

Supplementary Materials for

Matrix Fourier optics enables a compact full-Stokes polarization camera

Noah A. Rubin, Gabriele D'Aversa, Paul Chevalier, Zhujun Shi, Wei Ting Chen,
Federico Capasso*

*Corresponding author. Email: capasso@seas.harvard.edu

Published 5 July 2019, *Science* **365**, eaax1839 (2019)
DOI: [10.1126/science.aax1839](https://doi.org/10.1126/science.aax1839)

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 and S2
Caption for Movie S1
References

Other Supplementary Materials for this manuscript include the following:
(available at science.scienmag.org/content/365/6448/eaax1839/suppl/DC1)

Movie S1

Supplementary Materials for “Matrix Fourier optics enables a compact full-Stokes polarization camera”

Noah A. Rubin,¹ Gabriele D’Aversa,^{1, 2} Paul Chevalier,¹ Zhujun Shi,³ Wei Ting Chen,¹ and Federico Capasso^{1,*}

¹*Harvard John A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, MA 02138, USA*

²*Section de Physique, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

³*Department of Physics, Harvard University, Cambridge, MA 02138, USA*

(Dated: April 23, 2021)

CONTENTS

S1. Mathematical exposition	2
A. Beam splitting by diffraction gratings	2
B. Beam splitting by phase-only scalar gratings	3
C. Optimization of phase-only gratings	4
D. Parallel polarization analysis by matrix gratings	5
E. Parallel polarization analysis by linearly birefringent matrix gratings	7
1. Linear birefringence implies $ p_k\rangle = q_k^*\rangle$	8
2. Relation to past work on “Metasurface Polarization Optics” [7, 8]	9
3. A unitary, linearly birefringent matrix grating for polarization analysis can only have two orders acting as analyzers	10
S2. Optimization-based design of matrix gratings for parallel polarization analysis	14
A. Development of an optimization scheme	15
1. Parameterization of the grating unit cell	15
2. Optimization	16
B. Results for 11-element gratings	18
1. Progression of the optimization	18
2. Optimized results	19
C. Conversion to metasurface design	20
S3. Experimental characterization of gratings	20
A. Fabrication	20
B. Mueller matrix polarimetry	20
1. Comparison of optimized, simulated, and experimentally measured Mueller matrices	22
C. Measurement results for octahedron grating	23
D. Overall grating efficiency	24
S4. Polarization imaging	26
A. Optical design of the camera system	26
B. Angle-dependent calibration	27
C. “Checkerboard” image registration	28
D. Polarization image synthesis	29
E. Edge effects and the zero order	30
S5. Caption for Supplemental Movie S1	31
References	32

* capasso@seas.harvard.edu

S1. MATHEMATICAL EXPOSITION

Here, we attempt to provide a more fluid picture of the logic underpinning matrix diffraction and gratings described in the main text, offering clarity that space constraints there did not allow. In the course of this development, we prove several results that were stated in the main text.

This section is organized in order to make a parallel with the simpler case of scalar diffraction gratings which simply direct light in different directions in a way that is not polarization sensitive. We begin with these.

A. Beam splitting by diffraction gratings

A diffraction grating is an obstacle that varies periodically with space. Being periodic, light encountering a grating splits into a finite number of diffraction *orders*. If a diffraction grating having periodically-varying transmission function $\tilde{t}(x)$ is illuminated by a unit-amplitude plane wave normally incident on the grating, grating order k will propagate at an angle $\theta_k = \arcsin \frac{k\lambda}{d}$ where λ is the wavelength of the illuminating light and d is the periodicity of the grating. Additionally, each diffraction order has a complex-valued scalar weight a_k giving the amplitude of the electric field and thus the intensity and efficiency of order k . This weight is given by the coefficient of the term e^{-ikx} in the Fourier series of $t(x)$:

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikx} \tilde{t}(x) dx. \quad (\text{S1})$$

Here, we have taken $d = 2\pi$ for convenience.

A question which naturally arises is how a grating can be designed to split an incoming plane wave into a finite number of orders with equal intensity (or, more generally, some specified weighting) or, equivalently, to combine a finite number of beams into one plane wave. This question is of practical importance for optical system design. How can this be accomplished?

A grating that maps one input plane wave to one diffraction order will always be given by a linear phase ramp (a simple blazed grating), in accordance with the Fourier transform shift theorem, so this is a trivial example.

The next simplest example would be to consider a grating in which light is equally split into two diffraction orders, say (for sake of demonstration), the first and zeroth orders. Then, the required transmission hologram is straightforwardly given by the sum

$$t(x) = \frac{1}{\sqrt{2}}(1 + e^{ikx}). \quad (\text{S2})$$

By mathematical construction, if a grating having this transmission function is illuminated by a normally incident, unit-amplitude plane wave, exactly two diffraction orders of equal strength will emerge. However, it is worth noting that this $t(x)$ will not map the unit-amplitude plane wave into an output with uniform amplitude. Indeed, the power transmission function specified by $t(x)$ is now spatially-varying:

$$T(x) = t^*(x)t(x) = 1 + \cos kx. \quad (\text{S3})$$

$T(x)$ is plotted over one period of the grating d in Fig. S1. There is a symmetry that assures that $\int_0^d T(x) dx = 1$ (i.e., overall energy conservation). This was accounted for in Eq. S2 with the factor $1/\sqrt{2}$. However, at some points a transmission above 1 is required, while at others, a transmission below 1 is demanded. While energy is conserved overall, it is not conserved point-to-point: gain and loss may be locally required. That is, this mandates a grating in which energy is globally conserved, but a redistribution of light among different parts of the grating unit cell is required.

At this point, we may wonder what would happen if only loss is implemented (it is practically easier to modulate local absorption than gain). If Eq. S3 is scaled by a factor of $1/2$, it will attain a maximum value of one (no gain) and only loss will be demanded (this is also depicted in Fig. S1, by a dashed curve).

However, this has a price: Now $\int_0^d T(x) dx = 1/2$ and half of the incident energy must be absorbed by the grating to split the light into just two orders. If we generalize to the case of N orders of equal intensity, it is easy to see that the power scaling factor will become $1/N$ to ensure loss-only amplitude modulation. In order to send light in N directions with a grating, must we accept an efficiency of $\eta = 1/N$?

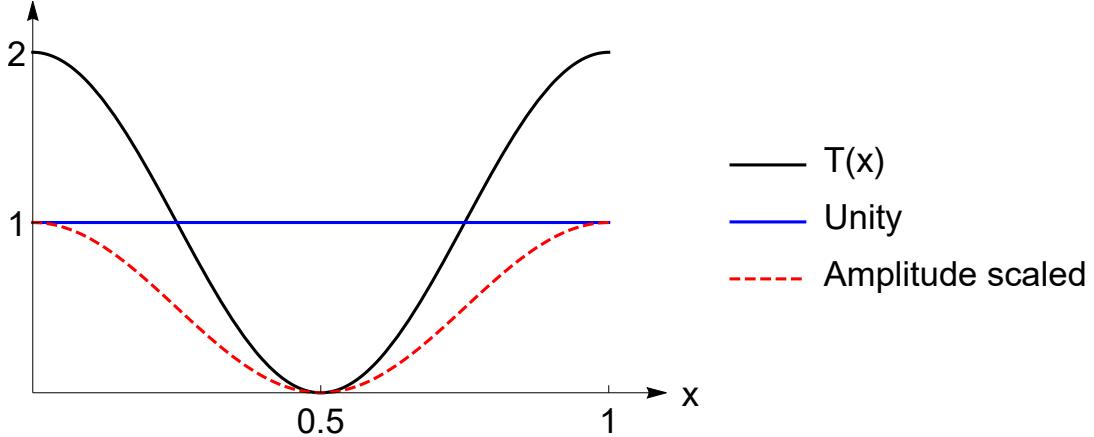


Figure S1. Transmission functions demanded by a grating producing exactly *two* non-zero diffraction orders. The x axis is in units of d , the grating period.

B. Beam splitting by phase-only scalar gratings

We next pose a somewhat related question. Is it ever possible to split light into a finite number of diffraction orders with 100% efficiency with *no* modulation of loss or gain? This would imply that the grating is *phase-only*, with $\tilde{t}(x) = e^{i\phi(x)}$. These gratings have been a subject of significant investigation in diffractive optics [27, 28, 53].

It is possible to prove that a phase-only grating may only have one or infinitely many diffraction orders, and so, a phase-only grating can never be the solution to the general beam splitting scenario laid out above. We prove this here, borrowing from [28] and [54]. The proof proceeds by contradiction.

In what follows, we assume that the grating is illuminated by a unit-amplitude plane wave so that the field following the grating is also $e^{i\phi(x)}$. Let the grating order coefficients (the Fourier coefficients) of the periodic grating be given by the set $\{a_k\}$. Furthermore, let us assume that all diffraction orders of the grating for which $a_k \neq 0$ lie between $k = m$ and $k = n$, with $n > m$. Suppose $n - m = N$. The phase-only grating can then be written as a Fourier expansion:

$$e^{i\phi(x)} = \sum_{k=m}^n a_k e^{ikx}. \quad (\text{S4})$$

Being phase-only, we have

$$e^{-i\phi(x)} e^{i\phi(x)} = 1 = \sum_{k=m}^n \sum_{\ell=m}^n a_\ell^* a_k e^{i(k-\ell)x}. \quad (\text{S5})$$

Clearly, the sum may possess no amplitude modulation (i.e., it may have no dependence on x). All exponential terms for which $k - \ell \neq 0$ must cancel. We can write a system of equations:

$$\begin{aligned} a_m^* a_n &= 0 \\ a_m^* a_{n-1} + a_{m+1}^* a_n &= 0 \\ a_m^* a_{n-2} + a_{m+1}^* a_{n-1} + a_{m+2}^* a_n &= 0 \\ &\vdots \\ a_m^* a_{m+2} + a_{m+1}^* a_{m+3} + \dots + a_{n-2}^* a_n &= 0 \\ a_m^* a_{m+1} + a_{m+1}^* a_{m+2} + \dots + a_{n-1}^* a_n &= 0 \\ \sum_{k=n}^m |a_k|^2 &= 1 \end{aligned} \quad (\text{S6})$$

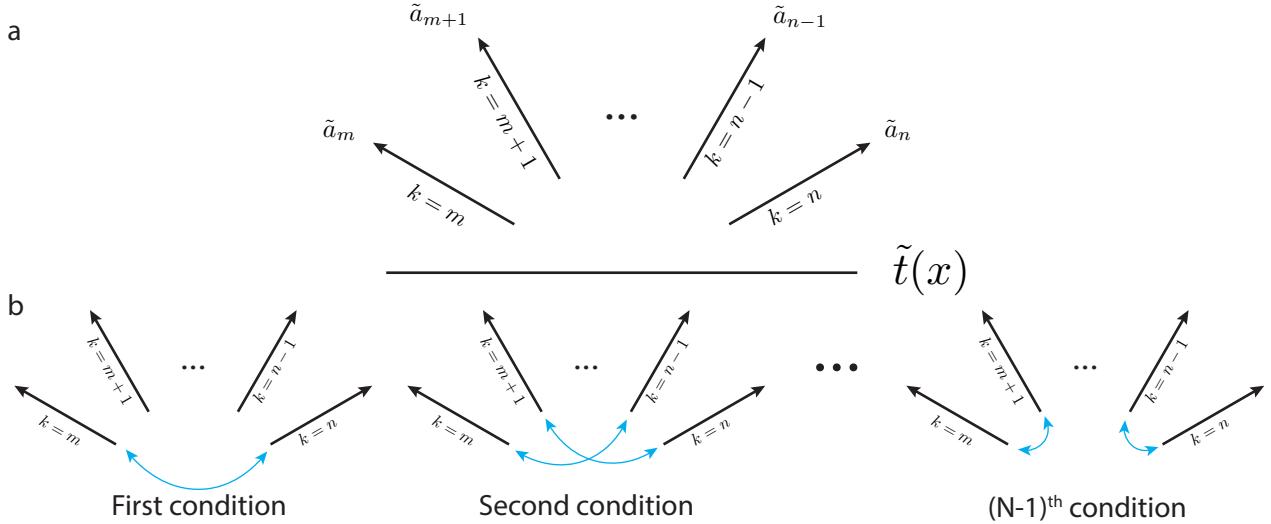


Figure S2. a. Schematic of diffraction into a finite number of orders between $k = m$ and $k = n$ by a diffraction grating. b. Each equation in Eq. S22 examines the all pairs of orders separated by a given integer distance, starting with N in the first equation and proceeding downwards with the base case being nearest-neighbor diffraction orders.

There are $N+1$ conditions that must be fulfilled. The first equation assures that the highest order exponential does not contribute to the sum (this will be a term that goes as $e^{iN\tilde{x}}$). The next equation covers terms having exponentials of order $N - 1$, and so forth. Finally, at the base of the system, we obtain an equation for “first-order” exponentials where “nearest neighbor” grating order coefficients are multiplied together. The last equation assures that the DC term has unit magnitude; this can be interpreted as a statement of energy conservation (i.e., Parseval’s theorem).

The first equation implies that one of a_m or a_n , the weights of the outermost diffraction orders, must be zero. Suppose we choose $|a_m| = 0$. The next equation implies $|a_{m+1}| = 0$, and the next that $|a_{m+2}| = 0$. As we trace through the equations, we will eventually conclude that all coefficients must be zero except for a_n . Therefore, for a phase-only grating (if a finite number of terms is assumed), only one can have an intensity differing from 0 (there can also be an infinite number of diffraction orders). This is the familiar case of a blazed grating with a linear phase profile in x . A phase-only grating may also possess an infinite number of diffraction orders whose a_k differ from zero; this case is not excluded by this proof since we assumed a finite number of orders between $k = n$ and $k = m$.

C. Optimization of phase-only gratings

The above proof by contradiction shows that beam-splitting with a diffraction grating can never be accomplished with 100% efficiency. But it has not answered the question of whether we can do better than $1/N$.

It stands to reason that since 100% of incident light passes through a phase-only grating, a high proportion of it could go into a limited set of diffraction orders with some leakage into all other diffraction orders. What is the best that a phase-only grating can do, concentrating as much light as possible in a limited set of desired diffraction orders while minimizing this leakage? This problem was solved for three orders in [53] but was stated and solved in complete generality by Romero & Dickey [27, 28].

First, we state the problem:

Problem. Given a set of diffraction orders whose indices are given in the set $\{\ell\}$, find the function $\phi(x)$ with period d that maximizes the efficiency functional $\eta(\phi(x))$ given by

$$\eta(\phi(x)) = \frac{\sum_{k \in \{\ell\}} |a_k|^2}{\sum_{k=-\infty}^{k=+\infty} |a_k|^2} \quad (S7)$$

under the constraints that the magnitude of coefficients a_k , and thus the power of the diffraction orders, are in a desired relative proportion. In the notation of [28], we insist that $|a_k| = c\gamma_k$ for a constant c and a specified set of weights $\{\gamma_k\}$. The $\{a_k\}$ are given by Eq. S1.

N	η_{phase}	η_{loss}
2	0.8106	0.500
3	0.9256	0.3333
4	0.9119	0.2500
5	0.9212	0.2000
6	0.8817	0.1667
7	0.9684	0.1429

Table S1. Comparison of diffraction efficiency for gratings that implement the optimized phase-only gratings defined by Eqs. S8 and S9 (η_{phase}) and those that implement loss-only modulation ($\eta_{\text{loss}} = 1/N$, as in Fig. S1) as a function of N , the number of diffraction orders into which light is directed. This assumes equal intensity on these orders for simplicity.

This is inherently a problem of variational calculus. Using the method of Lagrange multipliers (to address the constraints), it can be shown [28] that the optimum phase function has the form

$$t(x, \{\alpha_k\}, \{\mu_k\}) = e^{i\phi(x)} = \frac{s(x, \{\alpha_k\}, \{\mu_k\})}{|s(x, \{\alpha_k\}, \{\mu_k\})|} \quad (\text{S8})$$

with

$$s(x, \{\alpha_k\}, \{\mu_k\}) = \sum_{k \in \{\ell\}} \mu_k e^{i\alpha_k} e^{ikx}. \quad (\text{S9})$$

This gives an analytical form for the optimum. The weights $\{\mu_k\}$ and phases $\{\alpha_k\}$ must be determined based on the constraints of the original problem. Romero & Dickey showed that this results in a system of $2N - 3$ (non-linear) equations, where N is the number of diffraction orders to be controlled, which can be numerically solved with, e.g., Newton's method.

Eqns. S8 and S9 are notable because they take the problem of finding the optimum, continuous grating phase function $\phi(x)$, which is an infinite dimensional problem (defining the function at every point in the unit cell), and reduce it to one with a dimension of $2N - 3$ which is numerically easy to address.

In Table S1 we compare the efficiency η_{phase} obtained by a phase-only grating optimized using this approach with the efficiency $\eta_{\text{loss}} = 1/N$ that would be realized with loss-only modulation, as discussed above, for the problem of equally splitting light into N diffraction orders. We do not know of a general proof showing that $\eta_{\text{phase}} > \eta_{\text{loss}} \forall N$ — however, inspection of Table S1 suggests that this is indeed the case, and this appeals to intuition since a phase-only grating is lossless. It seems a phase-only grating can be more successful in splitting light than one designed with a more direct approach. Moreover, the phase-only approach is more compatible with readily accessible grating implementations, such as surface-relief gratings and, more recently, metasurfaces.

D. Parallel polarization analysis by matrix gratings

Here, we use the matrix approach introduced in the main text to examine gratings that implement polarization analyzers (polarizers) on their diffraction orders. In the main text, we focused on the design and demonstration of a “tetrahedron” grating in which four diffraction orders behave as analyzers whose polarizations uniformly sample the Poincaré sphere. This grating was the subject of main text Fig. 2. We focused on this tetrahedron grating in the main text because the tetrahedron configuration of polarization states implements optimal polarimetry with the minimum number of measurements [36] and thus enables the full-Stokes camera that is a major focus of this work. However, we also designed, fabricated, and tested a second grating, an “octahedron” grating, that implements six distinct polarization state analyzers in parallel on its orders. In this supplement, we discuss both the tetrahedron and octahedron gratings together.

