
Research Article

Gaurav Arya*, William F. Li, Charles Roques-Carmes, Marin Soljačić, Steven G. Johnson, and Zin Lin

End-to-end Optimization of Metasurfaces for Imaging with Compressed Sensing

Abstract: We present a framework for the end-to-end optimization of metasurface imaging systems that reconstruct targets using compressed sensing, a technique for solving underdetermined imaging problems when the target object exhibits sparsity (i.e. the object can be described by a small number of non-zero values, but the positions of these values are unknown). We nest an iterative, unapproximated compressed sensing reconstruction algorithm into our end-to-end optimization pipeline, resulting in an interpretable, data-efficient method for maximally leveraging metaoptics to exploit object sparsity. We apply our framework to super-resolution imaging and high-resolution depth imaging with a phase-change material. In both situations, our end-to-end framework discovers optimal metasurface structures for compressed sensing recovery, automatically balancing a number of complicated design considerations to select the best imaging measurement matrix from a complex, physically-constrained manifold with millions of dimensions. The optimized metasurface imaging systems are robust to noise, significantly improving over random scattering surfaces and approaching the ideal compressed sensing performance of a Gaussian matrix, showing how a *physical* metasurface system can demonstrably approach the *mathematical* limits of compressed sensing.

Keywords: Compressed sensing, End-to-end inverse design, Computational imaging

1 Introduction

In this paper, we combine a subwavelength flat-optics (“metasurface”) physical platform [10, 17, 27, 30] with a compressed sensing (CS) reconstruction algorithm [13] to end-to-end optimize the performance of underdetermined imaging systems. Underdetermined imaging systems, where the imaging sensor has fewer pixels than the target object, are becoming increasingly important as researchers attempt to extract more information from objects while keeping imaging systems compact, inexpensive, and fast. CS is a powerful, principled way of solving underdetermined imaging problems using a *sparsity* prior, where the object (or some transformation of it) is known to have a large number of nearly-zero values [11, 19, 61]. However, in CS problems, the parameters of the physical platform (which determine the “measurement matrix” of the system) have traditionally been assumed to be fixed. In this work, we show how to “end-to-end” optimize the metasurface imaging system to find the *physically realizable* measurement matrix that minimizes the CS reconstruction error. Thus, we propose a framework (Fig. 1, Section 2.1) for answering the following question: how can we maximally leverage metaoptics to exploit object sparsity? We embed an unapproximated, iterative ℓ_1 -regularized CS algorithm into our end-to-end pipeline, performing an adjoint sensitivity analysis of the Karush-Kuhn-Tucker (KKT) optimality conditions of the reconstructed CS solution for efficient backpropagation (Section 2.2). Like any classical CS algorithm, the reconstruction process is interpretable, data-efficient, and generalizable to data outside the training distribution: the *only* assumption made is sparsity, and no additional data is needed to formulate and solve the optimization problem. This distinguishes our approach from previous end-to-end metasurface [17, 35, 36, 59] and diffractive-

*Corresponding author: **Gaurav Arya**, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, e-mail: aryag@mit.edu

William F. Li, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Charles Roques-Carmes, Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA

Marin Soljačić, Research Laboratory of Electronics and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Steven G. Johnson, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Zin Lin, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA

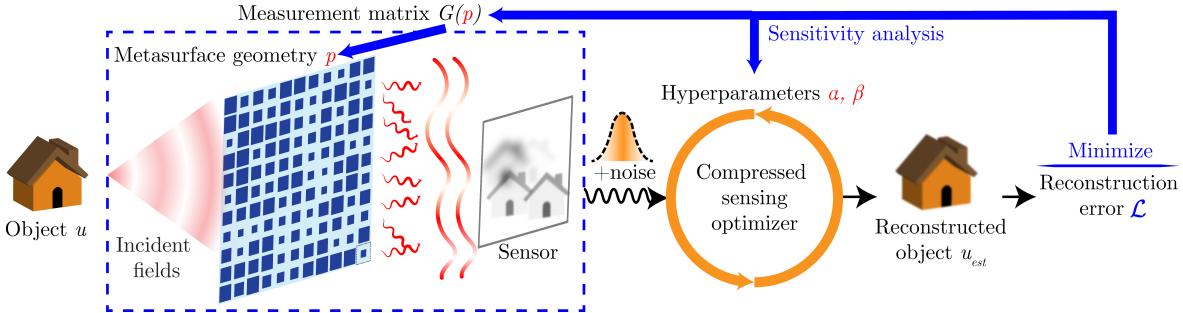


Fig. 1: The bilevel optimization problem for a metasurface imaging system in an underdetermined and potentially multi-channel imaging setting. A compressed sensing minimization problem is *nested* within the optimization pipeline. We find the optimal measurement matrix that is physically realizable by our metasurface optics, along with the best corresponding reconstruction hyperparameters, through end-to-end optimization of the reconstruction error.

optics [5, 14, 40, 51, 54, 60] optimization that considered overdetermined problems where CS was not necessary and/or took a data-intensive deep-learning approach to image reconstruction.

Our end-to-end optimization forms composite lens-like metasurface configurations that dramatically improve upon random scattering surfaces. For example, for pixel-limited imaging of sparse objects with a small sensor, we discover computational metasurface designs that approach the theoretically optimal performance of physically unrealizable Gaussian measurement matrices (Section 3.1). Importantly, metasurfaces exploit full-wave electromagnetic physics—such as multiple scattering, strong wavelength/polarization dispersion, and resonances [20, 38]. Even greater inference capabilities can potentially be unlocked by phase-change materials (PCMs) [24, 48, 50], which allow one to collect more data with a single device. We show that our end-to-end framework can maximize complex electromagnetic interactions in PCMs to tackle severely underdetermined multi-channel imaging problems (Section 3.2) such as high-resolution 3D imaging. Our work represents a new paradigm for enabling *maximal symbiosis* between metaoptics and CS. Our framework is also immediately applicable to diffractive optics, which are simply a special limiting case of metasurfaces (when the surface is locally uniform on the wavelength scale). These results open up new possibilities for high-resolution depth, spectral, and polarization imaging and beyond (Section 4). For example, in subsequent work we have applied our end-to-end CS framework to design 3D volumetric non-paraxial metasurfaces for detecting angles and wavelengths of sparse incoming beams, yielding order-of-magnitude improvements in resolution \times size [32].

Our metasurface physics platform is an ultra-thin nanostructured interface composed of an aperiodic array of subwavelength nanopillars, similar to structures previously fabricated for various metaoptics applications [16, 27, 28, 33, 35, 42]. In contrast to diffractive optical elements, which can be modeled by scalar diffraction theory [12], metasurfaces consist of nanoscale elements which require the simulation of “full-wave” Maxwell’s equations. This results in enhanced dispersion, which can be used to realize multi-functional devices [16, 44, 45]. Furthermore, metasurfaces can be fabricated over large scales with chip-scale fabrication techniques in various mature (e.g., titania [18] and silicon [4]) and emerging (e.g., phase-change materials (PCM) [24, 48, 50]) material platforms. The large number of degrees of freedom (up to $\gtrsim 10^6$ in this paper) and the rich physics at the nanoscale have spurred an emerging field of large-scale metasurface inverse design [34, 41, 43]. However, most metaoptics designs have been confined to optics-only applications, such as lensing, beam-steering, and holography [15]. In those applications, the desired image is prescribed *a priori*. In contrast, recent work has introduced *end-to-end* metaoptics inverse design [35, 36, 59], whereby metasurfaces are directly optimized to minimize the reconstruction error in overdetermined imaging settings, computationally discovering the best sensor image for subsequent reconstruction. This paper extends end-to-end metasurface optimization to the underdetermined regime, where sensor data is limited, for both single-channel (2D) imaging (Section 3.1) with a small sensor and multi-channel (3D) high-resolution depth imaging with a phase-change material (Section 3.2).

