

Supplement: Polarization Multi-Image Synthesis with Birefringent Metasurfaces

Dean Hazineh, Soon Wei Daniel Lim, Qi Guo, Federico Capasso, Todd Zickler

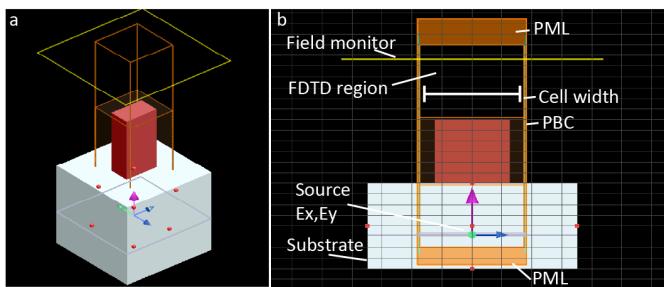


Fig. 1. Visualization of the cell FDTD calculation from a (a) 3D view and (b) 2D side-view. Field calculations are conducted on a grid of points within the FDTD region. PML refers to “perfectly matched layers” and PBC denotes periodic boundary conditions. We utilize the index of refraction for SiO_2 as the substrate and TiO_2 for the nanofin.

S1. GENERATION OF THE CELL DATASET

The nanofin cell dataset utilized in this work was generated by finite-difference time-domain (FDTD) field calculations with the commercial software, Ansys Lumerical. We highlight for the interested reader that there are several other free code packages that could alternatively be used. Specifically, a different method to solve Maxwell’s equations while assuming periodic boundary conditions that has gained significant attention recently in the metasurface community is the rigorous coupled wave analysis method (RCWA). Free and heavily validated RCWA code packages include [1], [2], [3].

A visualization of the FDTD calculations for a single cell instantiation is displayed in Fig. (1). The FDTD simulation region corresponds to a 3D (non-uniform) spatial grid of points, over which the electromagnetic fields are computed (displayed as the orange box in panels a-b). An ideal plane-wave source consisting of two linear polarization states, E_x and E_y , is injected from within the substrate. Notably, periodic boundary conditions for the simulation region are used transverse to the incident light while perfectly matched layers are used at the top and bottom boundaries. The cell width, as noted in the main paper, corresponds here to the width of the FDTD simulation region.

While the fields are calculated throughout the simulation region, we collect the Fourier-transformed complex fields at the “monitor” (denoted by the yellow square in panels a-b). It is positioned a few hundred nanometers above the nanofin to avoid near-field effects. The 2D fields recorded

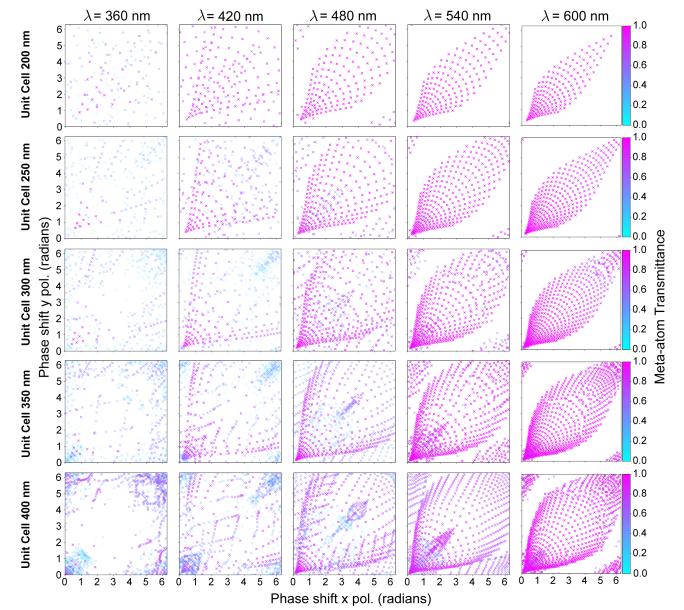


Fig. 2. Visualization of the nanofin optical response datasets for different cell widths (rows) and viewed for different wavelength slices (columns). The display is similar to Figure 1(b) of the main paper. Each point corresponds to a particular instantiation of the nanofin widths w_x and w_y but the point color (and transparency) map to the transmission percent. Complete coverage in the scatterplot would indicate that the library of cells can enable a decoupling of the ϕ_x and ϕ_y response.

at the monitor are then propagated to an axial point in the far field (a few micrometers above). We define the field at this centered, distant point to be the optical response of the cell. The amplitude transmittance and phase is normalized relative to the calculation with no nanostructure present.