Moreover, we wish to stress that the matrix Fourier optics approach presented in the main text is only valid in the paraxial (small angle) limit. Strictly speaking, when beams of different polarizations traveling at different angles interfere, the result is most generally a 3D electric field vector; if the beams are close in angle, the 2D Jones description of light’s polarization can still be used. Whether explicitly stated or not, this assumption underlies most all past work in spatially-varying polarization optics, including in metasurfaces.

This section will closely parallel the development of Sec. S1A. It is clear that the two problems are very similar. There, we sought to split light into just N diffraction orders. In this section, we wish to do the same, but with polarization in mind. In particular, the light on order k should take on an intensity dictated by Malus' Law in accordance with its preferred analyzer polarization $|q_k\rangle$.

More specifically, we can specify a set of diffraction orders $\{\ell\}$. For each order $k \in \{\ell\}$, we specify that its matrix Fourier coefficient \tilde{J}_k should take the form $\tilde{J}_k = a_k |p_k\rangle\langle q_k|$ where a_k is a scalar weight, $|q_k\rangle$ is the Jones vector of the polarization that the order analyzes for, and $|p_k\rangle$ is the output polarization of the order, as in the main text. 2×2 Jones matrix quantities are denoted with a tilde throughout the main text and this supplement.

For any choice of these parameters, we can follow the approach of the main text and immediately write the Fourier series

$$\tilde{J}(x, y) = \sum_{\vec{k} \in \ell} a_k |p_k\rangle\langle q_k| e^{i(k_x x + k_y y)}. \quad (\text{S10})$$

(For the sake of correspondence with the main text, we cast this problem in 2D, but the same arguments would all hold in 1D.)

By mathematical construction, this grating will have exactly N orders which each behave as polarization state analyzers for the proscribed polarization states $\{|q_k\rangle\}$.

At this point in Sec. S1A, we computed the power required by the computed $t(x)$ as $T = t^*t$. What is the equivalent for the matrix quantity $\tilde{J}(x, y)$? The equivalent transmission is given by

$$T = \frac{1}{2} \text{Tr}(\tilde{J}^\dagger \tilde{J}). \quad (\text{S11})$$

Here, Tr is the matrix trace. This T is a power transmission averaged over all possible incident polarizations on the Poincaré sphere. In fact, T is the expression for the upper-leftmost element of the Mueller matrix M corresponding to \tilde{J} [55] — this makes sense because this element of the Mueller matrix controls intensity transmission. It is also readily seen that if \tilde{J} is unitary (that is, if $\tilde{J}^\dagger \tilde{J} = \mathbb{1}$), $T = 1$ which is commensurate with the definition of a unitary matrix.

Next, in analogy to the two beam splitting example of Sec. S1A, we can compute the $\tilde{J}(x, y)$ by Eq. S10 for some specific example and examine what variation of T over the grating unit cell it necessitates. We examine two 2D examples:

1. **An ideal tetrahedron grating:** This grating sends all light into just four diffraction orders. Each diffraction order acts as an analyzer for one of the four Jones vectors corresponding to the Stokes polarization states of a tetrahedron inscribed in the Poincaré sphere on the orders $\{\ell\} = \{(0, -1), (0, 1), (1, 0), (0, -1)\}$. This is shown in the top left of Fig. 2(a) of the main text, and reproduced in Fig. S3.¹
2. **An ideal octahedron grating:** This grating sends all light into six diffraction orders. Each diffraction order acts as an analyzer for one of the six Jones vectors corresponding to the Stokes polarization states of an octahedron inscribed in the Poincaré sphere on the orders $\{\ell\} = \{(0, -1), (0, 1), (1, 0), (0, -1), (1, 1), (-1, -1)\}$. These polarization states are $|x\rangle$, $|y\rangle$, $|45^\circ\rangle$, $|135^\circ\rangle$, $|R\rangle$, and $|L\rangle$, perhaps the most familiar polarization ellipses. This is shown in the bottom left of Fig. 2(a) of the main text, and reproduced in Fig. S3.

The tetrahedron example was a focus of the main text, and the octahedron example was also realized and is discussed throughout this supplement.

In what follows, we will assume that for each order of these gratings, $|p_k\rangle = |q_k^*\rangle$, i.e., that the output polarization on the diffraction order is the same as the polarization analyzed for except with flipped handedness. In the main text it was stated that this restriction is implied by the use of linearly birefringent waveplate-like elements in the realization of $\tilde{J}(x, y)$. This will be proven below, but we already work under this assumption here.

For each of the two cases above, we can form the sum suggested by the Fourier expansion of Eq. S10.² Next, we can examine the spatially-dependent transmission (cf. Eqn. S11) that these would require. This has been done in

¹ The orientation of the tetrahedron is not unique, but throughout this work we choose one in which one vertex is at the north pole of the sphere, corresponding to circularly polarized light, and another vertex is aligned along the S_1 axis of the Poincaré sphere. These two choices are enough to constrain the exact polarization states of the tetrahedron.

² The last piece of this sum is the weight coefficients $\{a_k\}$. As will be explained below, insisting on *overall* energy conservation and equal intensity diffraction orders means that all the $\{a_k\}$ should be $1/\sqrt{2}$ in the tetrahedron case and $1/\sqrt{3}$ in the octahedron case.

Fig. S3 for both the perfect tetrahedron and octahedron cases. For each of these examples, the polarization-averaged transmission $\frac{1}{2}\text{Tr}(\tilde{J}^\dagger \tilde{J})$ is plotted, along with a blue plane representing unity. It can be shown that, in both examples,

$$\iint_{(x,y) \in \Omega} \frac{1}{2}\text{Tr}(\tilde{J}(x,y)^\dagger \tilde{J}(x,y)) dx dy = 1 \quad (\text{S12})$$

where Ω denotes one unit cell of the grating, meaning that energy is conserved overall. This is built into the construction of this problem by choice of the $\{a_k\}$, as we discuss in the next section, much like in the scalar case above where overall energy conservation was assured by energy conservation in the Fourier series. However, by inspection of the plots in Fig. S3, it can be seen that T is not unity everywhere in the unit cell. Just as in the simple scalar grating described above, amplitude gain is required at some places and loss at others in a way that preserves overall intensity but mandates a redistribution of intensity.

So far, this matrix diffraction example has been completely analogous to the simple scalar example of Sec. S1 A. In that discussion a proposed solution to this problem is loss-only modulation. That notion can be extended to this case — if $\tilde{J}(x,y)$ is scaled down by

$$\max_{(x,y) \in \Omega} \frac{1}{2}\text{Tr}(\tilde{J}(x,y)^\dagger \tilde{J}(x,y)) \quad (\text{S13})$$

each order will behave as an analyzer for the preferred polarization specified in the design with no local gain required, at the expense of an overall scaling in efficiency by the same factor.

These matrix examples are not analogous in every way to the simple scalar case of Sec. S1 A, however. Here, a full 2×2 matrix quantity must be implemented at each point of the grating unit cell, so the scalar quantity T cannot paint a complete picture. In other words, the polarization transforming properties of the required $\tilde{J}(x,y)$ must be implemented at each point to achieve the desired functionality on each diffraction order.

This is shown in Fig. S3 as well. For each grating, the required diattenuation (a unitless ratio) and retardance (an angle)³ are plotted at each point in the unit cell. These quantities vary significantly, and rapidly, point-by-point across the unit cell. It can be seen that to directly implement the Jones matrix necessitated by Eq. S10, a medium that can control overall transmission amplitude as well as precise polarization properties would be required at each point.

Perfection as mandated by the expansion in Eq. S10, then, requires agile, complex control of polarization and amplitude in a way that varies quickly with space over the periodic unit cell. This presents a challenge to practical implementation with real materials and real structures, beyond the realm of mathematics. In the scalar example of Sec. S1 A, we turned at this point to phase-only gratings, which could not solve the stated problem perfectly (in that some light had to leak into neighboring diffraction orders) but nonetheless could be implemented with practical diffraction gratings and could be shown to achieve high efficiencies in an optimized case.

Is there an equivalent in this matrix example, a mathematical simplification of $\tilde{J}(x,y)$ corresponding to gratings that easily realized in practice?

E. Parallel polarization analysis by linearly birefringent matrix gratings

In the next section, we consider matrix gratings in which $\tilde{J}(x,y)$ is constrained to act as a linearly birefringent, unitary matrix — that is, as a simple waveplate — everywhere. Structures of these kinds can be implemented simply with metasurfaces composed of dielectric pillars exhibiting form birefringence. If the dielectric has high index contrast with the surrounding medium, it may strongly confine light in its interior. If the pillar's transverse profile has two perpendicular mirror symmetry axes (as is the case with, e.g., a rectangle), two modes will propagate in the pillar which, due to the anisotropy, will propagate with different indices. As a result, the pillar can behave as a waveplate-like element whose Jones matrix can be written as

$$\tilde{J} = R(\theta) \begin{pmatrix} e^{i\phi_x} & 0 \\ 0 & e^{i\phi_y} \end{pmatrix} R(-\theta). \quad (\text{S14})$$

³ Diattenuation is the degree to which a polarization element acts like a polarizer. A diattenuation of 1 means that there is perfect transmission for one preferred polarization and perfect extinction for another. Retardance is the waveplate-like character of a Jones matrix which quantifies the amount by which one polarization is delayed relative to another. A general Jones matrix operator can have both of these properties at once, and not necessarily along the same eigen-axes. For more detail, see, e.g., [56].

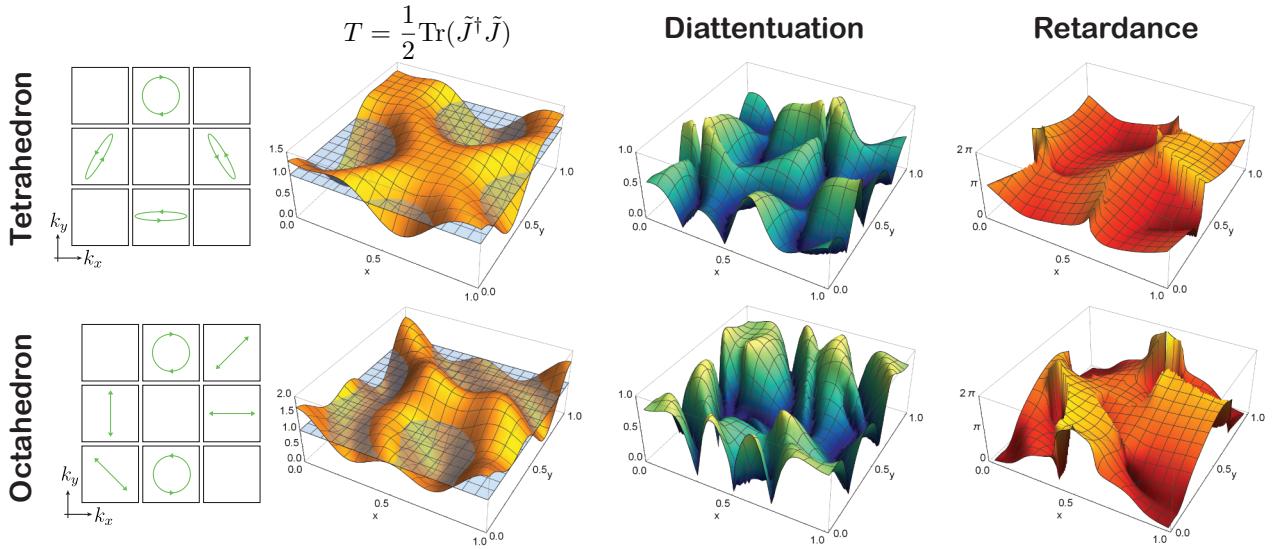


Figure S3. For each of the two cases in the main text, a *perfect* 2D matrix grating analyzing for a tetrahedron (top row) or octahedron (bottom row) of polarization states is synthesized by means of Eq. S10 – these polarization states and their corresponding diffraction orders are shown at the left, reproduced from main text Fig. 2(a). Each grating yields a $\tilde{J}(x, y)$ over its periodic 2D unit cell that can be analyzed to yield the three plots at the right. These are, from left to right, T (the polarization-averaged intensity transmittance given by Eqn. S11, with a blue surface showing unity), the required diattenuation, and the required retardance. In some sense, this figure is analogous to Fig. S1 in matrix space.

Here, $R(\theta)$ is a 2×2 rotation matrix by the angle θ corresponding to the physical angle by which the pillar element is rotated. The phases ϕ_x and ϕ_y can be continuously and independently controlled by adjusting the transverse dimensions of the pillar's cross-section. A library of pillar dimensions (w_x, w_y) and the phases ϕ_x, ϕ_y they impart can be assembled by full-wave simulation, and the rotation angle θ can take on any value between 0° and 360° . This has been extensively demonstrated by us and others elsewhere [7, 8, 38]. So, we do not dwell on this here except to say that if $\tilde{J}(x, y)$ is everywhere of the form of Eq. S14, there exists a convenient, established medium for its practical realization.

In this work, then, we insist that the Jones matrix $\tilde{J}(x, y)$ of the matrix grating should everywhere be of the form of Eq. S14. As described in the main text, Eq. S14 implies two specific restrictions:

1. $\tilde{J}(x, y)$ must be unitary everywhere. This implies that the local Jones matrix may only modify polarization states but not intensity. A unitary matrix may only enact phase changes and must preserve inner product between polarization states.
2. More specifically, $\tilde{J}(x, y)$ must be *linearly birefringent*. The eigenvectors of a unitary matrix are strictly orthogonal. However, these eigenvectors can, in general, be linear, circular, or elliptical. The requirement of linear birefringence means they must always be linear.

1. Linear birefringence implies $|p_k\rangle = |q_k^*\rangle$

It is simple to show that Eq. S14 will always describe a symmetric matrix for any choice of θ , ϕ_x , and ϕ_y because the eigenbasis of \tilde{J} will always be a rotated version of the $|x\rangle/|y\rangle$ system. In other words, the fact that \tilde{J} is linearly birefringent means that it is also symmetric in the $|x\rangle/|y\rangle$ polarization basis (the basis in which Eq. S14 is written). If everywhere $\tilde{J}(\vec{r})$ is of the form of Eq. S14, $\tilde{J}^T(x, y) = \tilde{J}(x, y)$. It is also true that if many matrices which are themselves symmetric are added together, the resultant sum will be symmetric as well.

A grating formed from a spatially-varying Jones matrix has diffraction orders whose behavior can be described by a Jones matrix which is a Fourier integral of $\tilde{J}(x, y)$ over the grating unit cell. This Fourier transform is a weighted sum of symmetric matrices — if $\tilde{J}(x, y)$ is everywhere symmetric, all \tilde{J}_k , the Jones matrices governing the behavior of the diffraction orders, must be as well.

If one of the \tilde{J}_k is to take the form of an analyzer, that is, if an order is to act like a polarizer in this work, it must be symmetric. Under what conditions is an analyzer Jones matrix symmetric? An analyzer Jones matrix is given by

$\tilde{J} = a|p\rangle\langle q|$. Parameterizing the output polarization as $|p\rangle = \begin{bmatrix} p_1 & p_2 \end{bmatrix}^T$ and the polarization being analyzed for as $|q\rangle = \begin{bmatrix} q_1 & q_2 \end{bmatrix}^T$, we have

$$J = a|p\rangle\langle q| = a \begin{pmatrix} q_1^* p_1 & q_2^* p_1 \\ q_1^* p_2 & q_2^* p_2 \end{pmatrix} \quad (\text{S15})$$

By inspection, this \tilde{J} is only symmetric if $q_1^* p_2 = q_2^* p_1$ or

$$\frac{q_1^*}{q_2^*} = \frac{p_1}{p_2}. \quad (\text{S16})$$

If $|q\rangle$ and $|p\rangle$ are both normalized Jones vectors, this requires that $|p_k\rangle = |q_k^*\rangle$ (since the ratio between their two components is identical in amplitude and phase).

2. Relation to past work on “Metasurface Polarization Optics” [7, 8]

On the Poincaré sphere this means that the vector corresponding to $|p_k\rangle$ must be mirrored about the equator with respect to $|q_k\rangle$ or, physically, that its polarization ellipse must be the same shape with the electric field rotating in an opposite sense (“flipped handedness”).

This conclusion has been identified in past work in which the goal was to impart two independent phase profiles on an arbitrary basis of polarization states [7, 8]. These past works took a scalar approach, designing devices in which output polarization is uniform over space when one of the polarization basis vectors is incident. That approach is a specific subcase of the matrix framework presented here: If the polarization basis of interest is given by $|\lambda^+\rangle$ and $|\lambda^-\rangle$ with $\langle\lambda^+|\lambda^-\rangle = 0$ (borrowing the notation of [7]), then the spatially-varying Jones matrix of those works is given by

$$\tilde{J}(x, y) = e^{i\phi^+(x, y)}|(\lambda^+)^*\rangle\langle\lambda^+| + e^{i\phi^-(x, y)}|(\lambda^-)^*\rangle\langle\lambda^-| \quad (\text{S17})$$

where $\phi^+(x, y)$ and $\phi^-(x, y)$ are the two phase profiles of interest to be experienced by the polarization basis $|\lambda^+\rangle$ and $|\lambda^-\rangle$ which are carried by the 0th order of diffraction. Eq. S17 captures the essential quality of the work in [7, 8] in that any polarization basis can experience independent phase profiles, which is a sub-case of the matrix approach of this work. Consequently, the proof that $|p_k\rangle = |q_k^*\rangle$ for an analyzer produced in the diffraction pattern of a linearly birefringent grating given here is more general.

A specific example of this is the so-called “geometric” or “Pancharatnam-Berry” phase approach to designing diffraction gratings (or holograms more generally) that are sensitive to circularly polarized light. Each circular polarization is converted to the other (equivalent to a mirroring about the equator). A geometric phase grating can be written as

$$\tilde{J}(x, y) = e^{+i\frac{2\pi \sin \alpha}{\lambda}}|R\rangle\langle L| + e^{-i\frac{2\pi \sin \alpha}{\lambda}}|L\rangle\langle R| \quad (\text{S18})$$

where α is the angle at which each circular polarization is redirected. This is a specific subcase of Eq. S17, which is itself a specific subcase of the matrix Fourier optics approach we present.

