Towards Mixed-State Coded Diffraction Imaging

Benjamin Attal and Matthew O'Toole

Abstract—Coherent diffraction imaging (CDI) is a computational technique for reconstructing a complex-valued optical field from an intensity measurement. The approach is to illuminate an object with a coherent beam of light to form a diffraction pattern, and use a phase retrieval algorithm to reconstruct the object's complex transmittance from the measurement. However, as the name implies, conventional CDI assumes highly coherent illumination. Recent works therefore extend CDI to account for partial coherence and imperfect detection, by modeling light as an incoherent mixture of multiple fields (*e.g.*, multiple wavelengths) and recovering each field simultaneously. In this work, we make strides towards the practical implementation and usage of *multi-wavelength* diffraction imaging. In particular, we provide novel analysis of the noise characteristics of multi-wavelength diffraction imaging, and show that it is preferable to coherent diffraction imaging under high signal-independent noise. Additionally, we present a compact coded diffraction imaging system and corresponding phase retrieval algorithms to robustly and simultaneously recover complex fields representing multiple wavelengths. Using a novel mixed-norm color prior, our prototype system reconstructs a larger number of multi-wavelength fields from fewer measurements than existing methods, and supports applications such as micron-scale optical path difference measurement via synthetic wavelength holography.

Index Terms—Phase Retrieval, Diffraction Imaging, Partial Coherence

1 Introduction

Coherent diffraction imaging (CDI) is a computational imaging technique for reconstructing the complex transmittance field of an object. Originally invented as an extension to X-ray crystallography in order to image non-crystalline samples [1], CDI illuminates a sample with X-rays (*i.e.*, high-energy photons) or another type of radiation (*e.g.*, visible light) to produce a diffraction pattern, and solves a phase retrieval problem to reconstruct the complex field from intensity-only measurements [2]. Extensions of CDI include coded diffraction pattern imaging (CDP) [3], [4], [5] and ptychography [6], [7], [8], [9], both of which solve an easier phase retrieval problem by capturing multiple diffraction patterns. Ptychography achieves this by mechanically moving an aperture stop [10] or changing the illumination angle [11], and CDP uses a spatial light modulator to code light before it reaches the sensor.

As the name suggests, these coherent diffraction imaging techniques often assume coherent light, and therefore decoherence can introduce unwanted complications [12]. Decoherence in scattering experiments can be expressed as the result of "mixing" multiple mutually-incoherent modes of light together. For example, when a light source consists of multiple wavelengths, the corresponding measurement is the sum of the intensities of the complex-valued fields associated with each wavelength.

Taking inspiration from work in mixed-state X-ray ptychography [12], we explore algorithms for reconstructing state mixtures from measurements captured with a near-field CDP system using visible light. Although potentially applicable to many forms of decoherence and optical setups, we primarily investigate the multi-wavelength phase retrieval problem. Instead of using a Bayer filter array or field-sequential imaging to capture color, we propose to capture multiple diffraction patterns that mix multiple wavelengths together, and solve a non-convex optimization problem to unmix the resulting images. We analyze the performance of mixed-state CDP, which we hope will lead to its practical usage and inspire future work that explores other forms of decoherence.

The contributions of this work include the following:

- a mixed-state CDP prototype that simultaneously recovers fields at multiple different wavelengths;
- a compressive mixed-norm color prior that aids in multi-wavelength recovery with fewer measurements than other techniques;
- an analysis of the behavior of mixed-state CDP under additive noise; and
- an application of mixed-state CDP to spatially-varying optical path difference (OPD) measurement via synthetic wavelength holography [13], [14], [15].

2 Related Work

2.1 Coherent Diffraction Imaging

CDI was first demonstrated by Miao *et al.* [1] in 1999. Based on techniques used in X-ray crystallography, this lensless microscopy technique requires illuminating an object with a planar field to produce a diffraction pattern. Reconstructing the complex field from the intensity measurements involves solving a phase retrieval problem. Unlike in-line or off-axis holography, CDI does not rely on interference between a scattered wave and reference wave, and therefore avoids the issues associated with conventional holography [16] (*e.g.*, mismatch in intensities of scattered and reference waves, lower spatial resolution due to axial and lateral vibrations).

There are a number of drawbacks associated with CDI techniques, however. First, it can require high dynamic range imaging to capture the diffraction pattern of an object, increasing acquisition time. Second, the phase retrieval

• *B. Attal and M. O'Toole are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213.*
E-mail: battal@andrew.cmu.edu

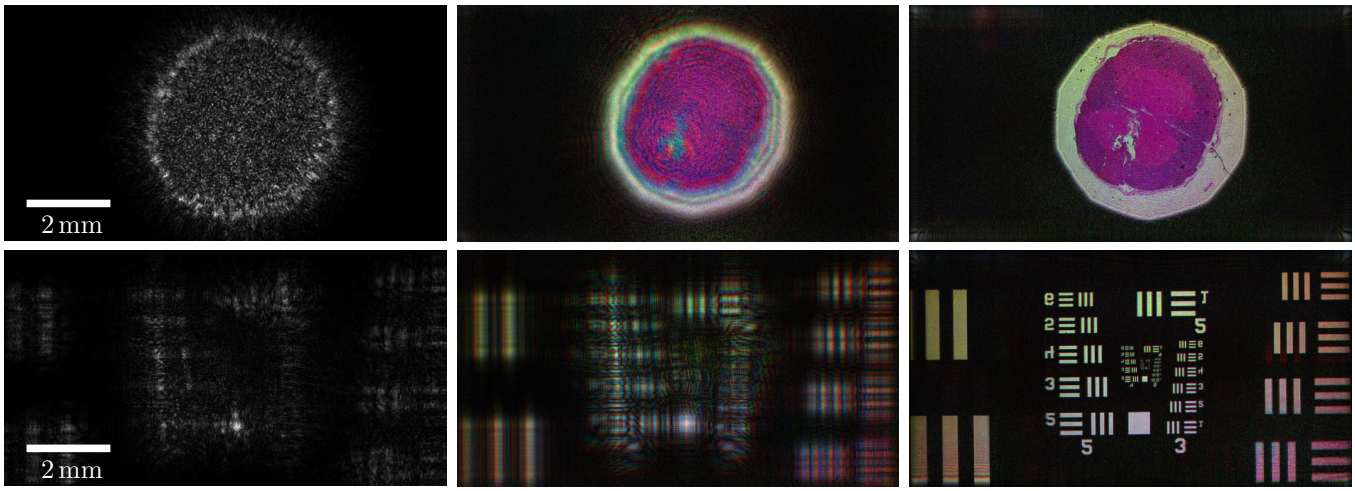


Fig. 1. **Mixed-state CDP.** Our proposed imaging technique can recover mutually incoherent modes from a partially coherent field. Here, we demonstrate the ability to simultaneously reconstruct three modes, each at a different wavelength: $638\,\mathrm{nm}$ (red), $520\,\mathrm{nm}$ (green), and $445\,\mathrm{nm}$ (blue). Each individual mode is fully coherent, but because there is no observable interference between the light at these different wavelengths, they are added together on an intensity basis. (*Left*) We show one of the raw captured images, which is a sum of the intensities of each color at the sensor plane. *(Center)* We show the amplitude of the reconstructed color modes at the SLM plane. *(Right)* We digitally refocus the captured field to the object plane. (*Top*) The first row represents a microscope slide of a rabbit spinal cord (48 measurements), and (*Bottom*) the second row is of a USAF resolution chart (48 measurements). Note that we do not make use of any filtering optics to reconstruct these images.

process is sensitive to noise and ambiguities. These drawbacks have been amended through a number of single- and multi-shot techniques involving ptychography and CDP, by increasing observation diversity.

2.2 Ptychography

Ptychography overcomes the disadvantages of conventional CDI by mechanically scanning the aperture or changing the illumination angle to produce multiple diffraction patterns. The object's field can be reconstructed more reliably from this set of images using basic phase retrieval reconstruction processes. As discussed by Pfeiffer [6], the theoretical concept of ptychography was introduced by Hegerl and Hoppe in 1969 [7], [8] and had been largely forgotten until its rediscovery in 2007 [9]. Since then, various forms of ptychography have been popularized, with applications including high-resolution microscopy [11].

2.3 Coded Diffraction Pattern Imaging

CDP extends CDI by using a sequence of known, but random, coded modulation patterns to encode additional information into the measured diffraction patterns. Zhang *et al.* [3] first demonstrated the ability to reconstruct complex-valued fields by placing a modulator in between the object and sensor. By shifting the modulator transversely, a set of K diffraction patterns are collected and can be used to robustly recover the object's field. Zhang *et al.* [4] also demonstrated the ability to capture complex-valued wave fields from a single diffraction pattern (typically referred to as coherent modulation imaging). The addition of a modulator provides two key advantages: (1) it reduces the dynamic range requirements of the sensor, and (2) it greatly facilitates the phase retrieval problem. While initially demonstrated for visible light, this technique is applicable to a broad range of radiation at all wavelengths, and was later demonstrated with X-ray sources by Zhang *et al.* [5]. This technique would later be extended with high-speed amplitude modulation [17], and used to develop new practical high-resolution wavefront sensing solutions [13], [18].

2.4 The Effect of Decoherence

The advances made in CDI depend on the ability to control the optical field, which typically assume a high degree of coherence in the diffraction images. As a result, decoherence is often unwanted, reducing the contrast of the diffraction patterns and significantly impacting the quality of the reconstruction results for conventional CDI.