Compressed sensing (CS) has become increasingly prevalent in computational imaging, whereby optical components are paired with a CS reconstruction algorithm, which in some sense finds the “sparsest” object that produces the observed image. A key contribution of our work is to perform end-to-end optimization with a *nested* CS minimization problem that is solved in a fully iterative manner (not approximated, unrolled, or truncated), as opposed to optimizing proxies for good CS performance [47, 62] or employing the

approximation of a differentiable unrolled network (a fixed number of iterations) [23, 40, 63]. We solve the full bilevel optimization problem by performing a sensitivity analysis of the KKT conditions [2, 9] associated with the CS convex-optimization problem, using an adjoint formulation to efficiently backpropagate the gradient (Section 2.2). Crucially, we extend previous work that performed hyperparameter optimization of CS [7] by also optimizing with respect to the metasurface parameters underlying the measurement matrix itself.

2 Method

2.1 The bilevel optimization problem

Our pipeline consists of a physics platform, which simulates light propagation through the metasurface to a grayscale sensor, and a compressed-sensing reconstruction algorithm, which solves a convex optimization problem (Fig. 1). We optimize this pipeline end-to-end by minimizing the expected relative mean-square error (MSE) of the reconstructions over a distribution of training objects. Given a training object u (flattened into a vector), a raw noisy image y is formed by:

$$y = G(p)u + \eta, \quad \eta \sim \mathcal{N}(0, \sigma), \quad (1)$$

where G is the measurement matrix of the imaging system, dependent on the metasurface geometry p (see below), and η is an additive Gaussian noise [51]. We now write a bilevel optimization problem for our end-to-end pipeline:

$$\min_{p, \alpha, \beta} \mathcal{L} = \left\langle \frac{|u - u_{\text{est}}|_2^2}{|u|_2^2} \right\rangle_{u, \eta}, \quad (2)$$

$$u_{\text{est}} = \arg \min_x |Gx - y|_2^2 + \alpha|\Psi x|_1 + \beta|x|_2^2. \quad (3)$$

The reconstructed object u_{est} is formed by solving a generalized Lasso problem [56] with an additional ElasticNet ℓ_2 term [65] (which accelerates convergence as explained below). Here, $\langle \cdot \rangle_{u, \eta}$ denotes averaging over multiple training objects and noise realizations, and Ψ is a *sparsifying* transformation under which we expect the object to have a large number of coefficients that are nearly zero. The ℓ_1 regularization term $|\Psi x|_1 = \sum_j |(\Psi x)_j|$ reflects this prior by encouraging sparsity of the vector Ψx [13]. In the imaging settings we treat in this paper, Ψ is the identity operator (ℓ_1 regularization), but we also develop our sensitivity analysis in Section 2.2 for the 2D/3D anisotropic gradient operator with periodic boundary conditions—total variation (TV) regularization [46]—to show the generality of our framework for arbitrary linear transformations Ψ . We solve Eq. (3) using the matrix-free fast iterative soft-thresholding algorithm (FISTA) [6, 26], which allows us to exploit the convolutional structure of the large matrix G [3, 35].

We work in the paraxial regime, where the response of the physics platform is characterized by a set of point spread functions (PSFs)—the 2D image of a point source on the optical axis—one for each depth and/or wavelength channel [22]. The computation of the PSFs consists of spherical wave propagation, followed by a simulation of light scattering at the metasurface, followed by a near-to-far field transformation to compute the electric field in the sensor plane. To model scattering at the metasurface, we use the locally periodic approximation [43], in which the large-area metasurface is locally approximated by sub-wavelength *periodic* unit cells, each containing a nanopillar. We use rigorous coupled-wave analysis (RCWA) [37] to simulate the complex-valued transmission coefficients of a library of unit cells, from which we construct Chebyshev polynomials that interpolate the transmission coefficients as a function of the nanopillar geometrical parameters [43]. This differentiable Chebyshev surrogate model allows for rapid simulation of metasurface scattering during optimization.

From these PSFs, we assemble the measurement matrix G . The action of G is to convolve each channel of the object with its corresponding PSF, sum up the resulting contributions to the electric field intensity in the sensor plane, and then crop the field onto the sensor’s area. Here, the number of sensor pixels is smaller than the number of pixels in the (potentially multi-channel) object, and thus G has fewer rows than columns. We add independent Gaussian noise to the sensor readings, which approximates Poisson shot noise

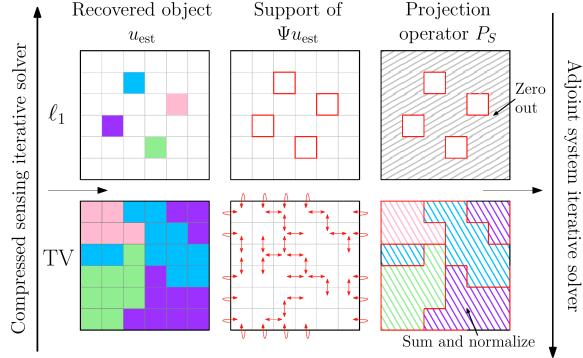


Fig. 2: The formation of the projection operator P_S , given the Lasso solution found in the forward pass. For simplicity we depict 6×6 2D objects, where different colors represent different intensity values, in the cases of both ℓ_1 and TV regularization. The support of Ψu_{est} is denoted in red (boxes around non-zero pixels for ℓ_1 regularization, arrows representing non-zero differences for TV regularization). The operator P_S is then used to set up the adjoint system, which is solved in the backward pass.

in the limit of large photon flux. The standard deviation σ of the Gaussian noise η is chosen to be a few percent of the mean image intensity, i.e. proportional to $|Gu|_1$.

In our bilevel formulation, the choice of ElasticNet reconstruction ($\beta > 0$) instead of pure Lasso ($\beta = 0$) serves a crucial purpose *during* the optimization. The initial metasurface geometry leads to a measurement matrix G that is highly unsuitable for compressed sensing; with no ℓ_2 term, the number of FISTA iterations required for convergence can be orders of magnitude larger for unoptimized metasurface geometries than for optimized geometries. Intuitively, for poorly performing G , nearby pixels are difficult to distinguish, so FISTA spends a large number of iterations making an arbitrary, badly conditioned decision about which non-zero entries to include. With the ℓ_2 term, the algorithm instead “gives up” on this futile task and spreads the values across many entries. In practice, we find that a large initial β makes the number of nested iterations roughly *independent* of G , fixing the slow-start issue. As we approach the optimized matrix, the need for the ℓ_2 term diminishes and the optimization chooses to shrink β to ≈ 0 , recovering the pure Lasso (see Fig. 3g).

2.2 Sensitivity Analysis of Compressed Sensing Reconstruction

For large-scale optimization ($\sim 10^6$ parameters) to be feasible, we must compute the sensitivity of the error \mathcal{L} with respect to the parameters p , α , and β . An adjoint method allows us to backpropagate gradients through our pipeline, at a computational cost similar to that of \mathcal{L} . For the majority of the pipeline, this can be done effortlessly using automatic differentiation tools [25].

However, differentiating through Eq. (3) requires special handling. Since we compute the reconstruction u_{est} (a vector of length $\approx 10^5$ in our optimizations) through an iterative solver, directly backpropagating through the solver’s iterations would require a prohibitive amount of memory. Furthermore, due to the non-differentiability of the ℓ_1 term $|\Psi u|_1$, the reconstruction u_{est} is itself not strictly differentiable with respect to G , α , and β .