Past works on metasurface polarization optics (e.g., refs. [7, 8]) have taken the mathematical assumption that when a fixed polarization state is incident, the polarization created across the metasurface is uniform from point-to-point, varying only in phase. Going one level of mathematical abstraction further, we can allow the polarization, the Jones vector, to vary from point-to-point. This means that the *polarization* of the far-field (or diffraction orders) can vary, too. This is the approach taken in recent work by us [22] and others [23] — the key here is that a *particular* incident polarization is assumed by the design (the few examples cited here are by no means a complete representation of the countless papers in this subject area).

The matrix Fourier optics formalism of this work assumes no particular incident polarization, and does not assume that the field created by the metasurface is uniform in polarization. It is based on Fourier analysis of Jones matrices, so its description of the diffraction pattern handles all possible incident polarizations at once.

We can classify past work in this domain into three mathematical categories. For simplicity, the description is cast in terms of gratings, but these labels apply to more general optical elements without spatial periodicity as well:

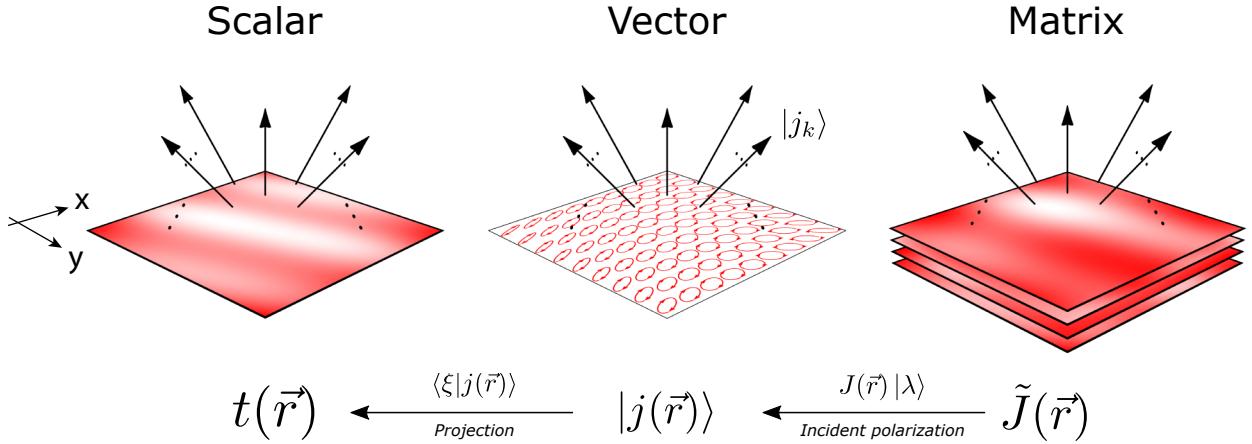


Figure S4. : A hierarchical view of polarization-dependent paraxial diffraction. In the scalar regime (top), a periodic electric field distribution $t(\vec{r})$ produces discrete orders with scalar weights $\{a_k\}$. In the vector regime (middle), the full polarization state $|j(\vec{r})\rangle$ is allowed to vary periodically with space and produces diffraction orders with characteristic Jones vectors (polarizations) $\{|j_k\rangle\}$. Finally, in the matrix regime, the Jones matrix of the grating may vary with space as $\tilde{J}(\vec{r})$ — in this case, the Fourier coefficients $\{\tilde{J}_k\}$ are themselves operators, encoding polarization-dependent behaviors. The vector field can be recovered from the matrix description if a particular polarization $|\lambda\rangle$ is incident, and a scalar field from a vector one if analyzed along a particular polarization $|\xi\rangle$.

- **Scalar:** The problem of designing a polarization-dependent optical element can be reduced to the scalar domain by considering that a particular incident polarization evokes a field that is uniform in its polarization over space, only differing in amplitude and phase. This allows us to ascribe a spatially-varying, *scalar* function $t(\vec{r})$ to the field. In the Fourier domain, this can be described by a set of *scalar* Fourier coefficients $\{a_k\}$.
- **Vector:** A particular incident polarization evokes a field with a spatially-varying Jones *vector* given by $|j(\vec{r})\rangle$. In the Fourier domain, diffraction orders are also described by Jones vector coefficients given by $\{|j_k\rangle\}$ that are themselves polarization ellipses.
- **Matrix:** This description, the subject of this work, is independent of any particular illuminating light and its polarization. A spatially-varying Jones *matrix* modifies incident polarization as a function of space according to the function $\tilde{J}(\vec{r})$. The diffraction orders, then, modify polarization as dictated by their Fourier coefficient Jones matrices $\{\tilde{J}_k\}$. This is a Fourier decomposition over the optical element's matrix-valued transmission function, rather than an electric field.

Finally, we note that these three regimes of polarization-dependent diffraction — scalar, vector, and matrix — are in fact intimately related, with the scalar and vector pictures as subcases of the matrix description. If a plane wave with Jones vector $|\lambda\rangle$ is normally incident on a matrix grating $\tilde{J}(\vec{r})$, a vector field $|j(\vec{r})\rangle = \tilde{J}(\vec{r})|\lambda\rangle$ results with Fourier coefficients $|j_k\rangle = \tilde{J}_k|\lambda\rangle$. If that vector field is projected (analyzed) along a particular polarization $|\xi\rangle$, a scalar field $t(\vec{r}) = \langle\xi|j(\vec{r})\rangle$ results with scalar Fourier coefficients $\tilde{a}_k = \langle\xi|j_k\rangle$. After all, a scalar field is described by a single spatially-varying complex number, a vector field by two, and a 2×2 Jones matrix element by four. This hierarchy is depicted in Fig. S4.

3. A unitary, linearly birefringent matrix grating for polarization analysis can only have two orders acting as analyzers

In the scalar regime, it was possible to prove that a phase-only grating may only have one or infinite diffraction orders. We have seen that for two specific examples (the octahedron and tetrahedron gratings) complex amplitude and polarization control is necessary for an exact solution, encapsulated in Fig. S3. Can a linearly birefringent matrix grating, which can be conveniently implemented practically, ever be the solution of the problem we wish to solve? Can a linearly birefringent matrix grating ever direct all light into a finite number of orders whose Fourier matrix coefficients J_k all behave as polarization state analyzers?

This proceeds in analogy to Sec. S1 A. Here, we assume a 1D matrix grating given by $\tilde{J}(x)$ (this can be generalized to 2D). Suppose all non-zero diffraction orders of the grating lie between $k = m$ and $k = n$, with $n > m$ with $n - m = N$. Each grating order k has an operator \tilde{J}_k which we insist takes the form of an analyzer so that we have

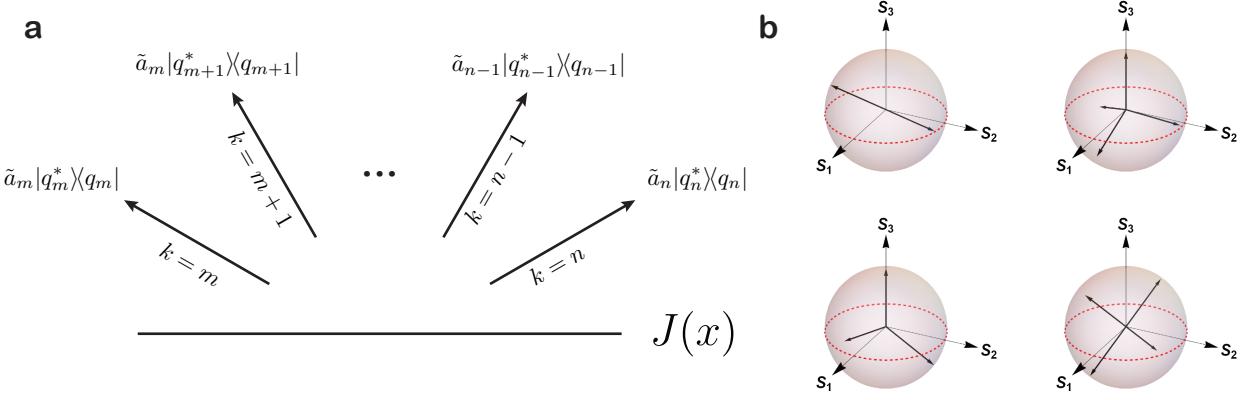


Figure S5. **a.** Schematic of parallel polarization analysis by a linearly birefringent matrix grating, using the same notation introduced in this section. Each order from $k = m$ to $k = n$ implements its own polarization analyzer matrix \tilde{J}_k . **b.** Each equation in Eq. S22 examines the all pairs of orders separated by a given integer distance, starting with N in the first equation and proceeding downwards with the base case being nearest-neighbor diffraction orders. **b.** Some examples of analyzer configurations $\{\vec{q}_k\}$ that obey the symmetry condition Eq. S29, including the tetrahedron of this work.

$$\tilde{J}_k = a_k |p_k\rangle\langle q_k|. \quad (\text{S19})$$

The matrix grating $\tilde{J}(x)$ is then defined by the set of scalar weights $\{a_k\}$, the set of output polarization states $\{|p_k\rangle\}$, and the set of analyzer polarization states $\{|q_k\rangle\}$. If we insist that the grating is linearly birefringent, it follows that $\{|p_k\rangle\} = \{|q_k^*\rangle\}$. Now, we can write a Fourier series for $\tilde{J}(x)$ as

$$\tilde{J}(x) = \sum_{k=m}^n a_k |q_k^*\rangle\langle q_k| e^{ikx}. \quad (\text{S20})$$

This grating must be *unitary everywhere*, and so we form the product $\tilde{J}^\dagger(x)\tilde{J}(x)$:

$$\tilde{J}^\dagger(x)\tilde{J}(x) = \sum_{k=m}^n \sum_{\ell=m}^n a_\ell^* a_k |q_\ell\rangle\langle q_\ell^*| q_k^* \langle q_k| e^{i(k-\ell)x} = \mathbb{1}. \quad (\text{S21})$$

Here $\mathbb{1}$ is the 2×2 identity matrix. As before, if the matrix product is independent of the spatial coordinate x , all exponential terms for which $k - \ell \neq 0$ must cancel. This permits a regrouping of the sum as

$$\begin{aligned}
 a_m^* a_n |q_m\rangle\langle q_m^*| q_n^* \langle q_n| &= 0 \\
 a_m^* a_{n-1} |q_m\rangle\langle q_m^*| q_{n-1}^* \langle q_{n-1}| + a_{m+1}^* a_n |q_{m+1}\rangle\langle q_{m+1}^*| q_n^* \langle q_n| &= 0 \\
 a_m^* a_{n-2} |q_m\rangle\langle q_m^*| q_{n-2}^* \langle q_{n-2}| + a_{m+1}^* a_{n-1} |q_{m+1}\rangle\langle q_{m+1}^*| q_{n-1}^* \langle q_{n-1}| + a_{m+2}^* a_n |q_{m+2}\rangle\langle q_{m+2}^*| q_n^* \langle q_n| &= 0 \\
 &\vdots \\
 a_m^* a_{m+2} |q_m\rangle\langle q_m^*| q_{m+2}^* \langle q_{m+2}| + a_{m+1}^* a_{m+3} |q_{m+1}\rangle\langle q_{m+1}^*| q_{m+3}^* \langle q_{m+3}| + \dots + a_{n-2}^* a_n |q_{n-2}\rangle\langle q_{n-2}^*| q_n^* \langle q_n| &= 0 \\
 a_m^* a_{m+1} |q_m\rangle\langle q_m^*| q_{m+1}^* \langle q_{m+1}| + a_{m+1}^* a_{m+2} |q_{m+1}\rangle\langle q_{m+1}^*| q_{m+2}^* \langle q_{m+2}| + \dots + a_{n-1}^* a_n |q_{n-1}\rangle\langle q_{n-1}^*| q_n^* \langle q_n| &= 0 \\
 &\sum_{k=m}^n |a_k|^2 |q_k\rangle\langle q_k| = \mathbb{1}.
 \end{aligned} \quad (\text{S22})$$

The imposing Eq. S22 is really no more complicated than Eq. S6 for phase-only gratings; now, it involves matrix quantities, increasing the bulk of the mathematical notation. Each equation in Eq. S22 is the coefficient of one exponential in the expansion Eq. S21 and can again be understood with the aid of Fig. S2(b). At the top level, we consider the term involving the most-distant-neighbors $k = m$ and $k = n$ and proceed downwards to the base case in which nearest-neighbors are considered. A schematic in Fig. S5(a) shows this pictorially.

Before considering the system Eq. S22 in its full complexity, it is worth remarking on the bottom line (reproduced here):

$$\sum_{k=m}^n |a_k|^2 |q_k\rangle\langle q_k| = \mathbb{1}. \quad (\text{S23})$$

This is the “DC” term of the expansion. Since this term has no space dependence, it must account for the entire sum. This equation is a matrix statement of Parseval’s equality from Fourier theory and is a statement of energy conservation in the matrix grating. We note that the outer product $|q_k\rangle\langle q_k|$ can be rewritten as [55]

$$|q_k\rangle\langle q_k| = \frac{1}{2}(\mathbb{1} + \vec{q}_k \cdot \vec{\sigma})\langle q_k|q_k\rangle. \quad (\text{S24})$$

Here, $\vec{q}_k = [q_k^{(1)} \ q_k^{(2)} \ q_k^{(3)}]^T$ is the Stokes state-of-polarization (SOP) corresponding to the Jones vector $|q_k\rangle$. The SOP is the last three elements of the Stokes vector and gives the coordinates of $|q_k\rangle$ ’s polarization state on the Poincaré sphere. The vector $\vec{\sigma} = [\sigma_1 \ \sigma_2 \ \sigma_3]^T$ is a vector where σ_k is the k^{th} Pauli matrix. Assuming that all the $|q_k\rangle$ are normalized so that $\langle q_k|q_k\rangle = 1 \forall k$, we can write from Eqs. S23 and S24

$$\sum_{k=m}^n \frac{|a_k|^2}{2}(\mathbb{1} + \vec{q}_k \cdot \vec{\sigma}) = \mathbb{1} \quad (\text{S25})$$

whose terms can be separated into the equations

$$\sum_{k=m}^n \frac{|a_k|^2}{2} = 1 \quad (\text{S26})$$

and

$$\sum_{k=m}^n \frac{|a_k|^2}{2} \vec{q}_k \cdot \vec{\sigma} = \mathbb{0} \quad (\text{S27})$$

where $\mathbb{0}$ is a 2×2 matrix of zeros.

Eq. S26 is a direct expression of energy conservation for this problem. In the case that all the orders are of equal strength (that is, that all $a_k = a$ with a constant), Eq. S26 implies that

$$|a|^2 = \frac{2}{N} \quad (\text{S28})$$

where N is the number of diffraction orders that are to behave as polarization analyzers with $\tilde{J}_k = a|q_k^*\rangle\langle q_k|$. For each order in this equal-intensity case, $|a|$ is the percent of incident power directed into the order when its preferred polarization is incident and provides an upper bound for I_q as defined in the text.⁴

Next we turn to Eq. S27, which can be further reduced to yield

$$\sum_{k=m}^n \frac{|a_k|^2}{2} \vec{q}_k = \vec{0} \quad (\text{S29})$$

with $\vec{0}$ being a three-element vector of zeros. This is a symmetry condition. It says that the “center of mass” of the analyzer Stokes vectors $\{\vec{q}_k\}$ as weighted by their peak intensities $\{a_k\}$ must be at the center of the Poincaré sphere.

⁴ It follows from Eq. S26 that for the tetrahedron grating (with $N = 4$) and the octahedron grating with ($N = 6$), the maximum value of I_q for any order is $1/2$ and $1/3$, respectively.

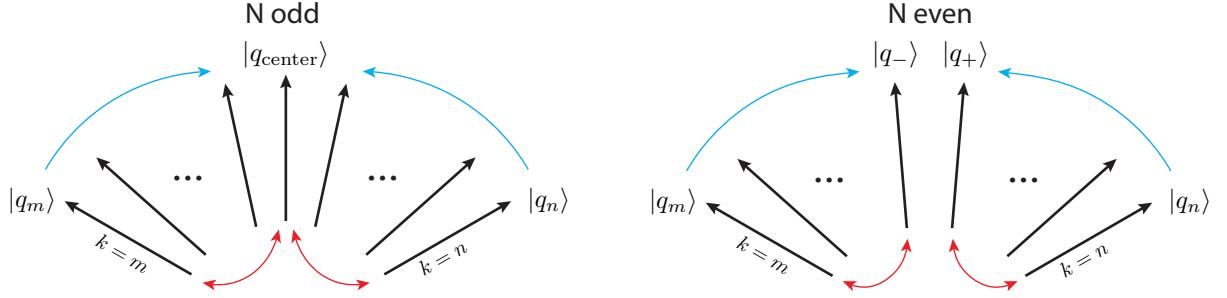


Figure S6. A contradiction preventing the solution of Eqs. S22 emerges at the midpoint of the diffraction pattern. The first equation requires that the outermost diffraction orders take on orthogonal analyzer polarizations $|q_n\rangle$ and $|q_m\rangle$. Each successive equation mandates that one diffraction order closer to the center from the left or right take on $|q_m\rangle$ or $|q_n\rangle$, respectively (blue arrows, top). At the equation in Eqs. S22 dealing with terms that are half the diffraction pattern apart, only two terms survive (red arrows) due to the orthogonality of analyzer vectors of the right and left halves of the diffraction pattern. In the case that N is odd, this equation mandates that the central order have an analyzer polarization state $|q_{\text{cent}}\rangle$ that takes on two states simultaneously (a contradiction). In the case that N is even, the inner two orders $|q_+\rangle$ and $|q_-\rangle$ are mandated by previous equations to be $|q_n\rangle$ and $|q_m\rangle$, but the next equation would require the opposite (also a contradiction). This midpoint contradiction can only be avoided in the case that $N = 2$ and $\langle q_m | q_n \rangle = 0$.