In the main paper, we utilized a cell size of 350 nm; however, we also investigated the optical response dataset for other cell dimensions ranging from 200 nm to 400 nm. The results of this test when conducting a coarse sweep over nanofin widths are visualized in Figure 2. While the different datasets are qualitatively similar, we empirically observe that changing the cell size has an effect similar to scaling the amount of phase-delay imparted by a given nanofin. We chose the 350 nm cell width as it presented sufficient decoupling of the two phases ϕ_x and ϕ_y in the mid-visible near $\lambda = 530$ nm. A cell size of 400 nm could also be effective, however, we prefer selection of the smallest

usable cell dimension to avoid the potential for non-zero diffraction orders.

S2. VALIDATION OF THE CELL DESIGN THEORY

In Section 3.1 of the main paper, we reviewed the cell design principle for metasurfaces. We consider the cells as independent building blocks and pre-compute their optical responses. By utilizing periodic boundary conditions in the calculations, we obtain an approximation to the true local optical response that is independent to the selection of nanostructures at other locations on the composite metasurface. Here, we demonstrate that this approximation is reasonable. We show that the PSFs computed for a metasurface using the cell model is almost equivalent to that obtained in the most general case where the field across the entire metasurface is solved for without partitioning.

We note, however, that it is generally computationally intractable to compute the fields across millimeter scale devices using a nanometer scale grid discretization. For this reason, we are only able to simulate the fields across $50 \mu\text{m}$ diameter metasurfaces, using the same FDTD software as is discussed in supplemental Section S1. Because we reduce the diameter of the metasurfaces relative to those considered in the main paper, we also reduce the lens-to-photosensor distances such that the designed f-number of each optic remains the same. This enables a better generalization of the findings.

For the FDTD calculations, we utilized 64 CPU cores and each full lens simulation took approximately 8 hours on a compute cluster. When calculating the fields across the full metasurface, we utilize perfectly matched layers for all the simulation boundaries (in contrast to the periodic boundary conditions that are used for the cell simulations). As a note, recent research on hardware and software acceleration has been leading to the development of specialized field solvers better suited to this task, a notable example is the recently released commercial software Tidy3D. In the future, simulations across millimeter or larger devices may become accessible.

We first review the analysis conducted for the task of steerable filters (similar to Figure 6 in the main paper). Utilizing the same optimization algorithm and target filter, we designed a $50 \mu\text{m}$ metasurface for infinity focus and monochromatic incident light of $\lambda = 532 \text{ nm}$. The optimized arrangement of nanofins can be seen in supplemental Figure 3a. In panel (b), we show the computed, modulated fields that are transmitted through the full metasurface in response to a normally-incident, linear polarized plane-wave, whose polarization angle is orientated at 45° with respect to the x-axis. The FDTD calculations are computed for the full metasurface without partitioning while the cell-based treatment stitches together the predicted, spatial modulation pattern based on the pre-computed dataset of cell optical responses. Notably, we observe excellent agreement in the phase predictions for the metasurface when assuming the cell-based treatment vs the more general but expensive full model. Qualitatively, the transmittance also has strong agreement although we observe more variations.

We now compare the predicted PSFs for both cases. The PSF calculations for the cell-based approach is done in the

