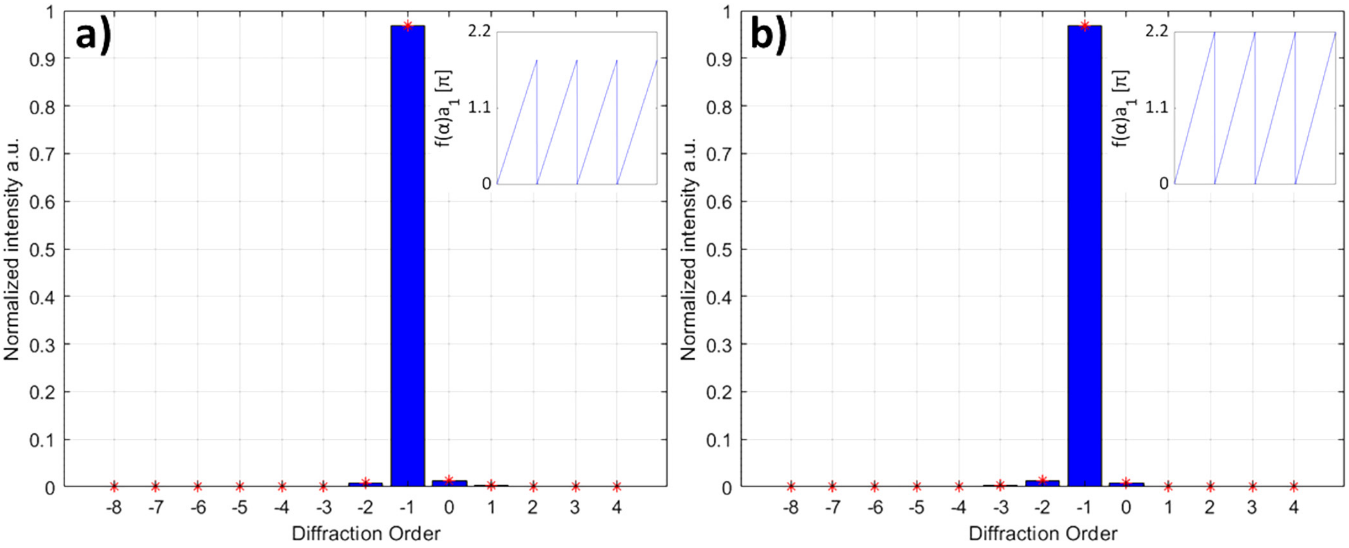

FIG. 22. Diffraction intensity distributions for phase mismatches of (a) −10% and (b) +10%. The insets show the corresponding groove profiles.

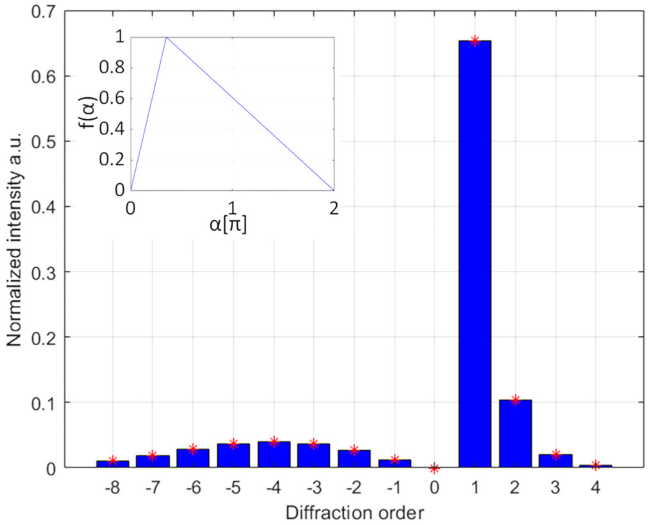

FIG. 23. Diffraction intensity distribution for a scalene triangular profile for s = 1.1. The inset shows a groove profile for 1 period.

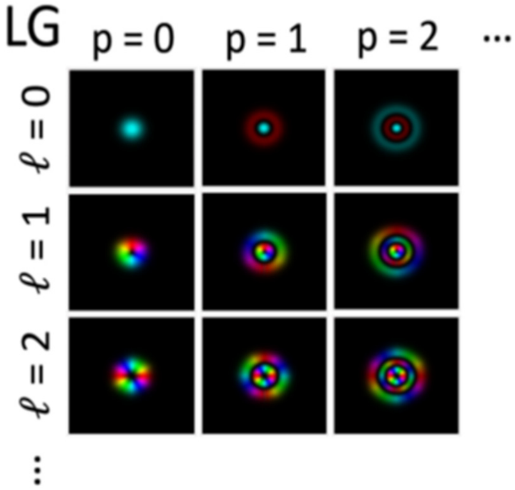

FIG. 24. Examples of Laguerre–Gauss beams with varying indices. The intensities and phase shifts of the wave functions are represented by their brightness and hue, respectively.

Laguerre–Gaussian beams are of greater interest than simpler Gaussian beams as they are solutions of the paraxial Helmholtz equation in cylindrical coordinates and are eigenstates of both the Fourier transform operation and OAM. In this way, they form a complete orthonormal basis characterized by two discrete quantum numbers p and ℓ, where ℓ is the azimuthal index or topological charge of OAM and p is a radial index, which defines the $(p+1)$ radial nodes in the intensity distribution. The wave function of a LG beam has the form[71]

$$\psi^{p}_{LG\ell}(\rho,\theta,z,t)=\frac{C_{\ell p}z_R}{\sqrt{z_R^2+z^2}}\left(\frac{\sqrt{2}\rho}{w(z)}\right)^{|\ell|}L_p^{|\ell|}\left(\frac{2\rho^2}{w^2(z)}\right)\exp(i(k_z z+\ell\theta-\omega t))$$
$$\times\exp\left(-\frac{\rho^2}{w^2(z)}+ik_z\frac{\rho^2}{2R(z)}\right)\exp(-i(2p+|\ell|+1)\xi(z)), \qquad (54)$$

where $L_p^{|\ell|}$ is the generalized Laguerre polynomial,[57] $C_{\ell p} = \sqrt{\frac{2^{|\ell|+1}p!}{(\pi(|\ell|+p)!)}}$ is a normalization factor, $w(z) = w_0\sqrt{1+\left(\frac{z}{z_R}\right)^2}$ is the beam waist radius along the propagation axis z, w_0 is the beam radius in focus, $z_R = \frac{k_z w_0^2}{2}$ is the Rayleigh range, $\xi(z) = \arctan\left(\frac{z}{z_R}\right)$ and $R(z) = z\left[1+\left(\frac{z_R}{z}\right)^2\right]$ is the radius of curvature of the complex wave front.

It is possible to demonstrate that the evolution of this kind of Gaussian beam along the optical axis is related only to the Gouy phase $\exp(-i(2p+|\ell|+1)\xi(z))$ and w_0, which makes it diffraction-shape-invariant, evolving only by the scale factor $\sqrt{1+\left(\frac{z}{z_R}\right)^2}$. This is a weaker condition for diffraction invariance than for Bessel beams, which are described in Sec. IV C, the difference being that Bessel beams are non-normalizable, and therefore, not exactly realizable experimentally. A series of simulated LG beams with varying indices are shown in Fig. 24.