Thibault and Menzel [12] categorizes decoherence in diffraction measurements into three groups: mixed states within the source (*e.g.*, due to the emission of partially coherent light), mixed states in the object (*e.g.*, due to the sample vibrating or changing over time), and mixed states from sensing (*e.g.*, due to the finite pixel area that spatially averages the incident light). In all of these scenarios, the optical fields can be expressed as an incoherent combination of multiple coherent fields.

It is therefore possible and relatively straightforward to generalize CDI and related techniques to reconstruct mixed states. For example, multi-wavelength phase retrieval algorithms (an example of a mixed-state algorithm) have been used with CDI to recover images at two distinct wavelengths by leveraging different spatial support constraints for each wavelength [19]. In X-ray ptychography, multi-wavelength reconstruction algorithms have also been used to account for the use of broadband illumination in diffraction imaging [20]. With respect to CDP, Dong *et al.* [21] demonstrated two-wavelength phase retrieval from only one coded diffraction pattern, and Gao and Cao [22] proposed a super-resolution technique that accounts for the finite size of pixels on a sensor.

In line with these prior works, this paper further explores the use of mixed-state algorithms for coded diffraction imaging. In particular, we provide a framework that can

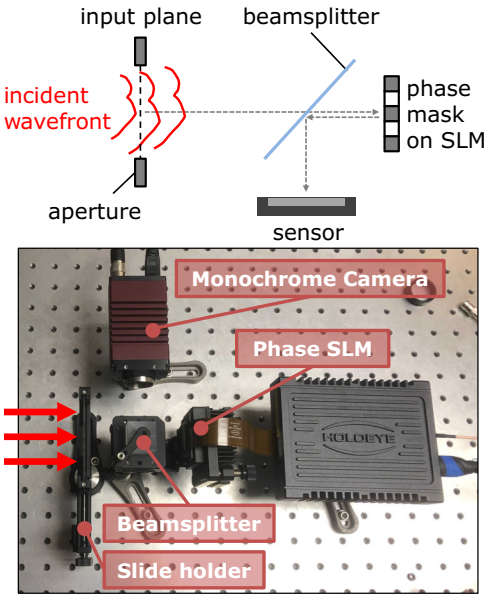


Fig. 2. **Illustration and photo of prototype system.** (*Top*) Similar to WISH [18], our mixed-state CDP system consists of only a few elements: a reflective phase-only spatial light modulator (SLM), a digital sensor, a 50R/50T beamsplitter. The input plane containing a microscope slide is back illuminated by a collimated RGB light source (not shown here). The field from the input plane propagates towards the SLM, the SLM modulates the phase of the field, the modulated wave propagates back towards the sensor, and the sensor measures the intensity of this wave. We capture multiple measurements of the field under different SLM patterns. (*Bottom*) Photo of our prototype system.

simultaneously recover mixed-states from few images by leveraging a compressive mixed-norm prior. Additionally, we provide a novel analysis of the noise characteristics of mixed-state CDP, which suggests that it is *preferable* to field-sequential CDP under high additive noise. While this work focuses primarily on our ability to simultaneously capture multi-wavelength fields from K diffraction patterns, we note that the same formulation can potentially be extended to model other sources of decoherence (*e.g.*, broadband illumination, birefringent or fluorescent samples, finite-sized photodiodes, temporally-multiplexed modulation patterns).

3 Coded Diffraction Pattern Imaging

Consider a CDP setup that consists of a spatial light modulator (SLM) and a camera aligned with a beamsplitter, as shown in Figure 2. The objective is to reconstruct a coherent field u incident on the SLM, from a collection of intensity measurements in response to different SLM patterns. The SLM modulates the phase of the field, and the field propagates to the sensor. Mathematically, the image formation model can be described as follows:

$$I^k \;=\; \left|P_z\left(\Phi^k_{SLM}\odot u\right)\right|^2 + \epsilon, \qquad (1)$$

where the complex vector $u \in \mathbb{C}^N$ represents the coherent field, the vector $\Phi^k_{SLM} \in \mathbb{C}^N$ is the k^{th} pattern on the SLM, and the operator $\odot$ is the Hadamard product representing element-wise multiplication between vectors. The linear operator $P_z : \mathbb{C}^N \to \mathbb{C}^M$ propagates the field from the SLM plane to the sensor plane, where z denotes the distance between the two planes. In addition, the measurements are affected by noise ϵ, including both shot and read noise; note that we do not model phase jitter associated with the SLM. The result is an intensity image $I^k \in \mathbb{C}^M$ on a M-pixel sensor in response to the k^{th} pattern on the SLM.

We formulate the phase retrieval problem to reconstruct the complex field u from K measurements as follows:

$$\arg\min_u \sum_{k=1}^{K}\left\|I^k - \left|P_z\left(\Phi^k_{SLM}\odot u\right)\right|^2\right\|^2 + \Gamma(u), \qquad (2)$$

where the function $\Gamma(u)$ is an optional prior on the field u, and the norm $\|\cdot\|$ is the Euclidean norm. This is known to be a non-linear and non-convex optimization problem, but many numerical algorithms for solving it exist (Section 3.3).

3.1 Mixed-State Image Formation Model

A fundamental assumption made in Equation (1) is that the field u is fully coherent. In this work however, we assume that the field is a mixture of mutually incoherent states, resulting in a superposition of the diffraction intensities at the sensor (*e.g.*, due to illuminating a sample with an RGB laser). We therefore extend the image formation model in Equation (1) to account for a combination of J distinct and mutually-incoherent fields:

$$I^k \;=\; \sum_{j=1}^{J}\left|P^j_z\left(\Phi^{j,k}_{SLM}\odot u_j\right)\right|^2 + \epsilon. \qquad (3)$$

Note that both the propagation operator P^j_z and SLM pattern $\Phi^{j,k}_{SLM}$ can depend on j. For example, when working with light at several different wavelengths, both the complex propagation operator and the phase modulation imparted by the SLM vary with wavelength.

As with Equation (2), we can compute a set of fields $\{u_j\}$ that satisfy our measurements by solving the following mixed-state optimization problem:

$$\arg\min_{\{u_j\}} \sum_{k=1}^{K}\left\|I^k - \sum_{j=1}^{J}\left|P^j_z\left(\Phi^{j,k}_{SLM}\odot u_j\right)\right|^2\right\|^2 + \Gamma\left(\{u_j\}\right). \qquad (4)$$

The prior $\Gamma(\{u_j\})$ can exploit correlations across the different fields and produce higher quality reconstructions for a given set of measurements.

3.2 Ambiguous Solutions in Mixed-State Sensing

Note that it is not always possible to uniquely recover the individual modes in Equation 4. Let $u_1, \ldots, u_n$ be modes with the *same* forward operator, such that $P^1_z = \cdots = P^n_z$ and $\Phi^{1,k}_{SLM} = \cdots = \Phi^{n,k}_{SLM}$ for all images k.

Suppose that $A^k = P_z \circ \Phi^k_{SLM}$ is the composite linear forward operator for image k, and let a^k_i be a row of this matrix corresponding to pixel i in image k. If we take $U = [u_1, \ldots, u_n]^T$ to be the matrix containing the modes as rows, then we can write the contribution of these modes to the measurement at pixel i in image k as:

$$\sum_{j=1}^{n}|u_j^T a^k_i|^2 = \|U a^k_i\|^2.$$

Any orthogonal transform V of the modes yields the same measurements since $\|V(Ua_i^k)\|^2 = \|Ua_i^k\|^2$. It is therefore only possible to recover a set of modes with the same forward operator up to an orthogonal transform ambiguity. Henceforth, we assume that the forward operators for each mode are distinct.

3.3 Reconstruction Algorithms

Our goal is to recover the states $\{u_j\}$ from the set of measurements $\{I^k\}$, by solving Equation (4). We assume that both the phase of the SLM patterns $\Phi_{SLM}^{j,k}$ and the propagation operator P_z^j are known and calibrated ahead of time, and that these operators are distinct for different modes j. Equation (4) is a non-linear and non-convex optimization problem. Luckily, the objective is generally well-behaved for similar problems [23]. Below, we make use of two algorithms to solve it.

3.3.1 Multi-Modal Gerchberg-Saxton

The classic Gerchberg-Saxton (GS) algorithm is a very popular phase retrieval technique. We leverage a generalized mixed-state version of the GS algorithm (Algorithm 1) (variants of which are presented in the literature [12], [21]) that follows four steps: (1) modulate the J wavefronts by each of the K SLM patterns and propagate the wavefronts to the sensor plane with operator P_z^j; (2) for each measurement, modify the average intensity of the J wavefronts with the measurement from the k^{th} SLM pattern; (3) propagate the wavefronts back to the SLM plane with operator P_{-z}^j and remove the phase modulation imparted by the SLM pattern; and (4) for every mode, compute the average of the K wavefronts. Note that step 2 ensures that the intensity of the light incident on the sensor agrees with the measurements. Despite being heuristic, GS is fast, highly memory efficient, and produces high-quality results. When $J = 1$, this method reduces to conventional CDP.

3.3.2 Multi-Modal Gradient Descent

A more principled approach to solve Equation (4) is to use first-order optimization techniques. Because optical wavefronts are complex-valued, we can use Wirtinger calculus to compute gradients and solve the phase retrieval problem. We rely on TensorFlow's AutoDiff to compute derivatives for our optimization problem. At high resolutions and for a large number of measurements, this procedure consumes a large amount of memory. However, it is a more principled and flexible framework that can support additional reconstruction priors.