In the case of equal intensities, the vectors $\{\vec{q}_k\}$ must be symmetrically distributed on the Poincaré sphere. A few examples of vector sets $\{\vec{q}_k\}$ that fulfill this condition are shown in Fig. S5(b). This makes sense intuitively: If the incident light is to be directed into a finite number of orders all of which behave as polarization state analyzers, the states analyzed by each order must be compatible with one another. If for some incident polarization one diffraction order is dark, the $|q_k\rangle$ of the other orders must assure that the lost light is directed there. These power and symmetry conditions Eqs. S26 and S29 are pre-conditions that must be satisfied.

Having examined just the DC term, we return to the rest of Eq. S22. The top level equation $N, a_m^* a_n |q_m\rangle \langle q_m^*| q_n \rangle \langle q_n^*| = 0$, which takes care of the exponential e^{iN_x} , mandates that $\langle q_m^* | q_n^* \rangle = 0$ (if orders m and n are to have non-zero weights $|a_m|$ and $|a_n|$). Equivalently, $\langle q_m | q_n \rangle = 0$ and *the polarization states $|q_m\rangle$ and $|q_n\rangle$ analyzed by the outermost diffraction orders must be orthogonal*.

The next equation, $N - 1$ which addresses the term $e^{i(N-1)x}$ reads

$$a_m^* a_{n-1} |q_m\rangle \langle q_m^*| q_{n-1} \rangle \langle q_{n-1}^*| + a_{m+1}^* a_n |q_{m+1}\rangle \langle q_{m+1}^*| q_n \rangle \langle q_n^*| = 0, \quad (\text{S30})$$

stipulating that a sum of two matrix quantities is 0. For this to be accomplished without the requirement that $|a_{n-1}| = 0$ and $|a_{m+1}| = 0$, there are two possibilities:

1. Both $\langle q_m^* | q_{n-1}^* \rangle = 0$ and $\langle q_{m+1}^* | q_n^* \rangle = 0$, meaning that $|q_{n-1}\rangle = |q_n\rangle$ and $|q_{m+1}\rangle = |q_m\rangle$ (since the previous equation mandated $\langle q_m | q_n \rangle = 0$).
2. The matrices $|q_m\rangle \langle q_{n-1}|$ and $|q_{m+1}\rangle \langle q_n|$ describe parallel matrices (that is, scaled versions of one another, so that they can add to 0 without both being zero). Since all the $\{|q_k\rangle\}$ are normalized, this implies that $|q_{n-1}\rangle = |q_n\rangle$ and $|q_{m+1}\rangle = |q_m\rangle$.

Both scenarios require that $|q_{n-1}\rangle = |q_n\rangle$ and $|q_{m+1}\rangle = |q_m\rangle$, so we accept this as a condition of solution. We then proceed to the next equation, which weights an exponential of order $N - 2$:

$$a_m^* a_{n-2} |q_m\rangle \langle q_m^*| q_{n-2} \rangle \langle q_{n-2}^*| + a_{m+1}^* a_{n-1} |q_{m+1}\rangle \langle q_{m+1}^*| q_{n-1} \rangle \langle q_{n-1}^*| + a_{m+2}^* a_n |q_{m+2}\rangle \langle q_{m+2}^*| q_n \rangle \langle q_n^*| = 0 \quad (\text{S31})$$

Since $|q_{n-1}\rangle = |q_n\rangle$ and $|q_{m+1}\rangle = |q_m\rangle$ and $\langle q_m | q_n \rangle = 0$ from the previous two equations, $\langle q_{m+1}^* | q_{n-1}^* \rangle = 0$ and the middle term can be eliminated. Now, the equation is exactly analogous to the last, and we can conclude that $|q_{n-2}\rangle = |q_n\rangle$ and $|q_{m+2}\rangle = |q_m\rangle$.

The equations proceed in this manner, with each equation successively mandating that a diffraction order one more index inwards from n must have its $|q_k\rangle = |q_n\rangle$ and one more inwards from m must have its $|q_k\rangle = |q_m\rangle$. To satisfy the Eqs. S22, the diffraction orders coming from the left side must all have the analyzer polarization $|q_m\rangle$ while from the right they must all have $|q_n\rangle$. Eventually, the right and left halves of the diffraction pattern meet.

In the case that N is odd, we arrive at a unique “central” diffraction order whose analyzer polarization state $|q_{\text{center}}\rangle$ does not clearly belong to the right or left half of the diffraction pattern. When we examine the part of Eq. S22 that concerns diffraction orders that are $(N - 1)/2$ integers apart, we arrive at an equation that simplifies to

$$a_n^* a_{\text{center}} |q_n\rangle \langle q_n^*| q_{\text{center}}^* \rangle \langle q_{\text{center}}| + a_{\text{center}}^* a_m |q_{\text{center}}\rangle \langle q_{\text{center}}^*| q_m^* \rangle \langle q_m| = 0. \quad (\text{S32})$$

Only these terms survive since the right and left halves of the diffraction pattern (Fig. S6, left) have orthogonal polarizations $|q_m\rangle$ and $|q_n\rangle$. By inspection, Eq. S32 requires a contradiction. Its solution would require that $|q_{\text{center}}\rangle$ is equal to both $|q_n\rangle$ and $|q_m\rangle$ simultaneously, an impossibility given that the first equation stipulates that $\langle q_n|q_m\rangle = 0$.

If instead N is even (Fig. S6, right), there is no clear “central” order, but instead a central pair of orders. In between the two is the place where the $|q_k\rangle = |q_n\rangle$ and $|q_k\rangle = |q_m\rangle$ regions meet. Since there is no central order, it would appear that in this case there is no contradiction. However, when we examine the part of Eq. S22 that concerns orders that are $N/2 - 1$ orders apart, we get an equation of the form

$$a_-^* a_m |q_-\rangle \langle q_m^*| q_m \rangle + a_n^* a_+ |q_n\rangle \langle q_n^*| q_+ \rangle = 0. \quad (\text{S33})$$

Here, the diffraction orders $-$ and $+$ are those just left and right of center, respectively, as depicted in Fig. S6, right (there is no one center order because N is even). All other terms are 0 in Eq. S33 since $\langle q_m|q_n\rangle = 0$, and the left and right halves must take on the analyzer polarizations of the outer orders $|q_m\rangle$ and $|q_n\rangle$. This same reason means that Eq. S33 contains a contradiction: Working inward from the outside, the equation weighting an exponential term one higher than that of Eq. S33 requires that $|q_-\rangle = |q_m\rangle$ and $|q_+\rangle = |q_n\rangle$, but Eq. S33 just the opposite, that $|q_-\rangle = |q_n\rangle$ and $|q_+\rangle = |q_m\rangle$. Once again a contradiction is encountered at the crossing over the midpoint of the diffraction pattern.

If there are any more than two orders with non-zero a_k , we will encounter this midpoint crossing problem depicted in Fig. S6. In essence, this stems from the mathematical fact that only two different polarization states can be orthogonal to one another. If, however, there are only two diffraction orders which act as analyzers for an orthogonal basis of polarizations $|q_m\rangle$ and $|q_m\rangle$ with $\langle q_n|q_m\rangle = 0$, then Eq. S22 can always be satisfied if the a_k of these orders are both equal to $1/2$ (which is required by the power and symmetry conditions Eqs. S26 and S29).⁵

We have shown by contradiction that if a linearly birefringent⁶, unitary matrix grating directs all incident light into a finite set of diffraction orders acting as polarization state analyzers, there may only be two such orders. These orders must be of equal strength and must analyze for orthogonal polarizations $|q_m\rangle$ and $|q_m\rangle$ such that $\langle q_n|q_m\rangle = 0$.

S2. OPTIMIZATION-BASED DESIGN OF MATRIX GRATINGS FOR PARALLEL POLARIZATION ANALYSIS

The key conclusion of Sec. S1 is that the *direct* implementation of gratings implementing parallel polarization analysis would necessitate complex, spatially-modulated amplitude and polarization behavior. In analogy with phase-only gratings in the scalar case, it is tempting to seek a solution that is unitary and linearly birefringent everywhere because this has a very straightforward, proven practical implementation in the form of dielectric metasurfaces which implement Jones matrices that are everywhere of the form of Eq. S14. In Sec. S1 it was also proven at some length that a linearly birefringent, unitary grating can never perfectly be the solution we seek — if all light is directed into a finite set of orders behaving as analyzers (polarizers), there can only be two such orders and they must analyze for orthogonal polarizations.

Similarly, in the scalar beam splitting problem, a phase-only grating can be shown to—mathematically—yield only one or infinite non-zero diffraction orders (Sec. S1B). If an *imperfect* solution is nonetheless accepted, with some light leaking into unwanted orders, optimized solutions with high efficiency can nonetheless be obtained (Sec. S1C, and [27, 28]). It is tempting to follow the same logic here. Can an optimized grating that is unitary and linearly birefringent everywhere implement analyzers for desired polarization states on a finite set of diffraction orders with high efficiency, minimizing leakage into other diffraction orders?

⁵ One may claim that in this derivation we have only considered cases in which all diffraction orders in the set between $k = m$ and $k = n$ have non-zero intensity, or $|a_k| \neq 0$. Can selectively “skipping” some diffraction orders (allowing some to have $|a_k| = 0$) sidestep the contradiction that occurs at the midpoint? However, by skipping some orders, the location of the midpoint simply moves, and a contradiction still occurs. Solution can only be obtained if all intermediate orders are “skipped” and we are left with just two orders.

⁶ If the grating is not constrained to be linearly birefringent, just unitary, we are free from the requirement that $|p_k\rangle = |q_k^*\rangle$ and the argument presented here breaks down. There are cases in which multiple (> 2) orders can behave as analyzers.

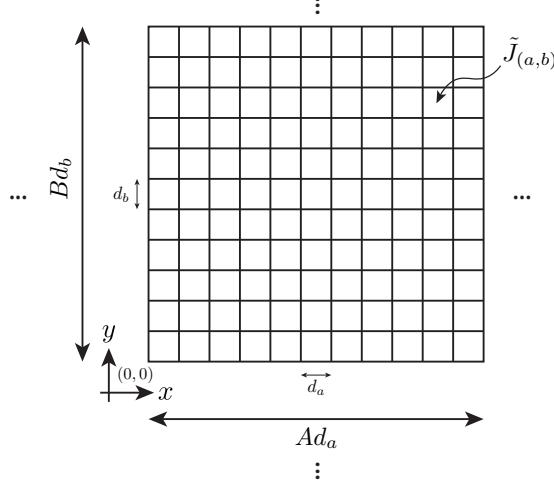


Figure S7. Schematic of the unit cell and its parameters. A square unit cell is shown here because the grating unit cells in this work are square, but in general $A \neq B$.

In the phase-only case, this was framed as an optimization under constraints with an analytical solution for the global optimum obtained using variational calculus (Sec. S1C). An analogous *analytical* approach applied to the (more complex) matrix case at hand is also possible. This will be the subject of future work.

In this work, we instead concoct a simple numerical scheme to search for an optimum. We begin by acknowledging that the scheme used here is not ideal — it may yield only a local maximum, and does not necessarily yield the best solution in a computationally sophisticated way. Best practices for grating optimization are not the goal of this work — the optimization is only run once, after all (computational efficiency is not of the utmost concern). What we instead seek to show here is that an awareness of the matrix picture of diffraction allows gratings with almost arbitrarily specified functions that admit straightforward design and realization.

A. Development of an optimization scheme

1. Parameterization of the grating unit cell

In this work, we work with rectangular unit cells (more specifically, square unit cells in the gratings we present). These gratings are to be realized with a 2D grid $A \times B$ individual structures, where A and B are integers. A discrete element at position (a, b) of the grating will have a local Jones matrix \tilde{J} that takes the form of Eq. S14. This local \tilde{J} can be defined by three parameters: ϕ_x (the phase imparted on light polarized along the local x axis of the element), ϕ_y (the phase imparted on light polarized along the local y axis of the element), and θ (the angular orientation of the element). Each of these could be parameterized by the continuously varying functions $\phi_x(x, y)$, $\phi_y(x, y)$, $\theta(x, y)$ over the grating unit cell for the purposes of optimization and later sampled at the center of each position (a, b) . Instead, we acknowledge the spatially discrete nature of the metasurface implementation from the start by parameterizing these three angles by a single number at each position in the $A \times B$ unit cell. This means that there will be $A \times B \times 3$ parameters in total to optimize, three for each discrete polarization element of the grating. A schematic of the unit cell and the meaning of these parameters are given in Fig. S7.

At each discrete position (a, b) of the grating unit cell, the Jones matrix $\tilde{J}_{(a,b)}$ is given by

$$\tilde{J}_{(a,b)} = R(-\theta_{(a,b)}) \begin{bmatrix} e^{i\phi_{x,(a,b)}} & 0 \\ 0 & e^{i\phi_{y,(a,b)}} \end{bmatrix} R(\theta_{(a,b)}). \quad (\text{S34})$$

Here, $\theta_{(a,b)}$, $\phi_{x,(a,b)}$, and $\phi_{y,(a,b)}$ are the grating parameters at the $(a, b)^{\text{th}}$ location in the grating unit cell, R is the 2×2 rotation matrix. The spatially-dependent, periodic Jones matrix may then be written as

$$\tilde{J}(x, y) = \left(\sum_{b=1}^B \sum_{a=1}^A \tilde{J}_{(a,b)} \text{rect}_{d_a, d_b}(x - (a - 1/2) \cdot d_a, y - (b - 1/2) \cdot d_b) \right) \circledast \text{III}_{A \cdot d_a, B \cdot d_b} \quad (\text{S35})$$

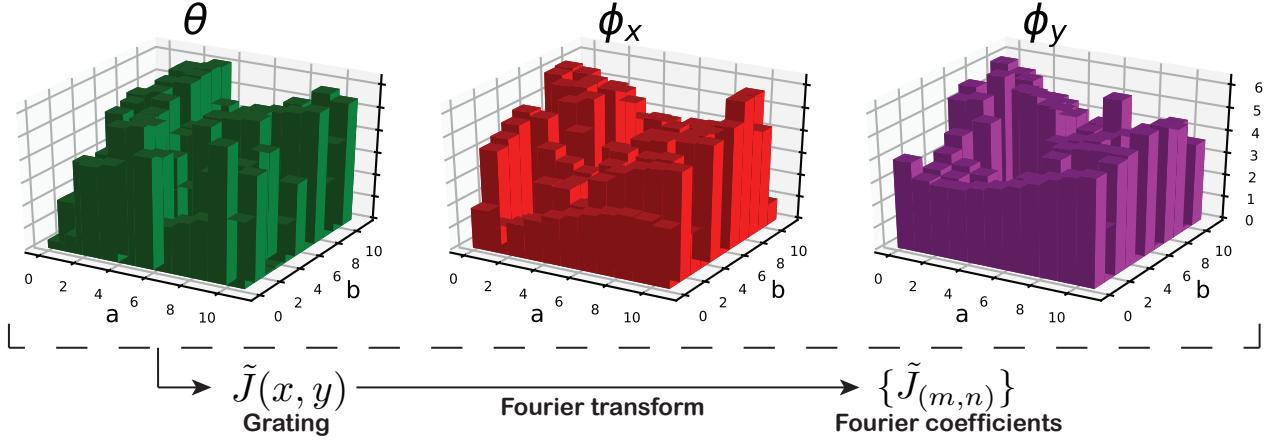


Figure S8. At each discrete location (a, b) in the $A \times B$ grating unit cell (shown here for an 11×11 square grating) we assume that there is a fixed value of θ , ϕ_x , and ϕ_y . These are the parameters to be optimized. Together these define the expression for the Jones matrix over the grating and the Jones matrices $\{\tilde{J}_{(m,n)}\}$ of the diffraction orders, the Fourier coefficients.

where d_a and d_b are the inter-element separations along the x and y directions, respectively, and $\text{rect}_{d_a, d_b}(x - (a - 1/2) \cdot d_a, y - (b - 1/2) \cdot d_b)$ is a function that is equal to 1 in a $d_a \times d_b$ rectangle about the point $((a - 1/2) \cdot d_a, (b - 1/2) \cdot d_b)$ and 0 everywhere else.⁷ The symbol $\text{III}_{A \cdot d_a, B \cdot d_b}$ denotes an infinite 2D array of Dirac delta functions separated by the dimensions of one unit cell, $A \cdot d_a$ and $B \cdot d_b$ in x and y respectively. Convolution (denoted by \circledast) of the expression in parentheses—which gives one unit cell of the grating—with $\text{III}_{A \cdot d_a, B \cdot d_b}$ is a way of expressing the infinite periodicity of the grating in 2D.

Given a grating which is fully defined by $\{\theta_{(a,b)}\}$, $\{\phi_{y,(a,b)}\}$, and $\{\phi_{x,(a,b)}\}$ (the set of these parameters at every discrete position in the unit cell), we are interested in $\tilde{J}_{(m,n)}$ where $\tilde{J}_{(m,n)}$ is the $(m, n)^{\text{th}}$ coefficient in the Fourier series of the periodic function $\tilde{J}(x, y)$. It is $\tilde{J}_{(m,n)}$ which gives the polarization-dependent behavior of diffraction order (m, n) . $\tilde{J}_{(m,n)}$ is given by the projection of $\tilde{J}(x, y)$ onto the $(m, n)^{\text{th}}$ grating order (its $(m, n)^{\text{th}}$ Fourier coefficient), found by computing the Fourier integral of Eq. S35:

$$\tilde{J}(k_x, k_y) = \frac{1}{2\pi} \iint_{-\infty}^{\infty} \tilde{J}(x, y) e^{i(k_x x + k_y y)} dx dy. \quad (\text{S36})$$

The spatial frequencies k_x and k_y can be parameterized by integer indices m and n as $k_x = 2\pi \frac{m}{Ad_a}$ and $k_y = 2\pi \frac{n}{Bd_b}$ (since $\mathcal{F}(\text{III}_{A \cdot d_a, B \cdot d_b}) = \frac{1}{Ad_a \cdot Bd_b} \text{III}_{\frac{1}{A \cdot d_a}, \frac{1}{B \cdot d_b}}$, i.e., $\tilde{J}(x, y)$ is a periodic grating), can be expanded using the convolution and shift theorems to yield

$$\tilde{J}_{(m,n)} = \frac{1}{2\pi} \frac{1}{AB} \sum_{b=1}^B \sum_{a=1}^A \tilde{J}_{(a,b)} \text{sinc} \frac{m}{A} \text{sinc} \frac{n}{B} e^{i2\pi(\frac{m(a-1/2)}{A} + \frac{n(b-1/2)}{B})} \quad (\text{S37})$$

where $\text{sinc}(x) = \sin(\pi x)/(\pi x)$.

