JATIS Journal of Astronomical Telescopes, Instruments, and Systems

RESEARCH PAPER

Microstructured vortex and azimuthal cosine phase mask design for high-contrast imaging

**Lorenzo König,[a,*] Niyati Desai,[a] Skyler Palatnick,[b] Olivier Absil,[c]
Dimitri Mawet,[a,d] Maxwell Millar-Blanchaer,[b] and Eugene Serabyn[a]**
[a]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, United States
[b]University of California, Department of Physics, Santa Barbara, California, United States
[c]Université de Liège, STAR Institute, Liège, Belgium
[d]California Institute of Technology, Department of Astronomy, Pasadena, California, United States

---

ABSTRACT. At the extreme contrast levels required to image Earth-like planets around Sun-like stars, the polarization dependence of the vector vortex coronagraph becomes a limiting factor, making wavefront control difficult to perform in both polarizations simultaneously. An alternative is to use a polarization-independent scalar vortex phase mask, but achromatizing scalar masks remains challenging. We investigate using metasurfaces to increase the bandwidth of scalar vortex phase masks. Our design shows an improvement of up to 2 orders of magnitude compared to a scalar vortex made of a helical-shaped dielectric substrate. However, the characteristic phase discontinuities of scalar vortex phase masks introduce phase artifacts and remain challenging to manufacture accurately. The cosine phase mask is an alternative approach to implementing a coronagraphic phase mask with a continuously varying azimuthal phase profile but without the phase jump of the scalar vortex. In addition, it requires smaller phase coverage. We therefore also investigate a metasurface implementation of the cosine mask. We present results obtained using rigorous coupled-wave analysis and finite-difference time-domain simulations and find that the phase jumps of a scalar vortex result in significant stellar leakage, which does not appear in the case of the cosine mask. We then present the simulated coronagraphic performance and residual chromaticity of both designs and discuss their advantages and drawbacks. We conclude that metasurface scalar vortex and cosine phase masks are promising coronagraphic phase masks in light of upcoming ground and space telescope missions combining deep contrast and insensitivity to low-order aberrations.

[DOI: 10.1117/1.JATIS.11.2.025002]

Keywords: metasurface; exoplanet; scalar vortex; phase mask; coronagraph

Paper 24212G received Dec. 12, 2024; revised Mar. 21, 2025; accepted Mar. 24, 2025; published Apr. 18, 2025.

---

1 Introduction

Detecting and characterizing exoplanets require starlight to be suppressed by several orders of magnitude. For instance, direct imaging of terrestrial planets around M-dwarfs requires contrasts on the order of 1e–8, and characterizing Earth-like planets around Sun-like stars requires contrasts of 1e–10. One promising approach to achieve this goal is to use a coronagraph. In particular, the vortex coronagraph is a good choice for this task because of its small inner working angle but also its ability to trade it for reduced sensitivity to low-order aberrations and stellar sizes.[1,2] For instance, a topological charge of 6 as used in this study is insensitive to the lowest

---

*Address all correspondence to Lorenzo König, lorenzo.koenig@jpl.nasa.gov

order aberrations tip/tilt, focus, astigmatism, and spherical aberration while retaining a small inner working angle of $2.4\,\lambda/\mathrm{D}$. The topological charge ℓ is the multiple of 2π that the phase acquires in a closed path around the central singularity. The vortex coronagraph has demonstrated excellent performance on major ground-based telescopes,[3–8] and recently, its performance has been pushed toward very high contrasts in the lab (down to 2.4e–10 for monochromatic light).[9] The use of efficient wavefront control techniques is mandatory at these high contrast levels. However, current vortex coronagraphs are vectorial in nature and therefore limited by their polarization properties because they imprint opposite phase ramps on the two circular polarizations. This makes wavefront control in both polarizations simultaneously challenging, even with multiple deformable mirrors. One option to avoid this issue is to split the polarizations, effectively doubling the number of optics. Alternatively, a polarizer-analyzer setup can be used to filter one polarization, at the expense of reducing the throughput by a factor of 2.

A potential alternative to the vector vortex mask is to use a polarization-independent, or scalar, vortex phase mask.[10,11] However, the simplest helical implementation of a scalar vortex mask is highly chromatic. So far, efforts toward increasing the bandwidth of scalar vortex phase masks have focused primarily on optimizing the design topography, showing promising results.[12–14] Here, we investigate metasurfaces as a technology to potentially increase the bandwidth of scalar phase masks.

Metasurfaces are thin structures consisting of subwavelength building blocks arranged in a single thin layer capable of providing partial amplitude and phase control.[15] By locally varying the geometry of their constituent building blocks, metasurfaces can imprint virtually any phase pattern on an incident beam. The advantage of using metasurfaces is that they provide more degrees of freedom with their geometric design parameters compared to traditional dielectric masks, which can be leveraged to increase their bandwidth of application, whereas their fabrication benefits from established microfabrication techniques. This makes metasurfaces a promising candidate to implement coronagraphic phase masks, such as the scalar vortex presented here.

The main characteristic of a vortex coronagraph is the azimuthal phase ramp introduced at the focal plane. The vector vortex implements the phase ramp using the geometric phase shifts of a spatially variant half-wave plate made of birefringent materials such as liquid crystal polymers,[16] subwavelength gratings,[17,18] or photonic crystals.[19] These implementations are based on the precise orientation of the fast axis of the half-wave plate resulting in an achromatic phase ramp, only limited by the polarization leakage originating from the mask's imperfect half-wave plate nature. By contrast, the scalar vortex uses longitudinal phase shifts to implement the phase ramp.[20] This means that the wrapping of the phase creates phase jumps of multiples of 2π, which translate into physical pattern discontinuities in a dielectric mask, or into metasurface patterns with abruptly changing geometries of the constituent building blocks. These discontinuities affect the quality of the phase ramp and, therefore, degrade the coronagraphic performance of the phase mask.

An alternative to the vortex phase mask is to use an azimuthal cosine phase mask topography,[21–23] which provides a continuous azimuthal phase profile without phase jumps. Figure 1 shows the phase profiles of different topographies of scalar vortex phase masks and the cosine phase mask. The cosine design is part of the same family of solutions to the coronagraphic problem as the vortex coronagraph and provides the same perfect starlight rejection.[22] The advantage of this design is twofold. First, the continuous phase pattern comes without the 2π phase jumps of the scalar vortex, avoiding the physical pattern discontinuities mentioned above. Second, the phase coverage of the cosine design is $2z_0 = 4.81\,\mathrm{rad} < 2\pi$, where z_0 is the first zero of the Bessel function of the first kind of order zero (J_0), reducing the required phase coverage of the phase mask by $\sim$25%. Combined with a metasurface implementation, the azimuthal cosine phase mask presented here is therefore a promising alternative to the scalar vortex coronagraph, relaxing manufacturing constraints due to its lower phase coverage and its continuously varying phase profile.