3.3.3 Mixed-Norm Color Prior

We propose a mixed-norm prior for our color experiments:

$$\Gamma(\{u_j\}) = \left\| \sqrt{\sum_{j=1}^{J} |(P_d^j u_j)_x|^2 + |(P_d^j u_j)_y|^2} \right\|_1 \quad (5)$$

The propagation operator P_d^j propagates the wave to the target image plane where the sample is *in-focus*, at a known distance d from the SLM. The operators $(\cdot)_x$ and $(\cdot)_y$ denote the partial derivatives (computed via forward finite differences) in the x and y directions, respectively. In the case of single-modal phase retrieval (*i.e.*, $J = 1$), Equation (5) reduces to a 2D total-variation (TV) regularizer. In practice, Equation (5) encourages group gradient sparsity, where the gradients are in the same position across all modes. This reduces the number of images required to capture the multi-wavelength fields (see Figure 5). We note that optimization with this prior is similar in spirit to the multiple-measurement-vector (MMV) problem described by Van den Berg and Friedlander [24].

ALGORITHM 1: Multi-Modal Gerchberg-Saxton Algorithm

Input: Measurements $\{I^k\}$, patterns $\{\Phi_{SLM}^{j,k}\}$, propagation operators $\{P_z^j\}$, max iterations *num_iters*
Output: Multi-modal fields $\{u_j\}$

```
// Initialize fields
{u_j} = PlanarField()
for n ∈ {1, ··· , num_iters} do
    // Step 1: Propagate field from SLM to
       sensor plane
    for j ∈ {1, ··· , J}, k ∈ {1, ··· , K} do
        z_{j,k} = P_z^j ( Φ_SLM^{j,k} ⊙ u_j )
    end
    // Step 2: Modify amplitudes to
       satisfy measurements
    for j ∈ {1, ··· , J}, k ∈ {1, ··· , K} do
        ẑ_{j,k} = √I^k ⊙ ( z_{j,k} / √(Σ_{j=1}^J |z_{j,k}|^2) )
    end
    // Step 3: Propagate field from sensor
       to SLM plane
    for j ∈ {1, ··· , J}, k ∈ {1, ··· , K} do
        û_{j,k} = ( Φ_SLM^{j,k} )^{-1} ⊙ ( P_{-z}^j ẑ_{j,k} )
    end
    // Step 4: Average estimates of fields
    for j ∈ {1, ··· , J} do
        u_j = (1/K) Σ_{k=1}^K û_{j,k}
    end
end
```

4 Performance Under Noise

The framework described in the previous section makes it possible to recover state mixtures from a superposition of intensity measurements. However, alternate solutions may be available, including field-sequential imaging (*i.e.*, illuminating the sample with one wavelength at a time) or sensing with a Bayer color filter array, reducing the problem back to conventional CDP. In the following section, we provide insight on the advantages of mixed-state CDP in the presence of read noise.

4.1 Theory

Here, we make use of our notation from the previous section. We assume for simplicity that the number of image pixels is equal to the number of entries in the complex field (*i.e.*, $M = N$) and that the number of unknowns $2JN$ (2 unknowns for each complex entry) is less than the number of knowns KN. Lastly, we assume that the forward model is

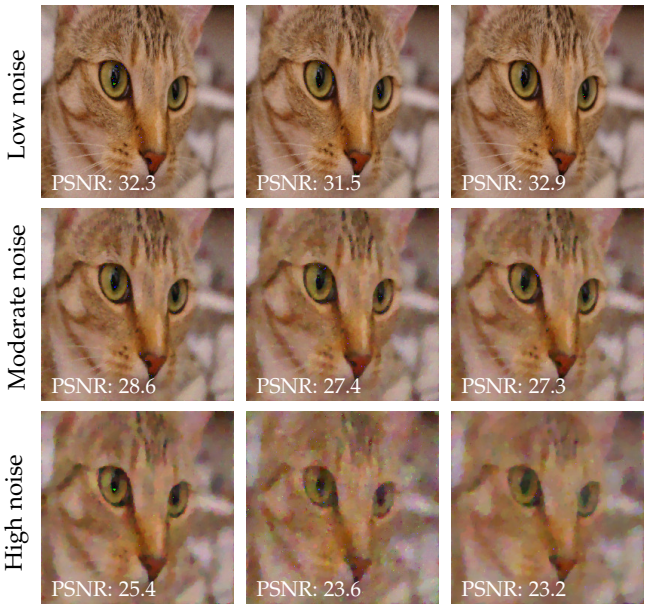


Fig. 3. **Performance under read noise**. We include representative reconstructions on a synthetic wavefront using **(a)** a mixed-state algorithm, **(b)** field-sequential imaging, and **(c)** Bayer imaging under three different read noise levels. 12 measurements were captured for each result.

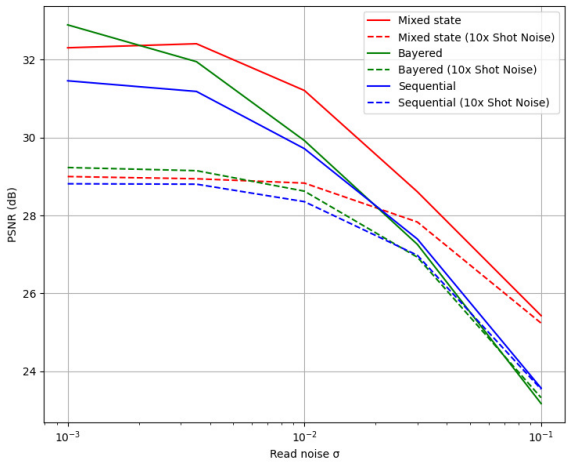


Fig. 4. **Quantitative performance under noise.** We show the performance of all methods (*Mixed-State*, *Sequential*, *Bayer*) under various amounts of signal independent read-noise. Our method (*Mixed-State*) consistently outperforms others for a moderate to large amount of noise. The results remain consistent with more $(10\times)$ signal-dependent shot noise.

injective. Let $(\hat{u}_1, \ldots, \hat{u}_J)$ be the ground truth modes, with ideal measurements $(\hat{I}^1, \ldots, \hat{I}^K)$. Specifically:

$$\hat{I}^k = \sum_{j=1}^{J} \left| P_z^j \left(\Phi_{SLM}^{j,k} \odot \hat{u}_j \right) \right|^2. \tag{6}$$

We can simplify this relation, writing

$$\hat{\mathbf{I}} = f(\hat{\mathbf{u}}) \tag{7}$$

where $\hat{\mathbf{I}} = (\hat{I}^1, \ldots, \hat{I}^K)$ is the vector containing all ideal measurements, and $\hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_J)$ contains the modes; the forward operator $f(\cdot)$ encapsulates SLM modulation, propagation, and intensity summation for all modes. For any input $\mathbf{u}$, we can write a linear approximation for f as

$$f(\mathbf{u} + \Delta\mathbf{u}) \approx f(\mathbf{u}) + \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u})\Delta\mathbf{u}. \tag{8}$$

Given a set of measurements with noise, $\hat{\mathbf{I}} + \epsilon$, a simplification of our optimization problem in Equation (4) with initial guess $\mathbf{u}$ is then

$$\arg\min_{\Delta\mathbf{u}} \left\| (\hat{\mathbf{I}} + \epsilon) - \left(f(\mathbf{u}) + \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u})\Delta\mathbf{u} \right) \right\|^2. \tag{9}$$

With a solution given by

$$\Delta\mathbf{u} = \frac{\partial f}{\partial \mathbf{u}}^{\dagger}(\mathbf{u})(\epsilon + \hat{\mathbf{I}} - f(\mathbf{u})), \tag{10}$$

we can use the above equation to derive how the recovered $\mathbf{u} + \Delta\mathbf{u}$ are affected by noise vectors ϵ about $\hat{\mathbf{I}}$.

Following the analysis of Schechner *et al.* [25], we assume Gaussian noise vectors ϵ with each entry $\epsilon_i \sim \mathcal{N}(0, \delta^2)$. We can then express the covariance matrix Σ of $\Delta\mathbf{u}$ as

$$\begin{aligned} \Sigma &= E\left[\left(\frac{\partial f}{\partial \mathbf{u}}^{\dagger}(\mathbf{u})\epsilon \right) \left(\frac{\partial f}{\partial \mathbf{u}}^{\dagger}(\mathbf{u})\epsilon \right)^T \right] \\ &= \delta^2 \frac{\partial f}{\partial \mathbf{u}}^{\dagger}(\mathbf{u}) \left(\frac{\partial f}{\partial \mathbf{u}}^{\dagger}(\mathbf{u}) \right)^T \\ &= \delta^2 \left(\frac{\partial f}{\partial \mathbf{u}}^{T}(\mathbf{u}) \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u}) \right)^{-1}. \end{aligned} \tag{11}$$

The mean squared error across all entries in the recovered modes $\mathbf{u} + \Delta\mathbf{u}$ is equal to the trace of this matrix over the number of unknowns. Specifically

$$\begin{aligned} MSE &= \frac{\delta^2}{2JN} \text{Trace} \left[\left(\frac{\partial f}{\partial \mathbf{u}}^{T}(\mathbf{u}) \frac{\partial f}{\partial \mathbf{u}}(\mathbf{u}) \right)^{-1} \right] \\ &= \frac{\delta^2}{2JN} \sum_{i=1}^{2JN} \frac{1}{\sigma_i^2}, \end{aligned} \tag{12}$$

where σ_i is the i^{th} singular value of the Jacobian $\frac{\partial f}{\partial \mathbf{u}}(\mathbf{u})$. If singular values are large, then this sum will be small, and $\mathbf{u} + \Delta\mathbf{u}$ will be more robust to noise.