LG beams are of interest to scientists working on magnetic materials and structured waves. For example, a LG wave function is functionally similar to a Landau state wave function.[72] By tuning a LG beam waist, it has been demonstrated experimentally that it is possible to couple them to Landau states.[73] Even though the generated LG beams were not pure, this experimental proof opens the possibility of observing transitions between the states. Furthermore, a LG beam has been used to demonstrate that it is possible to use paired S-CGHs for almost direct phase retrieval of EVBs (and of structured beams in general) in the Fraunhofer plane.[74] Pure LG beams are ideally generated using mixed S-CGHs,[75] as described in Sec. II F 1. The design and fabrication of mixed S-CGHs are reported in Fig. 25 for two experimental examples of LG beams with different characteristics. The first example [Figs. 25(a)–25(d)] shows a pure LG_0^{10} mode that has a simple circular structure. The second example [Figs. 25(e)–25(g)] shows two states with different OAM and p quantum numbers coherently summed together to give a superposition of LG modes with different radial and azimuthal indices. The phase in Fig. 25(f), which is the theoretical phase obtained by Fourier transforming the thickness profile of the hologram, illustrates the complexity of the beam. It can be considered as a proof of the power of amplitude and phase encoding in a single S-CGH for Laguerre–Gauss beam generation, and in general, EVB generation. In Fig. 25(d), there are no intensity ripples similar to those present in EVBs generated using a spiral design, as described in Sec. IV A 1. LG beam generation using different techniques has also been reported.[76]

LG beams are solutions of the paraxial Helmholtz equation in cylindrical coordinates, whereas Hermite–Gaussian (HG) beams

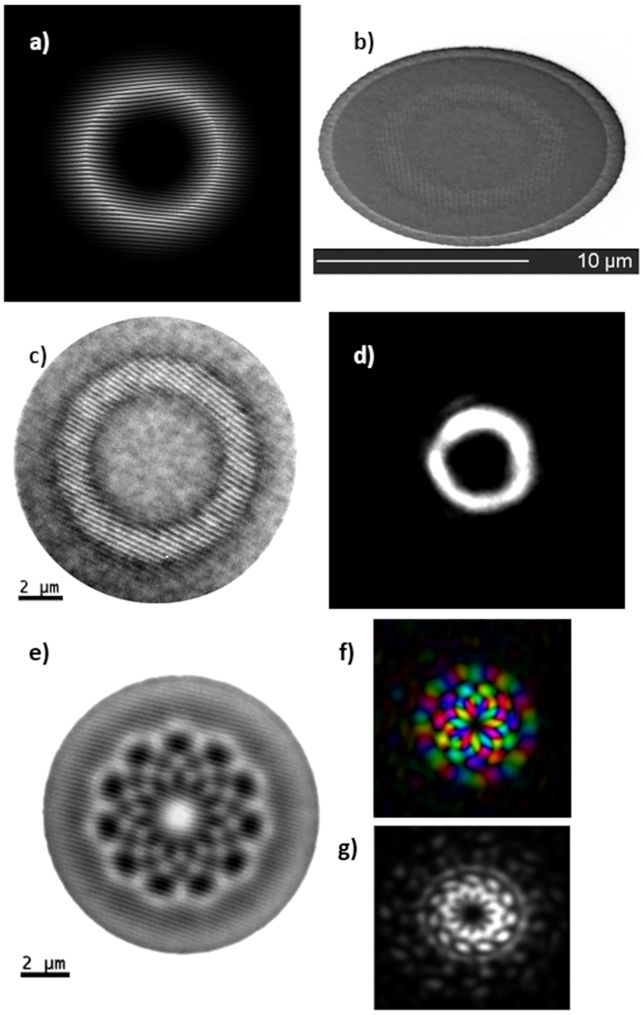

FIG. 25. Steps in the fabrication and validation of two LG beams: (a) Phase and amplitude utilized for fabrication, (b) tilted SEM image of the resulting mixed S-CGH (ion current: 300 pA; repetitions: 192; magnification: 3900×), (c) EFTEM map of the S-CGH, (d) diffraction image showing the "donut-like" shape of the generated EVB, (e) EFTEM map of superimposed LG beams, and (f) simulated amplitude and phase and (g) diffraction intensity of the 1st diffraction order.

are solutions of the same equation in Cartesian coordinates.[77] Although HG beams do not carry OAM, the first vortex beams generated by Allen *et al.* in 1992[77] were obtained by using a cylindrical lens to transform high-order HG modes into LG modes. In a TEM, it is possible to reproduce the effect of a cylindrical lens by increasing astigmatism. This approach has been exploited by Schattschneider *et al.*[78] to measure the OAM of an EVB and to measure the azimuthal (and radial) state for exact LG states.

B. Design and realization of a holographic OAM sorter

An interesting application of synthetic electron holograms is the development of a device that can be used to measure the OAM spectrum of an electron beam, referred to as an OAM sorter.[79,80] This device is composed primarily from two S-CGHs: an "unwrapper" S-CGH that unwraps an OAM-carrying electron beam and a "corrector" S-CGH that corrects the phase distortion introduced by the first S-CGH. The incoming electron beam contains the phase information of interest, after having interacted with a sample. The most straightforward example of an OAM-generating sample is the in-line S-CGH described in Sec. IV A 1. Figure 26 provides a schematic representation of the setup and transformations involved, including OAM generation, unwrapping, correction, and detection. In Fig. 26, an electron beam impinges on a generator S-CGH and is endowed with a spiraling phase shift with OAM = 1, corresponding to a 2π phase shift along one complete azimuthal path. The use of an in-line S-CGH simplifies the alignment of the beam on the sorter and excludes the effect of tilt (and off-axis aberrations).

The electron beam that is carrying OAM is directed onto the unwrapper S-CGH, which performs a conformal transformation from log polar to Cartesian coordinates. In this way, the phase information is unwrapped from an azimuthally varying arrangement to a linear arrangement, so that it is aligned along one Cartesian coordinate. The first S-CGH, or sorter 1 element, is a diffractive hologram. Therefore, the resulting pattern is found in a reciprocal plane. This unwrapping operation introduces a strong phase gradient. After the transformation, this additional phase must be removed. Therefore, the corrector is needed as an additional off-axis S-CGH. The final OAM spectrum is found in reciprocal space. After this correction operation, the OAM value can be found as an intensity spot, whose position from the center of the first diffraction order in reciprocal space is indicative of the magnitude of its OAM value. A calibration procedure using different OAM-generating S-CGHs with known OAM values is required, as each electron-optical configuration can introduce changes in rotation and magnification. Once a device is calibrated, a real sample, which imparts an unknown amount of OAM onto the electron beam, can be studied instead of the generator S-CGH. It is possible to measure the full OAM spectrum of a beam in one acquisition. Applications include measurements of the magnetic moments of dipoles,[80] as well as in EMCD[81,82] and plasmon characterization.[83]

S-CGH fabrication requires a knowledge of the functions that are to be encoded in the unwrapper S-CGH (Λ_1) and corrector S-CGH (Λ_2). The phase corresponding to the first element of the sorter is

$$\Lambda_1 = \varphi_0\, sign\left(\sin\left(2\pi a\left|y\arctan\left(\frac{y}{x}\right) + x\ln\left(\frac{\sqrt{x^2+y^2}}{b}\right) + x\right|\right)\right), \tag{55}$$

where a and b are parameters that are used to optimise the experimental efficiency, while *sign* denotes the sign function. The phase

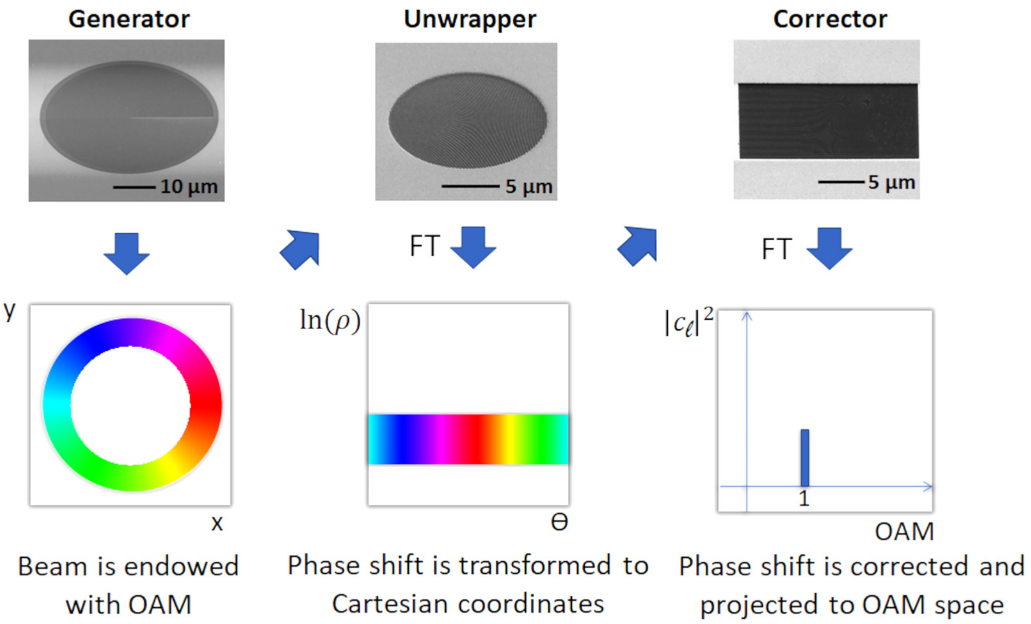
Generator
Unwrapper
Corrector
10 µm
5 µm
5 µm
FT
FT
y
x
$\ln(\rho)$
θ
$|c_\ell|^2$
1
OAM
Beam is endowed with OAM
Phase shift is transformed to Cartesian coordinates
Phase shift is corrected and projected to OAM space

