Measurement Matrix Design for Phase Retrieval Based on Mutual Information

Nir Shlezinger, *Member, IEEE*, Ron Dabora, *Senior Member, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—In phase retrieval problems, a signal of interest (SOI) is reconstructed based on the magnitude of a linear transformation of the SOI observed with additive noise. The linear transform is typically referred to as a measurement matrix. Many works on phase retrieval assume that the measurement matrix is a random Gaussian matrix, which, in the noiseless scenario with sufficiently many measurements, guarantees invertability of the transformation between the SOI and the observations, up to an inherent phase ambiguity. However, in many practical applications, the measurement matrix corresponds to an underlying physical setup, and is therefore deterministic, possibly with structural constraints. In this paper, we study the design of deterministic measurement matrices, based on maximizing the mutual information between the SOI and the observations. We characterize necessary conditions for the optimality of a measurement matrix, and analytically obtain the optimal matrix in the low signal-to-noise ratio regime. Practical methods for designing general measurement matrices and masked Fourier measurements are proposed. Simulation tests demonstrate the performance gain achieved by the suggested techniques compared to random Gaussian measurements for various phase recovery algorithms.**

***Index Terms*—Phase retrieval, measurement matrix design, mutual information, masked Fourier.**

I. Introduction

In a wide range of practical scenarios, including X-ray crystallography [1], diffraction imaging [2], astronomical imaging [3], and microscopy [4], a signal of interest (SOI) needs to be reconstructed from observations which consist of the magnitudes of its linear transformation with additive noise. This class of signal recovery problems is commonly referred to as *phase retrieval* [5]. In a typical phase retrieval setup, the SOI is first projected using a *measurement matrix* specifically designed for the considered setup. The observations are then obtained as noisy versions of the magnitudes of these projections. Recovery algorithms for phase retrieval received much research attention in recent years. Major approaches for designing phase retrieval algorithms include alternating minimization techniques [6], [7], methods based on convex relaxation, such as *phaselift* [8] and *phasecut* [9], and non-convex algorithms with a suitable initialization, such as *Wirtinger flow* [10], and *truncated amplitude flow (TAF)* [11].

The problem of designing the measurement matrix received considerably less attention compared to the design of phase retrieval algorithms. An important desirable property that measurement matrices should satisfy is a unique relationship between the signal and the magnitudes of its projections, up to an inherent phase ambiguity. In many works, particularly in theoretical performance analysis of phase retrieval algorithms [8], [10], [12], the matrices are assumed to be *random*, commonly with i.i.d. Gaussian entries. However, in practical applications, the measurement matrix corresponds to a fixed physical setup, so that it is typically a deterministic matrix, with possibly structural constraints. For example, in optical imaging, lenses are modeled using discrete Fourier transform (DFT) matrices and optical masks correspond to diagonal matrices [13]. Measurements based on oversampled DFT matrices were studied in [14], measurement matrices which correspond to the parallel application of several DFTs to modulated versions of the SOI were proposed in [8], and [15], [16] studied phase recovery using fixed binary measurement matrices, representing hardware limitations in optical imaging systems.

All the works above considered *noiseless observations*, hence, the focus was on obtaining uniqueness of the magnitudes of the projections in order to guarantee recovery, though the recovery method may be intractable [17]. When noise is present, such uniqueness no longer guarantees recovery, thus a different design criterion should be considered. Recovery algorithms as well as specialized deterministic measurement matrices were considered in several works. In particular, [18], [19] studied phase recovery from short-time Fourier transform measurements, [20] proposed a recovery algorithm and measurement matrix design based on sparse graph codes for sparse SOIs taking values on a finite set, [21] suggested an algorithm using correlation based measurements for flat SOIs, i.e., strictly non-sparse SOIs, and [22] studied recovery methods and the corresponding measurement matrix design for the noisy phase retrieval setup by representing the projections as complex polynomials.

A natural optimality condition for the noisy setup, without focusing on a specific recovery algorithm, is to design the

Manuscript received April 26, 2017; revised September 19, 2017; accepted September 19, 2017. Date of publication October 2, 2017; date of current version December 4, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Subhrakanti Dey. The work of R. Dabora is supported in part by the Israel Science Foundation under Grant 1685/16. This paper was presented in part at the 2017 International Symposium on Information Theory. *(Corresponding author: Nir Shlezinger.)*

N. Shlezinger and Y. C. Eldar are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: nirshlezinge@technion.ac.il; yonina@ee.technion.ac.il).