4.2 Field-Sequential CDP

In the field-sequential case, $\frac{\partial f}{\partial \mathbf{u}}(\mathbf{u})$ is a block-diagonal matrix, consisting of the Jacobians $\frac{\partial f_j}{\partial u_j}(u_j)$, where f_j is the composition of SLM modulation, propagation, and conversion of the complex field to intensity for jth mode. Each mode is modulated by K/J patterns, which is encapsulated by the forward operator f_j. In particular, the matrix takes the form

$$\begin{bmatrix} \frac{\partial f_1}{\partial u_1}(u_1) & 0 & \ldots & 0 \\ 0 & \frac{\partial f_2}{\partial u_2}(u_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{\partial f_J}{\partial u_J}(u_J) \end{bmatrix} \tag{13}$$

4.3 Mixed-State CDP

On the other hand, for mixed-state CDP, $\frac{\partial f}{\partial \mathbf{u}}(\mathbf{u})$ is no longer block-diagonal. Each mode j contributes to each measurement k via f_j^k, where $f_j^k(u_j) = |P_z^j(\Phi_{SLM}^{j,k} \odot u_j)|^2$ is the composition of SLM modulation, propagation, and conversion of the complex field to intensity for the j^{th} mode and the k^{th} SLM pattern. The matrix takes the form

$$\begin{bmatrix} \frac{\partial f_1^1}{\partial u_1}(u_1) & \frac{\partial f_1^2}{\partial u_2}(u_2) & \dots & \frac{\partial f_1^J}{\partial u_J}(u_J) \\ \frac{\partial f_2^1}{\partial u_1}(u_1) & \frac{\partial f_2^2}{\partial u_2}(u_2) & \dots & \frac{\partial f_2^J}{\partial u_J}(u_J) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_K^1}{\partial u_1}(u_1) & \frac{\partial f_K^2}{\partial u_2}(u_2) & \dots & \frac{\partial f_K^J}{\partial u_J}(u_J) \end{bmatrix} \tag{14}$$

This mixed-state matrix has Frobenius norm approximately $\sqrt{J}$ times the magnitude of the norm for the field-sequential CDP matrix, since for every measurement the Jacobians for all J modes are present. This suggests that the mixed-state CDP matrix has a larger response, on average, to any input than the field-sequential CDP matrix.

We note that a larger Frobenius norm does not necessarily imply smaller MSE. In particular, there is a danger of Equation (14) collapsing to a singular matrix with rank less than $2JN$. For example, this issue occurs when both (1) the forward operators for $n \geq 2$ modes are the same, and (2) the corresponding modes $u_1, \dots, u_n$ are linearly dependent. Thus, as the forward operators grow closer together, and the difference between the modes shrinks, at least one singular value of the mixed-state matrix will go to 0 and the MSE will increase.

The above suggests that our reconstruction procedure might fail, for instance, in settings involving broadband sources where mode varies continuously as a function of wavelength. In problems involving fewer wavelengths, however, where the wavelengths are sufficiently separated, we hypothesize that the larger response of the mixed-state CDP Jacobian is correlated with larger singular values, leading to a smaller mean-squared error via Equation (12).

4.4 Simulation

We verify this hypothesis empirically, with a simulated imaging system, various amounts of camera read noise, and two settings of shot noise. Specifically, we compare multi-wavelength diffraction imaging for three wavelengths (with mixed-norm prior) to two variants of CDP:

1) CDP using field-sequential color imaging + our mixed-norm prior
2) CDP using a Bayer color filter array + our mixed-norm prior

The simulation models a phase SLM and camera separated by $10\,\text{cm}$, and multi-wavelength illumination with wavelengths $638\,\text{nm}$, $520\,\text{nm}$, $445\,\text{nm}$, similar to our real setup in Figure 2. We create a multi-wavelength phase object from the *Cat* image shown in Fig. 3, where the phase for each wavelength ranges from 0 to 2π and is proportional to the intensity of the R, G, and B color channels in the image. We capture 12 diffraction patterns total for all methods.

As shown in Figure 3, our method strictly outperforms both coherent CDP variants for moderate to high noise read-noise regimes, producing sharper details with less chromatic aberration. It may be the case that adding more modes would further improve image quality, although we suspect that this is a trade-off, as more mixed-states will reduce diffraction pattern contrast. All methods perform comparably under low read-noise.

Additionally, we note that Bayer CDP is not as flexible as mixed-state CDP. While mixed-state CDP typically requires little to no modification of the imaging system to support recovery of multiple modes, the color filter array might need to be tailored to the specific application and wavelengths used. Further, as mentioned previously, Bayer filters may limit the lateral resolution of each recovered wavefront (though this did not appear to be a significant problem in our simulations).

5 Implementation

Given the algorithms and noise analysis described in the previous sections, we build a practical imaging prototype for mixed-state sensing. Below, we discuss implementation details of this system.

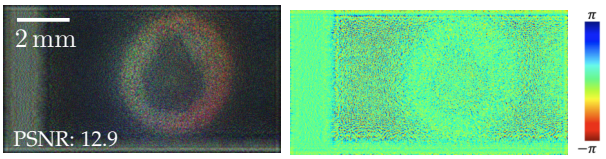
5.1 Hardware

Our prototype system consists of an SLM, a $25\,\text{mm}$ polarizing beamsplitter cube (Thorlabs PBS251), and a monochrome computer vision camera (Allied Vision Technologies Prosilica GT1910) with the bare sensor exposed; see Figure 2. The SLM is a Holoeye GAEA-2, with a spatial resolution of 4160×2464 and a $3.74\,\mu\text{m}$ pixel pitch. We only use an active area of 3000×1600 on the SLM, since parts of the SLM outside of this region contribute little light to the camera. Similar to WISH [18], we generate 320×171 phase patterns with a uniform random sampling of phases between 0 and π (with respect to $532\,\text{nm}$ light), and bilinearly upsample these patterns to the resolution of the SLM. Our camera's images have a maximum resolution of 1920×1080 with $5.5\,\mu\text{m}$ pixel pitch. Note that our camera is relatively low resolution compared to recent CDP setups [21], [18].

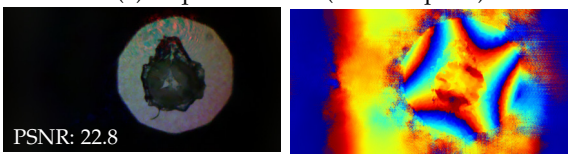
A RGB laser provides illumination at wavelengths $638\,\text{nm}$, $520\,\text{nm}$, and $445\,\text{nm}$, and the laser power at each wavelength is controlled with an Arduino Uno. Additionally, a pinhole spatial filter system (Thorlabs KT310) focuses the laser light with a microscope objective through a pinhole, and a $150\,\text{mm}$ lens (Thorlabs AC254-150-A-ML) collimates the light emerging from the pinhole. The collimated light then provides the back illumination for the sample being imaged.

5.2 Calibration

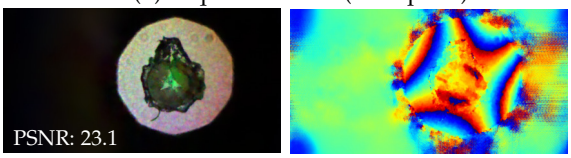
The multi-wavelength diffraction imaging procedures described in Section 3 are sensitive to erroneous calibration. For example, an incorrectly specified propagation distance will lead to an incorrect propagation operator P_z, and can degrade reconstruction. Furthermore, lateral shifts (in x, y) between the SLM and the sensor will manifest as a linear phase gradient applied to the reconstructed field. We follow procedure similar to Peng *et al.* [26] to calibrate our system, whose camera-in-the-loop calibration strategy tends to be robust to error in the initial parameters. In particular, we



(a) Sequential CDP (without prior)



(b) Sequential CDP (with prior)



(c) Mixed-state CDP (with prior)

Fig. 5. **Evaluation of mixed-norm color prior.** To test our proposed mixed-norm color prior, we perform multi-modal gradient descent on three test cases using very few measurements (3 images each). In each case, we refocus the recovered wave to the target plane. Note that the phase images are shown for the green color channel only. **Row 1:** Single-modal phase retrieval. We solve the phase retrieval one color channel at a time *without* our prior. However, because there is only one image per channel, the phase retrieval problem is ill-defined. **Row 2:** Single-modal phase retrieval with mixed-norm prior. Our proposed prior significantly improves reconstruction quality, as information across different wavelengths can be leveraged. **Row 3:** Multi-modal phase retrieval with mixed-norm prior. Measuring the response of multiple wavelengths at the same time encodes more information about the multi-wavelength field, and allows us to recover a higher quality reconstruction of the scene.

calibrate for propagation distance, lateral shift, and beam shape separately for each color. We use gradient descent on all parameters with a set of captured measurements for 64 random SLM patterns.

Because it is difficult to achieve perfect collimation for all wavelengths simultaneously, the shape of the input beam for all colors has a significant effect on image captures. Rather than approximate the beam as a mixture of Gaussians [26], we also solve for the full source incident on the SLM. It may be possible to perform calibration in tandem with sensing [27]; however, we find that it is more practical to calibrate for all parameters of the system once, and then re-use these parameters for all subsequent reconstructions.

5.3 Reconstruction Algorithm

We implement all algorithms using TensorFlow on a desktop computer with an 8-core Intel i7 CPU, 16 GB of RAM, and a single 2080 TI GPU. To avoid memory bottlenecks in our multi-modal gradient descent implementation, we split gradient computations over batches of measurements. Specifically, we compute gradients for each batch and sum the results, which gives the full gradient of the objective in Equation (4). For multi-modal GS, we similarly perform steps 1-4 (1) in batches, in order to avoid storing all high-resolution measurements on the GPU at once.

The propagation operators P_z^j are implemented using the angular spectrum method, which relies on 2D FFT operations to numerically propagate an optical field between two planes. For a more in-depth discussion of the angular spectrum method and numerical methods for wave propagation, we refer readers to Goodman [28] and Schmidt [29].