In this paper, we first describe the process of designing achromatic metasurface-based phase masks from a library of subwavelength building blocks (Sec. 2). We use this process to create an achromatic scalar vortex phase mask design and simulate the effect of its characteristic 2π phase jumps on the quality of the phase ramp (Sec. 3). We then present an azimuthal cosine phase mask

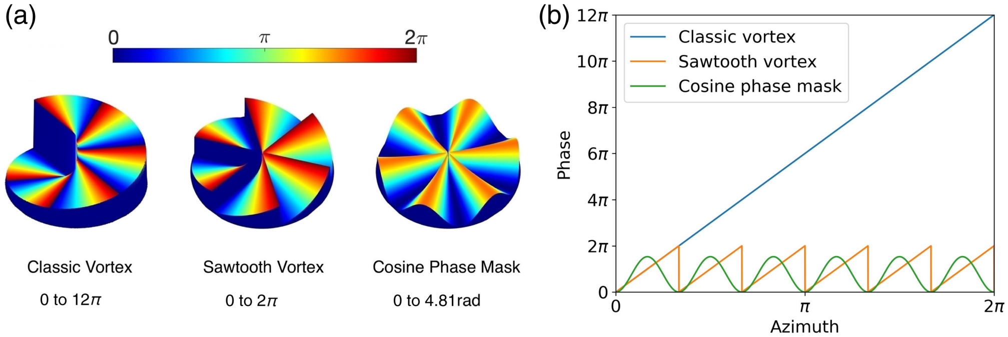

Fig. 1 (a) Scalar phase mask topographies for a charge-6 vortex. The classic vortex spans a continuous 0 to 12π in a single phase ramp. The sawtooth vortex breaks the phase ramp down into six parts covering 0 to 2π each, at the expense of increasing the number of phase jumps. The cosine phase mask has an overall smaller phase coverage of 0 to 4.81 rad and avoids any phase jumps (apart from the central singularity). (b) Azimuthal phase profiles of the three topographies.

design without the phase jumps of the scalar vortex and discuss its performance compared with the scalar vortex (Sec. 4). We conclude with the simulated performance of these 2 phase mask designs in a coronagraph (Sec. 5) and outline potential material choices and challenges when manufacturing these designs (Sec. 6).

2 Achromatic Scalar Metasurface

2.1 Calculating the Phase and Transmission Response

Metasurfaces are thin structures consisting of subwavelength building blocks enabling broad phase and amplitude control, which have been widely used to implement bulky optical components in a thin layer.[24–28] These applications rely on the engineering of the phase and transmission response of the individual metasurface building blocks. More recently, substantial effort has been put into increasing the bandwidth of these structures,[29,30] which is crucial for applications in high-contrast imaging. This is achieved by exploiting the increased degree of freedom of the various geometric parameters of their constituent building blocks, the aim of which is to provide an achromatic phase response over the desired bandwidth.

A metasurface is characterized by the geometric composition of its constituent building blocks. Polarization-independent metasurfaces are of particular interest because they enable the implementation of scalar phase masks, such as the scalar vortex phase mask. A scalar metasurface requires both polarizations of the incoming light to be affected in the same way, which can be achieved with building blocks that are symmetric along both orthogonal axes. Examples include circular nanopillars and square nanoblocks but also more complex shapes with holes or crosses.[30,31]

To accurately simulate the response of different metasurface geometries, we use the rigorous coupled-wave analysis[32] (RCWA) implementation RETICOLO.[33] This method allows the calculation of the phase and amplitude response of a metasurface pattern to an incident wave. RCWA requires periodic boundaries, which means that it describes accurately only locally periodic metasurface patterns. Therefore, the constituent metasurface building blocks of a larger pattern have to be sufficiently similar to their neighbors for RCWA to be valid. For regions of the metasurface pattern with steep phase gradients (e.g., the central singularity and phase jumps of a scalar vortex), this condition is no longer fulfilled; to describe the behavior of these regions correctly, we use the finite-difference time-domain[34] (FDTD) method, which allows simulating finite structures but requires larger computational resources. The simulation tools are discussed in more detail in Appendix A (Sec. 8).

We start with a simple metasurface based on square nanoblocks of different sizes. Although we apply our simulation tools to the design of a diamond metasurface optimized for a wavelength of 650 nm, they can be applied in the same way to other materials and wavelength ranges. We choose to use diamond because it is a material with interesting properties, including relatively high refractive index, high transparency across the visible and infrared bands, mechanical

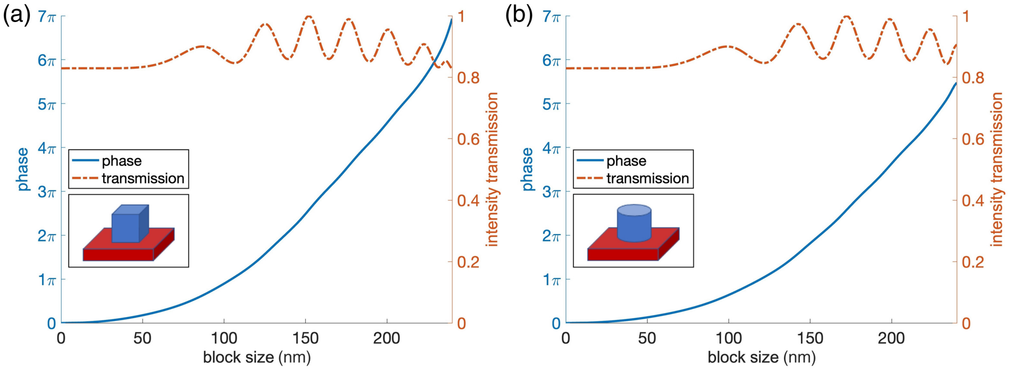

Fig. 2 Phase and transmission response of square (a) and circular (b) nanoblocks diamond metasurface at a wavelength of 650 nm. The metasurface height for this design is 1.6 μm with a periodicity of 240 nm.

stability in cryogenic environments, and one of the highest thermal conductivities of all materials, making a pattern etched into a single diamond substrate ideal for astronomical applications.[18,35] Figure 2 shows the simulated phase and transmission responses of a periodic array of square (a) and circular (b) diamond nanoblocks of different sizes at a fixed metasurface height. The periodicity Λ is chosen small enough to satisfy the subwavelength condition $\Lambda < \lambda/n$, where λ is the design wavelength and n is the refractive index of the substrate. The height is chosen as small as possible while maintaining the 2π phase coverage across the desired bandwidth, as explained below. The phase curve for square nanoblocks is steeper than for circular nanopillars and, therefore, results in a higher achievable phase coverage. This is because a square nanoblock fills the full unit cell, while this is not the case for a circular nanopillar.

2.2 Optimizing for Broadband Performance

Once the phase and amplitude response of a specific metasurface is obtained, its broadband performance must be assessed. The design of an achromatic scalar vortex phase mask requires a phase ramp of fixed topological charge across the desired bandwidth. A metasurface can reduce the total phase by wrapping it similar to the "sawtooth" vortex, which breaks a charge-ℓ phase ramp down into ℓ sectors each spanning 2π [Fig. 1(b)]. Figure 2 shows that 2π phase coverage can be achieved for many different ranges of block sizes at one individual wavelength. However, for an achromatic metasurface, the phase coverage must be 2π at all wavelengths within the band for a specific range of block sizes.

The important metric for a vortex phase mask is to obtain a phase ramp that corresponds to the topological charge of the vortex. For scalar vortex phase masks made of a helical-shaped dielectric substrate, the phase is highly chromatic because of the reciprocal wavelength dependence of the propagation phase. For a metasurface, the absolute phase response is chromatic as well, but the relative phase delay across the mask can be engineered to be much less chromatic.