FIG. 26. A sequence of holographic masks and transformations involved in the use of an OAM sorter.

corresponding to the second element of the sorter is

$$\Lambda_2 = \varphi_0 \; sign\left(\sin\left(2\pi ab\mathrm{exp}\left(-2\pi\frac{u}{a}\right)\cos\left(2\pi\frac{v}{a}\right)\right) + 2\pi cv\right), \qquad (56)$$

where $u = -a\ln\left(\sqrt{x^2 + y^2}/b\right)$, $v = a\arctan\left(\frac{y}{x}\right)$ and c is an additional scaling parameter. For the unwrapper shown in Fig. 26, the parameters were $a = 2$, $b = 0.01$, and $c = 0.6$. They can be tuned to match the relative S-CGH and holographic beam sizes. The peak-to-trough depth should maximize the diffraction efficiencies. Although the device can be used in any TEM, the electron-optical configuration is challenging and the use of free lens control and additional sets of lenses and apertures is recommended.

It is worth to point out that while the phase of the hologram is everywhere finite, the origin should be mapped on a segment where $u = -\infty$, meaning that the cusp in the origin has a gradient that is divergent. Of course, while designing the CGH, we are not able to capture this since we are bound to a maximum value of u due to the limitations imposed by the computer. In all virtual calculations we did, the origin is the only problematic point since the gradient remains finite and relatively small in all remaining pixels. The mapping is, therefore, everywhere correct except for the central pixel, but this effect is negligible with respect to the overall beam intensity.

C. Bessel beam

A third example of a possible application of S-CGHs is the generation of a non-diffractive Bessel beam. Bessel beams were first mathematically modeled by Durnin.[84] Experimentally, they were realized as photon quasi-Bessel beams, an approximation of Bessel beams, which had the same properties over finite distances.[85] Durnin and colleagues defined Bessel beams as beams "whose central maxima are remarkably resistant to the diffractive spreading commonly associated with all wave propagation."[86,87] A Bessel beam can be considered as a coherent superposition of conical plane waves, or as a set of plane waves propagating on a cone. Apart from being non-diffractive, they are also "self-healing," so that (apart from an overall decrease in intensity) they can recover their intensity profile. Moreover, a zeroth order Bessel beam has a smaller central spot diameter and longer depth of the field than other ordinary beams.[88]

In light optics, the generation of Bessel beams, or more precisely quasi-Bessel beams, has been achieved in many ways. The simplest approach is to use an annular slit or ring aperture.[84] This method works since the Fourier transform of a Bessel beam is a ring. A more efficient method is to use axicon lenses,[89–92] which remove the on-axis intensity oscillation, resulting in a smooth intensity variation in the beam propagation direction. Other methods are based on S-CGHs,[93] SLMs,[94,95] and cavities.[96,97]

In recent years, by taking inspiration from light optics, different methods have been adopted for the generation of electron quasi-Bessel beams. In 2014, Grillo *et al.* reported the use of an S-CGH to generate non-diffractive quasi-Bessel beams that were able to propagate for 0.6 m without noticeable spreading of their central maximum and could reconstruct.[37,98] Taking inspiration from the initial experiments by Durnin and colleagues used annular slits to generate quasi-Bessel beams.[99] In 2017, Zheng and colleagues used magnetic vortices with circular magnetic moment distributions, which are naturally present in soft magnetic thin films, as axicon lenses.[100] A generic Bessel beam wave function can be expressed in the form

$$\psi(\rho, \theta, z; t) = J_n(k_\rho \rho) e^{in\theta} e^{i(k_z z - \omega t)}, \tag{57}$$

where ρ, θ, z are cylindrical coordinates, J_n is the nth order Bessel function of the first kind, n is an integer, k_ρ and k_z are the transverse and longitudinal components of the wave vector, respectively, and $k^2 = k_\rho^2 + k_z^2 = \frac{2m\omega}{\hbar} = \left(\frac{2\pi}{\lambda_{dB}}\right)^2$, where m is the electron mass, $\hbar$ is the reduced Planck constant, and λ_{dB} is the electron's de Broglie wavelength. A Bessel wave function is a well-known non-normalizable solution of the Schrödinger equation of a free electron in cylindrical coordinates. From Eq. (57), it is possible to notice that the probability density is independent of both time and z, and is equal to $J_n^2(k_\rho \rho)$.

The phase S-CGH used by Grillo *et al.*[98] imprints on the transmitted beam the phase modulation

$$\varphi(\rho, \theta) = \varphi_0 sgn[\cos(k_\rho \rho + n\theta + g\rho\cos\phi)]. \tag{58}$$

The resulting off-axis Fresnel hologram has carrier frequency $g = \frac{2\pi}{\Lambda}$, where Λ is the grating spatial period. In this formula, the chosen profile shape was a squared one with argument $\alpha(\rho, \theta) = k_\rho \rho + n\theta + g\rho\cos\theta$, where n is the OAM topological charge. The $\alpha(\rho, \theta)$ that was used is similar to that in Eq. (52), i.e., the pitchfork design. The resulting quasi-Bessel beam was an OAM-carrying one. Figure 27 shows a fabricated phase S-CGH for quasi-Bessel-beam generation, the CGH that was used to produce it and an experimental diffraction image, in which it is possible to observe the generated quasi-Bessel beam. The typical dislocation of a pitchfork design is visible at the center of Fig. 27(b).

In a later paper,[37] by switching to a cosinusoidal profile and optimizing the fabrication procedure, Grillo and colleagues increased the transmission efficiency by 37 ± 3%. They pointed