For all experiments, we run multi-modal GS and multi-modal gradient descent until convergence (usually around 400 iterations). A forward-backward step of multi-modal GS lasts approx. $46\,\mathrm{ms}$ for a single 3200×3200 pixel coherent field. A gradient descent step takes $250\,\mathrm{ms}$ on a coherent field of the same size. Reconstruction time scales approximately linearly with the number of modes and number of measurements. As an example, single-modal GS requires 20 minutes (or 400 iterations) to reconstruct an optical field at one wavelength from 64 measurements, and multi-modal GS requires 1 hour to reconstruct a multi-wavelength field.

6 Results

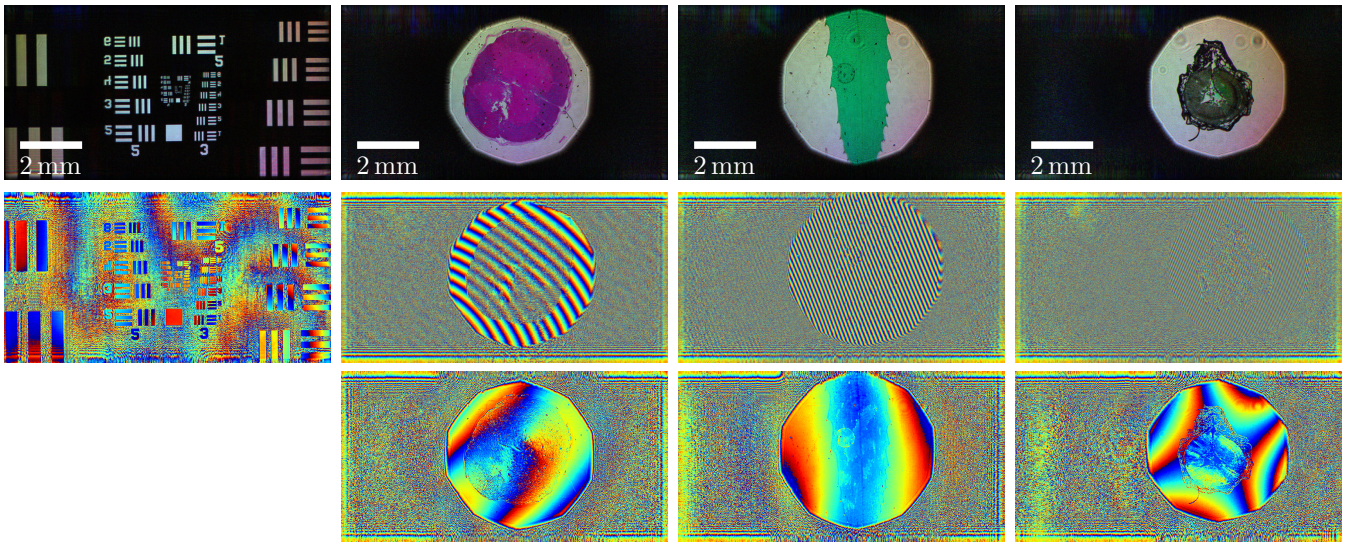
6.1 Multi-Wavelength Diffraction Imaging

Multi-wavelength diffraction imaging is a natural application of our framework, where the individual modes u_j in Equation (3) correspond to different wavelengths of light λ_j.

In this experiment, we use a USAF target and microscope slides as imaging samples. The microscope slides do not have uniform thickness, and thus exhibit steep phase gradients (Figure 6). Phase retrieval algorithms, such as GS [30] and gradient-descent [31] can struggle with high frequency phase. In order to simulate recovery of targets with smoother phase, we determine the linear phase gradient for each slide, and include the phase gradient in the SLM patterns used during the reconstruction process. This removes the undesirable phase gradients from the recovered fields. We also include results for the original, uncorrected fields in Figure 6.

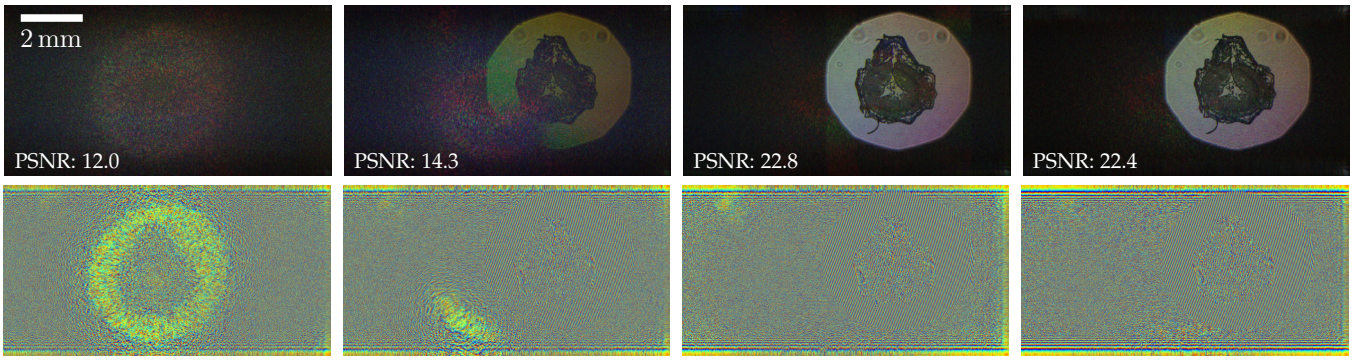
To generate PSNR values, we acquire a set of ground truth fields by solving the standard phase retrieval problem for each color independently, with 192 total measurements (Figure 6). We then compare our reconstruction of color to the standard model with each wavelength captured and reconstructed individually. Note that we compute PSNR for the *re-focused* fields at the target plane. Figure 5 demonstrates the effectiveness of the mixed-norm color prior, and enables our method to recover multi-wavelength fields with very few measurements (*e.g.*, 3 images). Even without this prior, Figure 7 demonstrates that the multi-modal measurements themselves encode more information and enable faster recovery of multi-wavelength fields compared to the baseline field-sequential CDP method. We show a more complete set of results and comparisons for our multi-modal gradient descent procedure in Figure 10, as well as results for our multi-modal Gerchberg Saxon procedure in Figure 11. We finally show results for single-image multi-wavelength reconstruction in Figure 8.

Our method generally converges to a reasonable field with very few measurements. We suspect that for fields with correlated modes, a good solution for a single color pushes the other colors towards good solutions. When the colors have vastly different intensities or contours (for example, when one color is completely blocked by a sample, while another passes through the sample), our method takes longer to converge. Additionally, the baseline field-sequential CDP



(a) USAF resolution target (b) Rabbit spinal cord (c) Hydrilla Verticillata leaf (d) Pine stem

Fig. 6. **Representative ground truth multi-wavelength fields.** We generate high-quality multi-wavelength fields by solving a phase retrieval problem one color at a time. A total of 288 images (96 per color channel) are used to reconstruct a high-quality multi-wavelength field of the USAF target **(a)**, and 192 images (64 per color channel) are used to reconstruct the microscope slides **(b)-(d)**. **Row 1:** Refocused intensity images of the recovered fields. **Row 2:** Phase images of the recovered fields. All phase images are shown for the green color channel only. The microscope slides do not have uniform depth, and contain a steep, approximately linear, phase gradient. We correct for these phase gradients in several of our results, in order to showcase recovery for smoother phase objects. **Row 3:** Phase image of the fields of microscope slides, after accounting for the linear phase gradients. Note the subtle features captured in the phase image of the Hydrilla Verticillata leaf, which are hidden in the intensity image.



(a) Sequential CDP ($K = 24$) (b) Sequential CDP ($K = 36$) (c) Sequential CDP ($K = 48$) (d) Mixed-State CDP ($K = 24$)

Fig. 7. **Comparison between sequential and mixed-state CDP. (a)-(c)** Results of running field-sequential CDP on the original microscope slides *without phase correction*, using 24, 36, and 48 measurements respectively. Each color channel is reconstructed separately using a third of the overall measurements. It is known that GS algorithms can struggle with fields steep phase gradients, as a result, more measurements are required to recover this wavefront accurately. **(d)** Results of running multi-modal GS on only 24 measurements. Even without the use of the mixed-norm color prior, our reconstruction procedure generates better reconstructions with fewer measurements than field-sequential CDP, suggesting that every measurement encodes more information about the unknown field.

performs better than our method in some cases (especially when the samples have smooth phase), but struggles when the sample has high frequency phase.

6.2 Spatially Varying OPD Measurement

Multi-wavelength wavefront sensing has recently been used for micron-scale depth ranging [13], [14], [15]. In previous work, however, the modes for different wavelengths are recovered one at a time. Here, we experiment with objects that produce spatially-varying optical path differences (OPDs) in the path of the color laser, and perform our mixed-state CDP approach for 3 wavelength recovery as before.

If the illuminating beam is approximately planar, then in a region close to the sample, the complex field due to a single wavelength λ can be modeled as

$$O_\lambda(x, y) \propto \exp\left(j2\pi \frac{d(x, y)}{\lambda}\right), \tag{15}$$

where $d(x, y)$ is the OPD for lateral coordinates x, y. The phase of the complex exponential in the equation above is proportional to optical path length. However, the challenge in trying to recover distance is that it wraps on the order of a single wavelength.