To see this, we first consider a dielectric mask with a helical phase ramp of a classic vortex, as shown in Fig. 1. The phase response at any azimuth θ on the mask is

$$\begin{aligned}\phi(\theta) &= 2\pi \cdot \frac{d(\theta)}{\lambda_n} + 2\pi \cdot \frac{h - d(\theta)}{\lambda} \\ &= 2\pi \cdot \frac{\theta h/2\pi}{\lambda/n} + 2\pi \cdot \frac{h - \theta h/2\pi}{\lambda} \\ &= 2\pi \cdot \frac{h}{\lambda} \cdot \left(1 + \frac{\theta \cdot (n-1)}{2\pi}\right)\end{aligned}$$

where $d(\theta)$ is the height profile of the mask, λ is the vacuum wavelength, $\lambda_n = \lambda/n$ is the wavelength at the refractive index of the mask n, and h is the maximum height of the helical ramp. The phase is seen to have a reciprocal λ-dependence. For a broadband vortex, the phase accumulated on a closed path around the optical axis needs to be constant: $\Delta\phi = \ell \cdot 2\pi$. For the classical scalar vortex on the other hand

$$\Delta\phi = \phi(\theta = 2\pi) - \phi(\theta = 0),$$
$$= 2\pi \cdot \frac{h}{\lambda} \cdot (n - 1).$$

The accumulated phase therefore also has a reciprocal dependence on the wavelength. For a sawtooth vortex, the phase ramp is broken into ℓ sectors with 2π phase coverage each, reducing the height h by a factor of ℓ, which partly reduces the chromaticity of the mask in a coronagraph.

For a metasurface-based scalar vortex phase mask, the relevant phase shift is acquired in a layer of constant thickness across the mask. The phase shift does not rely on the physical path difference in regions of different refractive index (air and dielectric), but on the phase shift acquired in this single layer, which is achieved by locally varying the effective refractive index, resulting in

$$\phi(\theta) = 2\pi \cdot \frac{h}{\lambda / n_{\mathrm{eff}}(\lambda, \theta)},$$

where h is the constant height of the metasurface layer and $\lambda / n_{\mathrm{eff}}$ the effective wavelength, dependent on the effective refractive index n_{eff} of the metasurface pattern at the wavelength of interest λ. Assuming that the phase ramp scales sufficiently linearly with θ, the relevant metric for the vortex effect is

$$\Delta\phi = \phi(\theta = 2\pi) - \phi(\theta = 0),$$
$$= \frac{2\pi h}{\lambda} \cdot (n_{\mathrm{eff}}(\lambda, \theta = 2\pi) - n_{\mathrm{eff}}(\lambda, \theta = 0)),$$
$$= \frac{2\pi h}{\lambda} \cdot \Delta n_{\mathrm{eff}}(\lambda).$$

The reciprocal dependence on the wavelength can now be compensated by engineering the dispersion of the effective refractive index of the metasurface such that its difference across the mask is proportional to the wavelength ($\Delta n_{\mathrm{eff}} = \lambda / h$). This is important because it does not require a material with positive dispersion (i.e., $n(\lambda) \propto \lambda$), which is challenging to achieve with high throughput. Instead, the design freedom of the metasurface can be used to engineer the needed wavelength dependence of the effective refractive index. Note that h and Δn_{eff} must be chosen such that the equation above is satisfied. h thereby determines the achievable phase coverage of the mask and therefore needs to be large enough to enable broadband 2π phase coverage.

For blocks much smaller than the wavelength, the relation between the filling factor and effective refractive index n_{eff} can be described semi-analytically with effective medium theory.[36] However, when the wavelength approaches the feature size of the metasurface, nonlinear effects become nonnegligible, and full wave propagation codes such as RCWA and FDTD must be used.

The key property of metasurfaces is that although the phase induced by the metasurface itself might be chromatic, the phase difference $\Delta\phi$ acquired along a closed path around the optical axis of the vortex phase mask can be engineered to be achromatic across a certain bandwidth by optimizing the size of the metasurface building blocks. This is summarized in Fig. 3, which shows the chromatic phase response for a square nanoblock geometry along with the effective refractive index n_{eff} at several wavelengths. Although the phase ϕ itself is chromatic [Fig. 3(b)], the resulting phase coverage $\Delta\phi$ is relatively achromatic [Fig. 3(d)] because the effective refractive index difference Δn_{eff} is proportional to the wavelength [Fig. 3(c)]. This is possible because the effective refractive index itself is highly chromatic [Fig. 3(a)].

3 Scalar Vortex Phase Mask Design

3.1 Metasurface Design

Using the results of the RCWA broadband optimization, the nominal phase pattern is discretized in a square lattice and each pixel is assigned a metasurface block returning the desired phase (at the central wavelength). This process results in the design shown in Fig. 4(a) for a sawtooth vortex with topological charge $\ell = 6$, wherein the block widths range from 190 to 230 nm (see Fig. 3). We optimized the design for a 20% bandwidth, centered at 650 nm.

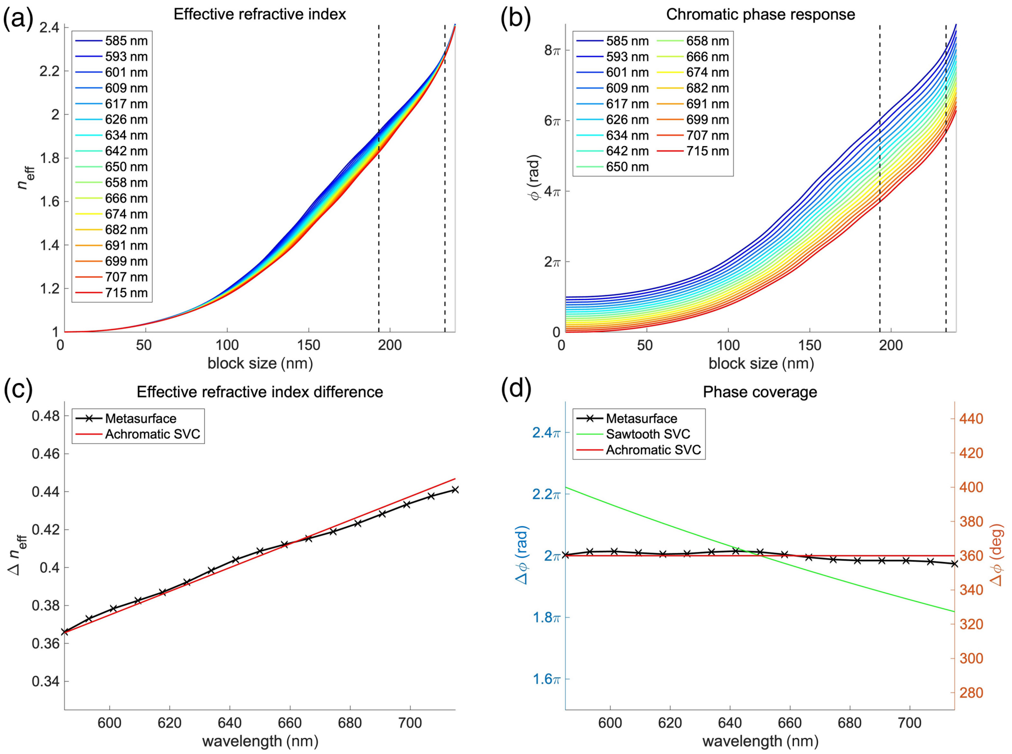

Fig. 3 Chromaticity of a metasurface consisting of square diamond nanoblocks. Effective refractive index (a) and phase response (b). A metasurface consisting of square nanoblocks with a height of 1.6 μm and a periodicity of 240 nm is used. The optimal region providing broadband 2π phase coverage is highlighted with dashed black lines. Effective refractive index difference Δn_{eff} (c) and phase difference $\Delta\phi$ (d) at the 2π phase discontinuities. Although n_{eff} and ϕ are chromatic, the acquired phase difference $\Delta\phi$ is relatively achromatic because the linear wavelength dependence of Δn_{eff} compensates for the $1/\lambda$ dependence of $\Delta\phi$. The targeted achromatic vortex case is shown as a red line, and the curve for a sawtooth scalar vortex (with six 2π phase jumps) made from a helical-shaped dielectric mask is shown as a green line.