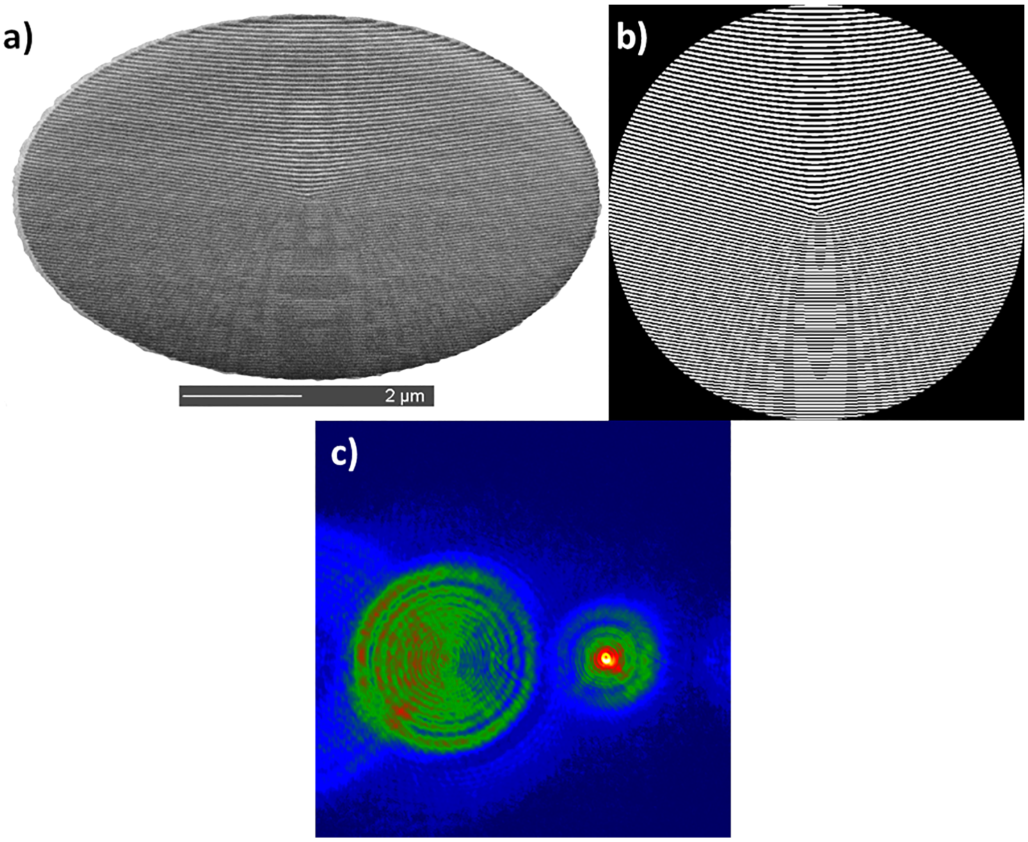

FIG. 27. (a) SEM image of a phase S-CGH used for the generation of quasi-Bessel beams with $\ell = 2$, recorded with the stage tilted to highlight the three-dimensional features. (b) Profile function used to obtain (a) by FIB milling. (c) Experimental diffraction image. The first diffraction order on the right shows a quasi-Bessel beam. The parameters used in Ref. 98 were used to create (b).

out possible application fields of quasi-Bessel beams generated using S-CGHs: smaller aperture radii are best suited for STEM, while larger radii are best suited for interferometry. Applications of such structured beams in electrons include classical techniques such as tomography[101] and strain mapping,[102] as well as conventional STEM, low dose STEM, and HR-STEM.[37,98,103,104]

D. C_S corrector

As a final example of beam shaping, we would like to show that one does not rely on the scheme of Sec. II E. A problem that has long affected electron microscopy is the presence of spherical (C_S) aberration in any magnetic lens that has cylindrical symmetry. Although a solution has been found by using a complicated set of multipoles,[105–107] it is interesting to determine whether one can produce an S-CGH that is able to compensate for spherical aberration by introducing, in the condenser aperture plane, an equal phase of opposite sign to that of the C_S aberration. The desired phase α is not known in the diffraction plane, but directly in the S-CGH plane. Therefore, the aim is to correct the C_S aberration in a STEM probe by using an aperture in the condenser plane. An early realization of an S-CGH able to compensate spherical aberration was realized by Shiloh *et. al.*[108] using on-axis correctors with wrapped phase (in here, they also demonstrated that it was possible to use S-CGHs to compensate also for other important aberrations such as two- and three-fold astigmatism). Later, different groups[109–111] have produced holograms using slightly different recipes. It is also worth mentioning that recently a novel proposal for shaping electron beams for Cs correction using optical fields has been reported.[112]

The general formula of the argument of the phase profile is

$$\alpha(\rho) = \frac{2\pi}{\lambda}\left(-\Delta f\rho^2 + \frac{1}{4}C_s\rho^4\right) + g\rho\cos(\theta). \tag{59}$$

The inline (on-axis) approach is recovered when $g = 0$. For correction to be applicable over a wide field (beyond a standard STEM probe), it is necessary to have a phase ranging over $4 - 6\pi$. One can use either a continuous slope with a thickness $t = \alpha$ or a discontinuous slope $t = Mod(\alpha, 2\pi)$ with 2π phase wraps. The first approach results in a thick membrane and significant absorption, while the second approach requires precise tuning of the discontinuities.

An inline version with a large value of Δf can be used to create many beams that are in focus at different values of the z coordinate.[113] Although any kind of groove, such as a sinusoid can be used, this approach has not been used so far [see Fig. 28(b)]. Here, we describe the off-axis approach, which allows excellent control of the phase by employing a thin membrane at the cost of spurious diffraction orders. Grillo *et al.*[94] explained how to remove such spurious orders by the smart use of optics components. Figure 28(a) shows a typical aberration function in the presence of defocus ($C_S = 0.5$ mm and $\Delta f = 40$ nm). The corresponding off-axis phase S-CGH and a realization are shown in Figs. 28(c) and 28(d), respectively.

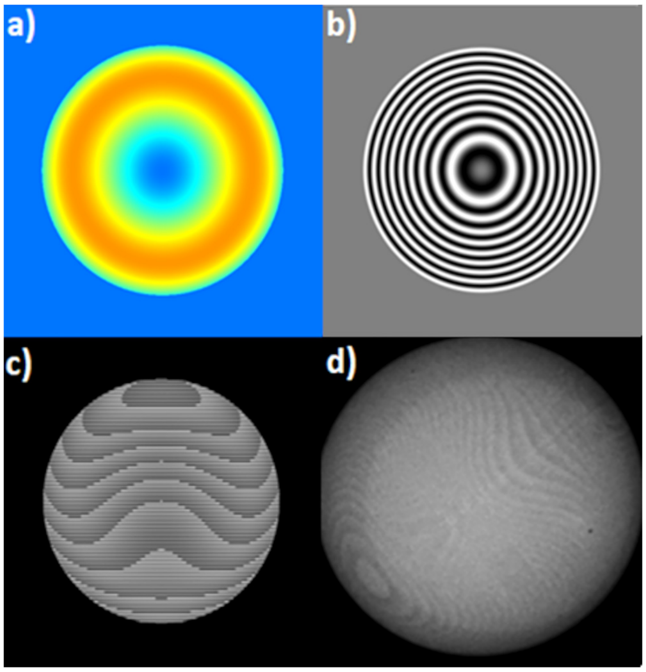

FIG. 28. Design of a holographic C_S corrector. (a) Desired phase plate, (b) inline S-CGH with a sinusoidal groove for $\Delta f = 400$ nm, (c) Off-axis S-CGH, and (d) realization in silicon nitride.

For the realization of the hologram, it is preferable to use a sinusoidal or a blazed groove shape, which ensures a smoother variation of the phase, as described in Sec. II G. For practical reasons, the carrier frequency must be quite large so that isolating a specific beam in the diffraction plane (with the desired C_S value) is easier. These constraints naturally lead to the use of very large holograms, resulting in challenges in fabrication and durability, as mentioned in Sec. III E.

V. CONCLUSIONS

In this Tutorial, we have reviewed the concept of "imaging" holography and explored synthetic holography, from the use of CGHs to simulate interference patterns to the possibility of engineering wave functions and new techniques in materials science. We have provided mathematical descriptions of the most commonly used groove profiles in amplitude and phase S-CGHs, discussing their efficiency, and the design of mixed phase and amplitude S-CGHs. We have described two fabrication techniques that can be used to manufacture S-CGHs, including their optimization and limitations. Finally, we have provided examples of possible uses of S-CGHs in the field of electron vortex beams. We have highlighted the fact that real phase S-CGHs are sensitive to imperfections introduced by fabrication, reducing their efficiency.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme (Grant No. 766970, project "Q-SORT;" Grant No. 856538, project "3D MAGiC;" Grant No. 823717, project "ESTEEM3," Grant No. 964591, project SMARTelectrons; Grant No. 101035013, Project "MINEON"), Deutsche Forschungsgemeinschaft (Project-ID 405553726—TRR 270), DARPA TEE program (Grant No. MIPR# HR0011831554), and from an Ontario Early Researcher Award (ERA) and a Canada Research Chair (CRC).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: MATHEMATICAL DEMONSTRATION FOR FOURIER COEFFICIENTS CALCULATIONS

This appendix contains the mathematical demonstration required to carry out the integrals to calculate the Fourier coefficients in Appendixes B and C.