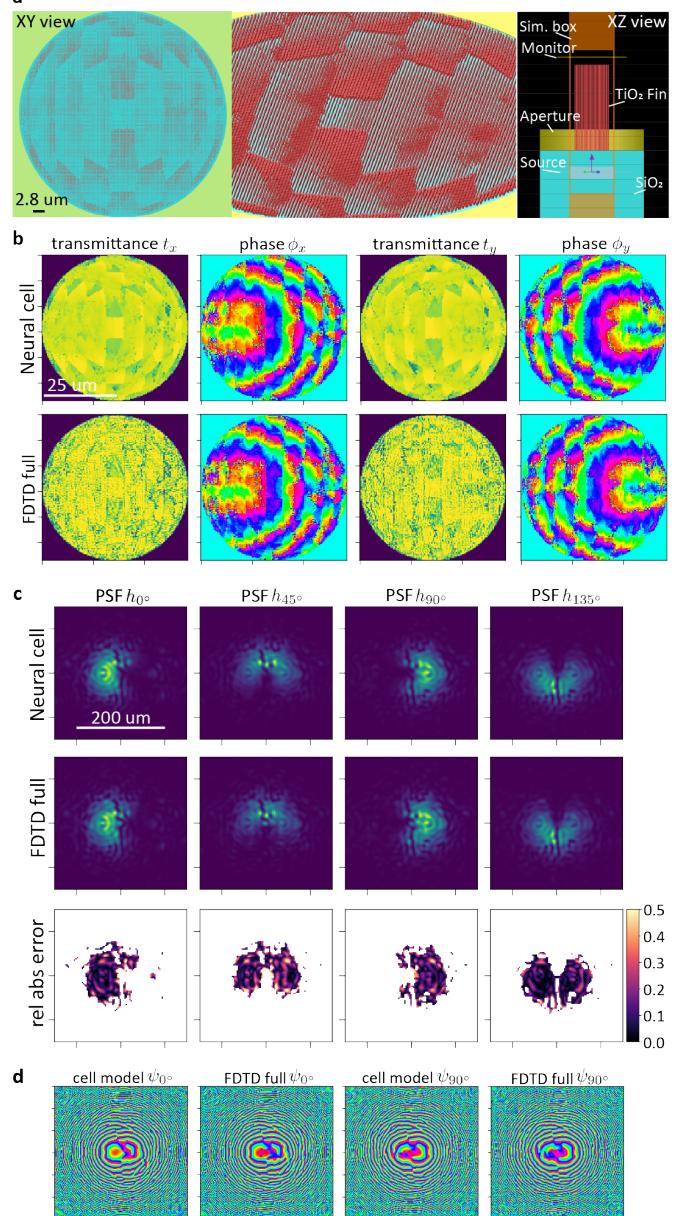


Fig. 3. A comparison of full-field FDTD calculations to the cell-based calculations presented in the main paper for a reduced size metasurface that implements the steerable Gaussian derivatives (similar to Figure 6 of the main paper). (a) The optimized $50 \mu\text{m}$ diameter metasurface loaded into the FDTD software. Details of the calculation are shown in the right-most panel. (b) The phase and transmittance just after the metasurface, in response to a normally incident plane wave of wavelength $\lambda = 532 \text{ nm}$. We refer to the mapping from the nanofin cell to the predicted optical response when utilizing the pre-trained MLP as the "neural cell" prediction. FDTD full refers to the direct calculation of the field when simulating the entire device. (c) Predicted intensity PSFs from the cell-based treatment and Fresnel propagation vs directly from FDTD. The relative absolute error is computed and shown only for pixels with an intensity of at least 5% the peak intensity. (d) The phase distribution at the output plane computed by both methods.

same manner as in the main paper. We first utilize the pre-trained MLP to map the metasurface cells to their local optical response. We then propagate the field defined by the collection of per-cell responses to the photosensor plane using the Fresnel integral. We do this calculation directly for