In theory, the RCWA simulations enable us to simulate the phase response of this scalar vortex metasurface design for different wavelengths. However, as mentioned, due to the periodic boundaries used in RCWA, the results are only accurate for regions of the metasurface with smooth phase profiles. The present design however features six large discontinuities at the locations of the 2π phase jumps and one at the central phase singularity. This is because the steep phase gradient in these regions breaks the local periodicity of the metasurface. Its behavior is therefore accurately described with FDTD, which does not require local periodic boundary conditions. Figure 4(b) shows the simulated phase response of the metasurface pattern at the design wavelength. The phase response at other wavelengths in the 20% bandwidth is similar because the design is achromatic. Although the overall helical phase ramp is clearly visible, the discontinuities of the pattern create significant artifacts in the phase response (rms of 0.209 rad). This is illustrated in more detail in Figs. 4(c) and 4(d), showing an azimuthal cut along the black circle in the phase map and the residuals with respect to the ideal phase ramp of a charge-6 vortex.

3.2 Effect of Phase Discontinuities

The FDTD simulations of the metasurface scalar vortex design show that the discontinuities can cause significant phase distortions. It is therefore important to understand the effect of such artifacts on the performance of the vortex phase mask. To do this, the discontinuities can be simulated with FDTD, whereas the regions of the mask with smooth phase gradients can be treated with RCWA. It is worth noting that FDTD could in principle be used to simulate the full phase mask; however, this comes at a greatly increased computational cost compared with RCWA (a simulation, as shown in Fig. 4, requires computational resources on the order of 1500 CPUh,

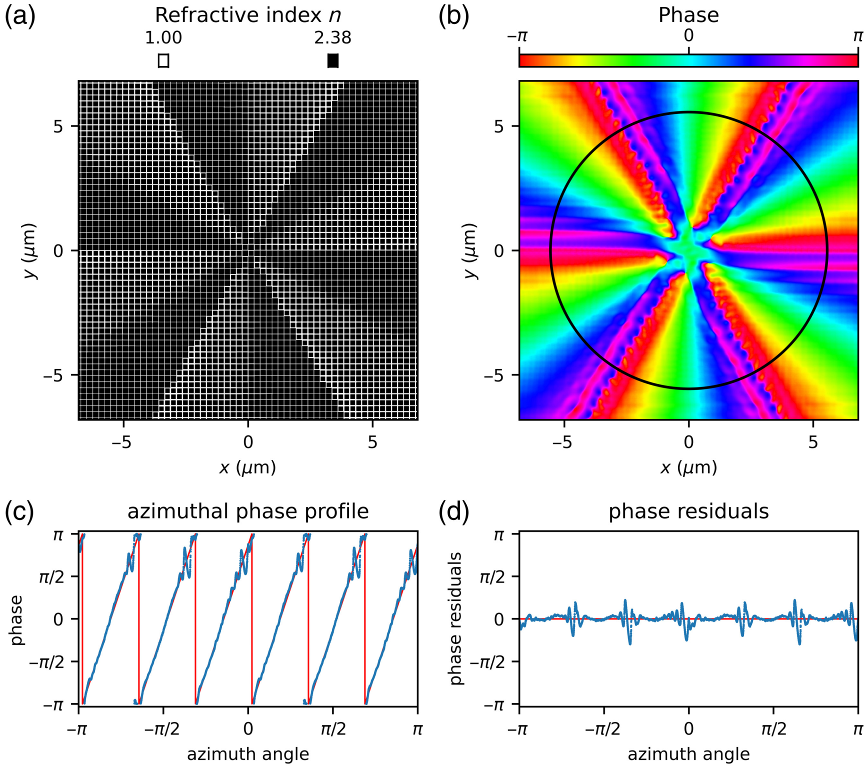

Fig. 4 Metasurface design for a sawtooth vortex phase mask of topological charge $\ell = 6$ (a) and phase response simulated with FDTD (b). This design uses square diamond nanoblocks of 1.6 μm height with a periodicity of 240 nm and is optimized for 20% bandwidth (585 to 715 nm). The azimuthal phase profile (c) is obtained along the cut shown as a black circle in the phase map. Its residuals (d) are small for regions with a local periodic pattern but higher close to the 2π phase discontinuities.

whereas the equivalent RCWA simulation needs only seconds on a modern laptop). We therefore choose to combine these two methods, which enables us to determine how much of an effect the phase artifacts have on the coronagraphic performance of the vortex. Note that the central discontinuity is small compared to the six phase discontinuities extending along the six wedges of the mask. Previous studies have shown that the central defect of a microstructured vortex phase mask is limited.[37,38] We therefore focus on the six extended pattern discontinuities of the metasurface vortex.

To quantify the effect of the discontinuities, FDTD simulations at relatively high resolution are carried out in a limited region around them. Figure 5 shows the result of this simulation, with the metasurface design [Fig. 5(a)], the phase response [Fig. 5(b)], and the phase profile perpendicular to the pattern discontinuity [Fig. 5(c)]. The phase profile shows significant oscillations within ± 2 μm of the pattern discontinuity. This represents a significant fraction of the full Airy disk for low f-numbers (for a charge-6 vortex at 650 nm, it covers 24% of the Airy disk at $f/40$, and 12% at $f/80$). These oscillations are high-frequency deviations from the desired phase response and should only have a small influence on the coronagraphic performance. However, because of the high-performance levels considered here, a coronagraphic propagation is needed to quantify the effect of these deviations. We do this by propagating a perfect vortex phase ramp and adding the phase artifact, which we approximate using a damped sine function $(0.7\,\sin(14x)\,\exp(-2|x|))$ adapted to the simulated phase response. The coronagraphic propagation is performed using the Fast Linearized Coronagraph Operator (FALCO)[39] software. For an $f/40$ system, this results in a leakage term of 6e–9 in an annular region ranging from 3 to $10\,\lambda/D$ in the coronagraphic image plane, representative of the discovery space of the coronagraph. This is on the same order of magnitude as the contrast of the presented metasurface designs (Sec. 5). The effect is stronger for smaller f-numbers and higher topological charge,

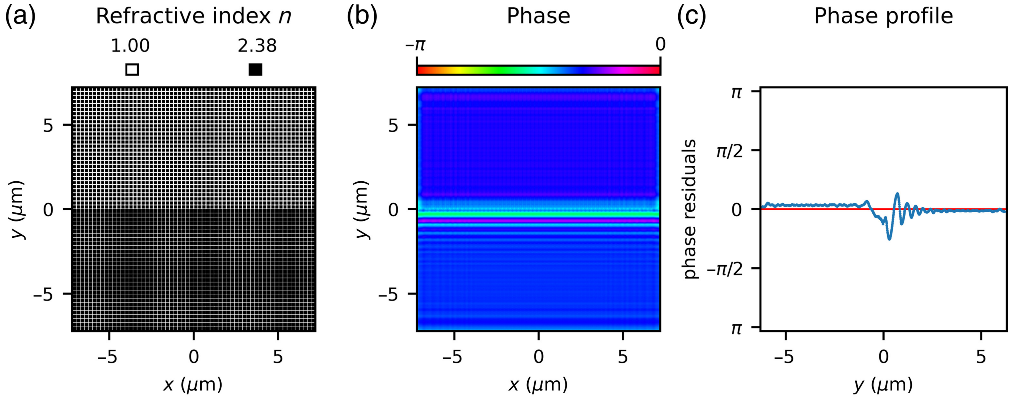

Fig. 5 FDTD simulation of pattern discontinuity around a 2π phase jump. Metasurface design (a), phase response (b), and averaged phase profile along the vertical direction (c). This design is simulated at a wavelength of $\lambda = 650$ nm.

as well as for shorter wavelengths, because the physical discontinuities cover a larger part of the Airy pattern.