Given a periodic function $f(\alpha(\vec{\rho}))$ its Fourier transform can be written as

$$F(K) = \int f(\alpha(\vec{\rho}))\exp(ik\vec{\rho})d\vec{\rho}, \tag{A1}$$

but $f(\alpha(\vec{\rho}))$ can also be developed as

$$f(\alpha(\bar{\rho})) = \sum_n c_n \exp(i\alpha(\vec{\rho}))_, \tag{A2}$$

where $c_n = \int f(\alpha)\exp(in\alpha)d\alpha$, so that the Fourier transform can be rewritten as

$$F(K) = \int \sum_n c_n \exp(i\alpha(\vec{\rho}))\exp(ik\vec{\rho})d\vec{\rho}_. \tag{A3}$$

We can then invert the sum and integral and obtain

$$F(K) = \sum_n c_n \int \exp(i\alpha(\vec{\rho}))\exp(ik\vec{\rho})d\vec{\rho}. \tag{A4}$$

In this expression, it is possible to notice that the terms in the integral are the diffraction orders. This relation holds in general as long as all the relevant $\exp(in\alpha(\bar{\rho}))$ are bandwidth-limited functions.

We can also calculate the squared modulus of the Fourier transform

$$|F(K)|^2 = \sum_n c_n^2 |\exp(i\alpha(\vec{\rho}))\exp(ik\vec{\rho})d\vec{\rho}|^2_, \tag{A5}$$

which tells us that, in the limit that no diffraction order overlap with each other, the $|c_n|^2$ coefficient of development of the function $f(\alpha(\vec{\rho}))$ indicate also the efficiency of the complete diffraction order.

APPENDIX B: CALCULATION FOR PHASE S-CGHs

This appendix contains in-depth calculations of the different profiles for phase S-CGHs.

1. Cosine profile

A sinusoidal/cosinusoidal profile can be described by the periodic function $f(\alpha) = \frac{1}{2}(1 + \cos(\alpha(\vec{\rho}))$, with a transmission function of the form

$$\begin{aligned} T(\vec{\rho}) &= e^{i\frac{\tilde{a}}{2}\cos(\alpha(\vec{\rho}))}e^{i\frac{\tilde{a}}{2}} = e^{i\frac{a_1}{2}\cos(\alpha(\vec{\rho}))}e^{-\frac{a_2}{2}\cos(\alpha(\vec{\rho}))}e^{i\frac{\tilde{a}}{2}} \\ &= e^{ia'_1\cos(\alpha(\vec{\rho}))}e^{-a'_2\cos(\alpha(\vec{\rho}))}e^{i\tilde{a}'}, \end{aligned} \tag{B1}$$

where the primed variables are used to simplify the calculations. $T(\vec{\rho})$ can be rewritten using the Jacobi-Anger expansion

$$T(\vec{\rho}) = e^{i\tilde{a}'}\sum_{n=-\infty}^{\infty}(i^n J_n(\tilde{a}'))e^{in\alpha(\vec{\rho})}, \tag{B2}$$

where the Fourier series coefficients of $T(\vec{\rho})$ are

$$\tau_n = i^n J_n(\tilde{a}')e^{i\tilde{a}'}. \tag{B3}$$

$|\tau_n|^2$ can be plotted with the complex argument Bessel function while an analytical approximation can be derived from Eq. (B1). The first term of $T(\vec{\rho})$ can be rewritten using the Jacobi–Anger expansion, while the second term can be expanded by making use of the approximation $a'_2 \ll 1$, resulting in the expression

$$T(\rho) \approx \exp(ia)\exp(ia_1\cos(\alpha))(1 - a_2\cos(\alpha)). \tag{B4}$$

Then,

$$T(\rho) \approx \exp(i\tilde{a}')\left[\sum_n(-i)^n \mathrm{J}_n(a'_1)\exp(in\alpha)\right](1 - a'_2\cos(\alpha)). \tag{B5}$$

If the Fourier coefficient is defined according to the expression

$$\tau_m = \int T(\rho)\exp(-im\alpha)d\alpha \tag{B6}$$

and we use the property of convolution

$$\begin{aligned} \tau_m = \exp(i\tilde{a}')&\left\{\int\left[\sum_n(-i)^n \mathrm{J}_n(a'_1)\exp(in\alpha)\right]\exp(-im\alpha)d\alpha\right\} \\ &\times\left\{\int(1 - a'_2\cos(\alpha))\exp(-im\alpha)d\alpha\right\}, \end{aligned} \tag{B7}$$

then we can develop the terms

$$\int \left[\sum_n (-i)^n \mathrm{J}_n(a'_1)\exp(in\alpha)\right]\exp(-im\alpha)d\alpha = \sum_n \delta_{m,n}\mathrm{J}_n(a'_1)(-i)^n, \tag{B8}$$

$$\begin{aligned}\int (1 - a_2\cos(\alpha))\exp(-im\alpha)d\alpha &= \int \left(1 - \frac{a'_2}{2}(\exp(i\alpha) + \exp(-i\alpha))\right) \\ &\quad \times \exp(-im\alpha)d\alpha \\ &= \delta_{m,0} - \frac{a'_2}{2}\left(\delta_{m,1} + \delta_{m,-1}\right),\end{aligned} \tag{B9}$$

to obtain the final coefficient in the form

$$\tau_m = (-i)^m(\mathrm{J}_m(a'_1))\left[\delta_{m,0} - \frac{a'_2}{2}(\delta_{m,1} + \delta_{m,-1})\right]\exp(i\tilde{a}'). \tag{B10}$$

If we now make use of discrete convolution according to the expression

$$\tau_m = \exp(i\tilde{a}')\sum_k (-i)^k(\mathrm{J}_k(a'_1))\left[\delta_{k,\, m-0} - \frac{a'_2}{2}(\delta_{k,\, m-1} + \delta_{k,m+1})\right], \tag{B11}$$

then

$$\tau_m = (-i)^m\mathrm{J}_m(a_1) - \frac{a_2}{2}((-i)^{m+1}\mathrm{J}_{m+1}(a'_1) + (-i)^{m-1}\mathrm{J}_{m-1}(a_1)]\exp(i\tilde{a}'), \tag{B12}$$

$$\tau_m = (-i)^m\left\{\mathrm{J}_m(a_1) + \frac{ia_2}{2}[\mathrm{J}_{m-1}(a_1) - \mathrm{J}_{m+1}(a_1)]\right\}\exp(i\tilde{a}').$$

The efficiency of the n^{th} diffraction order is proportional to $|\tau_n|^2$, where

$$|\tau_m|^2 = \left[J_m^2(a'_1) - \frac{a'^2_2}{4}(J_{m-1}(a'_1) - J_{m+1}(a'_1))^2\right]e^{-2a'_2}. \tag{B13}$$

2. Squared profile

A periodic grating function with a squared profile $f(\alpha) = \frac{1}{2}Sign|\sin(\alpha(\vec{\rho}))|$ has the transmission function

$$\begin{aligned}T(\vec{\rho}) &= e^{i\frac{\tilde{a}}{2}Sign(\sin(\alpha(\vec{\rho})))} = e^{i\tilde{a}'Sign(\sin(\alpha(\vec{\rho})))} \\ &= e^{ia'_1 Sign(\sin(\alpha(\vec{\rho})))}\, e^{-a'_2 Sign(\sin(\alpha(\vec{\rho})))},\end{aligned} \tag{B14}$$

where the $\tilde{a}$ is primed to take into account that the amplitude of $Sign(\sin(\alpha(\vec{\rho})))$ is half the peak-to-valley distance, simplifying the calculation. The Fourier coefficients of $T(\vec{\rho})$ can be calculated from the expression

$$\tau_n = \frac{1}{2\pi}\int_0^{2\pi} e^{i\tilde{a}'Sign(\sin(\alpha(\vec{\rho})))}e^{-in\alpha(\vec{\rho})}d\alpha. \tag{B15}$$

As a result of the properties of the *Sign* function, Eq. (B15) can be written

$$\begin{aligned}\tau_n &= \frac{1}{2\pi}\left[\int_0^{\pi} e^{i\tilde{a}'}e^{-in\alpha(\vec{\rho})}d\alpha + \int_0^{\pi} e^{-i\tilde{a}'}e^{-in\alpha(\vec{\rho})}d\alpha\right] \\ &= \begin{cases}\cos(\tilde{a}')\ for\ n = 0, \\ 0\ for\ n\ even, \\ \dfrac{2\sin(\tilde{a}')}{n\pi}\ for\ n\ \ odd,\end{cases}\end{aligned} \tag{B16}$$

such that, for example,

$$\begin{aligned}\tau_1 &= \frac{2\sin(\tilde{a}')}{\pi} \\ &= \frac{2}{\pi}\left[\sin\left(a'_1\right)\cosh\left(a'_2\right) + i\cos\left(a'_1\right)\sinh\left(a'_2\right)\right].\end{aligned} \tag{B17}$$

The efficiency of the first diffracted order is proportional to

$$|\tau_1|^2 = \frac{4}{\pi^2}[\sin^2(a'_1)\cosh^2(a'_2) + \cos^2(a'_1)\sinh^2(a'_2)]\,. \tag{B18}$$

The maximum is reached when $a'_1 \sim 1.57$ rad, so the optimal peak-to-valley phase difference is $\Delta\varphi \sim \pi$.