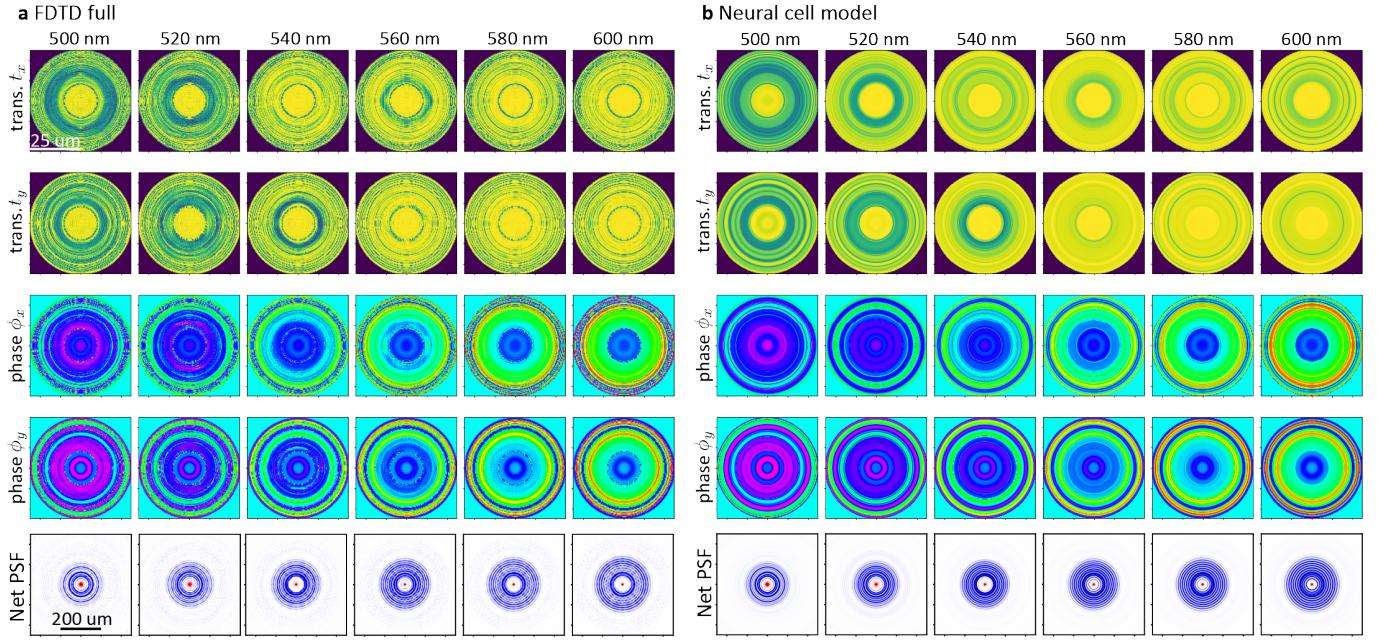


Fig. 4. Similar to supplemental Figure 3, we compare the cell-based approach (and the Fresnel propagated PSFs) to the equivalent calculation utilizing FDTD for the $50 \mu\text{m}$ metasurface shown in supplemental Figure 5.

h_{0° and h_{90° and then obtain h_{45° and h_{135° by computing the interference. This set of PSFs are shown in supplemental Figure 3c-d.

For the full FDTD case, we first directly solve for the field after the metasurface. We then use the FDTD software itself to propagate this field to the output plane; in doing so, a rigorous treatment of propagation is used which differs from the Fresnel integral and does not assume the paraxial approximation. This comparison thus also provides validation for our differential propagator and treatment of interference. The FDTD predicted PSFs are also displayed in panel c-d, and we find excellent agreement between the cell based calculations and the more rigorous FDTD treatment.

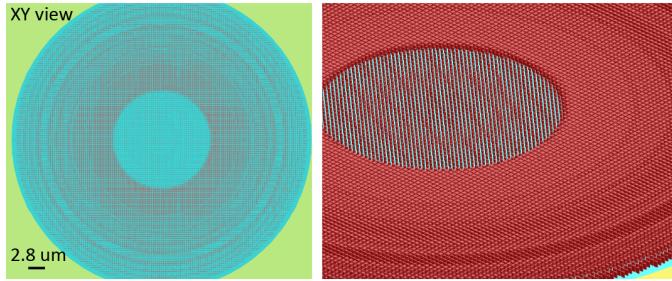


Fig. 5. Visualization of the optimized metasurface to be used in a broadband setting. This metasurface corresponds to the analyzed profiles in supplemental Figure 4.

We now explore the full-field analysis applied to a broadband case, shown in supplemental Figure 4. The technical details of the cell model and FDTD calculations mirror the above discussion, so we focus here on the results. Similar to Section 4.3 of the main paper, we consider the design of a metasurface that is used to approximate the Laplacian of Gaussian kernel for a wide range of incident wavelengths. For this design task, however, we find that it is suitable to

apply the filter-based objective given by Equation (10) in the main paper. We define the target filter kernel to have a width that increases with wavelength which matches the natural broadening of the PSF. The optimized nanofin metasurface is shown in Figure 5. Again, we compare the FDTD calculation (left panel of supplemental Figure 4) to the cell-based approach (right panel) and find strong agreement for the predicted transmittance, phase, and PSFs across the wavelength range.

S3. NEURAL OPTICAL MODEL EVALUATION

For gradient-based optimization of the metasurface, we require an efficient and differentiable approximation for the mapping between the nanostructure shape parameters and the optical response of the cell. As an alternative to the MLP, we also considered for this work elliptic radial basis function networks (ERBFs) and simple multivariate polynomial functions as applied in [4]. For all cases, we consider the input-output mapping depicted in Figure 2a of the main paper. To the best of our knowledge, ERBFs have not previously been explored in the context of this problem. First we review the performance of these alternative representations and after, we discuss auto-differentiable field solvers as a benchmark.