To surmount these hurdles, our starting point is the KKT conditions for the generalized Lasso [57]. We expect that Ψu_{est} is sparse; let S be the support of Ψu_{est} (the set of indices of non-zero values). We can then define P_S , the orthogonal projection operator onto the subspace of objects x that satisfy $\text{support } \Psi x \subseteq S$. When Ψ is the identity (ℓ_1 regularization), P_S restricts an object x to the indices in S , zeroing out the other values. When Ψ is the anisotropic gradient operator (TV regularization), P_S sums an object x over each of the piecewise constant regions induced by the sparse gradient Ψu_{est} , divides each region’s sum by the square root of the size of the region (the division ensures P_S is an orthogonal projection, i.e. $P_S = P_S^T$), and sets all elements in that region equal to this summed and normalized value. These two cases are illustrated

in Fig. 2. The KKT conditions then imply that

$$\underbrace{(P_S G^T G P_S^T + \beta I)}_A \underbrace{u_{\text{est}}}_x = \underbrace{P_S G^T y - \frac{1}{2} \alpha P_S \Psi^T \text{sgn } \Psi u_{\text{est}}}_b, \quad (4)$$

which is a square linear system with dimension equal to the rank of P_S . Intuitively, the hard work performed in solving Eq. (3) by convex optimization methods is to discover the support S . Once S (and the signs $\text{sgn } \Psi u_{\text{est}}$) of the elements of Ψu_{est} have been found in the forward pass, we know a subgradient of the non-differentiable ℓ_1 term $|\Psi u|_1$ at u_{est} . This information allows us to form an efficient backward pass.

Specifically, we can use Eq. (4) to compute the sensitivity of u_{est} . Given the gradient $\frac{\partial \mathcal{L}}{\partial u_{\text{est}}}$, we write $\frac{\partial \mathcal{L}}{\partial x} = P_S \frac{\partial \mathcal{L}}{\partial u_{\text{est}}}$. Given $\frac{\partial \mathcal{L}}{\partial x}$, we now apply the standard adjoint method for a linear system $Ax = b$ [53] to find $\frac{\partial \mathcal{L}}{\partial A}$ and $\frac{\partial \mathcal{L}}{\partial b}$ by solving a single linear system with left operator A^T (the adjoint system). Since $A = P_S G^T G P_S^T$ is symmetric, we can solve this system using the matrix-free conjugate-gradient method [49]. By backpropagating through the formation of A and b from G , α and β , we then find the desired sensitivities $\frac{\partial \mathcal{L}}{\partial G}$, $\frac{\partial \mathcal{L}}{\partial \alpha}$, and $\frac{\partial \mathcal{L}}{\partial \beta}$. In this manner, we can backpropagate the gradient through Eq. (3) and thus the entire pipeline. An important aspect of our algorithm is that it is “matrix-free” [31]: in metasurface optics, the measurement matrix is an enormous convolution with a Green’s function [22] that would be impractical to construct explicitly, but our algorithm is constructed to only employ the matrix implicitly as a linear operator (exploiting FFT-accelerated convolutions [52]).

We remark that the solution to the generalized Lasso is unique almost surely for the cases of ℓ_1 and TV regularization [1, 55], and hence A is non-singular. In our optimizations, we find the adjoint system to be well-conditioned (it is strictly easier than the convex-optimization forward problem). It is important to note that $\frac{\partial \mathcal{L}}{\partial G}$, $\frac{\partial \mathcal{L}}{\partial \alpha}$, and $\frac{\partial \mathcal{L}}{\partial \beta}$ are only *subgradients*, as the support of Ψu_{est} can change. However, the support of Ψu_{est} is locally constant almost everywhere [57]; in practice, we observe excellent results when plugging the subgradient $\frac{\partial \mathcal{L}}{\partial p}$ into standard gradient-based optimization algorithms. We employ the stochastic-gradient algorithm Adam [29] because our training objects u are drawn from a random distribution of sparse objects.

3 Results

3.1 Single-channel imaging with small sensor

We first apply our optimization technique to single-channel imaging of sparse objects in the visible regime. We consider the case of a *small sensor* that has fewer pixels (128×128) than the object we wish to reconstruct (256×256). This is an underdetermined situation in which a traditional lens would perform very poorly because the object would not fit on the sensor. However, working under the assumption that our object is sparse in the spatial domain (thus we set Ψ to the identity), one may hope for a “multiplexing” design that compresses the object onto the small sensor in such a way that the image can be deconvolved effectively through CS. It is unclear *a priori* what metasurface geometry would best realize this goal.

As shown in Fig. 3, the end-to-end optimization *spontaneously configures* an arrangement of lens-like focusing elements, leading to a number of sharp peaks in the far field (Fig. 3a). Over 250 iterations of optimization on training objects with 5% sparsity (the proportion of non-zero values), the metasurface—initialized as an array of identical nanonpillars—evolves into the structure shown in Fig. 3i. The relative RMSE (the square root of the relative MSE \mathcal{L}) converges to 7.8% (Fig. 3h). This is an order of magnitude improvement over a randomly structured metasurface (96% relative RMSE).

Effectively, the optimized metasurface system takes many copies of a sparse object (Fig. 3b shows a test object with 7% sparsity) and “interlaces” them on the smaller sensor (Fig. 3c). The object is faithfully recovered (18% relative RMSE, Fig. 3d) even though the small sensor image is further corrupted by noise (2%). Our optimization also automates the choice of the reconstruction-algorithm hyperparameters α and β . In particular, α varies significantly on a logarithmic scale (Fig. 3f) and settles at approximately 10^5 , while β shrinks to a negligible value that is orders of magnitude smaller than α (Fig. 3g).

We now discuss how the measurement matrix of our optimized system compares to the “gold standard” of compressed sensing, a Gaussian random matrix. Such a matrix is *unphysical* for our optical system, but by comparing to it we can characterize the performance of our system relative to *all* possible matrices,