4 Azimuthal Cosine Phase Mask Design

The extended phase discontinuities can potentially degrade the performance of the vortex phase mask at low f-numbers and high topological charges. An alternative is to use a phase mask topography without such discontinuities. The vortex phase mask is a special case in a more general set of solutions to the coronagraphic problem.[11,21–23] Another solution providing perfect starlight cancellation in the ideal case, similar to the vortex, is an azimuthal cosine phase mask, consisting of an azimuthally varying phase profile with a singularity at its center: $\phi(\theta) = z_j \cos(k \cdot \theta)$, where z_j is the zeros of the Bessel function of the first kind of order zero, k is the angular frequency (similar to the topological charge of a vortex), and θ is the azimuthal coordinate.[23] Note the reduced phase coverage for $j = 0$ of $2z_0 = 4.81$ compared with $2\pi = 6.28$ of the sawtooth vortex, which reduces the required height of the metasurface, and will be advantageous for manufacturing. The cosine phase mask provides perfect starlight cancellation in theory because it can be expressed as a superposition of vortex modes of even integer multiples of the topological charge, each individually providing perfect starlight cancellation.[13] Because of this behavior, the sensitivity of the cosine to low-order aberrations also follows the same power laws as for the vortex modes it can be decomposed into. A charge-6 cosine phase mask therefore conserves the properties of a charge-6 vortex mask[14] and, in particular, the low sensitivities to tip/tilt, which is a key requirement for future telescopes capable of resolving nearby stars.

The cosine phase mask topography does not have the extended phase discontinuities of a scalar vortex (except for the central singularity) and should therefore provide a smoother phase response than the scalar vortex phase mask. Following the process described in Sec. 3, we can create a metasurface design of the cosine phase mask and simulate its phase response with FDTD (Fig. 6). The azimuthal phase profile [Fig. 6(c), blue curve] shows that the phase response is very close to the theoretical phase profile (red curve), and does indeed not show the large artifacts visible in the scalar vortex design (rms of 0.086 rad, a reduction of a factor of ~3). The slight dips in the phase residuals [Fig. 6(d)] occur at the steepest phase gradients in the phase profile corresponding to regions where the metasurface parameters change rapidly, resulting in small deviations from the optimal phase response. Because of the overall reduced artifacts, the cosine phase mask is an interesting alternative for a metasurface implementation of a scalar coronagraphic phase mask.

5 Simulated Broadband Performance

To investigate the proposed designs in a coronagraphic setup, we simulate the coronagraphic performance with the FALCO software. We use a circular unobstructed entrance pupil with no wavefront errors and propagate the beam to the focal plane where we imprint the phase and

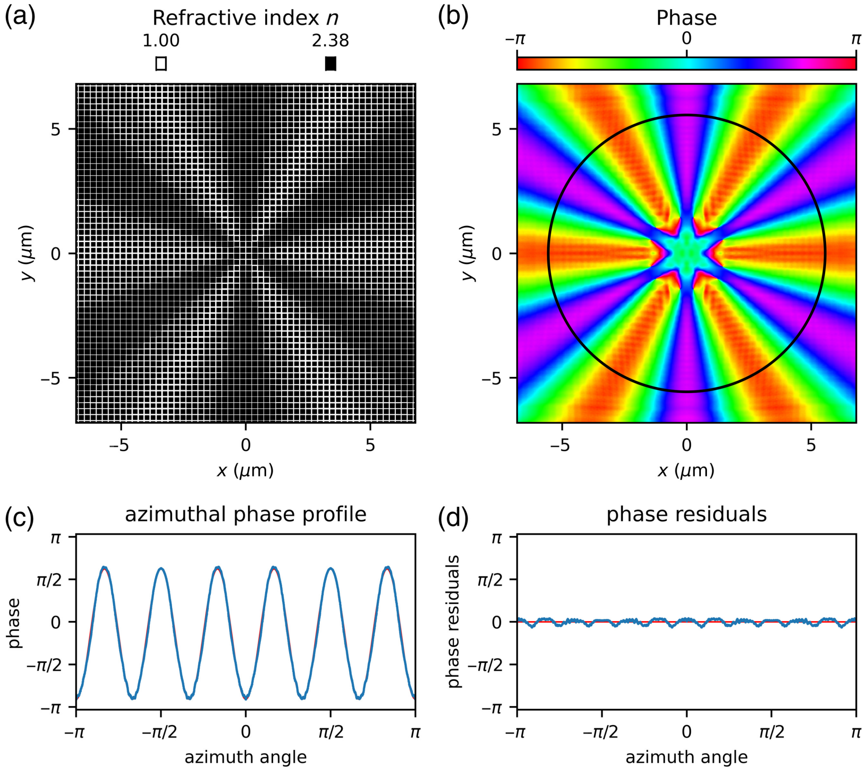

Fig. 6 Metasurface design for an azimuthal cosine phase mask of angular frequency $k = 6$ (a) and phase response simulated with FDTD (b). This design uses square diamond nanoblocks of 1.6 μm height with a periodicity of 240 nm and is optimized for 20% bandwidth (585 to 715 nm). The azimuthal phase profile (c) is obtained along the cut shown as a black circle in the phase map. Its residuals (d) are close to 0.

amplitude response of the metasurface mask onto the beam. Propagating to the next pupil plane, we use a Lyot stop undersized by 90%, which intercepts the light rejected by the focal plane mask, and propagate the residual fields to the final image plane. Finally, we compute the average contrast in an annular region from 3 to $10\,\lambda/D$, representative of the discovery space of the coronagraph.

Figure 7 shows the simulated average contrast in the final image plane for the metasurface scalar vortex and cosine phase mask presented in Secs. 3 and 4 (blue curves). The curves of the same phase topographies implemented in a helical/cosine-shaped dielectric substrate are shown as comparison (red curves). At the center wavelength, the shaped dielectric masks perform best

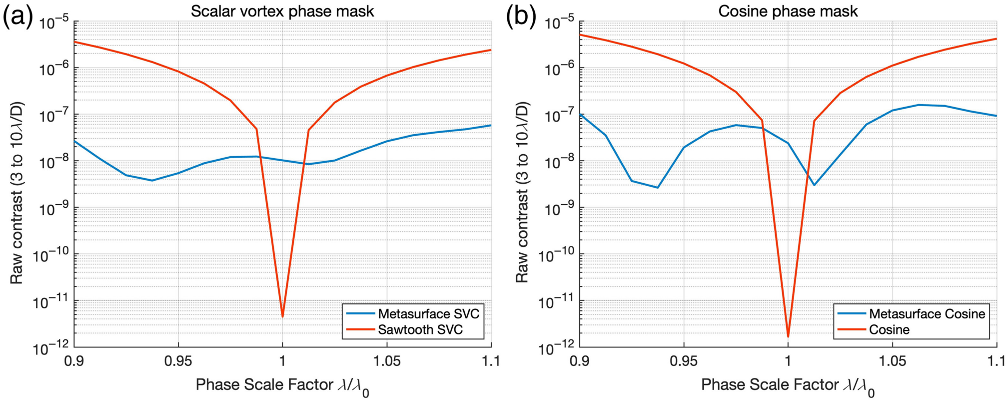

Fig. 7 Contrast curves (blue) for metasurface vortex (a) and cosine (b) designs. The contrast curves for a scalar vortex and cosine phase mask made of a shaped dielectric substrate are shown as a comparison (red).

(to the numerical noise floor on the order of 1e–12), but their performance quickly degrades off the central wavelength. Across the 20% bandwidth, both the vortex and cosine metasurface patterns perform significantly better than their shaped dielectric counterparts. The metasurface designs therefore provide a significant performance improvement of up to 2 orders of magnitudes across a broad band compared with their shaped dielectric counterparts.