Given sets of parameters $\{\theta_{(a,b)}\}$, $\{\phi_{x,(a,b)}\}$, and $\{\phi_{y,(a,b)}\}$ defining the linearly birefringent waveplate at each point, we can then compute the Fourier Jones matrix coefficient $\tilde{J}_{(m,n)}$ for any (m, n) . This is shown in Fig. S8.

2. Optimization

Eq. S37 permits analytical computation of the $\tilde{J}_{(m,n)}$ of the grating with knowledge of the $A \times B \times 3$ parameters that define it at each point. Next, we seek to address the inverse problem, that is to find the grating parameters that coax the $\{\tilde{J}_{(m,n)}\}$ to behave a certain way.

⁷ By using this rect box function in parameterizing the grating, we presume that light above every polarization element experiences a uniform Jones matrix and the transition between adjacent elements is instantaneous. A more realistic model could be developed taking into account the radiation patterns of the actual metasurface polarization elements, which we do not attempt here.

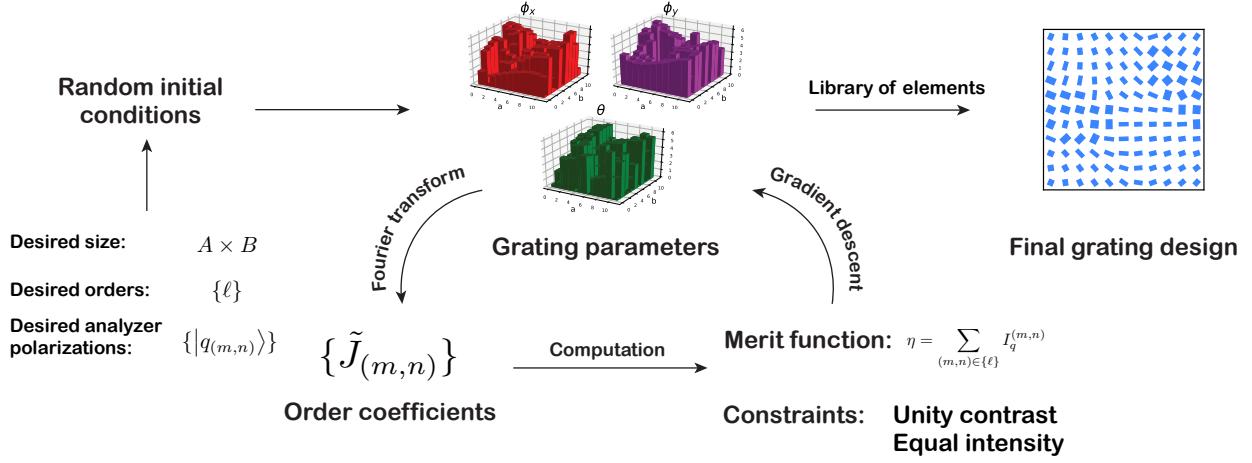


Figure S9. Overview of the optimization-based design flow for the gratings of this work described in Sec. S2 A.

First, we state the optimization problem we attempt to address here:

Problem. Given a set of grating orders $(m, n) \in \{\ell\}$ and a corresponding set of Jones vectors $\{|q_{(m,n)}\rangle\}$ for which these orders should act as polarization state analyzers, find the parameters $\{\theta_{(a,b)}\}$, $\{\phi_{x,(a,b)}\}$, and $\{\phi_{y,(a,b)}\}$ at each location (a, b) of an $A \times B$ unitary, linearly birefringent grating that maximize the energy transmission into the orders $\{\ell\}$ (minimizing leakage into all others) while mandating they behave as analyzers for the desired polarizations.

First, we define the quantity

$$I_q^{(m,n)} = \langle q_{(m,n)} | \tilde{J}_{(m,n)}^\dagger \tilde{J}_{(m,n)} | q_{(m,n)} \rangle. \quad (\text{S38})$$

$I_q^{(m,n)}$ gives the optical power (normalized to the incident power) on order (m, n) when its preferred polarization state $|q_{(m,n)}\rangle$ is incident.

We also define the polarization states $\{|q_{(m,n)}^\perp\rangle\}$ which are the polarizations orthogonal to those being analyzed, such that $\langle q_{(m,n)}^\perp | q_{(m,n)} \rangle = 0 \quad \forall (m, n)$. That allows us to define the intensity

$$I_{q,\perp}^{(m,n)} = \langle q_{(m,n)}^\perp | J_{(m,n)}^\dagger J_{(m,n)} | q_{(m,n)}^\perp \rangle. \quad (\text{S39})$$

$I_{q,\perp}^{(m,n)}$ gives the optical power (normalized to the incident power) on order (m, n) when $|q_{(m,n)}^\perp\rangle$ is incident; ideally, this would be 0.

Next, we define the *contrast* of order (m, n) as

$$\xi_{(m,n)} = \frac{I_q^{(m,n)} - I_{q,\perp}^{(m,n)}}{I_q^{(m,n)} + I_{q,\perp}^{(m,n)}}. \quad (\text{S40})$$

Ideally, this would be 1 for each order of interest so that the preferred polarization $|q_{(m,n)}\rangle$ is perfectly transmitted and the orthogonal polarization $|q_{(m,n)}^\perp\rangle$ is extinguished, in accordance with Malus' Law.

Finally, we define a merit function

$$\eta = \sum_{(m,n) \in \{\ell\}} I_q^{(m,n)}. \quad (\text{S41})$$

We perform a gradient descent optimization of the $A \times B \times 3$ parameters to maximize the merit function Eq. S41. This is done under the constraints (numerically using the method of Lagrange multipliers) that

$$\xi_{(m,n)} = 1 \quad \forall (m, n) \in \{\ell\} \quad (\text{S42})$$

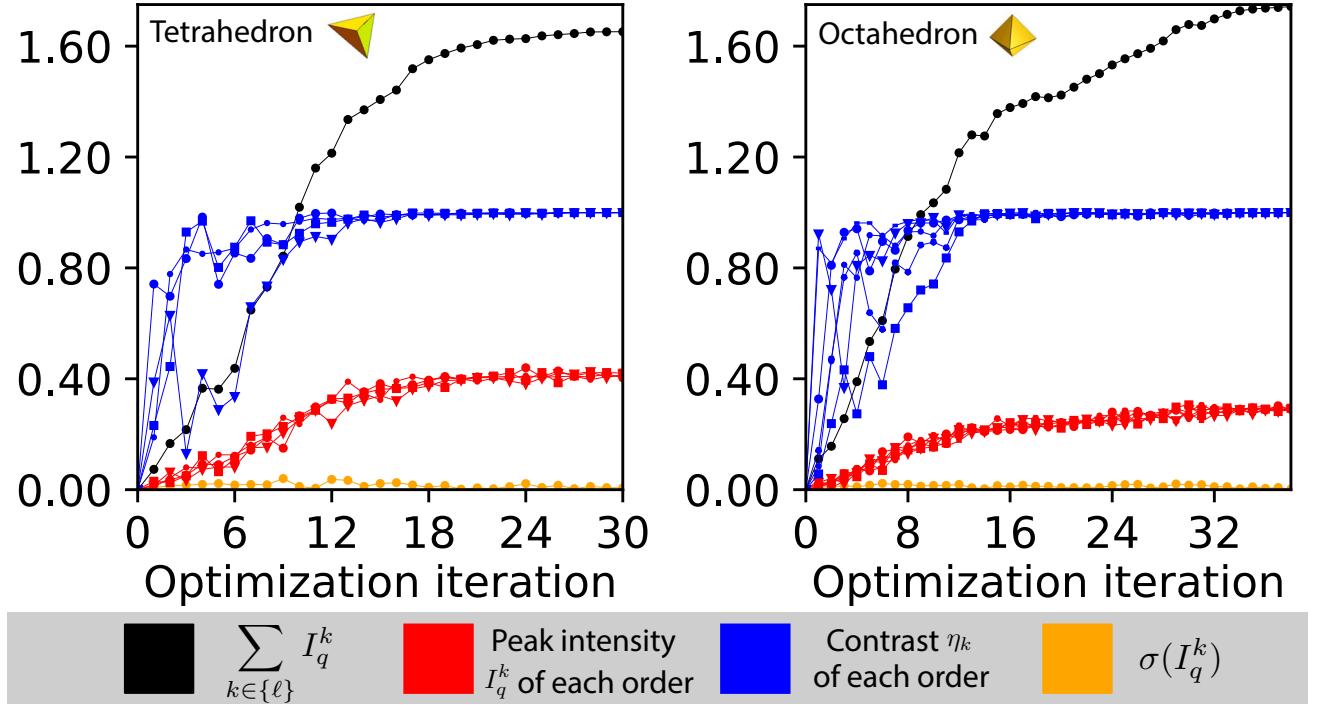


Figure S10. Progression of the optimization of the tetrahedron (left) and octahedron (right) grating unit cells. In each, the black line represents the merit function, $\sum_{k \in \{\ell\}} I_q^k$, which is the sum of the intensities observed on the set of diffraction orders $\{\ell\}$ when their designed analyzer polarization states are individually incident. Symmetry constrains the maximum achievable value of this merit function to 2 (Eq. S28). The red curves show each individual I_q^k for each diffraction order in the set with different symbols. The blue curves show the polarization contrast for each diffraction order in the set with different symbols; each of these may achieve a value of at most 1. Finally, a yellow curve shows the standard deviation of the set of $\{I_q^k\}$ for the diffraction orders of interest. Ideally, this is suppressed to 0.

and

$$\sigma(\{I_q^{(m,n)}\}) = 0 \quad \forall \quad (m, n) \in \{\ell\} \quad (\text{S43})$$

where σ denotes the standard deviation operator.

In words, optimizing the merit function Eq. S41 maximizes the total energy directed into the diffraction orders of interest and minimizes that leaked into other orders. The first constraint Eq. S42 forces the $\tilde{J}_{(m,n)}$ of the orders of interest into having the behavior of a polarization state analyzer for the desired polarization state by maximizing contrast for its desired polarization state. The second constraint Eq. S43 enforces that a situation will not occur in which one diffraction order is especially strong while others are very weak, encouraging an equal distribution of energy among the diffraction orders of interest.

Once the parameters of the unit cell have been optimized, $\tilde{J}(x, y)$ can be constructed by Eq. S35 and the polarization-behavior of the diffraction orders $\tilde{J}_{(m,n)}$ examined by Eq. S37. We note that this procedure is purely a numerical scheme—it is not specific with regard to wavelength or any real optical material. Beyond the notion of the plane wave expansion, this representation of the grating does not contain any physics.

B. Results for 11-element gratings

1. Progression of the optimization

This is performed for the tetrahedron and octahedron grating unit cells whose ideal, desired characteristics were described in Sec. S1D. The tetrahedron and octahedron gratings have four and six diffraction orders of interest, respectively, that act as polarization analyzers. Only the tetrahedron grating was presented in the main text.

Each begins with an 11×11 unit cell where θ , ϕ_x , and ϕ_y take on random initial values at each point in the array. The optimization then proceeds as described above (Fig. S9). At each iteration, the merit function and constraints

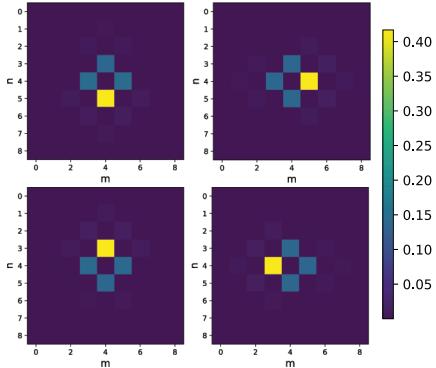


Figure S11. Calculated far-fields of the mathematically optimized 11×11 tetrahedron grating when the four polarizations of interest are incident. For each diffraction order, the fraction of incident power directed there is given. It can be seen that in each case one order is very bright with most of the remaining intensity distributed among the other three orders, as desired. These plots show the innermost 64 orders.

can be computed and graphed to visualize the optimization's progress. This is done for both gratings in Fig. S10. This continues until an arbitrary tolerance criterion is reached.

For the tetrahedron and octahedron gratings this occurs after 30 and 38 iterations, respectively. It is unsurprising that the octahedron grating, which mandates more constraints (six vs. four orders of interest) takes longer to optimize.

2. Optimized results

The optimization yields $\{\theta_{(a,b)}\}$, $\{\phi_{x,(a,b)}\}$, and $\{\phi_{z,(a,b)}\}$. These can be translated into a spatially-varying Jones matrix by Eq. S35 and the Fourier coefficient matrices $\tilde{J}_{(m,n)}$ can be found with Eq. S37. This can be done for many grating orders at once. For a particular incident polarization $|\lambda\rangle$, the power of each order can be calculated as $\langle\lambda|\tilde{J}_{(m,n)}^\dagger\tilde{J}_{(m,n)}|\lambda\rangle$. This permits plotting of the diffraction pattern predicted for the incident polarization $|\lambda\rangle$. For the optimized tetrahedron grating this has been done analytically in Fig. S11. Each of the four plots in Fig. S11 corresponds to the diffraction pattern (the inner 64 diffraction orders) observed when each of the four preferred polarizations in $\{|q_{(m,n)}\rangle\}$ are incident. It can be seen that in each case, the corresponding diffraction order is very bright with most of the remaining intensity distributed equally into the other three design orders. This equal distribution is expected for the tetrahedron because its symmetry means that the projection of one vertex onto any of the others is identical. It can be seen that over 40% of the incident light is directed into a diffraction order when its preferred polarization is incident (recall from Eq. S28 that the maximum in the tetrahedron case is 50%). Leakage into orders outside of the central four is minimized.

If we want to assign a single number quantifying the efficiency of the optimized grating, we can do so by computing the fraction of the incident power contained in the diffraction orders of interest as a function of incident polarization. This has been done for the optimized 11-element tetrahedron and octahedron gratings with the results plotted as a function of incident polarization state on the Poincaré sphere in Fig. S12. It can be seen that for all possible incident polarizations, over 82% (or 87%) of power will be directed into the orders of interest in the tetrahedron (or octahedron) case. This high prediction is a testament to the success of this very simple optimization scheme and supports the mathematical notion that if some light is allowed to leak into unwanted orders, a high efficiency can

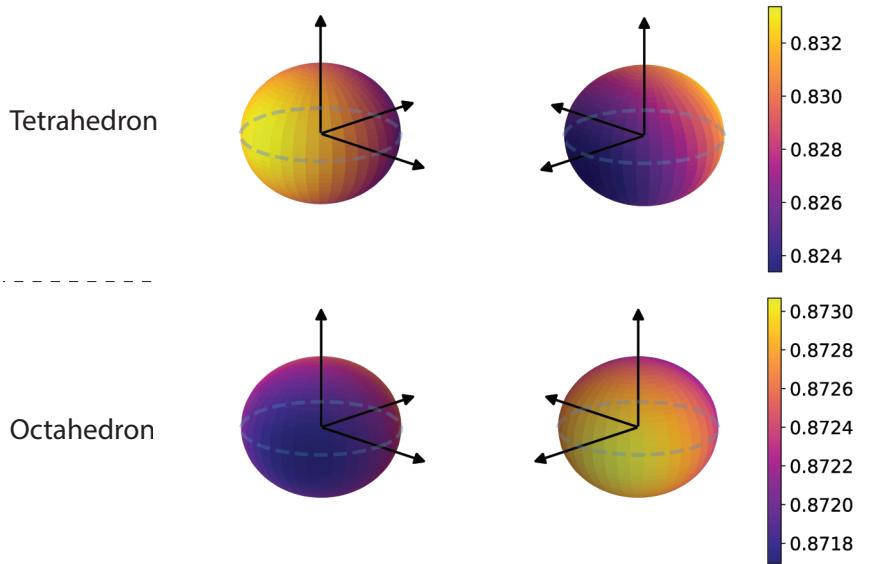


Figure S12. The mathematically optimized 11×11 tetrahedron and octahedron gratings are used to compute the fraction of power confined to the diffraction orders of interest as a function of the incident polarization state. The results are shown on the Poincaré spheres (two views are shown for each case for clarity). In both cases, well over 80% of incident power is confined to the desired orders for all possible incident polarization states.

be obtained with a strictly *unitary*, linearly birefringent grating even though these gratings cannot be the perfect, analytical solution. This parallels the picture of scalar, phase-only gratings presented in Sec. S1C.

C. Conversion to metasurface design

Once the grating is optimized by the scheme of Sec. S2A, we have the optimized sets $\{\theta_{(a,b)}\}$, $\{\phi_{x,(a,b)}\}$, and $\{\phi_{y,(a,b)}\}$ for each location in the unit cell of the diffraction grating. At each location, the geometry of a physical element implementing the desired ϕ_x and ϕ_y can be located in a material platform and wavelength of interest (TiO_2 and 532 nm in this work) and rotated by the angle θ . If many elements are simulated using full-wave numerical simulations (using, e.g., the RCWA or FDTD techniques), a library of elements can be assembled for which ϕ_x^{actual} , ϕ_y^{actual} , and overall transmission along x and y (t_x and t_y) are known for each element geometry. Given a required ϕ_x/ϕ_y pair, the quantity

$$|t_x e^{i\phi_x^{\text{actual}}} - e^{i\phi_x}|^2 + |t_y e^{i\phi_y^{\text{actual}}} - e^{i\phi_y}|^2, \quad (\text{S44})$$

which is a sort of distance in complex space between desired and actual element performances, can be computed for each element in the library. The element which minimizes this distance is chosen. This has become a commonly-used technique in the metasurface literature [8]. This can be done for each site in the grating unit cell, and the grating unit cell can be tessellated many times in the x and y directions to create the final grating.