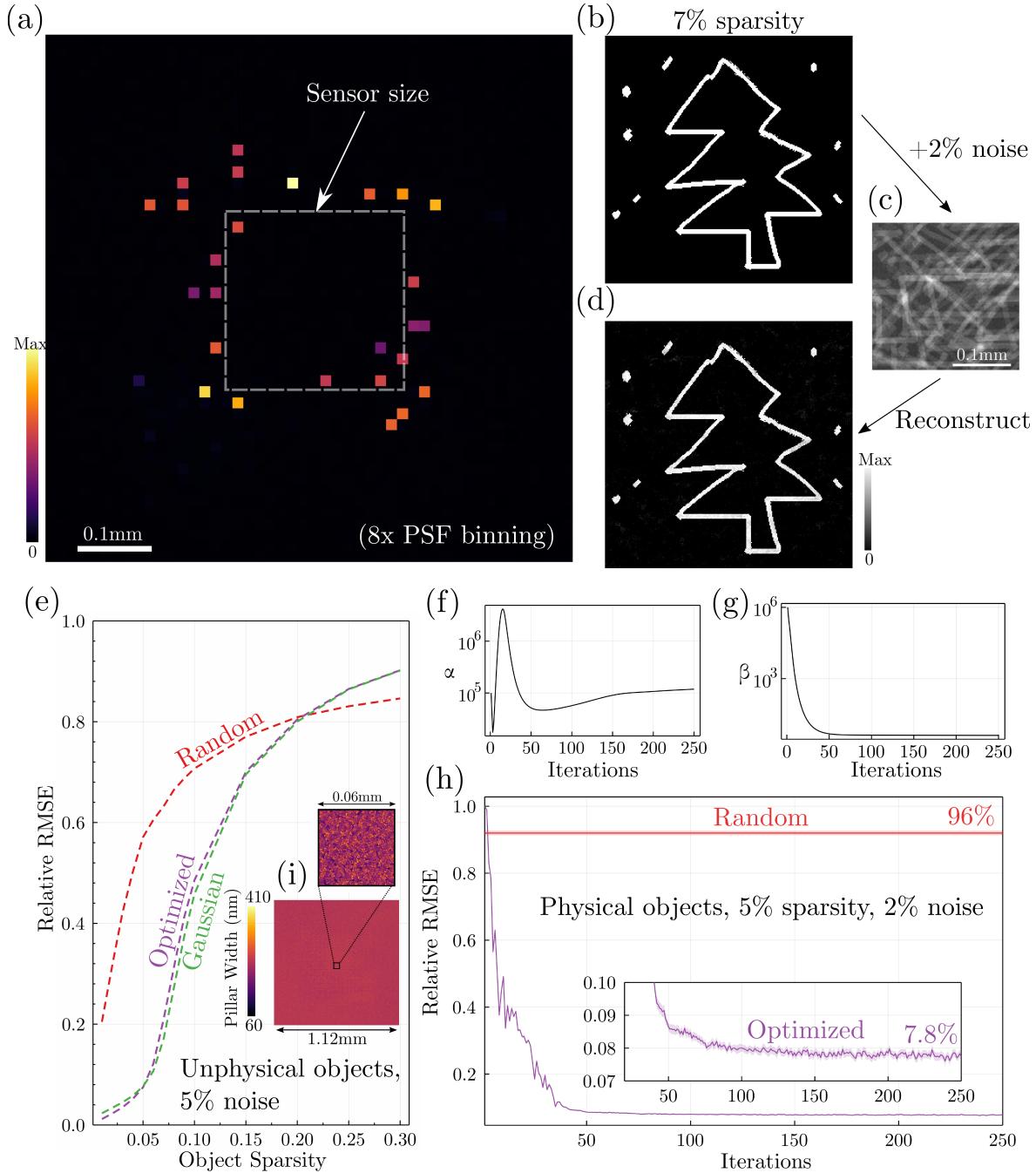


Fig. 3: Metasurface 2D imager with compressed sensing. (a) The optimized point spread function, a multi-focal metasurface with 27 foci (8x binned to emphasize lens positions). The relative size of the small sensor is shown; the multi-focal design spans beyond it so that the edges of the larger object are captured on the sensor. (An animation of the optimization is provided as supplementary material.) (b) Ground truth 256×256 sparse object (7% sparsity) with wavelength 550nm. (c) Image formed from ground truth object by the optimized design, with additive Gaussian noise (2%). (d) CS reconstruction of the ground truth object from the noisy image. (e) Reconstruction error on *unphysical objects* whose non-zero values are drawn from a distribution with mean 0 as a function of object sparsity. Shown for the optimized design, a random design, and the ideal Gaussian baseline under the same noise setting (5%). (f) Evolution of α during optimization. (g) Evolution of β during optimization. (h) Convergence of the reconstruction error during training on random objects with 5% sparsity. The objects have 3250 positive values uniformly sampled from the volume, with intensity values uniformly drawn from the interval $[0.8, 1.2]$. We also plot the error of a random metasurface. (i) The $1.12\text{mm} \times 1.12\text{mm}$ TiO_2 metasurface (refractive index $n \approx 2.4$) on a silica substrate, consisting of a grid of 2048×2048 sub-wavelength square unit cells, where each unit cell has width $\approx 470\text{nm}$. The square pillar in each cell has a height of 600nm, and width ranging from $\approx 60\text{nm}$ to $\approx 410\text{nm}$. The metasurface is depicted as a heatmap of the pillar widths, and we also zoom in on a $0.06 \times 0.06\text{mm}$ section of the metasurface.

whether physical or not. Representing Gaussian matrices would require a prohibitive amount of memory, so we use partial Gaussian circulant matrices as a computationally tractable equivalent [58]. As shown in Fig. 3e, a random metasurface significantly underperforms a Gaussian matrix. Remarkably, however, the optimized metasurface matches the performance of the Gaussian matrix throughout the entire error-sparsity “phase transition” [8] (the upswing segments of the error-sparsity relations in Fig. 3e).

We qualify this apparent optimality result by noting that the comparison in Fig. 3e is made with *unphysical* objects. (In the following, “unphysical” objects have mean zero and can have negative values; “physical” objects have only non-negative values.) On physical objects, optical systems are less robust to noise. This is due to the non-negativity of the measurement matrix of the metasurface system, which is a fundamental optical limitation. Intuitively, in the case of physical objects and measurement matrices, light can only be added to the sensor and not subtracted, making it harder to form sharp noise-robust features. This limitation is not faced by the Gaussian measurement matrices used as benchmarks in Fig. 3e, which have negative values and are therefore unphysical (not realizable with optics).

Mathematically, recalling our noise model (Section 2.1), the worse performance on physical objects manifests as a relative increase in the mean intensity magnitude of the image relative to unphysical objects (which include negative “intensities”), quantified by the following ratio:

$$\text{Image Mean Gap of } G = \frac{\langle |Gu|_1 \rangle_{\text{phys.}}}{\langle |Gu|_1 \rangle_{\text{unphys.}}} \cdot \frac{\langle |Xu|_1 \rangle_{\text{unphys.}}}{\langle |Xu|_1 \rangle_{\text{phys.}}}, \quad (5)$$

where G is the measurement matrix of the metasurface system, X is a Gaussian matrix, and $\langle \cdot \rangle$ denotes averaging over particular distributions of physical and unphysical objects. The Image Mean Gap characterizes the ability of the metasurface system to form sharp noise-robust image features on *physical* non-negative objects. From another perspective, the Image Mean Gap measures the “noisiness” of the PSF, i.e. its deviation from a sharply focused lens $G = I$.

We can use the Image Mean Gap to link the results of Fig. 3e to the physical case. We consider our training object distribution, where objects have 5% sparsity and their non-zero values are drawn uniformly from the interval [0.8, 1.2]. In this setting, the Image Mean Gap is ≈ 5.4 for a random metasurface and ≈ 2.1 for the optimized structure. This poorer Image Mean Gap, combined with the random metasurface’s poorer performance on unphysical objects in Fig. 3e, explains the random metasurface’s very poor performance on physical objects (96% relative RMSE). The Image Mean Gap of ≈ 2.1 ($2\times$ improvement) for the optimized structure is significantly better: it implies our metasurface system performs equivalently to an *unphysical* Gaussian matrix system that has $\approx 2.1\times$ more noise. It is unclear if the Image Mean Gap can be reduced even further while maintaining the design’s optimality on unphysical objects. One might hope that a reduced number of foci in the design, leading to less overlap of the object “copies” on the sensor, would lead to sharper image features on non-negative objects and thus a better Image Mean Gap. But this could come at the expense of optimality on unphysical objects due to reduced multiplexing of the object on the sensor, illustrating a tradeoff between the two factors. We leave more refined lower bound analyses for future work.

For physical imaging situations, we emphasize the importance of sparsity of the *measurement matrix*, which corresponds to lens-like PSFs in the convolutional model. Indeed, under a non-negativity constraint, one can show that *sparse* random matrices can perform well [21], and furthermore that convolution with certain random sparse patterns can lead to excellent performance [39]. Without any human input, the end-to-end optimization forms a *sparse* PSF that fits this characterization with optimal choices of the number and positions of lenses. In doing so, the optimization forms a measurement matrix with a reduced Image Mean Gap (≈ 2.1) and a near-exact match with a Gaussian matrix on unphysical objects, therefore finding a *physical* metasurface geometry that significantly improves over random scattering patterns.