ERBF networks are reviewed in [5], [6] and may be considered as a particular class of neural networks consisting of a single hidden layer. Each neuron in the hidden layer parameterizes a radial basis function, which in this case corresponds to a three-dimensional elliptic Gaussian potential (assuming the set of three inputs w_x , w_y , and λ). The neurons then compute the Euclidean distance of the inputs and its weights followed by the activation, in contrast to neurons in a typical MLP which utilize the dot-product between inputs and weights. The number of neurons in the

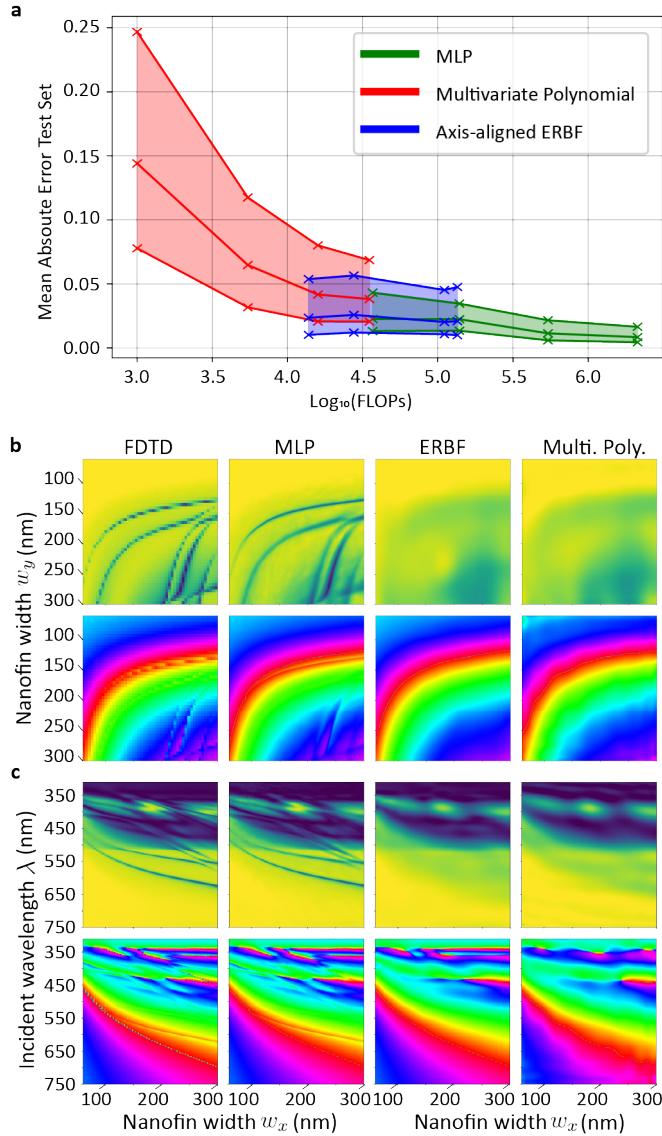


Fig. 6. A comparison of the MLP to other implicit representations. (a) We train several models of different sizes and evaluate the performance of each on a test-set of nanofin cells. In each case, we compute the predicted complex optical response for two basis polarization states and evaluate the mean absolute error relative to the true optical response. FLOPs refers to the number of floating point operations required per model evaluation. (b-c) This figure is analogous to Figure 2b-c of the main paper. We display the model predictions for slices through the dataset at (b) a fixed wavelength of $\lambda = 532 \text{ nm}$ and (c) at a fixed nanofin width $w_y = 180 \text{ nm}$. The points are queried at 5x the resolution of the training data.

hidden layer and thus the number of basis functions used to represent the data is a degree of freedom. The neuron's standard deviations and center coordinates, in addition to the weights and bias of a linear output projection layer, are all trainable parameters that may be updated by stochastic gradient descent.

We tested axis-aligned ERBFs with a number of neurons between 512 and 5000 to represent the nanofin cell dataset. Increasing the number of neurons beyond this point became impractical from a training time perspective. While we also explored the more general case of learnable covariance matrices rather than axis-aligned standard deviations, these

networks required a similar number of floating point operations (FLOPs) as the MLP but took orders of magnitude longer to train.

Alternatively, the multivariate polynomial formulation is relatively standard and the mapping is cast into a linear, matrix form. The coefficient matrix is updated by the method of least squares and we consider a separate matrix for each of the six output parameters. We tested polynomial orders up to 24, above which became impractical due to memory limitations. We also chose not to explore higher polynomial orders utilizing a stochastic gradient descent training scheme.