3. Triangular profile

We now consider a triangularly-shaped profile; specifically, an isosceles triangle that can be described by the profile function $f(\alpha) = \frac{1}{\pi}(Sign(\sin(\alpha(\vec{\rho}))))(\pi - Mod(\alpha(\vec{\rho}),\ 2\pi))$, with the transmission function

$$T(\vec{\rho}) = e^{i\tilde{a}\frac{1}{\pi}(Sign(\sin(\alpha(\vec{\rho}))))(\pi - Mod(\alpha(\vec{\rho}),2\pi))}, \tag{B19}$$

where $Mod(a, b)$ is the remainder after dividing a by b. The Fourier coefficients can be calculated from the integral

$$\tau_n = \frac{1}{2\pi}\int_0^{2\pi} e^{i\tilde{a}\frac{1}{\pi}(Sign(\sin(\alpha(\vec{\rho}))))(\pi - Mod(\alpha(\vec{\rho}),2\pi))}e^{-in\alpha(\vec{\rho})}d\alpha, \tag{B20}$$

resulting in the expression

$$\tau_n = \frac{-i(a_1 + ia_2)[(-1)^{n+1} + e^{i(a_1+ia_2)}]}{(a_1^2 + 2ia_1a_2 - a_2^2 - n^2\pi^2)}, \tag{B21}$$

from which the generic efficiency of the n^{th} diffracted order is proportional to

$$|\tau_n|^2 = \frac{(a_1^2 + a_2^2)[1 + 2(-1)^{n+1} e^{-a_2}(\cos(a_1)) + e^{-2a_2}]}{[a_1^4 + 2a_1^2 a_2^2 + a_2^4 + n^4\pi^4 - 2n^2 a_1^2 \pi^2 + 2n^2 a_2^2 \pi^2]}. \tag{B22}$$

4. Blazed profile

A specific example of a triangular profile is a blazed profile, which is similar to that of a sawtooth blade. The profile function is now simply $f(\alpha) = \frac{1}{2\pi}(Mod(\alpha(\vec{\rho}), 2\pi))$, while the transmittance function is

$$T(\vec{\rho}) = e^{i\tilde{a}\frac{1}{2\pi}(Mod(\alpha(\vec{\rho}),2\pi))}. \tag{B23}$$

As before, the Fourier coefficients are

$$\begin{aligned}\tau_n &= \frac{1}{2\pi}\int_0^{2\pi} e^{i\tilde{a}\frac{1}{2\pi}(Mod(\alpha(\vec{\rho}),2\pi))} e^{in\alpha(\vec{\rho})} d\alpha \\ &= \frac{-i(-1 + e^{i\tilde{a}})}{(\tilde{a} + 2\pi n)},\end{aligned} \tag{B24}$$

such that

$$|\tau_n|^2 = \frac{(1 + e^{-2a_2} - 2\cos(a_1)e^{-a_2})}{[(a_1 + 2\pi n)^2 + a_2^2]}. \tag{B25}$$

APPENDIX C: CALCULATIONS FOR AMPLITUDE S-CGHS

For most grating profiles, the calculations are relatively simple as they just involve calculating the squared modulus of the Fourier coefficients of the profile functions. However, some cases require full calculations.

1. Squared profile with an arbitrary duty cycle

For a square profile with an arbitrary duty cycle, the profile function between 0 and 2π is

$$f(\alpha) = \begin{cases} 1 & for\ 0 < \alpha \leq\ 2\pi D, \\ 0 & for\ 2\pi D < \alpha \leq 2\pi, \end{cases} \tag{C1}$$

where the duty cycle D is constant, with $0 < D < 1$. The fundamental frequency associated with $f(\alpha)$ is $\omega_0 = 1$ since the period is 2π. The Fourier coefficients are

$$\tau_0 = \frac{1}{2\pi}\int_0^{2\pi} f(\alpha) d\alpha = \frac{1}{2\pi}\int_0^{D^*2\pi} 1\, d\alpha = D, \tag{C2}$$

$$\begin{aligned}\tau_{n\neq 0} &= \frac{1}{2\pi}\int_0^{2\pi} f(\alpha) e^{in\omega_0\alpha} d\alpha = \frac{1}{2\pi}\int_0^{D*2\pi} 1 e^{in\alpha} d\alpha \\ &= \frac{1}{2\pi}\frac{1}{in}\left[e^{inD2\pi} - 1\right] = \frac{1}{2}\frac{e^{inD\pi}}{in\pi}\left[e^{inD\pi} - e^{-inD\pi}\right] \\ &= \frac{e^{inD\pi}}{n\pi}\sin(nD\pi) = De^{inD\pi}\text{sinc}(nD\pi),\end{aligned} \tag{C3}$$

Therefore,

$$|\tau_n|^2 = \begin{cases} D^2 & for\ n = 0, \\ D^2\text{sinc}^2(nD\pi) & for\ n \neq 0. \end{cases} \tag{C4}$$

REFERENCES

[1] D. Gabor, Nature **161**, 777–778 (1948).

[2] C. T. Koch and A. Lubk, Ultramicroscopy **110**, 460–471 (2010).

[3] T. Latychevskaia, P. Formanek, C. T. Koch, and A. Lubk, Ultramicroscopy **110**, 472–482 (2010).

[4] H. Lichte and M. Lehmann, Rep. Prog. Phys. **71**, 016102 (2008).

[5] A. Claverie, *Transmission Electron Microscopy in Micro-Nanoelectronics* (John Wiley & Sons, 2013).

[6] E. Voelkl, Ultramicroscopy **110**, 199–210 (2010).

[7] M. Takeda, H. Ina, and S. Kobayashi, J. Opt. Soc. Am. **72**, 156 (1982).

[8] B. R. Brown and A. W. Lohmann, Appl. Opt. **5**, 967 (1966).

[9] A. W. Lohmann and D. P. Paris, Appl. Opt. **6**, 1739 (1967).

[10] L. B. Lesem, P. M. Hirsch, and J. A. Jordan, Commun. ACM **11**, 661–674 (1968).

[11] V. Grillo and E. Rotunno, Ultramicroscopy **125**, 97–111 (2013).

[12] C. B. Burckhardt, Appl. Opt. **9**, 695 (1970).

[13] C. Neipp, C. Pascual, and A. Beléndez, J. Phys. D: Appl. Phys. **35**, 957–967 (2002).

[14] C. Neipp, I. Pascual, and A. Belendez, Opt. Express **10**, 1374 (2002).

[15] A. H. Tavabi, H. Larocque, P.-H. Lu, M. Duchamp, V. Grillo, E. Karimi, R. E. Dunin-Borkowski, and G. Pozzi, Phys. Rev. Res. **2**, 013185 (2020).

[16] A. H. Tavabi, P. Rosi, E. Rotunno, A. Roncaglia, L. Belsito, S. Frabboni, G. Pozzi, G. C. Gazzadi, P.-H. Lu, R. Nijland, M. Ghosh, P. Tiemeijer, E. Karimi, R. E. Dunin-Borkowski, and V. Grillo, Phys. Rev. Lett. **126**, 094802 (2021).