3.2 High-resolution 3D imaging with PCM

To show the flexibility of our framework, we now consider a multi-channel 64-depth imaging problem in the mid-infrared regime. In particular, we endeavor to image ℓ_1 -sparse, volumetric distributions of point sources at very fine depth resolutions. This leads to a severely underdetermined problem where an object with discretization $32 \times 32 \times 64$ depths is captured on a 96×96 sensor. To alleviate this problem, we incorporate a *reconfigurable* material (PCM), GSS4T1, which can undergo a rapid phase change from amorphous to crystalline states via electrical switching [64]. Thus, a GSS4T1 metasurface allows us to obtain two images from a single structure design with two material phases of refractive index contrast $\Delta n = 1.40$ (Fig. 5f). This

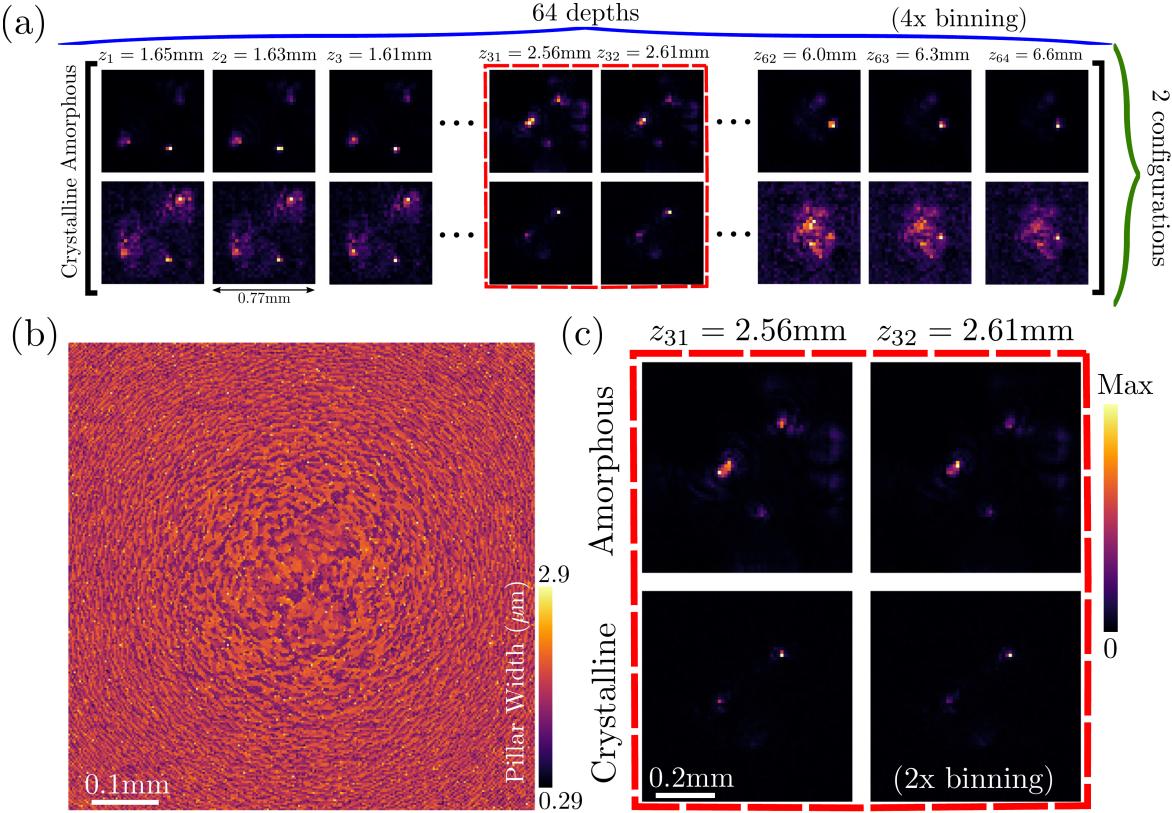


Fig. 4: Design of GSS4T1 3D imager with compressed sensing. **(a)** Optimized point spread functions, which exhibit high-frequency features, depth sensitivity, and phase change contrast (4x binned to emphasize lens positions). **(b)** The $0.77\text{mm} \times 0.77\text{mm}$ GSS4T1 metasurface on a CaF_2 substrate, consisting of a grid of 256×256 sub-wavelength square unit cells,. Each unit cell has width $\approx 3\mu\text{m}$ and index phase contrast $\Delta n = 1.40$. The square pillar in each cell has a height of $1.1\mu\text{m}$, and width ranging from $\approx 0.29\mu\text{m}$ to $\approx 2.9\mu\text{m}$. The metasurface is depicted as a heatmap of the pillar widths. **(c)** Zoom-in of PSFs at adjacent depths, illustrating the depth sensitivity and phase change contrast.

presents a fascinating problem for end-to-end optimization: the fineness of the required depth resolution necessitates high depth sensitivity, and the metasurface design must also ensure contrast between the images corresponding to the two material phases (but the same geometry) so that the second shot provides useful additional information. Our two-shot scheme using a PCM is significantly faster than multi-shot schemes that use digital micromirror devices or spatial light modulators [61].

In this setting, the optimization produces the metasurface shown in Fig. 4b. The PSFs of the optimized metasurface exhibit high-frequency lens-like features (Fig. 4a), necessary for good lateral resolution. Moreover, the PSFs demonstrate spatial dispersion and phase change contrast (Fig. 4c) despite significant physical challenges. For instance, the thickness of the GSS4T1 metasurface is a tenth of the wavelength in the mid-infrared regime, limiting the amount of wave scattering and dispersion available at the interface. Over the depth range, we notice dramatic evolution in the point-spread function, as various lenses of the multi-focal design come in and out of focus. However, physical limitations prevent the design from achieving such dramatic shifts between every pair of adjacent depths. Instead, the optimization tends to judiciously shift lens positions between adjacent depths by a few pixels, thereby shifting the images of adjacent channels on the sensor sufficiently to achieve the desired z -resolution. Crucially, the optimized design achieves these goals for the images of both configurations of the GSS4T1 metasurface. Furthermore, it achieves high contrast between configurations for the same depth channel, thus collecting *complementary* information from the two shots.

While these qualitative considerations give us a clue of the 3D imaging system’s resolving power, we now analyze the system’s CS performance on complex sparse scenes. In Fig. 5g, we plot the reconstruction error as a function of object sparsity on *unphysical objects* for the end-to-end optimized design, a random metasurface, and a partial Gaussian circulant baseline. With 5% noise, we observe that the optimized structure closely (but not perfectly) matches the performance of an ideal Gaussian matrix, while substantially outperforming random metasurfaces (e.g. $\approx 3\times$ smaller RMSE at 3% sparsity). Now, as in Section 3.1, we study the performance gap due to the non-negativity of the imaging system. At 3% object sparsity, the Image Mean Gap (Eq. (5)) between unphysical objects and non-negative training objects with intensity values in the range [0.8, 1.2] is 46 for a random structure and 18 for an optimized structure. Once again, the Image Mean Gap is worse for a random structure ($\times 2.5$), showing how the end-to-end optimization (which is trained on physical non-negative objects) favors PSFs that are more lens-like and lead to sharper image features. The lens-like features of the optimized PSFs (Figs. 4a and 4c) are not perfectly focused, however. This is reflected in the Image Mean Gap of 18 for the optimized structure, which is still relatively large; we expect that incorporating more complex nanophotonic structures than a single-layer metasurface will alleviate this issue (Section 4).