The test performance and computational complexity defined by the number of floating point operations is displayed in Figure 6 for the different models. For the MLP and the ERBF networks, the same set of test cells in the dataset were withheld during training. We find that the MLP substantially outperforms the other two representations in terms of achievable accuracy. While the largest MLP (1024 neurons in each hidden layer) required an order of magnitude more FLOPs per inference than the other largest models tested, this difference proved unimportant as the MLP is still efficient enough to be used in the optimization of a 2 mm metasurfaces with a single desktop GPU (RTX 3090). Moreover, the smallest MLP (e.g. 256 neurons) also outperforms the other representations. The model predictions when reproducing slices of the dataset are visualized in panels b-c.

Lastly, we compare the computational complexity of these approximate models to the direct field calculations evaluated utilizing an auto-differentiable field solver. Here, we consider the Tensorflow implementation of rigorous coupled wave analysis (RCWA) released in [2]. RCWA solves Maxwell's equations in the Fourier domain and formulates the problem as an eigenequation. For small cell sizes and periodic boundary conditions, RCWA is generally more computationally efficient than FDTD. The accuracy of the calculations depends on the grid discretization used when defining the simulation and on the number of Fourier modes applied to the solution; typically 81 or more Fourier modes are required to obtain converged results.

For this study, we consider a discretization of the nanofin cell into a 512x512 grid. We then consider 49, 81, and 121 Fourier modes. Using the Tensorflow profiler, we then estimate the number of floating point operations required per RCWA cell evaluation (and per wavelength) to be approximately, 363M, 1.62B, and 5.38B. Although memory bottlenecks are generally the limiting factor for applying auto-differentiable field solvers to inverse design problems, we summarize that RCWA would be several orders of magnitude more expensive as compared to the MLP from the perspective of number of FLOPs.

S4. MINIMUM-BIAS REGULARIZATION

We now discuss the motivation for the particular form of the bias regularization term introduced in Equation (11) of the main paper. The inspiration comes from analysis of the per-pixel "signal-to-bias" ratio metric introduced by the authors in [7], which we have rewritten in a generalized form in Equation (9) of the main paper. Since computing

the per-pixel ratio may be numerically unstable, we instead consider differential distances $D(\cdot)$ between the two vectors/matrices:

$$D(|H\alpha|, H|\alpha|) \equiv D\left(\left|\sum_c \alpha_c h_c\right|, \sum_c |\alpha_c| h_c\right), \quad (1)$$

where we are utilizing the notation $|X| = \sqrt{X \circ X}$ to denote a per-element absolute value utilizing the Hadamard product \circ , c enumerates the four polarization channels $c \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, α_c denote a scalar constant used in the digital synthesis, and h_c is a PSF from the set.

We find that it is insightful to first consider the difference d of the squared vectors¹ which may be expanded via:

$$\begin{aligned} d &= \left(\sum_c \alpha_c h_c\right)^2 - \left(\sum_c |\alpha_c| h_c\right)^2 \\ &= \sum_i \sum_j (\alpha_i \alpha_j - |\alpha_i| |\alpha_j|) h_i \circ h_j \\ &= 0 + \sum_{i \neq j} (\alpha_i \alpha_j - |\alpha_i| |\alpha_j|) h_i \circ h_j \\ &= \sum_{\text{sign}(i) \neq \text{sign}(j)} -2|\alpha_i| |\alpha_j| h_i \circ h_j. \end{aligned}$$

From the last line, we then observe that the L1-norm of this difference vector d , $\|d\|_1$, corresponds to the regularization that is used in the main paper. Specifically, summing over elements of d yields the overlap integral for PSFs that are combined with digital weights of opposite sign.

We now show that this masked regularization occurs up to a normalization term when considering the non-squared vector difference for d . In other words, variations of the masked orthogonality emerge when D in Equation (1) corresponds to standard vector norms:

$$\begin{aligned} d &= \left|\sum_c \alpha_c h_c\right| - \sum_c |\alpha_c| h_c \\ &= \frac{(\sum_c \alpha_c h_c)^2 - (\sum_c |\alpha_c| h_c)^2}{\sum_c \alpha_c h_c + \sum_c |\alpha_c| h_c} \\ &= \frac{\sum_{\text{sign}(i) \neq \text{sign}(j)} -2|\alpha_i| |\alpha_j| h_i \circ h_j}{\sum_c \alpha_c h_c + \sum_c |\alpha_c| h_c} \end{aligned}$$

In the last line, the denominator is a purely positive vector. If we compute $\|d\|_1$ in this case, we again obtain the overlap integral for pairs of PSFs but with a per-pixel weighting dependent on the intensity distributions.