[17] R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Springer Science & Business Media, 2011).

[18] K. Iakoubovskii, K. Mitsuishi, Y. Nakayama, and K. Furuya, Phys. Rev. B **77**, 104102 (2008).

[19] P. Pozzi, *Giulio; Hawkes, Advances in Imaging and Electron Physics—Particles and Waves in Electron Optics and Microscopy* (Academic Press, 2016).

[20] H. Kohl and L. Reimer, *Transmission Electron Microscopy* (Springer New York, New York, 2008).

[21] V. Grillo, G. Carlo Gazzadi, E. Karimi, E. Mafakheri, R. W. Boyd, and S. Frabboni, Appl. Phys. Lett. **104**, 043109 (2014).

[22] T. R. Harvey, J. S. Pierce, A. K. Agrawal, P. Ercius, M. Linck, and B. J. McMorran, New J. Phys. **16**, 093039 (2014).

[23] S. Bhattacharyya, C. T. Koch, and M. Rühle, Ultramicroscopy **106**, 525–538 (2006).

[24] R. Shiloh, Y. Lereah, Y. Lilach, and A. Arie, Ultramicroscopy **144**, 26–31 (2014).

[25] L. Grünewald, D. Gerthsen, and S. Hettler, Beilstein J. Nanotechnol. **10**, 1290–1302 (2019).

[26] A. Auslender, M. Halabi, G. Levi, O. Diéguez, and A. Kohn, Ultramicroscopy **198**, 18–25 (2019).

[27] M. Schowalter, J. T. Titantah, D. Lamoen, and P. Kruse, Appl. Phys. Lett. **86**, 112102 (2005).

[28] A. Harscher and H. Lichte, in *ICEM14, Cancun, Mexico* (Institute of Physics Publishing, 1998), Vol. 31, p. 553.
[29] M. Wanner, D. Bach, D. Gerthsen, R. Werner, and B. Tesche, Ultramicroscopy **106**, 341–345 (2006).
[30] A. Sanchez and M. A. Ochando, J. Phys. C: Solid State Phys. **18**, 33–41 (1985).
[31] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy* (Springer U.S., Boston, 2009).
[32] F. Venturi, "New approaches for phase manipulation and characterisation in the transmission electron microscope," Ph.D. thesis (University of Nottingham, 2018).
[33] V. Grillo, E. Karimi, R. Balboni, G. C. Gazzadi, F. Venturi, S. Frabboni, J. S. Pierce, B. J. McMorran, and R. W. Boyd, Microsc. Microanal. **21**, 503–504 (2015).
[34] B. J. McMorran, A. Agrawal, I. M. Anderson, A. A. Herzing, H. J. Lezec, J. J. McClelland, and J. Unguris, Science **331**, 192–195 (2011).
[35] P. Pellat-Finet, Opt. Lett. **19**, 1388 (1994).
[36] A. Lubk, K. Vogel, D. Wolf, F. Röder, L. Clark, and J. Verbeeck, in *European Microscopy Congress 2016 Proceedings* (Wiley-VCH Verlag GmbH & Co. KgaA, Weinheim, Germany, 2016), pp. 705–706.
[37] V. Grillo, J. Harris, G. C. Gazzadi, R. Balboni, E. Mafakheri, M. R. Dennis, S. Frabboni, R. W. Boyd, and E. Karimi, Ultramicroscopy **166**, 48–60 (2016).
[38] H. Lichte, P. Formanek, A. Lenk, M. Linck, C. Matzeck, M. Lehmann, and P. Simon, Annu. Rev. Mater. Res. **37**, 539–588 (2007).
[39] W. Koch, A. Lubk, F. Grossmann, H. Lichte, and R. Schmidt, Ultramicroscopy **110**, 1397–1403 (2010).
[40] K. Y. Bliokh, I. P. Ivanov, G. Guzzinati, L. Clark, R. Van Boxem, A. Béché, R. Juchtmans, M. A. Alonso, P. Schattschneider, F. Nori, and J. Verbeeck, Phys. Rep. **690**, 1–70 (2017).
[41] A. Lubk, "Chapter Four - Electron Optics in Phase Space," in *Advances in Imaging and Electron Physics* (Elsevier, 2018), Vol. 206, pp. 105–140.
[42] E. J. Kirkland, *Advanced Computing in Electron Microscopy* (Springer Nature, 2020).
[43] E. Bolduc, N. Bent, E. Santamato, E. Karimi, and R. W. Boyd, Opt. Lett. **38**, 3546 (2013).
[44] W.-H. Lee, Appl. Opt. **18**, 3661 (1979).
[45] W.-H. Lee, Appl. Opt. **13**, 1677 (1974).
[46] L. A. Giannuzzi, *Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice* (Springer Science & Business Media, 2004).
[47] T. Schachinger, A. Steiger-Thirsfeld, S. Löffler, M. Stöger-Pollach, S. Schneider, D. Pohl, B. Rellinghaus, and P. Schattschneider, in *European Microscopy Congress 2016 Proceedings* (Wiley-VCH Verlag GmbH & Co. KgaA, Weinheim, Germany, 2016), pp. 717–718.
[48] E. Mafakheri, A. H. Tavabi, P.-H. Lu, R. Balboni, F. Venturi, C. Menozzi, G. C. Gazzadi, S. Frabboni, A. Sit, R. E. Dunin-Borkowski, E. Karimi, and V. Grillo, Appl. Phys. Lett. **110**, 093113 (2017).
[49] W. Hu, K. Sarveswaran, M. Lieberman, and G. H. Bernstein, IEEE Trans. Nanotechnol. **4**, 312–316 (2005).
[50] M. Häffner, A. Haug, A. Heeren, M. Fleischer, H. Peisert, T. Chassé, and D. P. Kern, J. Vac. Sci. Technol., B **25**, 2045 (2007).
[51] S. Hettler, L. Radtke, L. Grünewald, Y. Lisunova, O. Peric, J. Brugger, and S. Bonanni, Micron **127**, 102753 (2019).
[52] L. Clark, A. Béché, G. Guzzinati, A. Lubk, M. Mazilu, R. Van Boxem, and J. Verbeeck, Phys. Rev. Lett. **111**, 064801 (2013).
[53] J. Verbeeck, A. Béché, K. Müller-Caspary, G. Guzzinati, M. A. Luong, and M. Den Hertog, Ultramicroscopy **190**, 58–65 (2018).
[54] P. Thakkar, V. A. Guzenko, P.-H. Lu, R. E. Dunin-Borkowski, J. P. Abrahams, and S. Tsujino, J. Appl. Phys. **128**, 134502 (2020).
[55] B. J. McMorran, A. Agrawal, P. A. Ercius, V. Grillo, A. A. Herzing, T. R. Harvey, M. Linck, and J. S. Pierce, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **375**, 20150434 (2017).
[56] M. Uchida and A. Tonomura, Nature **464**, 737–739 (2010).
[57] J. Verbeeck, H. Tian, and P. Schattschneider, Nature **467**, 301–304 (2010).
[58] A. Béché, R. Van Boxem, G. Van Tendeloo, and J. Verbeeck, Nat. Phys. **10**, 26–29 (2014).
[59] G. Pozzi, P.-H. Lu, A. H. Tavabi, M. Duchamp, and R. E. Dunin-Borkowski, Ultramicroscopy **181**, 191–196 (2017).
[60] J. Harris, V. Grillo, E. Mafakheri, G. C. Gazzadi, S. Frabboni, R. W. Boyd, and E. Karimi, Nat. Phys. **11**, 629–634 (2015).
[61] H. Larocque, I. Kaminer, V. Grillo, G. Leuchs, M. J. Padgett, R. W. Boyd, M. Segev, and E. Karimi, Contemp. Phys. **59**, 126–144 (2018).
[62] R. Shiloh, P.-H. Lu, R. Remez, A. H. Tavabi, G. Pozzi, R. E. Dunin-Borkowski, and A. Arie, Phys. Scr. **94**, 034004 (2019).
[63] A. Béché, R. Winkler, H. Plank, F. Hofer, and J. Verbeeck, Micron **80**, 34–38 (2016).
[64] C. W. Johnson, D. H. Bauer, and B. J. McMorran, Appl. Opt. **59**, 1594 (2020).
[65] E. Karimi, G. Zito, B. Piccirillo, L. Marrucci, and E. Santamato, Opt. Lett. **32**, 3053 (2007).
[66] R. D. Guenther, *Modern Optics* (Wiley, 1990).
[67] L. G. Gouy, C. R. Acad. Des. Sci. Paris **110**, 1251 (1890).
[68] S. Feng and H. G. Winful, Opt. Lett. **26**, 485 (2001).
[69] T. D. Visser and E. Wolf, Opt. Commun. **283**, 3371 (2010).
[70] A. M. Yao and M. J. Padgett, Adv. Opt. Photonics **3**, 161 (2011).
[71] S. M. Lloyd, M. Babiker, G. Thirunavukkarasu, and J. Yuan, Rev. Mod. Phys. **89**, 035004 (2017).
[72] K. Y. Bliokh, P. Schattschneider, J. Verbeeck, and F. Nori, Phys. Rev. X **2**, 41011 (2012).
[73] P. Schattschneider, T. Schachinger, M. Stöger-Pollach, S. Löffler, A. Steiger-Thirsfeld, K. Y. Bliokh, and F. Nori, Nat. Commun. **5**, 4586 (2014).
[74] F. Venturi, M. Campanini, G. C. Gazzadi, R. Balboni, S. Frabboni, R. W. Boyd, R. E. Dunin-Borkowski, E. Karimi, and V. Grillo, Appl. Phys. Lett. **111**, 223101 (2017).
[75] F. Venturi, R. Balboni, G. C. Gazzadi, M. Campanini, E. Karimi, V. Grillo, S. Frabboni, and R. W. Boyd, in *European Microscopy Congress 2016 Proceedings* (Wiley-VCH Verlag GmbH & Co. KgaA, Weinheim, Germany, 2016), pp. 709–710.
[76] C. W. Johnson, J. S. Pierce, R. C. Moraski, A. E. Turner, A. T. Greenberg, W. S. Parker, and B. J. McMorran, Opt. Express **28**, 17334 (2020).
[77] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, Phys. Rev. A **45**, 8185–8189 (1992).
[78] P. Schattschneider, M. Stöger-Pollach, and J. Verbeeck, Phys. Rev. Lett. **109**, 084801 (2012).
[79] G. C. G. Berkhout, M. P. J. Lavery, J. Courtial, M. W. Beijersbergen, and M. J. Padgett, Phys. Rev. Lett. **105**, 153601 (2010).
[80] V. Grillo, A. H. Tavabi, F. Venturi, H. Larocque, R. Balboni, G. C. Gazzadi, S. Frabboni, P.-H. Lu, E. Mafakheri, F. Bouchard, R. E. Dunin-Borkowski, R. W. Boyd, M. P. J. Lavery, M. J. Padgett, and E. Karimi, Nat. Commun. **8**, 15536 (2017).
[81] E. Rotunno, M. Zanfrognini, S. Frabboni, J. Rusz, R. E. Dunin Borkowski, E. Karimi, and V. Grillo, Phys. Rev. B **100**, 224409 (2019).
[82] M. Zanfrognini, E. Rotunno, J. Rusz, R. E. Dunin Borkowski, E. Karimi, S. Frabboni, and V. Grillo, Phys. Rev. B **102**, 184420 (2020).
[83] M. Zanfrognini, E. Rotunno, S. Frabboni, A. Sit, E. Karimi, U. Hohenester, and V. Grillo, ACS Photonics **6**, 620–627 (2019).
[84] J. Durnin, J. Opt. Soc. Am. A **4**, 651 (1987).
[85] J. Durnin, J. J. Miceli, and J. H. Eberly, Phys. Rev. Lett. **58**, 1499–1501 (1987).
[86] J. Durnin, J. J. Miceli, and J. H. Eberly, Phys. Rev. Lett. **66**, 838 (1991).
[87] D. McGloin and K. Dholakia, Contemp. Phys. **46**, 15–28 (2005).
[88] Y. Lin, W. Seka, J. H. Eberly, H. Huang, and D. L. Brown, Appl. Opt. **31**, 2708 (1992).
[89] J. H. McLeod, J. Opt. Soc. Am. **44**, 592 (1954).
[90] Z. Bin and L. Zhu, Appl. Opt. **37**, 2563 (1998).
[91] T. Tanaka and S. Yamamoto, Opt. Commun. **184**, 113–118 (2000).
[92] A. Thaning, Z. Jaroszewicz, and A. T. Friberg, Appl. Opt. **42**, 9 (2003).
[93] A. Vasara, J. Turunen, and A. T. Friberg, J. Opt. Soc. Am. A **6**, 1748 (1989).
[94] J. A. Davis, E. Carcole, and D. M. Cottrell, Appl. Opt. **35**, 599 (1996).