We also perform a visual test of the system using a physical ground truth object with ≈ 1000 non-zero values ($\approx 1.5\%$ sparsity) where patterns of varying structure have been hidden at 3 of the 64 depths. This ground truth object differs substantially from the objects in the training distribution which have uniformly distributed non-zero values across all depths. Nevertheless, the optimized imaging system accurately recovers the support and forms a reconstruction with 18% relative RMSE (Fig. 5c), despite the image being corrupted with 2% noise, showing how the system generalizes to out-of-distribution sparse data. In comparison to the optimized system, a random metasurface system achieves a much less accurate reconstruction of the same object (59% relative RMSE, Fig. 5d), and fails to accurately recover the support. We also note the significant improvement over *single-shot* designs where the GSS4T1 metasurface is kept in only its crystalline state. On the same ground truth object with 3% sparsity (Fig. 5a), an end-to-end optimized single-shot design obtains a significantly worse reconstruction (74% relative RMSE, Fig. 5e). This is due in large part to the delayed sparsity-error phase transition for the two-shot design (the rightward shift and slower incline of the sparsity-error relations in Fig. 5g), emphasizing the crucial role of the material index contrast and the ability of our end-to-end optimization to exploit it.

4 Conclusion and Outlook

Our results show how end-to-end optimization can navigate complicated, physically-constrained optimization spaces to discover measurement matrices with improved compressed-sensing performance. In certain cases where there are enough physical degrees of freedom for a given task (such as 2D imaging with a small sensor), the optimization discovers a measurement matrix that approaches the performance of a Gaussian

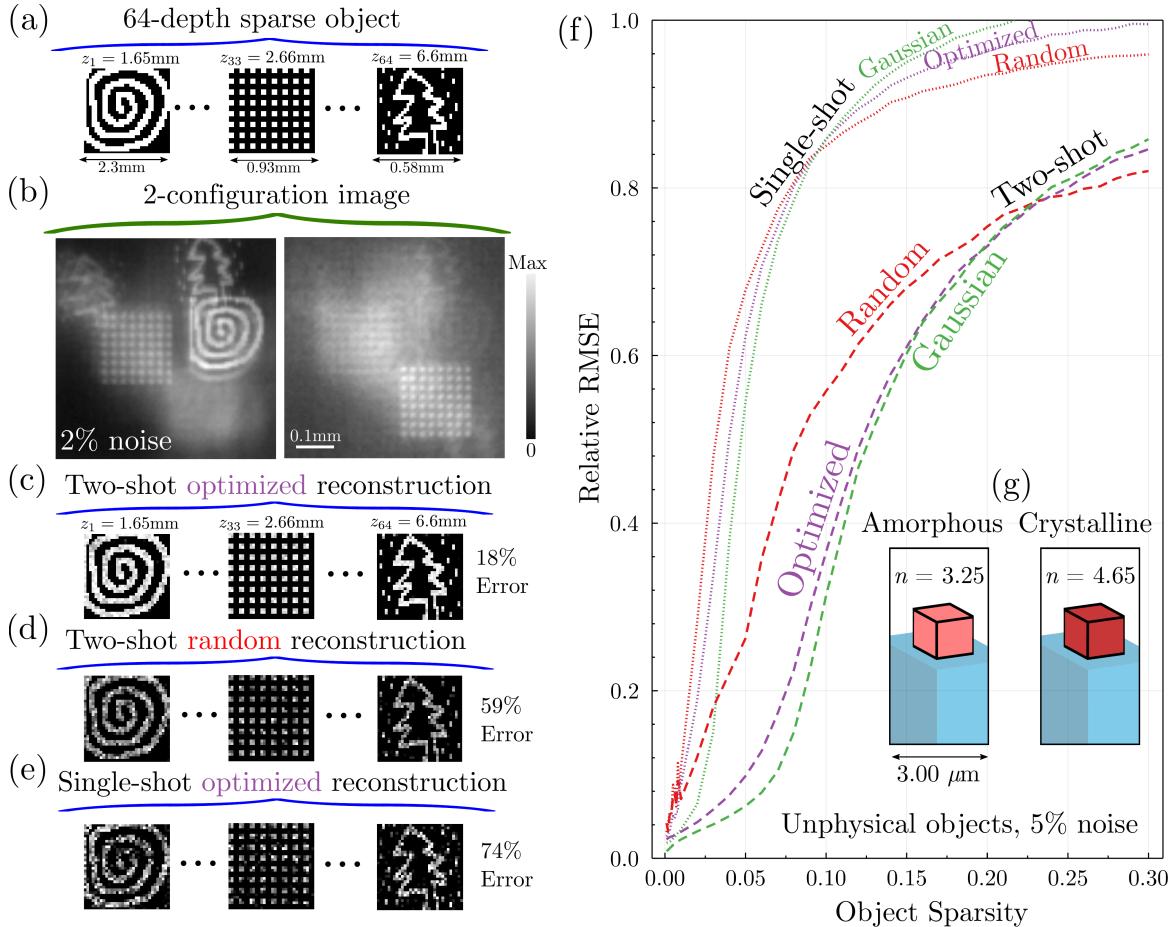


Fig. 5: Analysis of GSS4T1 3D imager with compressed sensing. (a) A 64-depth sparse ground truth test object ($\approx 1.5\%$ sparsity, or 1000 non-zero values) with wavelength $3.2\mu\text{m}$, and the depth channels inversely spaced from 1.65mm to 6.6mm . Patterns of varying structure are “hidden” in three of the 64 channels, plotted in grayscale. (b) The two images formed of the ground truth object by the amorphous and crystalline metasurface states of the optimized two-shot structure, with additive Gaussian noise (1%). (c) CS reconstruction of the ground truth object for the optimized two-shot system (18% relative RMSE). (d) CS reconstruction of the ground truth object for a two-shot system with a random metasurface (59% relative RMSE). (e) CS reconstruction of the ground truth object for an optimized single-shot system (74% relative RMSE). (f) Reconstruction error on *unphysical objects* with mean 0 as a function of object sparsity for the optimized design, a random design, an ideal Gaussian baseline, and the corresponding single-shot designs under the same noise setting (5%). (The single-shot end-to-end design is directly optimized for the single shot case, and the single-shot Gaussian baseline has half as many rows as the two-shot matrix.) (g) GSS4T1 metasurface pillar in amorphous ($n = 3.25$) and crystalline ($n = 4.65$) states.

matrix for *noise-robust* compressed sensing. Even under severe physical constraints as in the case of a deeply sub-wavelength PCM interface, end-to-end optimization can balance a complicated set of considerations to obtain measurement matrices which significantly outperform random two-shot designs and optimized single-shot designs for high-resolution 3D imaging. We expect that our technique—which, in generality, allows one to pair any measurement system with an ℓ_1 -regularized CS minimization algorithm and optimize the system *end-to-end*—will find use in a wide range of CS recovery systems whose design involves a large number of degrees of freedom.

In the design of such systems, it is crucial to take into account the type of prior imposed on the object. We note that our end-to-end framework supports any linear sparsifying transformation Ψ ; we need only be equipped with the projection operator P_S to perform the sensitivity analysis. Future work may build upon our method by performing end-to-end optimization of physical systems that incorporate more complicated non-linear priors that cannot be straightforwardly put in the form of the generalized Lasso.

Looking forward, as one leverages richer physics to extract more and more information from a scene with limited sensor data, we anticipate a growing need to cope with severely underdetermined problems. Our end-to-end framework, which supports arbitrary multi-channel (and multi-configuration) settings, paves the way for underdetermined nanophotonics imaging systems that can extract hyperfine depth, spectral, and/or polarization information from a single shot. The ultra-compact all-dielectric metasurface designs we have presented are directly amenable to large-scale fabrication platforms [33] and may enable the realization of metasurface cameras for imaging sparse scenes (e.g., locating stars in the night sky). We are particularly excited about future applications of our technique to thermal imaging, where one may endeavour to construct a full spectral imager that can reconstruct both emissivity and temperature profiles from a single shot. Furthermore, by incorporating multi-layered metasurfaces and photonic crystal slabs, one may exploit even richer *non-local* physics and develop imaging systems whose measurement matrix differs substantially in structure from a simple convolutional model. In this case, there are few conventional principles to guide the design; it is our hope that end-to-end optimization with compressed sensing will allow one to exploit the full power of nanophotonics in challenging underdetermined imaging problems.