For both cases, the contrast varies with wavelength, but the variations are more pronounced for the cosine case. This is due to the nature of the residual phase and transmission variations with wavelength, which seem to have a larger impact on the coronagraphic performance of the cosine design than the vortex. The mean contrast is 2e–8 for the vortex and 6e–8 for the cosine. This appears counterintuitive following the discussion on the effect of phase discontinuities, but it can partly be explained by the contrast difference of the respective shaped dielectric designs (red curves in Fig. 7, average contrast of 1.0e–6 for the vortex versus 1.6e–6 for the cosine). The remaining difference could be due to the different sensitivities to deviations from the optimal mask response. The large variations in the contrast for the cosine design indicate a higher sensitivity to these deviations, which can be explained by the fact that the overall phase coverage of the cosine mask is lower, and a phase defect of a given magnitude results in a higher relative deviation from the ideal phase. In addition, the transmission of the metasurface masks is chromatic as well, resulting in a variable spatial transmission. Overall, it appears that although deviations from the optimal phase and amplitude response are present for both designs, the vortex phase mask is less sensitive to these deviations than the cosine and, hence, results in a deeper contrast.

It is important to note that although the metasurface designs provide good broadband phase control, their transmission is still chromatic. Therefore, the resulting performance shown in Fig. 7 is limited by both phase and transmission deviations from the nominal phase mask topography. However, it is possible to reduce the transmission variations by introducing an antireflective layer on top of the metasurface.[40] Its material and thickness would still have to be optimized, but its refractive index should lie in between the index of the incident medium and the substrate, similar to thin antireflective coatings. In addition, an antireflective coating will be needed on the backside of the substrate to reduce unwanted ghost reflections.

The current designs using square metasurface building blocks of varying sizes reach contrasts on the order of 1e–8. More complex shapes could be considered as pathways to increase the performance of metasurface masks by exploiting their increased design freedom (at the cost of higher complexity of optimization).[41–43] Another option would be to combine metasurfaces with shaped dielectric masks to compensate for their chromaticity. These strategies could push the performance of metasurface masks to even better contrasts.

6 Fabrication Challenges

The designs considered in this study are optimized specifically for a diamond metasurface in the visible wavelength range (centered at 650 nm). Due to the scale invariance of Maxwell's equations, the methods described can be readily applied to other wavelength ranges. The main difference is the appropriate choice of materials for different wavelength ranges. We chose diamond as a material for our study because of its excellent material properties over a broad wavelength range and its previous use for microstructured coronagraphic masks.[18] Diamond is transparent over a broad range of wavelength bands, from the ultraviolet to the mid-infrared, and has a high refractive index reducing the height necessary to gain 2π phase control. However, diamond microfabrication of these particular metasurface designs might be challenging due to the small features required for achromatic performance. Therefore, we discuss the suitability of other material platforms for different wavelength ranges below.

6.1 Ultraviolet

Diamond in its single-crystalline form is transparent down to 225 nm but is hard to manufacture at the required small feature sizes, and no diamond metasurfaces at these wavelengths have been manufactured to our knowledge. SiO_2 (JGS1) is also transparent down to 175 nm but has a relatively low refractive index requiring high aspect ratios. Hafnium oxide (HfO_2) is one of the few materials that have been used to successfully manufacture transmissive metasurfaces in the ultraviolet.[44] Although the small feature sizes make manufacturing of these metasurfaces challenging, this could be the best approach to manufacture the proposed metasurface designs in the ultraviolet.

6.2 Visible

In the visible wavelength range, titanium dioxide (TiO_2) or silicon nitride (Si_3N_4) has been used to implement high-transmission metasurface optics. These implementations typically use a fused silica substrate (SiO_2) with metasurface blocks on top which has the advantage of acting as an etch stop. This guarantees a uniform height everywhere on the mask, which is more challenging to achieve using a metasurface etched into a single substrate. The advantage of using a SiO_2 substrate is that the subwavelength condition $\Lambda < \lambda/n$ is partially relaxed because of the lower refractive index of the substrate. However, for a metasurface layer with a higher refractive index than the substrate, this comes at the expense of additional resonances resulting in highly nonlinear phase and amplitude response of the metasurface, typically leading to additional variation in transmission. Single crystal diamond could also be used for the visible range but would need dedicated process development to enable the required small feature sizes. Therefore, TiO_2 or Si_3N_4 constitutes the most promising fabrication approach in the visible range.

6.3 Infrared

Longer wavelengths are typically advantageous for manufacturing metasurface designs because they relax the subwavelength condition. In the near-infrared, silicon is a promising material beyond 1.2 μm because of established fabrication techniques inherited from the semiconductor industry. Monolithic metasurfaces for coronagraphic applications made of crystalline silicon and metasurfaces made of amorphous silicon on fused silica have been reported.[45,46] However, the high refractive index contrast results in large transmission variability across the mask and versus wavelength. This challenge could be addressed with an antireflective layer on top of the metasurface blocks and on the back of the substrate. Despite this drawback, silicon is the most promising material platform for manufacturing the proposed designs in the near-infrared. For the mid-infrared, diamond remains an excellent choice because of its high transmission, mechanical stability in cryogenic environments, and its past use in fabricating microstructured coronagraphic phase masks. Overall, to manufacture the proposed metasurfaces, it would likely be easiest to start from the infrared because of the relaxed subwavelength limit.

7 Conclusion

The detection and spectroscopy of Earth-like planets will require direct imaging instruments with a high-performance coronagraph, which provides high throughput, such as the scalar vortex coronagraph. This scalar architecture is in particular more suited for wavefront control compared with the established vector vortex coronagraph as it imprints the same phase ramp regardless of polarization and, therefore, does not require the polarizations to be split. Here, we have presented metasurface designs capable of achromatizing the scalar vortex coronagraph. Metasurfaces can achromatize a vortex because of their ability to correct for the wavelength dependence of the propagation phase. We have presented a metasurface-based scalar vortex design showing 2e–8 raw contrast in an annular region from 3 to 10 λ/D, making it a promising candidate for observing rocky planets around M-dwarfs with future ground-based telescopes. We also showed the effect of the phase discontinuities due to the 2π phase wrapping of the scalar vortex, which leads to a contrast floor of 6e–9 for an $f/40$ system. We then also investigated a metasurface design implementing an azimuthal cosine phase mask reaching a performance of the same order of magnitude (6e–8 raw contrast from 3 to 10 λ/D). The raw contrasts achieved in numerical simulations by these designs are similar to other achromatization strategies applied to the scalar vortex.[12,14] However, the proposed designs show reduced sensitivity to low-order aberrations compared with other scalar vortex topographies.[14] This is explained in more detail in Appendix B (Sec. 9) and illustrated in Table 1, which compares different scalar vortex types by their chromaticity and sensitivity to aberrations. Although optimized design topographies (e.g., the wrapped vortex or the dimpled vortex) can improve the broadband contrast of the classic and sawtooth vortex, they do not address the issue of sensitivity to low-order aberrations

Table 1 Comparison of scalar vortex types by chromaticity and sensitivity to aberrations. The chromaticity of the classic and sawtooth vortex can be improved by optimizing the design topography azimuthally (wrapped vortex) or radially (dimpled vortex), although not addressing the sensitivity to low-order aberrations. The metasurface vortex provides a similar degree of achromaticity but reduces the sensitivity to aberrations.