Wu *et al.* [13] solve this problem by combining the complex fields at 2 wavelengths to create a larger "synthetic" wavelength, which can be used to unwrap optical path lengths. Here, we provide a simple extension that combines ≥ 2 wavelengths, similar to [14], [15] . In particular, consider

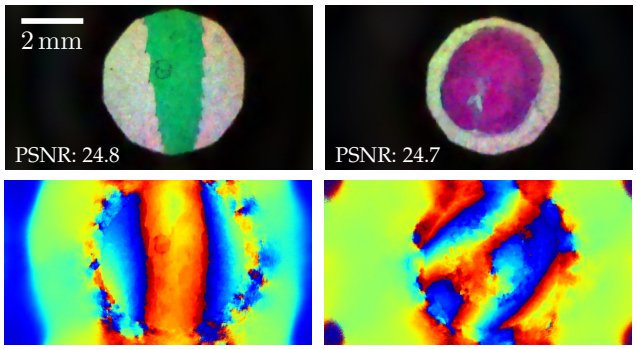
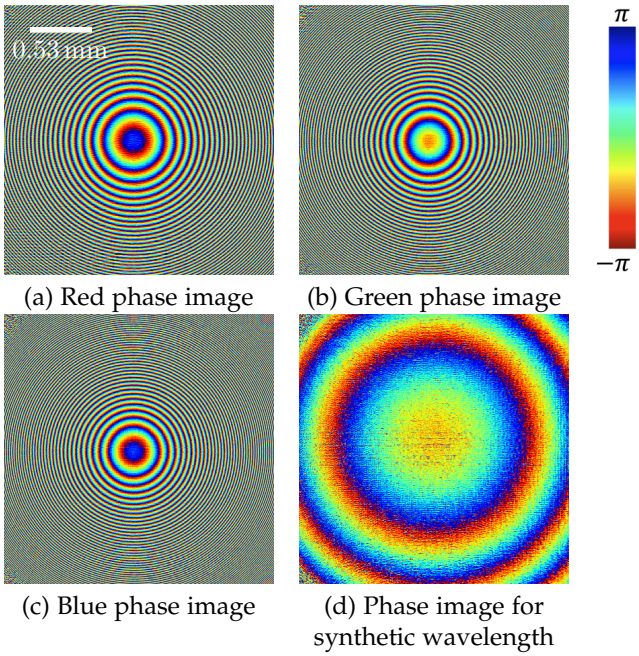


Fig. 8. **Single-shot mixed-state CDP with mixed-norm prior.** Only one image is used to reconstruct the entire color wavefront. Some detail is lost in single-image recovery, but dominant edges in amplitude and phase are preserved. We suspect that it is possible to improve reconstruction quality with stronger priors, by exploiting temporal information or employing optical components which further distinguish the forward operators for each mode (such as diffraction gratings).



(a) Red phase image (b) Green phase image

(c) Blue phase image (d) Phase image for synthetic wavelength

Fig. 9. **OPD measurement results.** A $150\,\mathrm{mm}$ lens, reconstructed from 32 measurements. **(a)-(c)** Phase from multi-modal GS for red, green, and blue light. **(d)** The phase for a synthetic wavelength of $31.7\,\mu\mathrm{m}$.

J wavelengths $(\lambda_1, \ldots, \lambda_J)$, and the product:

$$\prod_{j=1}^{J} O_{\lambda_j}(x,y)^{c_j} \propto \exp\left(j2\pi d(x,y) \sum_{j=1}^{J} \frac{c_j}{\lambda_j}\right) \tag{16}$$

where c_j are arbitrary integer values. The effective synthetic wavelength in this case is $\Lambda = (\sum_{j=1}^{J} \frac{c_j}{\lambda_j})^{-1}$, where larger synthetic wavelengths correspond to less phase wrapping. Therefore, in order to maximize range, we can maximize the wavelength Λ. A challenge with this general formulation is that every time we take a product with one mode, we add a small amount of phase noise. Therefore, choosing a good set of c_j amounts of trading off between range and the magnitude of the coefficients c_j.

In our case, we have $\lambda_1 = 638\,\mathrm{nm}$, $\lambda_2 = 520\,\mathrm{nm}$, and $\lambda_3 = 445\,\mathrm{nm}$. To find c_1, c_2, c_3, we perform a brute force search for the largest synthetic wavelength with $-5 \leq c_j \leq 5$. This yields $c_1 = -1, c_2 = 2, c_3 = -1$ and a synthetic wavelength of $31.7\,\mu\mathrm{m}$, $50\times$ the largest optical wavelength. Using this synthetic wavelength, we show the partially unwrapped surface of a lens in Figure 9. We note that because the synthetic wavelength is still quite small, the recovered OPDs remain wrapped.

7 Limitations and Conclusion

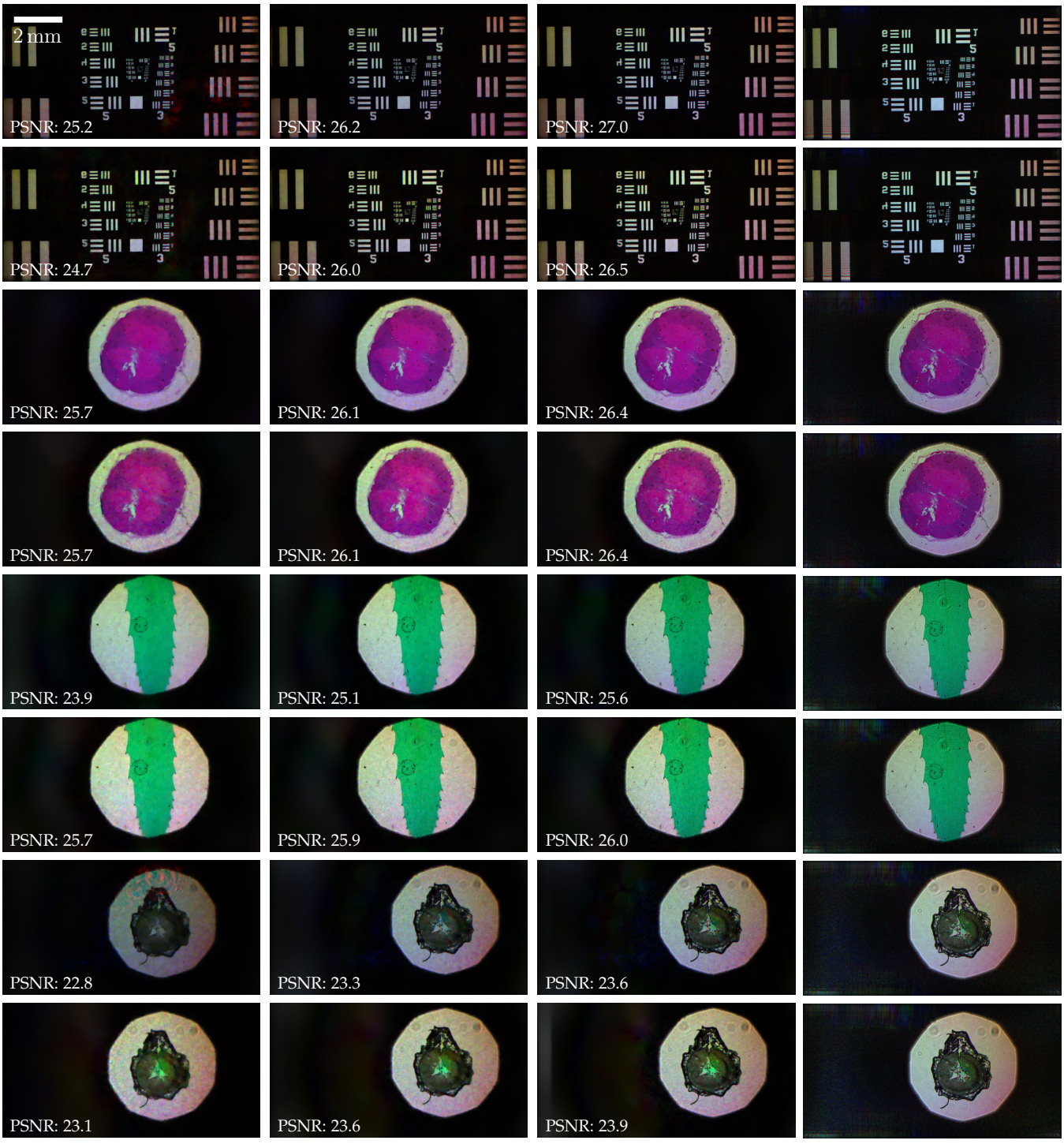
In this work, we proposed a framework for mixed-state coded diffraction imaging. We presented a set of robust mixed-state phase retrieval algorithms, analyzed their performance under various amounts of signal independent read noise, and implemented a practical mixed-state diffraction imaging sensing prototype. Using this prototype, we demonstrated the two example applications of multi-wavelength diffraction imaging and spatially varying OPD measurement.

One limitation of all mixed-state methods is that, without additional assumptions, they can only recover mixed-states uniquely if the forward operators acting on each state are distinct (Section 3.2). Further, we provide novel noise analysis which suggests that mixed-state methods fail when the forward operators for each mode, and the modes themselves, grow closer together (Section 4.3). While we show empirically that this is not a problem for multi-wavelength diffraction imaging when the wavelengths are sufficiently separated (Section 4.4), it could be a challenge for other potential applications. However, for these applications it may be possible to introduce additional optical components (diffraction gratings, birefrigent optics, spinning diffusers), which further "separate" the forward operators for each state. In addition, for problems involving broadband illumination where the source varies continuously with wavelength, one may be able to introduce additional compressive priors to make reconstruction tractable.

While spatially varying OPD measurement is an interesting application of our framework, larger unwrapped ranges are required for most practical applications. However, although the experimental results for OPD measurement presented in this paper are limited, with more robust calibration, and some tuning of the each wavelength, it should be possible to achieve far larger synthetic wavelengths that are less affected by noise [13]. For example, this can be achieved using two (or more) wavelengths, where one wavelength is close to an integer multiple of another.