Research Funding: This work was supported in part by the Simons Foundation and by the U.S. Army Research Office through the Institute for Soldier Nanotechnologies under Award No. W911NF-18-2-0048.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Data availability: Data underlying the results presented in this paper may be obtained from the authors upon reasonable request. The code used is publicly available at <https://github.com/gaurav-arya/ImagingOpt.jl>.

References

- [1] Ali Alnur, Tibshirani Ryan J. The Generalized Lasso Problem and Uniqueness // *Electronic Journal of Statistics*. 2019. 13, 2. 2307 – 2347.
- [2] Amos Brandon, Kolter J Zico. Optnet: Differentiable optimization as a layer in neural networks // *International Conference on Machine Learning*. 2017. 136–145.
- [3] Antipa Nick, Kuo Grace, Heckel Reinhard, Mildenhall Ben, Bostan Emrah, Ng Ren, Waller Laura. DiffuserCam: Lensless single-exposure 3D imaging // *Optica*. Jan 2018. 5, 1. 1–9.
- [4] Arbab Amir, Horie Yu, Bagheri Mahmood, Faraon Andrei. Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission // *Nature nanotechnology*. 2015. 10, 11. 937–943.
- [5] Baek Seung-Hwan, Ikoma Hayato, Jeon Daniel S, Li Yuqi, Heidrich Wolfgang, Wetzstein Gordon, Kim Min H. Single-shot Hyperspectral-Depth Imaging with Learned Diffractive Optics // *Proc. IEEE International Conference on Computer Vision (ICCV)*. October 2021.
- [6] Beck Amir, Teboulle Marc. A fast iterative shrinkage-thresholding algorithm for linear inverse problems // *SIAM Journal on Imaging Sciences*. 2009. 2, 1. 183–202.
- [7] Bertrand Q., Klopfenstein Q., Blondel M., Vaiter S., Gramfort A., Salmon J. Implicit differentiation of Lasso-type models for hyperparameter optimization // *Proceedings of the 37th International Conference on Machine Learning*. 2020.
- [8] Blanchard Jeffrey D. Toward deterministic compressed sensing // *Proceedings of the National Academy of Sciences*. 110, 4. 2013. 1146–1147.
- [9] Blondel Mathieu, Berthet Quentin, Cuturi Marco, Frostig Roy, Hoyer Stephan, Llinares-López Felipe, Pedregosa Fabian, Vert Jean-Philippe. Efficient and modular implicit differentiation // *arXiv preprint arXiv:2105.15183*. 2021.

- [10] *Bomzon Ze'ev, Kleiner Vladimir, Hasman Erez.* Pancharatnam–Berry phase in space-variant polarization-state manipulations with subwavelength gratings // Optics letters. 2001. 26, 18. 1424–1426.
- [11] *Boominathan Vivek, Robinson Jacob T., Waller Laura, Veeraraghavan Ashok.* Recent advances in lensless imaging // Optica. Jan 2022. 9, 1. 1–16.
- [12] *Born Max, Wolf Emil.* Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of light. 2013.
- [13] *Candes Emmanuel J., Wakin Michael B.* An Introduction To Compressive Sampling // IEEE Signal Processing Magazine. 2008. 25, 2. 21–30.
- [14] *Chang Julie, Wetzstein Gordon.* Deep optics for monocular depth estimation and 3d object detection // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. 10193–10202.
- [15] *Chen Hou-Tong, Taylor Antoinette J, Yu Nanfang.* A review of metasurfaces: physics and applications // Reports on Progress in Physics. 2016. 79, 7. 076401.
- [16] *Chen Wei Ting, Zhu Alexander Y, Sanjeev Vyshakh, Khorasaninejad Mohammadreza, Shi Zhujun, Lee Eric, Capasso Federico.* A broadband achromatic metalens for focusing and imaging in the visible // Nature Nanotechnology. 2018. 13, 3. 220–226.
- [17] *Del Hougne Philipp, Imani Mohammadreza F, Diebold Aaron V, Horstmeyer Roarke, Smith David R.* Learned integrated sensing pipeline: reconfigurable metasurface transceivers as trainable physical layer in an artificial neural network // Advanced Science. 2020. 7, 3. 1901913.
- [18] *Devlin Robert C, Khorasaninejad Mohammadreza, Chen Wei Ting, Oh Jaewon, Capasso Federico.* Broadband high-efficiency dielectric metasurfaces for the visible spectrum // Proceedings of the National Academy of Sciences. 2016. 113, 38. 10473–10478.
- [19] *Duarte Marco F., Davenport Mark A., Takhar Dharmpal, Laska Jason N., Sun Ting, Kelly Kevin F., Baraniuk Richard G.* Single-pixel imaging via compressive sampling // IEEE Signal Processing Magazine. 2008. 25, 2. 83–91.
- [20] *Engelberg Jacob, Levy Uriel.* The advantages of metaleenses over diffractive lenses // Nature Communications. Apr 2020. 11, 1. 1991.
- [21] *Gilbert Anna, Indyk Piotr.* Sparse Recovery Using Sparse Random Matrices // Proceedings of the IEEE. 98, 6. 04 2010. 937–947.
- [22] *Goodman Joseph W.* Introduction to Fourier optics, 3rd ed. // CO: Roberts & Co. Publishers. 2005. 1.
- [23] *Gregor Karol, LeCun Yann.* Learning Fast Approximations of Sparse Coding // Proceedings of the 27th on International Conference on Machine Learning. Madison, WI, USA: Omnipress, 2010. 399–406. (ICML'10).
- [24] *He Qiong, Sun Shulin, Zhou Lei.* Tunable/Reconfigurable Metasurfaces: Physics and Applications // Research. 2019. 2019.
- [25] *Innes Mike, Edelman Alan, Fischer Keno, Rackauckas Chris, Saba Elliot, Shah Viral B, Tebbutt Will.* A Differentiable Programming System to Bridge Machine Learning and Scientific Computing. 2019.
- [26] *Kamilov Ulugbek S.* A Parallel Proximal Algorithm for Anisotropic Total Variation Minimization // IEEE Transactions on Image Processing. 2017. 26, 2. 539–548.
- [27] *Khorasaninejad Mohammadreza, Chen Wei Ting, Devlin Robert C, Oh Jaewon, Zhu Alexander Y, Capasso Federico.* Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging // Science. 2016. 352, 6290. 1190–1194.
- [28] *Khorasaninejad Mohammadreza, Zhu Alexander Yutong, Roques-Carmes Charles, Chen Wei Ting, Oh Jaewon, Mishra Ishan, Devlin Robert C., Capasso Federico.* Polarization-insensitive metalenses at visible wavelengths // Nano letters. 2016. 16, 11. 7229–7234.
- [29] *Kingma Diederik P, Ba Jimmy.* Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014.
- [30] *Lalanne Philippe, Astilean Simion, Chavel Pierre, Cambril Edmond, Launois Huguette.* Blazed binary subwavelength gratings with efficiencies larger than those of conventional échelette gratings // Optics letters. 1998. 23, 14. 1081–1083.
- [31] *Langville Amy N., Meyer Carl D.* Google's PageRank and Beyond. Princeton: Princeton University Press, 2011.
- [32] *Li William F, Arya Gaurav, Roques-Carmes Charles, Lin Zin, Johnson Steven G, Soljačić Marin.* Transcending shift-invariance in the paraxial regime via end-to-end inverse design of freeform nanophotonics // Optics Express. In Press.
- [33] *Li Zhaoyi, Pestourie Raphaël, Park Joon-Suh, Huang Yao-Wei, Johnson Steven G., Capasso Federico.* Inverse design enables large-scale high-performance meta-optics reshaping virtual reality // arXiv preprint arXiv:2104.09702. 2021.
- [34] *Lin Zin, Groever Benedikt, Capasso Federico, Rodriguez Alejandro W, Lončar Marko.* Topology-optimized multilayered metaoptics // Physical Review Applied. 2018. 9, 4. 044030.
- [35] *Lin Zin, Pestourie Raphaël, Roques-Carmes Charles, Li Zhaoyi, Capasso Federico, Soljačić Marin, Johnson Steven G.* End-to-end metasurface inverse design for single-shot multi-channel imaging // arXiv preprint arXiv:2111.01071. 2021.
- [36] *Lin Zin, Roques-Carmes Charles, Pestourie Raphaël, Soljačić Marin, Majumdar Arka, Johnson Steven G.* End-to-end nanophotonic inverse design for imaging and polarimetry // Nanophotonics. 2021. 10, 3. 1177–1187.
- [37] *Liu Victor, Fan Shanhui.* S4: A free electromagnetic solver for layered periodic structures // Computer Physics Communications. 2012. 183, 10. 2233–2244.
- [38] *Mait Joseph N., Athale Ravindra A., Gracht Joseph van der, Euliss Gary W.* Potential Applications of Metamaterials to Computational Imaging // Frontiers in Optics / Laser Science. 2020. FTu8B.1.
- [39] *Marcos Diego, Lasser Theo, López Antonio, Bourquard Aurélien.* Compressed imaging by sparse random convolution // Opt. Express. Jan 2016. 24, 2. 1269–1290.