Scalar vortex type	Achromatic	Insensitive to aberrations
Classic vortex	✗	✗
Sawtooth vortex[13]	✗	✗
Wrapped vortex[12]	✓	✗
Dimpled vortex[14]	✓	✗
Metasurface vortex	✓	✓

caused by the inherent chromaticity of their phase response. Instead, the phase response of the metasurface designs is achromatic by design and therefore closer to the desired achromatic phase response of a perfect charge-6 vortex, making it less sensitive to low-order aberrations. Therefore, the presented metasurface designs are promising candidates for use on current and future ground-based telescopes, as well as future space telescopes. Finally, we have outlined possible material choices for the implementation of these designs across different wavelength ranges. Manufacturing these designs will be challenging but will benefit from fabrication processes of microstructured optics and the semiconductor industry and will likely start with a silicon infrared mask given the high degree of maturity of the relevant manufacturing processes.

8 Appendix A: Simulation Tools

This section provides additional detail about the software tools used throughout this study. First, we use the RCWA implementation RETICOLO to obtain the phase and amplitude response of a periodic array of metasurface building blocks. The software returns the complex amplitude of the electric field from which we compute the phase and amplitude of the transmitted wave. Because of the subwavelength condition, only the zeroth order propagates. A reference simulation propagating through an air layer of the same thickness is used for normalization. The local periodicity required by RCWA is fulfilled for regions of the metasurface masks without steep phase gradients.

For the simulation of aperiodic patterns such as the phase transitions and the central singularity, we use the FDTD implementation pyMEEP. We use a simulation cell containing the metasurface pattern surrounded by perfectly matched layers[47] on all sides to minimize reflections at the cell boundaries. The perfectly matched layers are artificial nonphysical layers and are therefore not included in the figures. We measure the response of the metasurface device to a plane wave monitoring the complex electric field, from which we compute the phase and amplitude of the transmitted wave.

Finally, we used the coronagraphic propagation software FALCO to simulate the coronagraphic performance of the proposed phase masks. FALCO uses a multi-scale Matrix Fourier transform approach to propagate through the coronagraph, which provides the necessary accuracy for high-contrast applications at the 1e–12 level. FALCO returns the coronagraphic image obtained for a given focal plane mask defined by its local phase and amplitude response, which we then use to compute the contrast in the desired region of interest.

9 Appendix B: Sensitivity to Low-Order Aberrations

This section provides a sensitivity analysis of the presented metasurface designs to low-order aberrations and compares them to previous scalar vortex design topographies. Aberrations impact the coronagraphic performance of a vortex phase mask, but depending on its charge,

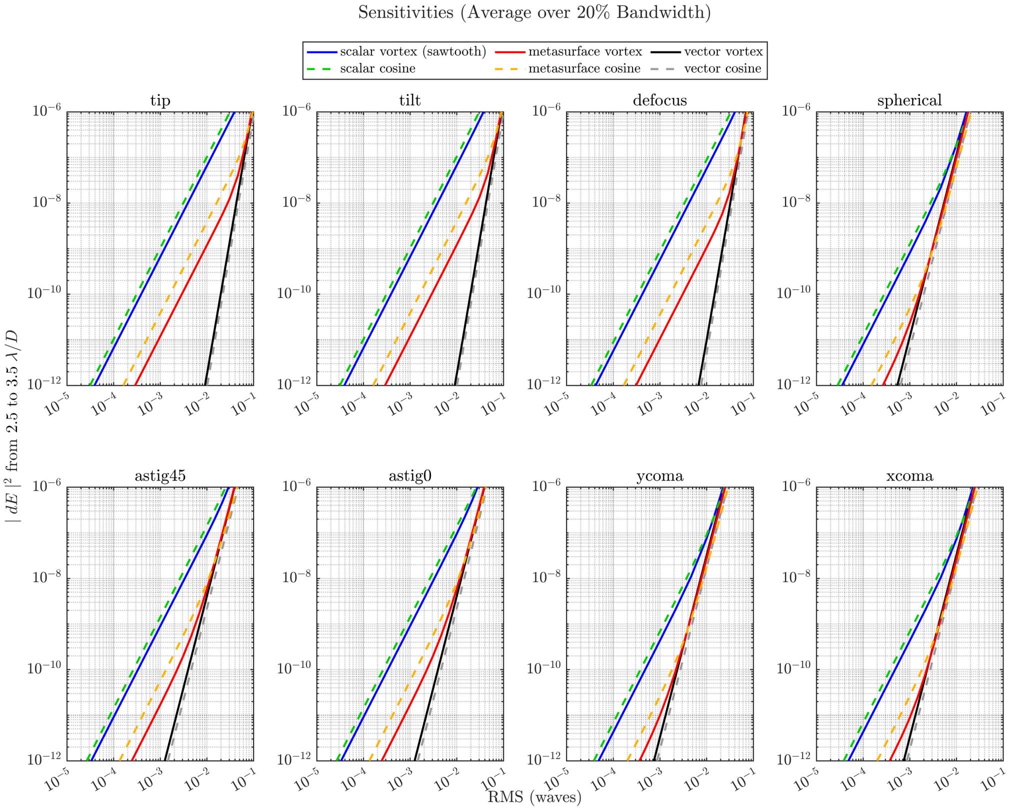

Fig. 8 Sensitivity to low-order aberrations of charge-6 metasurface phase masks. The change in intensity $|dE|^2$ in an annular region from 2.5 to 3.5 λ/D is shown against the injected aberration RMS for eight different low-order aberrations. The analysis has been carried out for the metasurface vortex (red) and cosine (orange) phase mask and is compared with the sawtooth scalar vortex (green) and a scalar cosine mask (orange). In addition, the limit of a perfect vector vortex and cosine mask is shown (black and gray).

certain low-order aberrations are known to have a significantly smaller effect on the coronagraphic performance (they are in the null space of the coronagraph).[48] For scalar vortex phase masks, this is only valid at the design wavelength, and the sensitivity degrades quickly when increasing the bandwidth because of their chromaticity.[14] The metasurface designs presented here partly restore the behavior of a perfect vortex mask because they are close to achromatic by design. This is illustrated in Fig. 8, which shows the effect of different low-order aberrations on the change in intensity $|dE|^2$ in a representative region of interest of the coronagraphic image plane (an annulus from 2.5 to 3.5 λ/D). Although previous scalar vortex topographies (green and blue curve) show high sensitivity to aberrations, the metasurface designs (red and orange) show reduced sensitivity approaching the behavior of a perfect achromatic vortex and cosine (black and gray curve). This is because of their phase response closely approaching the achromatic phase response of a perfect vortex.

Disclosures

Part of this work builds on two previously published proceedings articles (Refs. 49 and 45). The authors declare no conflicts of interest.

Code and Data Availability

The work in this paper uses the open source software RETICOLO,[33] MEEP,[34] and FALCO,[39] which are available online. The detailed code is available upon reasonable request to the authors.

Acknowledgments

Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (Grant No. 80NM0018D0004). L.K. is supported by an appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory, California Institute of Technology, administered by Oak Ridge Associated Universities under contract with NASA. O.A. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 819155).