S5. THE MEASUREMENT OPERATOR AND SIMULATED NOISE

As introduced in Section 3.4 of the main text, we define a noisy measurement operator $\Gamma(\cdot)$ which maps the photons at the photosensor plane to detected electrical signal. This measurement model is defined according to the EMVA standard [8] via,

$$\Gamma(X) = \text{Round}[(\mathcal{P}(qX) + \mathcal{N}(\mu_d, \sigma_d)) k], \quad (2)$$

1. While this type of vector difference is not used in standard distance metrics, it is helpful to consider as a start given the abnormal usage of the absolute value

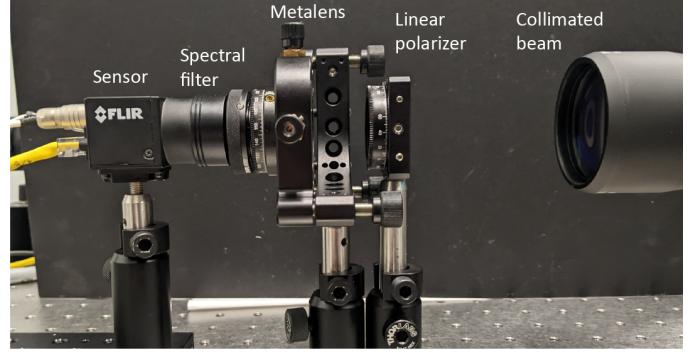


Fig. 7. Experimental setup used to measure the polarization encoded PSFs, discussed in Section 4.4 of the main paper. The polarization camera used is the FLIR Blackfly BFS-PGE-51S5; the global coordinate frame is defined relative to the photosensor, such that the x- and y-axis directions aligns with the 0° and 90° nanowire polarizers on the detection pixels. The linear polarizer at the entrance of the optical system is orientated at 45° .

where \mathcal{P} denotes the Poisson and \mathcal{N} the normal distribution, q and k are the detector quantum efficiency and gain, and μ_d and σ_d parameterize the dark noise. For the simulations in this work, we model the noise properties of the BFS-PGE-51S5 photosensor, which is the polarizer-mosaicked photosensor used in the experiment (and shown in supplemental Figure 7). EMVA technical specifications for this photosensor are available from the manufacturer (although reported only for 525 nm incident light); we utilize the values $q = 0.24$, $\mu_d = 2.45 e^-$ and inverse gain $k^{-1} = 0.18$ for all wavelengths and polarizations.

When rendering scenes via Equation (2) of the main paper, we desire a rescaling of the scene irradiance $\mathcal{I}_c(u, v)$ such that the per-channel intensity produced at the photosensor by an ideal lens corresponds to a noisy measurement with a specified peak signal-to-noise ratio (PSNR). We identify this scaling factor using the following relation [8]:

$$N_{\text{photons}}(\text{PSNR}) = \frac{\text{PSNR}^2}{2q} \left(1 + \sqrt{1 + \frac{4(\sigma_d^2 + \sigma_q/k^2)}{\text{PSNR}^2}}\right). \quad (3)$$

Here, N_{photons} refers to the peak number of photons which sets the maximum value of \mathcal{I}_c . The variable σ_q corresponds to quantization noise which in this case is defined by a 12 bit ADC conversion. By utilizing supplemental equations (2)-(3), we are able to predict the effectiveness of designed multi-image synthesis systems which are generally well-known to be sensitive to measurement noise.

S6: METASURFACE NANOFABRICATION

The metasurface design is written into 600 nm thick ZEP520A positive electron beam resist (Zeon Specialty Materials Inc.) using electron beam lithography (Elionix HS-50 50 kV). The resist voids are back-filled with amorphous titanium dioxide using atomic layer deposition (Cambridge NanoTech Savannah) and the excess titanium dioxide is etched back using inductively-coupled plasma reactive ion etching (Oxford PlasmaPro 100 Cobra). The electron beam resist is removed by overnight immersion in Remover PG (Kayaku Advanced Materials). An opaque gold mask (2