[95] J. A. Davis, E. Carcole, and D. M. Cottrell, Appl. Opt. **35**, 593 (1996).
[96] A. J. Cox and D. C. Dibble, J. Opt. Soc. Am. A **9**, 282 (1992).
[97] K. Uehara and H. Kikuchi, Appl. Phys. B Photophys. Laser Chem. **48**, 125–129 (1989).
[98] V. Grillo, E. Karimi, G. C. Gazzadi, S. Frabboni, M. R. Dennis, and R. W. Boyd, Phys. Rev. X **4**, 011013 (2014).
[99] K. Saitoh, K. Hirakawa, H. Nambu, N. Tanaka, and M. Uchida, J. Phys. Soc. Jpn. **85**, 043501 (2016).
[100] C. Zheng, T. C. Petersen, H. Kirmse, W. Neumann, M. J. Morgan, and J. Etheridge, Phys. Rev. Lett. **119**, 174801 (2017).
[101] P. A. Midgley and R. E. Dunin-Borkowski, Nat. Mater. **8**, 271–280 (2009).
[102] G. Guzzinati, W. Ghielens, C. Mahr, A. Béché, A. Rosenauer, T. Calders, and J. Verbeeck, Appl. Phys. Lett. **114**, 243501 (2019).
[103] E. Rotunno, A. H. Tavabi, E. Yucelen, S. Frabboni, R. E. Dunin Borkowski, E. Karimi, B. J. McMorran, and V. Grillo, Phys. Rev. Appl. **11**, 044072 (2019).
[104] S. Hettler, L. Grünewald, and M. Malac, New J. Phys. **21**, 033007 (2019).
[105] M. Haider, H. Rose, S. Uhlemann, B. Kabius, and K. Urban, J. Electron Microsc. **47**, 395–405 (1998).
[106] M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius, and K. Urban, Nature **392**, 768–769 (1998).
[107] O. L. Krivanek, N. Dellby, and A. R. Lupini, Ultramicroscopy **78**, 1–11 (1999).
[108] R. Shiloh, R. Remez, and A. Arie, Ultramicroscopy **163**, 69–74 (2016).
[109] V. Grillo, A. H. Tavabi, E. Yucelen, P.-H. Lu, F. Venturi, H. Larocque, L. Jin, A. Savenko, G. C. Gazzadi, and R. Balboni, Opt. Express **25**, 21851 (2017).
[110] M. Linck, P. A. Ercius, J. S. Pierce, and B. J. McMorran, Ultramicroscopy **182**, 36–43 (2017).
[111] R. Shiloh, R. Remez, P.-H. Lu, L. Jin, Y. Lereah, A. H. Tavabi, R. E. Dunin-Borkowski, and A. Arie, Ultramicroscopy **189**, 46–53 (2018).
[112] A. Konečná and F. J. G. de Abajo, Phys. Rev. Lett. **125**, 030801 (2020).
[113] J. Verbeeck, H. Tian, and A. Béché, Ultramicroscopy **113**, 83–87 (2012).