S3. EXPERIMENTAL CHARACTERIZATION OF GRATINGS

An important part of this work is the claim that particular *behaviors*, being encapsulated in Jones matrices, can be embedded onto the orders of a diffraction grating allowing for arbitrary polarization-dependent functionality in parallel. In this work we focus on gratings whose orders act like polarizers, or polarization analyzers. In Sec. S1 we examined the mathematical restrictions of this problem and showed that optimization is necessary, and in Sec. S2 we discuss an optimization-based design scheme for these gratings. Here, we discuss how these gratings, once realized, can be tested to verify the extent to which the polarization-dependent behavior of each diffraction order matches the design.

Strictly speaking, this would involve the measurement of the Jones matrix of each diffraction order. The Jones matrix (and the Jones vector), however, are direct descriptions of the electric field which contain phase information of signals varying at optical frequencies. As such, they do not admit straightforward measurement. As a proxy, we measure the Mueller matrix [55] of each diffraction order because the Mueller formalism deals with beam powers which are directly measurable.

A. Fabrication

The gratings shown in this work are based on dielectric metasurfaces consisting of rectangular TiO_2 pillars which operate across the visible range (we work at $\lambda = 532$ nm here). Here, we describe the fabrication of these gratings. We do so extremely briefly because these gratings have been documented extensively elsewhere [38], and, more importantly, the essence of this work does not rely on any particular material, wavelength, or modality by which polarization-dependence is realized. Any polarization-dependent obstacle could be described by the matrix formalism of this work.

The gratings are first written into an electron beam resist using a 125 keV electron beam lithography system. The resist is developed, and TiO_2 is deposited using atomic layer deposition (ALD) filling in the voids in the resist from the sidewalls. In addition to filling these voids, this creates an excess layer of TiO_2 on top of the pillars. This remaining oxide is etched away using a Cl_2/BCl_3 chemistry. Finally, excess electron beam resist is removed with a solvent, yielding free-standing pillars of TiO_2 which are roughly 600 nm in height.

In Fig. S13 we provide additional zoomed-out SEM images of the octahedron and tetrahedron gratings.

B. Mueller matrix polarimetry

The process of measuring a sample's Mueller matrix is known as Mueller matrix polarimetry [24]. For each grating order of interest (four in the tetrahedron case, six in the octahedron case) we perform Mueller matrix polarimetry to

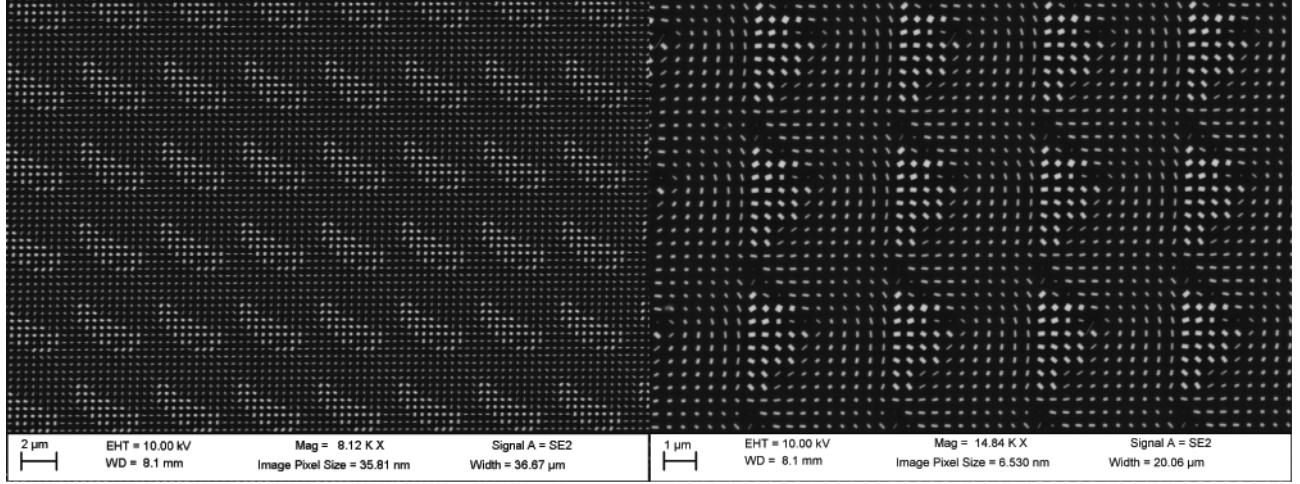


Figure S13. Additional scanning electron micrographs of the octahedron (left) and tetrahedron (right) gratings discussed in the main text.

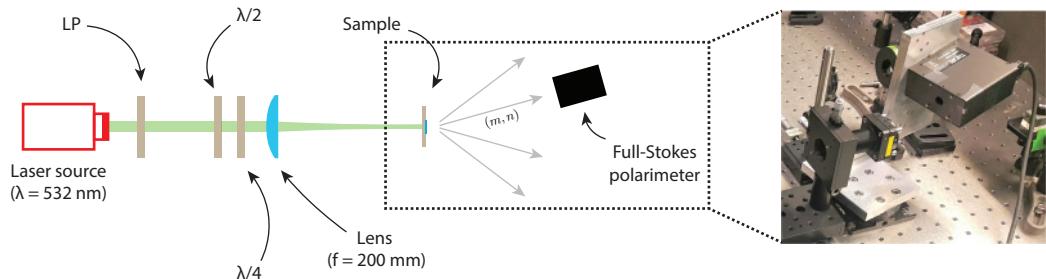


Figure S14. Mueller matrix polarimetry: optical setup for measuring the Mueller matrix $\tilde{M}_{(m,n)}$ associated with grating order (m,n) . Inset: the full-Stokes polarimeter must “orbit” the sample on a spherical mount in order to see each grating order at normal incidence.

measure the experimental Mueller matrix $\tilde{M}_{(m,n)}$ of order (m,n) .

The process involves illuminating the grating with a laser beam in a number of polarization states and observing the polarization and intensity of the outgoing beam with a full-Stokes polarimeter. If the polarization state of the incoming light is known, the transforming properties of the grating order can be deduced.

We now describe this in more detail, referring to the optical setup pictured in Fig. S14. A collimated laser beam ($\lambda = 532$ nm) is passed through a linear polarizer (LP) to set a definite starting polarization state. This LP is aligned to the preferential polarization of the laser mode to maximize output power. Next, the beam encounters a half-wave plate followed by a quarter-wave plate. The angular orientations θ_{HWP} and θ_{QWP} of the waveplates are adjustable. K pairs of quarter and half-wave plate angular orientations are selected. The waveplates visit these angular orientations and the polarization state of the beam produced is recorded by a commercial full-Stokes polarimeter (the Stokes vector of the beam produced in the k^{th} orientation is denoted by \vec{S}_k). We refer to this procedure as the *calibration* of the experiment, and the results are stored in a calibration matrix \tilde{C} :

$$\tilde{C} = \begin{bmatrix} | & | & | \\ \vec{S}_1^{\text{in}} & \vec{S}_2^{\text{in}} & \dots & \vec{S}_K^{\text{in}} \\ | & | & | \end{bmatrix}. \quad (\text{S45})$$

Here, \tilde{C} is a $4 \times K$ matrix. In our measurement, $K = 16$ and so 16 waveplate configurations are chosen so that the output beam covers the Poincaré sphere more-or-less uniformly.

Next, the beam is allowed to illuminate the metasurface grating sample. Although the gratings are designed for normally incident plane-wave illumination, the beam first passes through a lens. Since the lens’ focal length is very

long (200 mm), the incident light is still plane wave-like. However, the lens means all incident light is confined to the boundaries of the grating (which at 500 $\mu\text{m} \times 500 \mu\text{m}$ in size is quite small) so that measurements of the grating's efficiency can be made.

The full-Stokes polarimeter is placed so that it stares into grating order (m, n) . It is particularly important that the polarimeter stares straight into the diffraction order of interest to render an accurate measurement. A mount was machined (inset, Fig. S14) that allows the polarimeter to move between diffraction orders while keeping the grating centered.

Next, the waveplates again visit all K orientations and the grating is illuminated with all of the polarizations recorded in \tilde{C} . The Stokes vector output on order (m, n) in response to the k^{th} input polarization is recorded as $\tilde{S}_k^{(m,n)}$ and all of these Stokes vectors are stored in the output matrix $\tilde{O}_{(m,n)}$:

$$\tilde{O}_{(m,n)} = \begin{bmatrix} | & | & | \\ \tilde{S}_1^{(m,n)} & \tilde{S}_2^{(m,n)} & \dots & \tilde{S}_K^{(m,n)} \\ | & | & & | \end{bmatrix}. \quad (\text{S46})$$

Again, $\tilde{O}_{(m,n)}$ is a $4 \times K$ matrix. We note that it is the Mueller matrix $\tilde{M}_{(m,n)}$ of the grating order, the quantity we wish to measure, that maps the k^{th} input polarization \tilde{S}_k^{in} to the recorded polarization $\tilde{S}_k^{(m,n)}$:

$$\tilde{M}_{(m,n)} \tilde{S}_k^{\text{in}} = \tilde{S}_k^{(m,n)}. \quad (\text{S47})$$

We can express this for all grating orders at once in matrix form:

$$\tilde{M}_{(m,n)} \tilde{C} = \tilde{O}_{(m,n)}. \quad (\text{S48})$$

Since the matrix \tilde{C} is not square (unless $K = 4$), it cannot be inverted. We must use its (right) pseudo-inverse given by $\tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1}$ where T denotes a matrix transposition. Then, we can write

$$\tilde{M}_{(m,n)} = \tilde{O}_{(m,n)} \tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1}. \quad (\text{S49})$$

Eq. S49 represents the Mueller matrix of order (m, n) determined in the least-squares sense from the experimental data contained in \tilde{C} and $\tilde{O}_{(m,n)}$.

1. Comparison of optimized, simulated, and experimentally measured Mueller matrices

The properties of each diffraction order can be compared in three domains: from the optimized spatially-varying $\tilde{J}(x, y)$ (obtained as described in Sec. S2), from a full-wave simulation, and from experimental measurement (as described above).

From a Fourier transform of the analytical expression for $\tilde{J}(x, y)$ from the optimization, the Jones matrix of any desired diffraction order $\tilde{J}_{(m,n)}$

The grating unit cells in this work are simulated using finite difference time-domain (FDTD) simulation. A monitor is placed several wavelengths above the grating structure. The Jones vector of each physical diffraction order can be computed from the far-field projection of this monitor. If two-simulations are run — one for $|x\rangle$ polarized illumination and one for $|y\rangle$ polarized illumination — the Jones matrix $\tilde{J}_{(m,n)}$ can be determined. The Jones matrix is easy to directly reconstruct in the simulation case because phase information is preserved in the numerical simulation.

From experiment, however, we obtain a Mueller matrix $\tilde{M}_{(m,n)}$, not a Jones matrix. For comparison, the Jones matrices obtained from the numerical optimization and full-wave simulation can be transformed into an equivalent Mueller matrix up to an overall phase using the identity [55]

$$\tilde{M}_{j+1,k+1} = \frac{1}{2} \text{Tr}(\tilde{J} \sigma_k \tilde{J}^\dagger \sigma_j). \quad (\text{S50})$$

Here, j and k are indices that run from 0 to 3 to cover all elements of the 4×4 Mueller matrix and σ_i is the i^{th} Pauli matrix (with σ_0 being the 2×2 identity matrix).

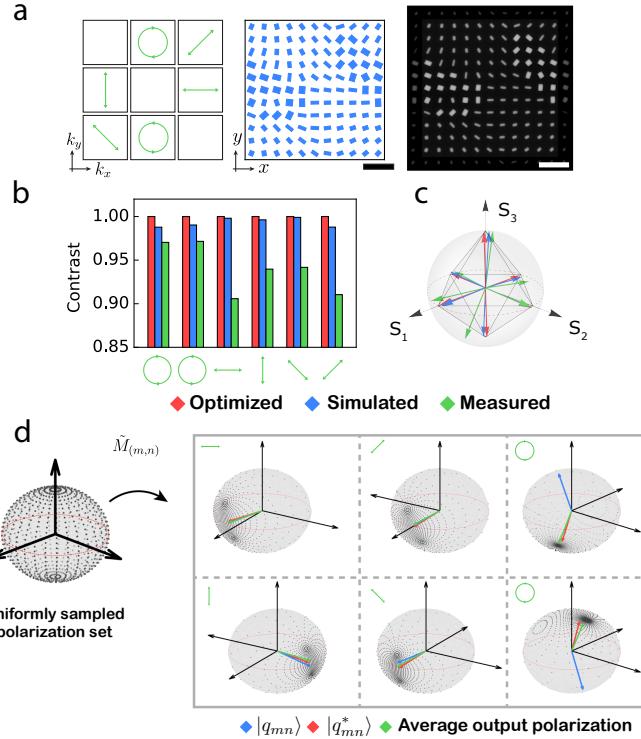


Figure S15. Results for a grating that analyzes an octahedron of polarization states on six inner diffraction orders. **a.** On the left is a map of the diffraction orders and the polarization states they analyze in k -space. These correspond to an octahedron inscribed in the Poincaré sphere and are the six common polarizations $|x\rangle$, $|y\rangle$, $|45^\circ\rangle$, $|135^\circ\rangle$, $|R\rangle$, and $|L\rangle$. The right two panels show the grating unit cell optimized to implement this behavior in design (center) and as a scanning electron micrograph (SEM) of the grating as-fabricated. For both, the scale bar is 1 μm . **b.** The polarization contrast of each order is shown. Each order is labeled by the polarization ellipse it analyzes, and results from the analytical grating design (as-optimized, red), a full-wave simulation of the grating (blue), and experiment (green) are given. **c.** The polarizations analyzed by each order (for which they have the contrasts given in (b)) of the octahedron grating are shown on the Poincaré sphere alongside a octahedron indicating the desired analyzer polarizations as predicted by the optimization, a full-wave simulation, and as-measured. **d.** Simulated output polarization as a function of input. Each sphere shows how the sampled incident Poincaré sphere distorts when operated on by the Mueller matrices of the six diffraction orders of interest (as in main text Fig. 2(e)).

C. Measurement results for octahedron grating

In main text Fig. 2 we describe only the tetrahedron grating. This was done in the interest of space since it is the tetrahedron that is used in the camera application. As a second example, we created an octahedron grating using the same procedure. It is discussed alongside the tetrahedron grating throughout this supplement. The octahedron grating analyzes six different polarization states on its diffraction orders, these being the six common polarizations $|x\rangle$, $|y\rangle$, $|45^\circ\rangle$, $|135^\circ\rangle$, $|R\rangle$, and $|L\rangle$ that correspond to the vertices of an octahedron inscribed in the Poincaré sphere.

Fig. S15 presents the same experimental results for this octahedron grating that were presented for the tetrahedron grating in main text Fig. 2. In Fig. S15(a), a map is given of the diffraction orders of interest in k -space labeled with the polarization states these orders analyze. A schematic and SEM image of the grating as-fabricated are also shown.

Mueller matrix polarimetry is performed on this grating with the results shown in Fig. S15(b) and (c). Each group of bars in (b) gives results for one of the six orders, and each group of bars is labeled by the polarization it analyzes for. In (b), the contrast of the grating orders is given and, as for the tetrahedron grating, exceeds 90% in all cases. The polarization states analyzed for are given on the sphere to the right of (c), and, as for the tetrahedron grating, these correspond closely to the vertices of the octahedron. The efficiency of the octahedron is described in Sec. S3D.

In Fig. S15(c), the polarization transfer characteristic of each diffraction order is shown by operating on a uniformly sampled set of 1024 possible input polarizations. Points are strongly attracted in a particular direction showing that the orders do indeed act as strong polarizers. For all six orders, the average output polarization state overlaps closely with the analyzer state mirrored about the equator of the Poincaré sphere, as for the tetrahedron case in main text Fig. 2(e), supporting the result of Sec. S1E3 mandating that $|p_k\rangle = |q_k^*\rangle$.

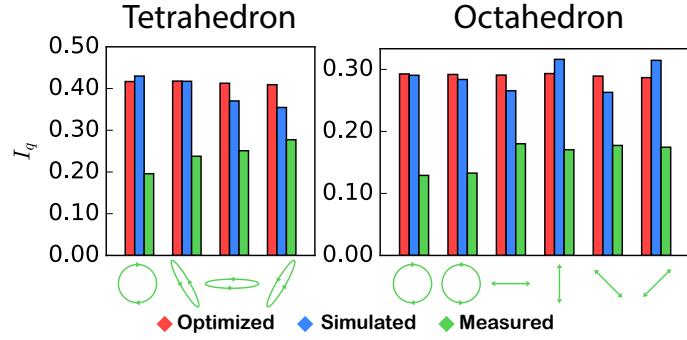


Figure S16. For all four (or six) orders of interest of the tetrahedron (or octahedron) grating, we plot I_q , the power (normalized to that of the incident beam) observed on the diffraction order when its preferred polarization is incident. Each group of bars represents one order and results are given for the numerical prediction from the optimized $\tilde{J}(x, y)$ (red), a full-wave simulation (blue), and the experimental measurement (green). Each group of bars is labeled by the polarization ellipse that the order analyzes.