- [40] *Markley Eric, Liu Fanglin Linda, Kellman Michael, Antipa Nick, Waller Laura.* Physics-Based Learned Diffuser for Single-shot 3D Imaging // NeurIPS 2021 Workshop on Deep Learning and Inverse Problems. 2021.
- [41] *Molesky Sean, Lin Zin, Piggott Alexander Y, Jin Weiliang, Vucković Jelena, Rodriguez Alejandro W.* Inverse design in nanophotonics // Nature Photonics. 2018. 12, 11. 659–670.
- [42] *Mueller JP Balthasar, Rubin Noah A, Devlin Robert C, Groever Benedikt, Capasso Federico.* Metasurface polarization optics: independent phase control of arbitrary orthogonal states of polarization // Physical Review Letters. 2017. 118, 11. 113901.
- [43] *Pestourie Raphaël, Pérez-Arcibia Carlos, Lin Zin, Shin Wonseok, Capasso Federico, Johnson Steven G.* Inverse design of large-area metasurfaces // Opt. Express. Dec 2018. 26, 26. 33732–33747.
- [44] *Roques-Carmes Charles, Lin Zin, Christiansen Rasmus E, Salamin Yannick, Kooi Steven E, Joannopoulos John D, Johnson Steven G, Soljačić Marin.* Toward 3D-Printed Inverse-Designed Metaoptics // ACS Photonics. January 2022. 9, 1. 43–51.
- [45] *Rubin Noah A, D'Aversa Gabriele, Chevalier Paul, Shi Zhujun, Chen Wei Ting, Capasso Federico.* Matrix Fourier optics enables a compact full-Stokes polarization camera // Science. 2019. 365, 6448.
- [46] *Rudin Leonid I., Osher Stanley, Fatemi Emad.* Nonlinear total variation based noise removal algorithms // Physica D: Nonlinear Phenomena. 1992. 60, 1. 259–268.
- [47] *Satat Guy, Tancik Matthew, Raskar Ramesh.* Lensless Imaging With Compressive Ultrafast Sensing // IEEE Transactions on Computational Imaging. 2017. 3, 3. 398–407.
- [48] *Shalaginov Mikhail, An Sensong, Zhang Yifei, Yang Fan, Su Peter, Liberman Vladimir, Chou Jeffrey, Roberts Christopher, Kang Myungkoo, Rios Ocampo Carlos, Du Qingyang, Fowler Clayton, Agarwal Anuradha, Richardson Kathleen, Baleine Clara, Zhang Hualiang, Hu Juejun, Gu Tian.* Reconfigurable all-dielectric metalens with diffraction-limited performance // Nature Communications. 2021. 12, 1. 1–8.
- [49] *Shewchuk Jonathan R.* An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. USA: Carnegie Mellon University, 1994.
- [50] *Simpson Robert E, Cao Tun.* Phase Change Material Photonics // arXiv preprint arXiv:2006.14441. 2020.
- [51] *Sitzmann Vincent, Diamond Steven, Peng Yifan, Dun Xiong, Boyd Stephen, Heidrich Wolfgang, Heide Felix, Wetzstein Gordon.* End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging // ACM Transactions on Graphics (TOG). 2018. 37, 4. 1–13.
- [52] *Smith Steven W.* The scientist and engineer's guide to digital signal processing. 1997.
- [53] *Strang Gilbert.* Computational Science and Engineering. 2019.
- [54] *Sun Qilin, Zhang Jian, Dun Xiong, Ghanem Bernard, Peng Yifan, Heidrich Wolfgang.* End-to-end learned, optically coded super-resolution spad camera // ACM Transactions on Graphics (TOG). 2020. 39, 2. 1–14.
- [55] *Tibshirani Ryan J.* The lasso problem and uniqueness // Electronic Journal of Statistics. 2013. 7, none. 1456 – 1490.
- [56] *Tibshirani Ryan J., Taylor Jonathan.* The solution path of the generalized lasso // The Annals of Statistics. 2011. 39, 3. 1335 – 1371.
- [57] *Tibshirani Ryan J, Taylor Jonathan.* Degrees of freedom in lasso problems // The Annals of Statistics. 2012. 40, 2. 1198 – 1232.
- [58] *Tropp J.A., Wakin M.B., Duarte M.F., Baron D., Baraniuk R.G.* Random Filters for Compressive Sampling and Reconstruction // 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. 3. 2006. III–III.
- [59] *Tseng Ethan, Colburn Shane, Whitehead James, Huang Luocheng, Baek Seung-Hwan, Majumdar Arka, Heide Felix.* Neural nano-optics for high-quality thin lens imaging // Nature Communications. 2021. 12, 1. 6493.
- [60] *Tseng Ethan, Mosleh Ali, Mannan Fahim, St-Arnaud Karl, Sharma Avinash, Peng Yifan, Braun Alexander, Nowrouzezahrai Derek, Lalonde Jean-Francois, Heide Felix.* Differentiable compound optics and processing pipeline optimization for end-to-end camera design // ACM Transactions on Graphics (TOG). 2021. 40, 2. 1–19.
- [61] *Willett Rebecca M, Marcia Roummel F, Nichols Jonathan M.* Compressed sensing for practical optical imaging systems: a tutorial // Optical Engineering. 2011. 50, 7. 072601.
- [62] *Yanny Kyrollos, Antipa Nick, Liberti William, Dehaeck Sam, Monakhova Kristina, Liu Fanglin Linda, Shen Konlin, Ng Ren, Waller Laura.* Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy // Light: Science & Applications. Oct 2020. 9, 1. 171.
- [63] *Zhang Jian, Ghanem Bernard.* ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. 1828–1837.
- [64] *Zhang Yifei, Chou Jeffrey B, Li Junying, others .* Broadband transparent optical phase change materials for high-performance nonvolatile photonics // Nature Communications. 2019. 10, 1. 1–9.
- [65] *Zou Hui, Hastie Trevor.* Regularization and Variable Selection via the Elastic Net // Journal of the Royal Statistical Society. Series B (Statistical Methodology). 2005. 67, 2. 301–320.