Despite these limitations, we believe that the algorithms, imaging prototype, and analysis presented here are important tools that can help make mixed-state CDP more practically useful. In particular, a novel insight of our work is that, not only is mixed-state CDP *possible* given enough measurements, but it is *preferable* in high-noise regimes. Combining more than three modes may further improve performance, though at the cost of increased demands on computational resources. As one increases the number of modes, it may also be possible to investigate optimal multiplexing for mixed-state CDP similar to Schechner *et al.* [25], an interesting direction for future work.

Code and data relevant to this paper are available on our project website: https://imaging.cs.cmu.edu/mixedstate/



(a) 3 images (1 per channel) (b) 6 images (2 per channel) (c) 12 images (4 per channel) (d) Ground truth

Fig. 10. **Comparison between sequential CDP (rows 1, 3, & 5) and mixed-state CDP (rows 2, 4, & 6) for gradient descent with priors.** In all cases, the color fields converge to a reasonable solution when using the smallest number of measurements used (*i.e.*, 3 measurements). Both approaches work reasonably well with our prior, though multi-modal gradient descent performs better.

Acknowledgments

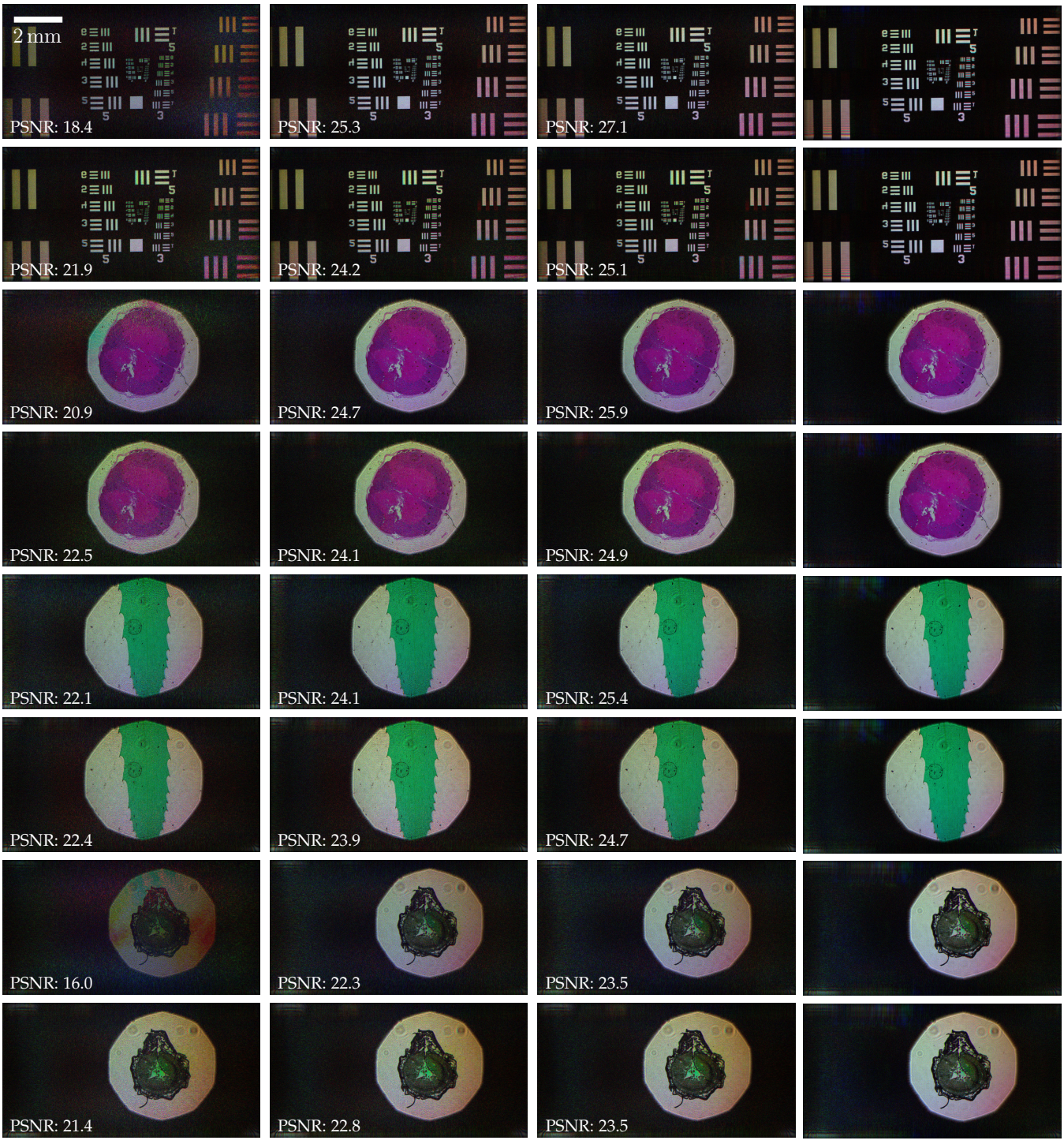
We thank the following for helpful discussions: Wei-Yu Chen, Anat Levin, Aswin Sankaranarayanan, Gordon Wetzstein, and the anonymous reviewers. The authors acknowledge the support of an Uber Fellowship, and NSF Grants IIS-1900821 and IIS-2008464.

References

[1] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, no. 6742, pp. 342–344, 1999.

[2] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[3] F. Zhang, G. Pedrini, and W. Osten, "Phase retrieval of arbitrary complex-valued fields through aperture-plane modulation," *Physical Review A*, vol. 75, no. 4, p. 043805, 2007.



(a) 24 images (8 per channel) (b) 36 images (12 per channel) (c) 48 images (16 per channel) (d) Ground truth

Fig. 11. **Comparison between sequential CDP (rows 1, 3, & 5) and mixed-state CDP (rows 2, 4, & 6) when using the GS algorithm.** No priors are used. In all cases, color fields converge to a reasonable solution, though requires more measurements to produce results comparable to those constructed with the mixed-norm color prior.

[4] F. Zhang and J. Rodenburg, "Phase retrieval based on wave-front relay and modulation," *Physical Review B*, vol. 82, no. 12, p. 121104, 2010.

[5] F. Zhang, B. Chen, G. R. Morrison, J. Vila-Comamala, M. Guizar-Sicairos, and I. K. Robinson, "Phase retrieval by coherent modulation imaging," *Nature communications*, vol. 7, no. 1, pp. 1–8, 2016.

[6] F. Pfeiffer, "X-ray ptychography," *Nature Photonics*, vol. 12, no. 1, pp. 9–17, 2018.

[7] W. Hoppe, "Beugung im inhomogenen primärstrahlwellenfeld. i. prinzip einer phasenmessung von elektronenbeungungsinterferenzen," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 25, no. 4, pp. 495–501, 1969.

[8] R. Hegerl and W. Hoppe, "Dynamische theorie der kristallstrukturanalyse durch elektronenbeugung im inhomogenen primärstrahlwellenfeld," *Berichte der Bunsengesellschaft für physikalische Chemie*, vol. 74, no. 11, pp. 1148–1154, 1970.

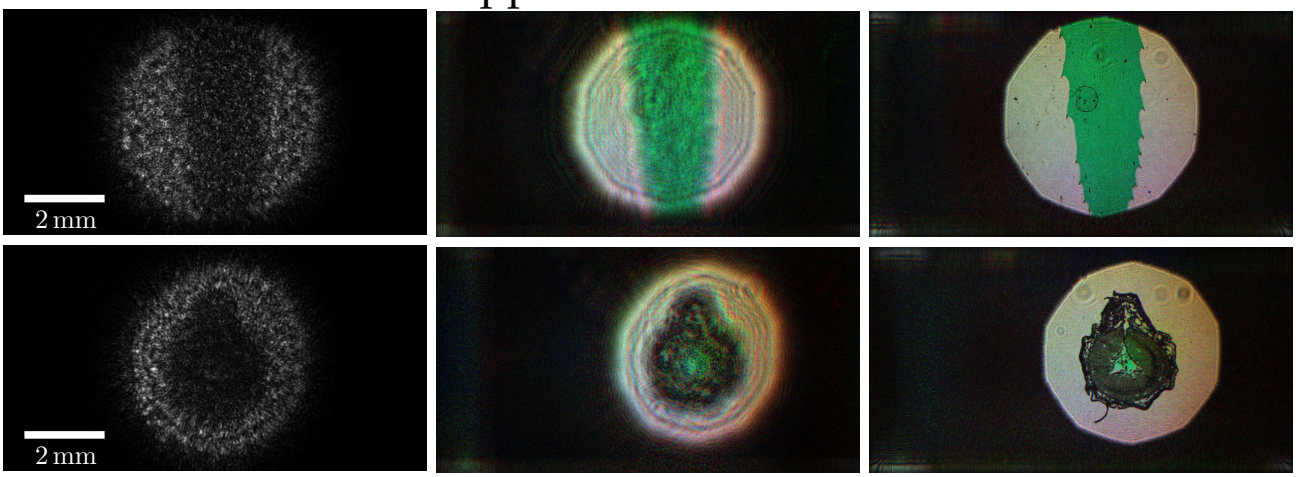
[9] J. Rodenburg, A. Hurst, and A. Cullis, "Transmission microscopy without lenses for objects of unlimited size," *Ultramicroscopy*, vol. 107, no. 2-3, pp. 227–231, 2007.