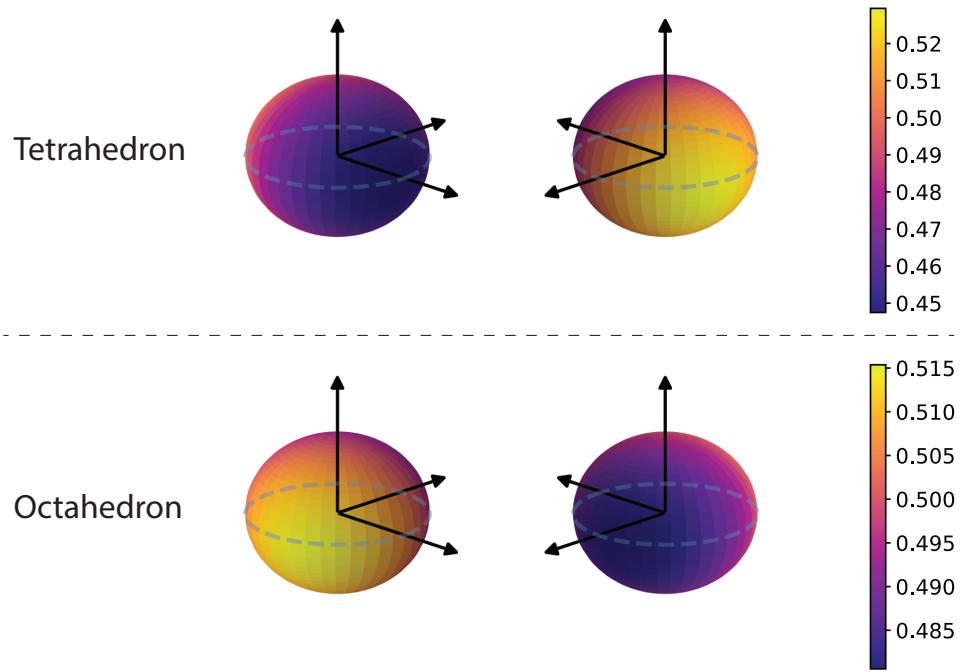


Figure S17. Percentage of incident power directed into the four or six diffraction orders of interest, as a function of incident polarization on the Poincaré sphere for the gratings experimentally measured in this work. For both gratings this efficiency is above 45% for all possible incident polarizations and averages around 50%. Two views are shown of each sphere for clarity.

D. Overall grating efficiency

The question of diffraction efficiency is one that naturally arises in this work. The gratings here are polarization-dependent, so “efficiency” cannot be captured with one single number because the amount of incident light concentrated in the grating orders of interest is polarization-dependent.

We examine efficiency by looking at two different quantities, represented in Figs. S16 and S17. First, it is interesting to examine the intensity observed (as a fraction of the intensity of the incident light) on each diffraction order of interest when the order’s preferred polarization is incident. This is the quantity we defined as $I_q^{(m,n)}$ in Eq. S38. We note that since there are many orders acting as polarization state analyzers in parallel, $I_q^{(m,n)}$ cannot be 1 since intensity must be shared in accordance with Malus’ Law. In fact, for the tetrahedron case (four analyzers) and octahedron case (six analyzers), each $I_q^{(m,n)}$ can take a maximum value of 1/2 and 1/3, respectively—this is governed by Eq. S28.

$I_q^{(m,n)}$ can be computed from the first row of the Mueller matrix. This has been done for the tetrahedron and

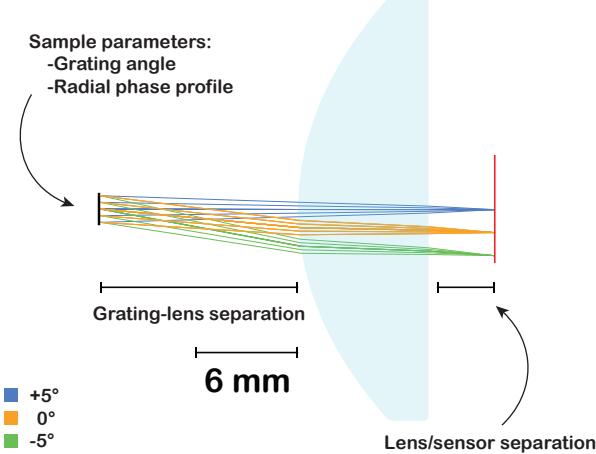


Figure S18. A side-view schematic of the camera considered here with all adjustable parameters labeled. Different colors here correspond to parallel ray bundles incident at different angles on the grating sample which are labeled in the key at bottom left. These are deflected from their original orientations before the sample by the grating (so that, e.g., $+5^\circ$ becomes nearly horizontal). The grating vector here is oriented at 45° so that one grating order is directed towards the corner of the imaging sensor. Not shown: 10 nm bandpass filter at $\lambda = 532$ nm in front of the imaging system.

Camera Parameter	Optimized Value
Grating angle	7.28°
Radial phase profile	$(-231 \frac{\text{rad.}}{\text{mm.}^2})\rho^2$
Grating-lens separation	11.5 mm
Lens-sensor separation	3.9 mm

Table S2. Optimized values of the variable parameters of the camera shown in Fig. S18. Here, ρ is the radial coordinate measured from the center of the lens in mm.

octahedron gratings in Fig. S16 for the grating as optimized (from $\tilde{J}(x, y)$), from simulation, and as measured using Mueller matrix polarimetry. We see that the optimization and simulation predict something within 5-10% of perfection (close to 40% for the tetrahedron and 30% for the octahedron grating). In experimental reality, between one-half and two-thirds of this (depending on the order) is observed. (As in Fig. 2(c) of the main text, it is not erroneous that in some cases simulation surpasses the numerical optimization in Fig. S16 — the grating orders are coupled, so defects detracting from the performance of one order can benefit another.)

A second way of quantifying the efficiency of the grating is shown in Fig. S17. The power in order (m, n) when a given polarization is incident can be determined by operating on the Stokes vector \vec{S}_{inc} with unit power and taking the first element of the resulting output Stokes vector. These can be summed for all orders of interest to get a measure of efficiency. This has been done in Fig. S17 as a function of all possible incident polarizations on the Poincaré sphere. We see that for both gratings efficiency is approximately 50% with about 1% variation as the incident polarization state varies.

An overall efficiency of 50%, a 3 dB loss, is sufficient to enable practical applications (as demonstrated by this work). However, it is clear that there is a discrepancy between the observed efficiency and what would be expected. In Fig. S16, the green bars are lower than the red and blue ones, and the 50% efficiency in Fig. S17 is lower than predicted numerically in Fig. S12. We expect that most of this discrepancy stems from fabrication imperfection. Small discrepancies in the fabrication propagate through the Fourier transform in a complicated way that couples all grating orders together.

Moreover, the mathematical formulation behind the matrix approach of this work assumes an amplitude-transmittance model. Light is imagined to impinge on the grating, which is imagined to be very thin, and pick up a particular mask of a structured aperture (which in this case is matrix valued). Backward propagating light is not considered by this design. This neglects the full physical complexity of the grating system and is only an approximation of the full solution which is described by the rigorous coupled wave analysis (RCWA). Analytical application of RCWA to these matrix gratings will be a subject of future work.

S4. POLARIZATION IMAGING

A. Optical design of the camera system

This work has so far been concerned with the design of gratings that split light based on its polarization. If the intensity of these diffraction orders is measured, the grating alone is sufficient to determine the polarization of the incident light as a full-Stokes polarimeter. This would permit the determination of the polarization state of a laser beam. For a variety of practical applications [41, 42]—especially in remote sensing and machine vision—an imaging polarimeter, that is, a polarization camera, that permits characterization of the Stokes vector of the light forming every pixel of a photographic image is of far more interest.

Conceptually, the idea is simple. The gratings presented here can split light based on its polarization state, so if the grating were incorporated into an imaging system, could it form four images of a scene on its diffraction orders that could later be combined pixel-wise in order to find a Stokes-vector polarization image of a scene in post-processing? The question then becomes how to incorporate the matrix gratings of this work into an imaging system to do this.

There is no one unique approach here. The design of an imaging system around the grating could certainly become as complicated as camera objective lens design itself if a fully rigorous, performance-oriented, and aberration-correcting engineering challenge is undertaken [57]. That is not the goal of the present work. We strive to show here that a *simple* full-Stokes polarization imaging system can be realized where all necessary polarization optics is handled by a single diffraction grating and to demonstrate that this system can be used as a compact sensor in practical applications.

To that end, we settle on a basic imaging architecture that includes just three elements—those being a grating sample, an off-the-shelf catalog lens, and a CMOS imaging sensor, in that order. The camera we seek to design in this work is *infinite conjugate*, i.e., it does not object an image at a given distance, but rather images objects that are very far away (measured relative to the focal length of any optics in the imaging system) so that ray bundles emanating from any point on the object can be thought of as parallel. Here, we set the goal that the camera should image over a 10° ($\pm 5^\circ$) field-of-view (FOV), a modest goal that alleviates some constraints while still allowing for the imaging of real objects. For this imaging system, the FOV is the range of incident angles of ray bundles that can be imaged onto the sensor.

The tetrahedron grating from the main text is used in the camera. Its four diffraction orders implement analyzers that are sufficient for full-Stokes determination. Since the Stokes vector has four elements, four images are the minimum necessary for full-Stokes measurement. When mounted in the camera, the grating is turned 45° so that the orders deflect toward the corners of the imaging sensor, rather than toward the top, bottom, and sides.

In the optical design procedure, only one grating order is considered. If all aspects of the imaging system are symmetric about the optical axis, the imaging system will work identically for the three other orders by default in a way that is merely rotated by 90° , 180° , or 270° . The goal of the design is to take parallel ray bundles from the entire FOV ($\pm 5^\circ$) and focus them as well as possible within the bounds of one quadrant of the imaging sensor—the final result, the camera presented in the main text, is depicted in Fig. S18. As is shown there, we have four degrees of freedom to exploit in accomplishing this task (once an off-the-shelf lens is chosen). These are the distance between the grating sample and the lens, the distance between the back plane of the lens and the imaging sensor, the grating angle (and thus period) of the sample, and an overall, azimuthally symmetric phase profile (that is not polarization-dependent) imparted on top of the grating. This overall phase profile produces a weak lensing effect that can aid in imaging.

A ray trace of the system can be performed and these parameters can be varied by an optimization algorithm so that the -5° , 0° , and $+5^\circ$ ray bundles (or fields, in the terminology of ray tracing) are focused as well as possible at the plane of the imaging sensor under the constraint that they occupy the bounds of one quadrant. This “goodness” of focusing can be defined in terms of how many rays in the bundle fall within the Airy disk.

This optimization can be performed for many off-the-shelf lenses. The lens ultimately chosen (and depicted in Fig. S18) was ThorLabs part #AL2520. This is an aspheric lens with a focal length of $f = 20$ mm (as of this writing listed here) made of the high-index optical glass S-LAH64. This lens was found to produce a high quality of focus throughout the FOV—it is perhaps unsurprising that an aspheric lens was found to be advantageous here. The complete optimized parameters of the system designed around the AL2520 asphere are listed in Table S2. It is curious that, though the focal length of the asphere is 20 mm, it is only placed 3.9 mm from the imaging sensor in the final camera for image formation. This makes some intuitive sense: The imaging is not occurring in the center of the lens, but rather off-center where the gradient is steeper, decreasing the apparent focal length. We also note that, by inspection of Fig. S18, the camera is roughly telecentric in image space.

Not shown in Fig. S18 is a 10 nm bandpass filter at $\lambda = 532$ nm in front of the camera. Everything in this work, including this imaging, is monochrome. The grating is inherently dispersive and the inclusion of color complicates this work. We note, however, that if an imaging sensor with a color filter is used, the approach presented here could be used for spectropolarimetry, or color polarization imaging (the different colors would form different images due to

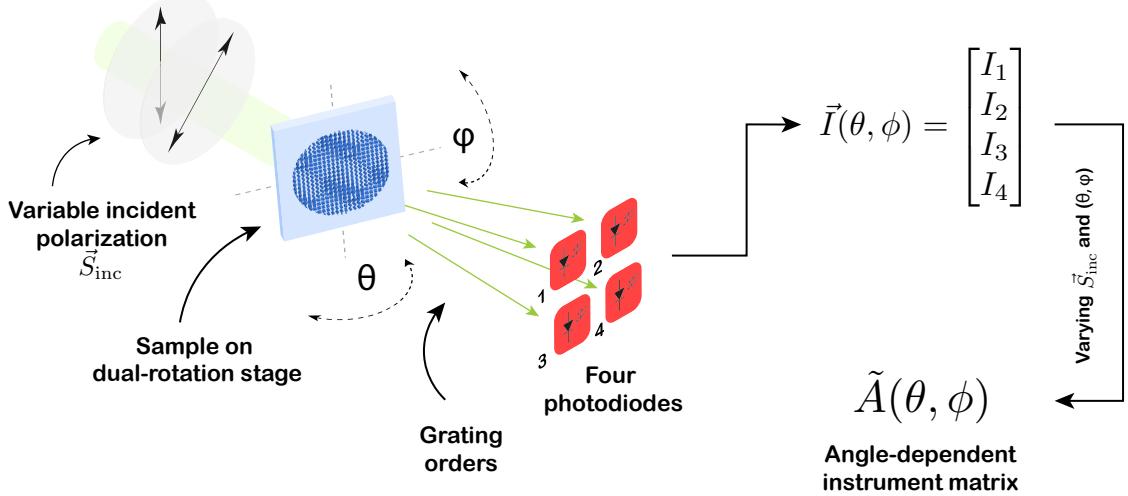


Figure S19. Angle-dependent calibration: The grating sample used in imaging is placed on a dual-rotation stage in front of a laser beam such that its orientation relative to the beam (θ, ϕ) can be adjusted. Light from the four diffraction orders to be used in imaging impinges on four photodiodes which report the intensity of the orders. The polarization of the laser beam is adjusted using waveplates and these photodiodes monitored, allowing for determination of the polarimetric instrument matrix \tilde{A} . This process is repeated at many different angular configurations (θ, ϕ) to determine the angle-dependent instrument matrix $\tilde{A}(\theta, \phi)$.

the dispersive nature of the grating). This is not attempted in the present work.

A second element not depicted in Fig. S18 is an aperture whose size can be adjusted which is placed in front of the grating at some distance. This is used to control the FOV of the camera so that images do not overlap on the sensor and explains why the polarization images presented here are circular. We note that this aperture is used differently than apertures in traditional photographic imaging systems, since it is not in a Fourier plane and does not control the f number of the system. The aperture here possibly causes some mechanical vignetting which we have not addressed in this work.

A 12-bit CMOS sensor is used here in order to increase the dynamic range of the system which is particularly helpful for outdoor imaging.

More sophisticated and nuanced approaches to designing polarization imaging systems based on these gratings will be the subject of future work. The simplicity of the system enabled by this grating and the practical results it yields (with no traditional polarization optics) are the main focus here.

B. Angle-dependent calibration

The output of the camera is four copies of an image, each analyzed along different polarizations. What are these polarization states? These must be known to produce the pixel-wise polarization image $\vec{S}(x, y)$. Each pixel of the polarization image is actually a super-pixel of four pixels (how these are grouped and identified is known as image registration, and will be addressed in the next section). The intensity at one pixel can be written as a vector $\vec{I} = [I_1 \ I_2 \ I_3 \ I_4]^T$ where I_1, I_2, I_3 , and I_4 are the pixel intensities from the individual sub-images. These intensities are the result of a projection of the incident light's polarization, \vec{S}_{inc} , onto the characteristic polarization of the quadrant k , \vec{S}_k . These characteristic polarizations can be grouped into the rows of a matrix \tilde{A} so we can write

$$\vec{I} = \begin{bmatrix} - & \vec{S}_1 & - \\ - & \vec{S}_2 & - \\ - & \vec{S}_3 & - \\ - & \vec{S}_4 & - \end{bmatrix} \vec{S}_{\text{inc}} = \tilde{A} \vec{S}_{\text{inc}}. \quad (\text{S51})$$

In polarimetry, \tilde{A} is known as the instrument matrix. If it is known, we can straightforwardly write that $\vec{S} = \tilde{A}^{-1} \vec{I}$. The process of determining this \tilde{A} is known as polarimetric calibration.

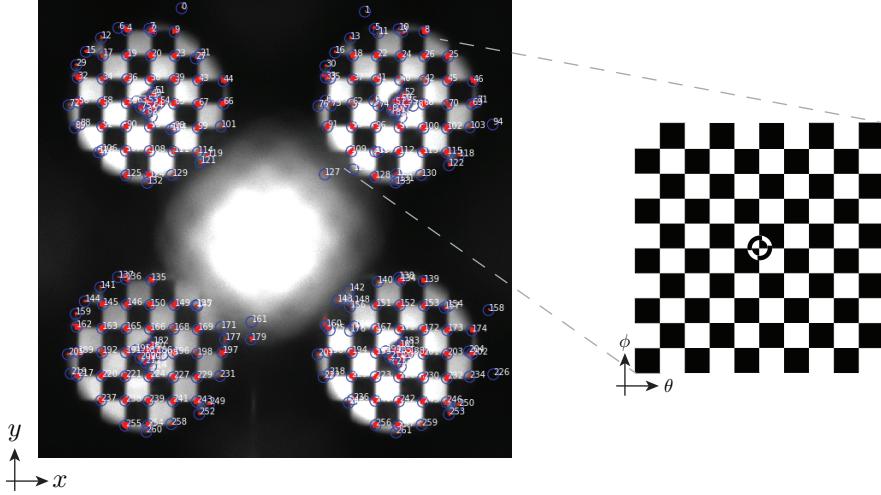


Figure S20. Checkerboard image registration: A checkerboard pattern in which each division is 1° of angle in each direction is used as a calibration target for the camera. Corners are identified in the image and labeled, used to yield optimized functions mapping between pixel location (x, y) and angular coordinates (θ, ϕ) of light incident on the pixel.

Calibration of the polarization camera presented in this work means determining the characteristic polarization of each quadrant. The camera presented here is infinite conjugate (i.e., it images objects that are imagined to be infinitely far away) and each pixel of the image corresponds to light incident from a given angular direction. Being a diffractive object made of nanostructures, it is possible to imagine that the characteristic polarization is a function of incident angle and so the instrument matrix itself must be a function of angle $\tilde{A}(\theta, \phi)$. We treat this empirically by utilizing a calibration procedure that can absorb angle-dependent effects.