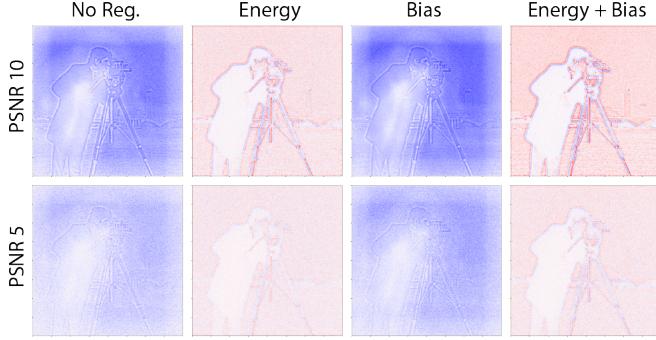


Fig. 8. Similar to the ablation study presented in Figure 5 of the main text. Here, we compare the image synthesis performance of optimized metasurfaces when designed with and without the regularization terms. For visual simplicity, we target a symmetric Laplacian of Gaussian kernel. Images are rendered using PSFs computed over a larger simulation area than that used during optimization (see Section 4 introduction for more details). Regularization has a substantial effect and the energy term enforces a spatially compact PSF. When paired with the bias term, the quality of synthesized images are improved. The importance of minimum-bias solutions can diminish with increasing SNR.

- [4] E. Tseng, S. Colburn, J. Whitehead, L. Huang, S.-H. Baek, A. Majumdar, and F. Heide, "Neural nano-optics for high-quality thin lens imaging," *Nature communications*, vol. 12, no. 1, pp. 1–7, 2021.
- [5] C. S. K. Dash, A. K. Behera, S. Dehuri, and S.-B. Cho, "Radial basis function neural networks: a topical state-of-the-art survey," *Open Computer Science*, vol. 6, no. 1, pp. 33–63, 2016. [Online]. Available: <https://doi.org/10.1515/comp-2016-0005>
- [6] I. Kononenko and M. Kukar, "Chapter 11 - artificial neural networks," in *Machine Learning and Data Mining*, I. Kononenko and M. Kukar, Eds. Woodhead Publishing, 2007, pp. 275–320. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9781904275213500113>
- [7] A. W. Lohmann and W. T. Rhodes, "Two-pupil synthesis of optical transfer functions," *Appl. Opt.*, vol. 17, no. 7, pp. 1141–1151, Apr 1978. [Online]. Available: <https://opg.optica.org/ao/abstract.cfm?URI=ao-17-7-1141>
- [8] E. M. V. Association, "Emva standard 1288: Standard for characterization of image sensors and cameras," online, 2010, www.emva.org.

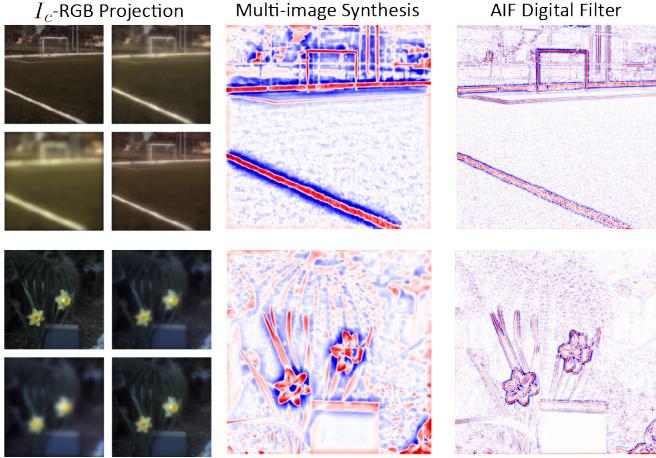


Fig. 9. Additional examples of the broadband, rendered and synthesized images produced by the optimized nanofin metasurface, equivalent to Figure 8d in the main paper but for different test scenes. See the caption and text for more details.

mm diameter) is deposited around the metasurface by positive tone photolithography in S1813 photoresist (Kayaku Advanced Material) with direct laser writing (Heidelberg MLA150), followed by electron beam evaporation of 5 nm of chromium and 200 nm of gold (Denton E-beam Evaporator). Residual photoresist is removed by overnight immersion in Remover PG (Kayaku Advanced Materials).

REFERENCES

- [1] V. Liu and S. Fan, " S^4 : A free electromagnetic solver for layered periodic structures," *Computer Physics Communications*, vol. 183, no. 10, pp. 2233 – 2244, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010465512001658>
- [2] S. Colburn and A. Majumdar, "Inverse design and flexible parameterization of meta-optics using algorithmic differentiation," *Communications Physics*, vol. 4, no. 65, 2021.
- [3] J. P. Hugonin and P. Lalanne, "Reticolo software for grating analysis," 2023.