[10] H. M. L. Faulkner and J. Rodenburg, "Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm," *Physical review letters*, vol. 93, no. 2, p. 023903, 2004.
[11] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution fourier ptychographic microscopy," *Nature photonics*, vol. 7, no. 9, p. 739, 2013.
[12] P. Thibault and A. Menzel, "Reconstructing state mixtures from diffraction measurements," *Nature*, vol. 494, no. 7435, pp. 68–71, 2013.
[13] Y. Wu, F. Li, F. Willomitzer, A. Veeraraghavan, and O. Cossairt, "Wished: Wavefront imaging sensor with high resolution and depth ranging," in *2020 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2020, pp. 1–10.
[14] F. Li, F. Willomitzer, M. M. Balaji, P. Rangarajan, and O. Cossairt, "Exploiting wavelength diversity for high resolution time-of-flight 3d imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2193–2205, 2021.
[15] F. Willomitzer, P. V. Rangarajan, F. Li, M. M. Balaji, M. P. Christensen, and O. Cossairt, "Fast non-line-of-sight imaging with high-resolution and wide field of view using synthetic wavelength holography," *Nature communications*, vol. 12, no. 1, pp. 1–11, 2021.
[16] T. Latychevskaia, "Holography and coherent diffraction imaging with low-(30–250 ev) and high-(80–300 kev) energy electrons: History, principles, and recent trends," *Materials*, vol. 13, no. 14, p. 3089, 2020.
[17] M. Li, L. Bian, and J. Zhang, "Coded coherent diffraction imaging with reduced binary modulations and low-dynamic-range detection," *Optics Letters*, vol. 45, no. 16, pp. 4373–4376, 2020.
[18] Y. Wu, M. K. Sharma, and A. Veeraraghavan, "Wish: wavefront imaging sensor with high resolution," *Light: Science & Applications*, vol. 8, no. 1, pp. 1–10, 2019.
[19] E. Malm, E. Fohtung, and A. Mikkelsen, "Multi-wavelength phase retrieval for coherent diffractive imaging," *Optics Letters*, vol. 46, no. 1, pp. 13–16, 2021.
[20] Y. Yao, Y. Jiang, J. Klug, Y. Nashed, C. Roehrig, C. Preissner, F. Marin, M. Wojcik, O. Cossairt, Z. Cai *et al.*, "Broadband x-ray ptychography using multi-wavelength algorithm," *Journal of Synchrotron Radiation*, vol. 28, no. 1, pp. 309–317, 2021.
[21] X. Dong, X. Pan, C. Liu, and J. Zhu, "Single shot multi-wavelength phase retrieval with coherent modulation imaging," *Optics Letters*, vol. 43, no. 8, pp. 1762–1765, 2018.
[22] Y. Gao and L. Cao, "Generalized optimization framework for pixel super-resolution imaging in digital holography," *Optics Express*, vol. 29, no. 18, pp. 28 805–28 823, 2021.
[23] Y. Li, C. Ma, Y. Chen, and Y. Chi, "Nonconvex matrix factorization from rank-one measurements," *arXiv preprint arXiv:1802.06286*, 2018.
[24] E. Van den Berg and M. P. Friedlander, "Sparse optimization with least-squares constraints," *SIAM Journal on Optimization*, vol. 21, no. 4, pp. 1201–1229, 2011.
[25] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "Multiplexing for optimal lighting," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 8, pp. 1339–1354, 2007.
[26] Y. Peng, S. Choi, N. Padmanaban, J. Kim, and G. Wetzstein, "Neural holography," in *ACM SIGGRAPH 2020 Emerging Technologies*, 2020, pp. 1–2.
[27] E. Bostan, R. Heckel, M. Chen, M. Kellman, and L. Waller, "Deep phase decoder: Self-calibrating phase microscopy with an untrained deep neural network," *arXiv preprint arXiv:2001.09803*, 2020.
[28] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.
[29] J. Schmidt, "Numerical simulation of optical wave propagation with examples in matlab." Society of Photo-Optical Instrumentation Engineers, 2010.
[30] G.-z. Yang, B.-z. Dong, B.-y. Gu, J.-y. Zhuang, and O. K. Ersoy, "Gerchberg–saxton and yang–gu algorithms for phase retrieval in a nonunitary transform system: a comparison," *Applied optics*, vol. 33, no. 2, pp. 209–218, 1994.
[31] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 277–299, 2015.

Benjamin Attal is a Ph.D. candidate at the Carnegie Mellon University (CMU) Robotics Institute advised by Professor Matthew O'Toole. He received his B.S./M.S from Brown University in 2018/19. He is a recipient of the Uber Fellowship at Carnegie Mellon University in 2021. His current interests include computational photography, wavefront sensing, and neural representations for inverse rendering.

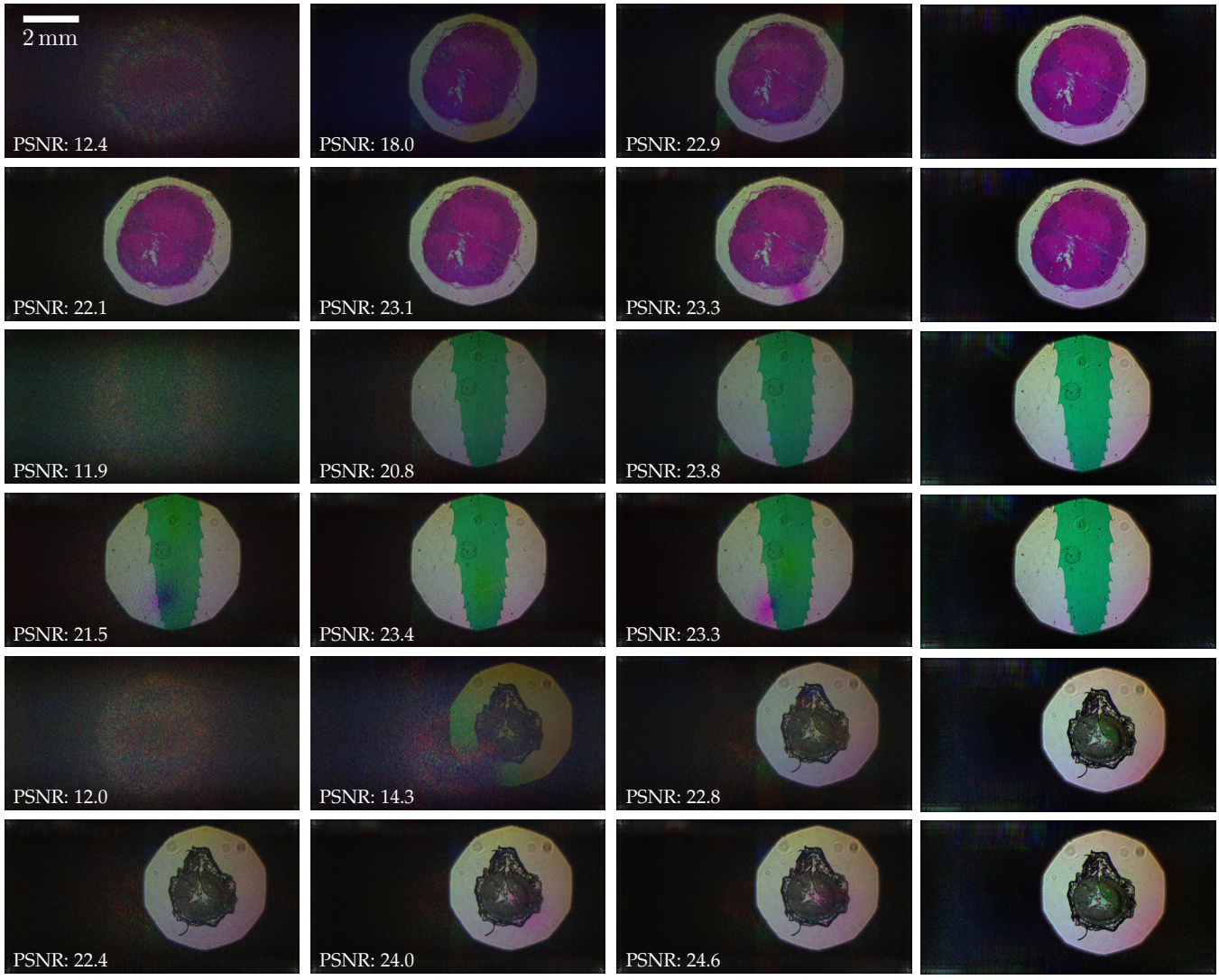
Matthew O'Toole is an Assistant Professor with the Robotics Institute and Computer Science Department at Carnegie Mellon University (CMU). He earned his Ph.D. from the University of Toronto and was previously a Banting postdoctoral fellow with the Computational Imaging group at Stanford University. He is the recipient of the ACM SIGGRAPH 2017 Outstanding Thesis Honorable Mention award, and runnerup best paper awards at CVPR 2022, CVPR 2014, and ICCV 2007.

Supplemental Results



(a) Intensity measurement (b) Recovered wavefront (c) Refocused wavefront

Fig. 12. **High-quality reconstruction results. (a)** Example of a grayscale measurement, consisting of a superposition of diffraction intensities for different modes. **(b)** Recovered color field at the SLM plane. We employ multi-modal GS to recover these fields using 48 images (16 per channel). **(c)** Color field refocused by numerically propagating wave to target plane.



(a) 24 images (8 per channel) (b) 36 images (12 per channel) (c) 48 images (16 per channel) (d) Ground truth

Fig. 13. **Comparison between sequential CDP (rows 1, 3, & 5) and mixed-state CDP (rows 2, 4, & 6) using GS.** No priors and no phase correction are used. Multi-modal reconstructions are typically of higher quality than the single-modal reconstructions in most of the above examples.