References

1. D. Mawet et al., "Annular groove phase mask coronagraph," *Astrophys. J.* **633**(2), 1191–1200 (2005).
2. G. Foo, D. M. Palacios, and G. A. Swartzlander, "Optical vortex coronagraph," *Opt. Lett.* **30**, 3308–3310 (2005).
3. C. Delacroix et al., "A diamond AGPM coronagraph for VISIR," *Proc. SPIE* **8446**, 84468K (2012).
4. D. Mawet et al., "L'-band AGPM vector vortex coronagraph's first light on VLT/NACO: discovery of a late-type companion at two beamwidths from an F0V star," *Astron. Astrophys.* **552**, L13 (2013).
5. D. Defrère et al., "L'-band AGPM vector vortex coronagraph's first light on LBTI/LMIRCam," *Proc. SPIE* **9148**, 91483X (2014).
6. E. Serabyn et al., "The WM Keck Observatory infrared vortex coronagraph and a first image of HIP 79124 B," *Astron. J.* **153**(1), 43 (2017).
7. D. Mawet et al., "The vector vortex coronagraph: laboratory results and first light at Palomar Observatory," *Astrophys. J.* **709**, 53 (2009).
8. J. Kühn et al., "An H-band vector vortex coronagraph for the Subaru Coronagraphic Extreme Adaptive Optics system," *Publ. Astron. Soc. Pac.* **130**, 035001 (2018).
9. J. Llop-Sayson et al., "Vector vortex coronagraph experiments in vacuum towards 10^{-10} contrast," *Proc. SPIE* **13092**, 130921Y (2024).
10. G. A. Swartzlander, Jr., "Achromatic optical vortex lens," *Opt. Lett.* **31**(13), 2042–2044 (2006).
11. G. Ruane et al., "Scalar vortex coronagraph mask design and predicted performance," *Proc. SPIE* **11117**, 454–469 (2019).
12. R. Galicher et al., "A family of phase masks for broadband coronagraphy: example of the wrapped vortex phase mask theory and laboratory demonstration," *Astron. Astrophys.* **635**, A11 (2020).
13. N. Desai et al., "Laboratory demonstration of the wrapped staircase scalar vortex coronagraph," *J. Astron. Telesc. Instrum. Syst.* **9**(2), 025001 (2023).
14. N. Desai et al., "Benefits of adding radial phase dimples on scalar coronagraph phase masks," *J. Astron. Telesc. Instrum. Syst.* **10**(1), 015001 (2024).
15. N. Yu and F. Capasso, "Flat optics with designer metasurfaces," *Nat. Mater.* **13**(2), 139–150 (2014).
16. N. Tabirian, H. Xianyu, and E. Serabyn, "High efficiency broadband liquid crystal polymer vector vortex waveplates," in *Proc. IEEE Aerosp. Conf.*, IEEE, pp. 1–7 (2017).
17. Z. Bomzon et al., "Radially and azimuthally polarized beams generated by space-variant dielectric subwavelength gratings," *Opt. Lett.* **27**, 285–287 (2002).
18. C. Delacroix et al., "Laboratory demonstration of a mid-infrared AGPM vector vortex coronagraph," *Astron. Astrophys.* **553**, A98 (2013).
19. N. Murakami et al., "Coronagraph focal-plane phase masks based on photonic crystal technology: recent progress and observational strategy," *Proc. SPIE* **8442**, 844205 (2012).
20. S. Khonina et al., "The phase rotor filter," *J. Mod. Opt.* **39**(5), 1147–1154 (1992).
21. F. Hénault, "Strehl ratio: a tool for optimizing optical nulls and singularities," *J. Opt. Soc. Am. A* **32**(7), 1276–1287 (2015).
22. F. Hénault, A. Carlotti, and C. Vérinaud, "Analysis of nulling phase functions suitable to image-plane coronagraphy," *Proc. SPIE* **9912**, 99126K (2016).
23. F. Hénault, "Analysis of azimuthal phase mask coronagraphs," *Opt. Commun.* **423**, 186–199 (2018).
24. M. Khorasaninejad et al., "Metalenses at visible wavelengths: diffraction-limited focusing and subwavelength resolution imaging," *Science* **352**(6290), 1190–1194 (2016).
25. M. I. Shalaev et al., "High-efficiency all-dielectric metasurfaces for ultracompact beam manipulation in transmission mode," *Nano Lett.* **15**(9), 6261–6266 (2015).
26. A. Arbabi et al., "Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission," *Nat. Nanotechnol.* **10**(11), 937–943 (2015).
27. N. Yu et al., "A broadband, background-free quarter-wave plate based on plasmonic metasurfaces," *Nano Lett.* **12**(12), 6328–6333 (2012).
28. N. A. Rubin et al., "Imaging polarimetry through metasurface polarization gratings," *Opt. Express* **30**, 9389–9412 (2022).

29. W. T. Chen, A. Y. Zhu, and F. Capasso, "Flat optics with dispersion-engineered metasurfaces," *Nat. Rev. Mater.* **5**(8), 604–620 (2020).
30. S. Shrestha et al., "Broadband achromatic dielectric metalenses," *Light Sci. Appl.* **7**(1), 85 (2018).
31. S. Palatnick et al., "Prospects for metasurfaces in exoplanet direct imaging systems: from principles to design," *Proc. SPIE* **12680**, 226–239 (2023).
32. M. Moharam and T. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," *J. Opt. Soc. Am.* **71**(7), 811–818 (1981).
33. J. P. Hugonin and P. Lalanne, "Reticolo software for grating analysis," arXiv:2101.00901 (2005).
34. A. F. Oskooi et al., "Meep: a flexible free-software package for electromagnetic simulations by the FDTD method," *Comput. Phys. Commun.* **181**(3), 687–702 (2010).
35. C. Delacroix et al., "Design, manufacturing, and performance analysis of mid-infrared achromatic half-wave plates with diamond subwavelength gratings," *Appl. Opt.* **51**, 5897–5902 (2012).
36. X. Zhang and Y. Wu, "Effective medium theory for anisotropic metamaterials," *Sci. Rep.* **5**(1), 7892 (2015).
37. L. König et al., "Optimal design of the annular groove phase mask central region," *Opt. Express* **30**, 27048–27063 (2022).
38. N. Roy et al., "Photonic-structure optimization using highly data-efficient deep learning: application to nanofin and annular-groove phase masks," *Phys. Rev. A* **109**, 013514 (2024).
39. A. E. Riggs et al., "Fast linearized coronagraph optimizer (FALCO) i: a software toolbox for rapid coronagraphic design and wavefront correction," *Proc. SPIE* **10698**, 878–888 (2018).
40. O. Koksal et al., "Antireflective multi-dielectric metasurfaces operating in the visible," *Laser Photon. Rev.* **18**(1), 2300656 (2024).
41. N. Song et al., "Broadband achromatic and polarization insensitive focused optical vortex generator based on metasurface consisting of anisotropic nanostructures," *Front. Phys.* **10**, 846718 (2022).
42. Z. H. Jiang et al., "Tailoring dispersion for broadband low-loss optical metamaterials using deep-subwavelength inclusions," *Sci. Rep.* **3**(1), 1571 (2013).
43. J. T. Heiden and M. S. Jang, "Design framework for polarization-insensitive multifunctional achromatic metalenses," *Nanophotonics* **11**(3), 583–591 (2022).
44. C. Zhang et al., "Low-loss metasurface optics down to the deep ultraviolet region," *Light Sci. Appl.* **9**(1), 55 (2020).
45. L. König et al., "Design and prototyping of broadband metasurface scalar phase masks for high-contrast imaging," *Proc. SPIE* **13100**, 1310024 (2024).
46. S. Palatnick et al., "Achromatizing photolithographically patterned metasurfaces with arbitrary, variable unit cell size," *Opt. Express* **32**, 47057–47075 (2024).
47. A. Oskooi and S. G. Johnson, "Distinguishing correct from incorrect PML proposals and a corrected unsplit PML for anisotropic, dispersive media," *J. Comput. Phys.* **230**(7), 2369–2377 (2011).
48. G. Ruane et al., "Vortex coronagraphs for the Habitable Exoplanet Imaging Mission concept: theoretical performance and telescope requirements," *J. Astron. Telesc. Instrum. Syst.* **4**(1), 1 (2018).
49. L. König et al., "Metasurface-based scalar vortex phase mask in pursuit of 1e–10 contrast," *Proc. SPIE* **12680**, 240–247 (2023).

Biographies of the authors are not available.