This angle-dependent calibration is depicted in Fig. S19 and is conceptually very similar to the grating measurements discussed in Sec. S3B and the main text. A collimated laser ($\lambda = 532$ nm) beam whose polarization state \vec{S}_{inc} can be adjusted with waveplates is incident on the grating. The grating is mounted on two rotation stages that are arranged so that the sample is located at the intersection point of the axes of these rotation stages. When the angles of these stages (θ, ϕ) are rotated it is as if light is incident on the grating from a different angular direction, as in the photography case. The grating is placed in its mount in the same orientation as it is placed in the camera system.

At a given orientation of the stages (θ, ϕ) the polarization of the laser beam can be adjusted by waveplates to visit a number of fixed configurations where the \vec{S}_{inc} is known (with a full-Stokes polarimeter, as in Sec. S3B). The response of the imaging grating orders to the incident polarization is measured with four silicon photodiodes. If this is done for many incident polarizations, the instrument matrix at the configuration (θ, ϕ) can be found in a least squares sense by inversion (again as in Sec. S3B).

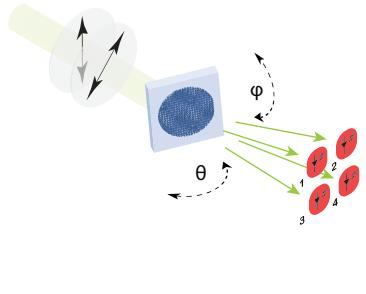
This process is repeated at many angular positions (θ, ϕ) and the results are interpolated to yield the angle-dependent instrument matrix $\tilde{A}(\theta, \phi)$. To be used in polarization image reconstruction pixel-by-pixel, the (θ, ϕ) corresponding to each pixel must be known. This is described in the next section.

C. “Checkerboard” image registration

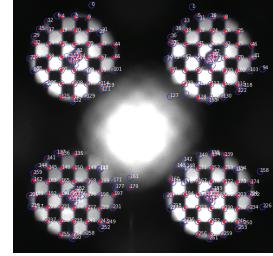
Since the camera is infinite-conjugate, each pixel (in each of the four quadrants) measures parallel ray bundles emanating from a specific angular direction (θ, ϕ). But, given a pixel coordinate (x, y) in the polarization image, how can we know its corresponding (θ, ϕ)? For that matter, since we have four copies of the image on the sensor, how do we align these to each other so that we know which collections of four pixels on the sensor belong together? Once these questions are answered, we can form the polarization image as $\vec{S}(x, y) = \tilde{A}^{-1}(\theta, \phi)\vec{I}(x, y)$ — knowing the angular coordinates at each point permits use of the correct inverse angle-dependent instrument matrix $\tilde{A}^{-1}(\theta, \phi)$ and correct image registration allows for the collection of the vector $I(x, y)$ at each point.

We address both of these questions with one procedure which is depicted in Fig. S20. A checkerboard pattern is prepared in which each division corresponds to 1° of angle in the θ or ϕ directions when placed at a 14 inch viewing

Angle-dependent calibration



Checkerboard image registration



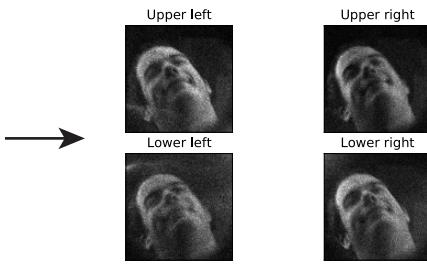
$$\rightarrow \tilde{A}(\theta, \phi) \quad \begin{matrix} \phi(x, y) \\ \theta(x, y) \end{matrix} \leftarrow$$

$$\downarrow \quad \quad \quad \downarrow$$

$$\tilde{A}(x, y)$$



Raw acquisition



Segmented image $\vec{I}(x, y)$

$$\longrightarrow \vec{S}(x, y) = \tilde{A}^{-1}(x, y) \vec{I}(x, y)$$

Polarization image

Figure S21. Synthesis of a full-Stokes polarization image. Angle-dependent calibration and checkerboard image registration yield the instrument matrix $\tilde{A}(x, y)$ as a function of pixel position (x, y) (top). When a raw acquisition is taken, it can be segmented into four subimages and the Stokes vector $\vec{S}(x, y)$ computed as a function of position to form a final polarization image.

distance from the sample, and the center is well marked. This checkerboard is shown in the right inset of Fig. S20.⁸ The checkerboard is printed, mounted 14 inches from the camera, and illuminated with a green LED flashlight. A raw image of the checkerboard is acquired which is shown on the left of Fig. S20.⁹

This image undergoes adaptive thresholding to increase the contrast of the checkerboard. Next, the Harris corner detection algorithm is applied to the image and the coordinates of corner locations are noted and indexed (shown with red dots in Fig. S20). These locations are labeled with angular coordinates (θ, ϕ) manually and a curve fitting algorithm is used to generate four functions — one for each quadrant of the imaging sensor — that map from pixel location (x, y) to angular coordinate (θ, ϕ) .

These functions also contain the center position (x_0, y_0) of each quadrant, information that can be used to superimpose the images on top of one another as a simple translation. In this way the images are very simply registered to one another. Future work will address distortions that could occur between different quadrants.

D. Polarization image synthesis

The angle-dependent calibration yields the function $\tilde{A}(x, y)$, the instrument matrix as a function of the angle of the incident ray bundle. The checkerboard image registration yields the functions $\theta(x, y)$ and $\phi(x, y)$ that show what angle in the field-of-view each pixel coordinate (x, y) corresponds to. Together, this yields the instrument matrix as a function of pixel coordinate $\tilde{A}(x, y)$. This is shown at the top of Fig. S21.

When a raw acquisition is taken, the four quadrants can be registered to one another (the centers of these images are also known from the checkerboard image registration). This segmented image yields the four-element intensity

⁸ The checkerboard here only appears as square and regular because it covers a narrow field-of-view ($\pm 5^\circ$). Over a larger field-of-view it would appear considerably more distorted.

⁹ The checkerboard in Fig. S20 appears to be out of focus. The checkerboard acquisition shown here was taken after outdoor photography — the configuration of the optics in the camera that renders faraway objects and the checkerboard (which is 14 inches away) in focus differs. However, the checkerboard acquisition must reflect the configuration of the camera when the images of interest were taken to assure proper image registration. This defocus does not present a problem for registration.

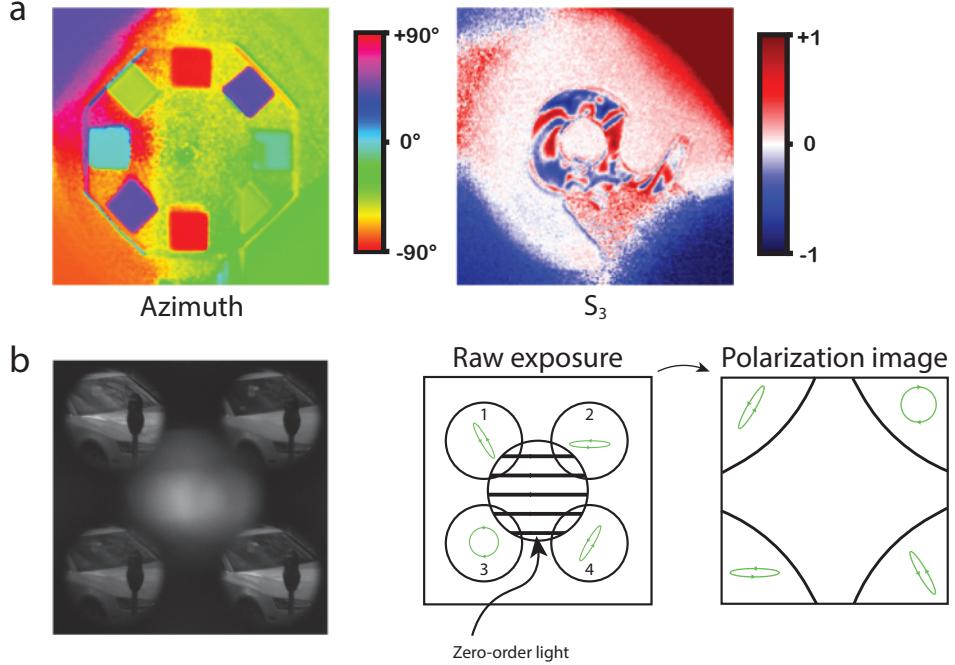


Figure S22. **a.** Examples of edge artifacts in the corners of two example images from the main text. Areas of the image that are dark have definite polarization information that is not just noise. **b.** This is explained by the fact that the undiffracted zero order affects all four corners of the image through the four different polarization channels. As a result, the Stokes vector predicted at the corners takes on each of the four analyzer vectors, explaining the artifacts in (a).

vector $\vec{I}(x, y)$ where each element is one pixel drawn from the same place on all four quadrants. Finally, using the results of the calibration, the polarization image is calculated as

$$\vec{S}(x, y) = \tilde{A}^{-1}(x, y)\vec{I}(x, y). \quad (\text{S52})$$

This is shown in the bottom of Fig. S21.

E. Edge effects and the zero order

Some of the polarization images displayed in the main text (in Figs. 4 and 5) show a definite pattern to the background information contained in dark parts of the image. What one would assume would be random noise—the Stokes vector computed from dark pixels—seemingly has some well-behaved structure. This is evident in Fig. S22(a) in which two examples have been taken from the main text. As a first example, the azimuth image of the polarizer wheel displays the correct polarizer orientations, but the azimuth value of the background seems to have a definite value that varies by region. As a second example, the S_3 image of a tape dispenser drawn from main text Fig. 5 seems to have definite values in the corners.

This has a simple explanation: Interference from the undiffracted, zero-order light. Some light does not interact with the grating and passes straight through, experiencing neither the grating nor the weak lens phase profile. This forms an unfocused background (it is not correctly imaged by the aspheric lens, and does not experience the metasurface's lens phase profile). This can be seen in the example raw exposure given on the left of Fig. S22. The zero order bleeds into the four images, but does so in a different corner in each case. Each order has a characteristic polarization. The result is the corners of the final processed polarization image $\vec{S}(x, y)$ simply take on the Stokes vectors of the polarization analyzed by the grating. In each corner region, only one quadrant contributes any signal, so the result of the inversion Eq. S52 there is the characteristic Stokes vector of the quadrant. This is shown in the two drawings on the right of Fig. S22(b).

This explains the artifacts in Fig. S22(a). In the azimuth image, we would expect the bottom-right and top-left

corners to have $\pm 45^\circ$ azimuth and the bottom left corner to have 0° azimuth. This is observed.¹⁰ In the S_3 image of the tape dispenser, it is seen that the upper right corner predicts circular polarization, and the other three corners have an S_3 of the opposite helicity.

This is a major area for improvement in future prototypes and is the most significant systematic error present in the camera. The inclusion of a Fourier plane—or even a physical block on the back of the lens—would enable filtering out of the zero order, eliminating this problem. Moreover, the imaging system could be optimized to keep the zero order away from the imaging light.

S5. CAPTION FOR SUPPLEMENTAL MOVIE S1

Outdoor, real-time imaging of cars on Oxford Street in Cambridge, Massachusetts. Custom software for real-time polarization imagery was developed. Four panels depict polarization information. Clockwise from top left: Raw exposure with four sub-images, S_0 , azimuth angle, and degree of polarization (DOP). Automobiles (especially their windshields) can be observed to display distinct signatures.

¹⁰ The upper right quadrant, which is dominated by the circular polarization channel, has an azimuth angle that stems from the imperfections of the analyzer - the azimuth angle of circularly polarized light is undefined, but experimental imperfections give the “circular” polarization a definite azimuth value, kicking it off the pole of the Poincaré sphere.

-
- [1] G. Cincotti, *IEEE Journal of Quantum Electronics* **39**, 1645 (2003).
- [2] E. Hasman, G. Biener, A. Niv, V. Kleiner, *Progress in Optics* **47**, 215 (2005).
- [3] A. W. Lohmann, *Applied Optics* **4**, 1667 (1965).
- [4] T. Todorov, L. Nikolova, N. Tomova, *Applied Optics* **23**, 4309 (1984).
- [5] T. Todorov, L. Nikolova, K. Stoyanova, N. Tomova, *Applied Optics* **24**, 785 (1985).
- [6] L. Nikolova, P. Ramanujam, *Polarization Holography* (Cambridge University, 2009).
- [7] J. P. B. Mueller, N. A. Rubin, R. C. Devlin, B. Groever, F. Capasso, *Physical Review Letters* **118**, 113901 (2017).
- [8] A. Arbabi, Y. Horie, M. Bagheri, A. Faraon, *Nature Nanotechnology* **10**, 937 (2015).
- [9] C. Oh, M. J. Escuti, *Optics Letters* **33**, 2287 (2008).
- [10] J. Kim, et al., *Optica* **2**, 958 (2015).
- [11] J. W. Goodman, *Introduction to Fourier Optics* (Roberts & Company Publishers, 2005).
- [12] R. C. Jones, *Journal of the Optical Society of America* **31**, 488 (1941).
- [13] I. Moreno, M. J. Yzuel, J. Campos, A. Vargas, *Journal of Modern Optics* **51**, 2031 (2004).
- [14] I. Moreno, C. Iemmi, J. Campos, M. J. Yzuel, *Optics Express* **19**, 4583 (2011).
- [15] W. Urbańczyk, *Optica Acta* **33**, 53 (1986).
- [16] J. P. McGuire, R. A. Chipman, *Journal of the Optical Society of America A* **7**, 1614 (1990).
- [17] See supplementary materials.
- [18] Z. Bomzon, V. Kleiner, E. Hasman, *Optics Letters* **26**, 1424 (2001).
- [19] M. Berry, *Proceedings of the Royal Society* **392**, 45 (1984).
- [20] E. Hasman, V. Kleiner, G. Biener, A. Niv, *Applied Physics Letters* **82**, 328 (2003).
- [21] D. Lin, P. Fan, E. Hasman, M. L. Brongersma, *Science* **345**, 298 (2014).
- [22] N. A. Rubin, et al., *Optics Express* **26**, 21455 (2018).
- [23] R. Zhao, et al., *Light: Science & Applications* **7**, 95 (2018).
- [24] R. M. A. Azzam, *Journal of the Optical Society of America A* **33**, 1396 (2016).
- [25] N. Yu, et al., *Science* **334**, 333 (2011).
- [26] P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, R. Devlin, *Optica* **4**, 139 (2017).
- [27] L. A. Romero, F. M. Dickey, *Journal of the Optical Society of America A* **24**, 2280 (2007).
- [28] L. A. Romero, F. M. Dickey, *Progress in Optics* **54**, 319 (2010).
- [29] A. Pors, M. G. Nielsen, S. I. Bolzhevolyi, *Optica* **2**, 716 (2015).
- [30] W. T. Chen, et al., *Nanotechnology* **27**, 224002 (2016).
- [31] S. Wei, Z. Yang, M. Zhao, *Optics Letters* **42**, 1580 (2017).
- [32] E. Arbabi, S. M. Kamali, A. Arbabi, A. Faraon, *ACS Photonics* **5**, 3132 (2018).
- [33] T. Todorov, L. Nikolova, *Optics Letters* **17**, 358 (1992).
- [34] A. Martínez, *Science* **362**, 750 (2018).
- [35] R. Azzam, I. Elminyawi, A. El-Saba, *Journal of the Optical Society of America A* **5**, 681 (1988).
- [36] D. S. Sabatke, et al., *Optics Letters* **25**, 802 (2000).
- [37] J. S. Tyo, *Applied Optics* **41**, 619 (2002).
- [38] R. C. Devlin, M. Khorasaninejad, W.-T. Chen, J. Oh, F. Capasso, *Proceedings of the National Academy of Sciences* **113**, 10473 (2016).
- [39] A. Espinosa-Soria, F. J. Rodríguez-Fortuño, A. Griol, A. Martínez, *Nano Letters* **17**, 3139 (2017).
- [40] W. Wu, Y. Yu, W. Liu, X. Zhang, *Nanophotonics* **8**, 467 (2019).
- [41] J. S. Tyo, D. L. Goldstein, D. B. Chenault, J. A. Shaw, *Applied Optics* **45**, 5453 (2006).
- [42] F. Snik, et al., *Proceedings of the SPIE* **9099** (2014).
- [43] S. G. Demos, R. R. Alfano, *Applied Optics* **36**, 150 (1997).
- [44] J. S. Tyo, M. P. Rowe, E. N. Pugh, Jr., N. Engheta, *Applied Optics* **35**, 1855 (1996).
- [45] Y. Cui, R. M. A. Azzam, *Optics Letters* **21**, 89 (1996).
- [46] M. W. Kudenov, M. J. Escuti, E. L. Dereniak, K. Oka, *Applied Optics* **50**, 2283 (2011).
- [47] L. B. Wolff, T. E. Boult, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 635 (1991).
- [48] S. Rahmann, N. Canterakis, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2001), pp. 149–155.
- [49] A. Kadambi, V. Taamazyan, B. Shi, R. Raskar, *Proc. IEEE International Conference on Computer Vision* (2015), pp. 3370–3378.
- [50] N. M. Garcia, I. de Erausquin, C. Edmiston, V. Gruev, *Optics Express* **23**, 14391 (2015).
- [51] A. Cerjan, S. Fan, *Physical Review Letters* **118**, 253902 (2017).
- [52] C. Menzel, C. Rockstuhl, F. Lederer, *Physical Review A* **82**, 053811 (2010).
- [53] F. Gori, et al., *Optics Communications* **157**, 13 (1998).
- [54] F. Gori, *Diffractive Optics and Optical Microsystems*, S. Martellucci, A. N. Chester, eds. (Springer US, 1997), pp. 3–22.
- [55] J. N. Damask, *Polarization Optics in Telecommunications* (Springer, 2005).
- [56] R. A. Chipman, W.-S. T. Lam, G. Young, *Polarized Light and Optical Systems* (CRC Press, 2019).
- [57] R. E. Jacobsen, G. G. Attridge, S. F. Ray, N. R. Axford, *The Manual of Photography: Photographic and Digital Imaging* (Focal Press, 2000)