R. Dabora is with the Department of Electrical and Computer Engineering, Ben-Gurion University, Be'er-Sheva 8410501, Israel (e-mail: ron@ee.bgu.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2017.2759101

measurement matrix to minimize the achievable mean-squared error (MSE) in estimating the SOI from the observations. However, in phase retrieval, the SOI and observations are not jointly Gaussian, which makes computing the minimum MSE (MMSE) for a given measurement matrix in the vector setting very difficult. Furthermore, even in the linear non-Gaussian setting, a closed-form expression for the derivative of the MMSE exists only for the scalar case [23], which corresponds to a single observation. Therefore, gradient-based approaches for MMSE optimization are difficult to apply as well.

In this work we propose an alternative design criterion for the measurement matrix based on maximizing the mutual information (MI) between the observations and the SOI. MI is a statistical measure which quantifies the "amount of information" that one random variable (RV) "contains" about another RV [24, Ch. 2.3]. Thus, maximizing the MI essentially maximizes the statistical dependence between the observations and the SOI, which is desirable in recovery problems. MI is also related to MMSE estimation in Gaussian noise via its derivative [25], and has been used as the design criterion in several problems, including the design of projection matrices in compressed sensing [26] and the construction of radar waveforms [27], [28].

In order to rigorously express the MI between the observations and the SOI, we adopt a Bayesian framework for the phase retrieval setup, similar to the approach in [29]. Computing the MI between the observations and the SOI is a difficult task. Therefore, to facilitate the analysis, we first restate the phase retrieval setup as a linear multiple input-multiple output (MIMO) channel of extended dimensions with an additive Gaussian noise. In the resulting MIMO setup, the channel matrix is given by the row-wise Khatri-Rao product (KRP) [30] of the measurement matrix and its conjugate, while the channel input is the Kronecker product of the SOI and its conjugate, and is thus non-Gaussian for any SOI distribution. We show that the MI between the observations and the SOI of the original phase retrieval problem is equal to the MI between the input and the output of this MIMO channel. Then, we use that fact that for MIMO channels with additive Gaussian noise, the gradient of the MI can be obtained in closed-form [31] *for any arbitrary input distribution*. We note that a similar derivation cannot be carried out with the MMSE design criterion since: 1) Differently from the MI, the MMSE for the estimation of the SOI based on the original observations is not equal to the MMSE for the estimation of the MIMO channel input based on the output; 2) For the MIMO setup, a closed-form expression for the gradient of the MMSE exists only when the input is Gaussian, yet, the input is non-Gaussian for any SOI distribution due its Kronecker product structure.

Using the equivalent MIMO channel with non-Gaussian input, we derive necessary conditions on the measurement matrix to maximize the MI. We then obtain a closed-form expression for the optimal measurement matrix in the low signal-to-noise ratio (SNR) regime when the SOI distribution satisfies a symmetry property, we refer to as Kronecker symmetry, exhibited by, e.g., the zero-mean proper-complex (PC) Gaussian distribution. Next, we propose a practical measurement matrix design by approximating the matrix which maximizes the MI for any arbitrary SNR. In our approach, we first maximize the MI of a MIMO channel, derived from the phase retrieval setup, after relaxing the structure restrictions on the channel matrix imposed by the phase retrieval problem. We then find the measurement matrix for which the resulting MIMO channel matrix (i.e., the channel matrix which satisfies the row-wise KRP structure) is closest to the MI maximizing channel matrix obtained without the structure restriction. With this approach, we obtain closed-form expressions for general (i.e., structureless) measurement matrices, as well as for constrained settings corresponding to masked Fourier matrices, representing, e.g., optical lenses and masks. The substantial benefits of the proposed design framework are clearly illustrated in a simulations study. In particular, we show that our suggested practical design improves the performance of various recovery algorithms compared to using random measurement matrices.

The rest of this paper is organized as follows: Section II formulates the problem. Section III characterizes necessary conditions on the measurement matrix which maximizes the MI, and studies its design in the low SNR regime. Section IV presents the proposed approach for designing practical measurement matrices, and Section V illustrates the performance of our design in simulation examples. Finally, Section VI concludes the paper. Proofs of the results stated in the paper are provided in the appendix.

II. Problem Formulation

A. Notations

We use upper-case letters to denote RVs, e.g., X, lower-case letters for deterministic variables, e.g., x, and calligraphic letters to denote sets, e.g., $\mathcal{X}$. We denote column vectors with boldface letters, e.g., $\mathbf{x}$ for a deterministic vector and $\mathbf{X}$ for a random vector; the i-th element of $\mathbf{x}$ is written as $(\mathbf{x})_i$. Matrices are represented by double-stroke letters, e.g., $\mathbb{M}$, $(\mathbb{M})_{i,j}$ is the (i,j)-th element of $\mathbb{M}$, and $\mathbb{I}_n$ is the $n \times n$ identity matrix. Hermitian transpose, transpose, complex conjugate, real part, imaginary part, stochastic expectation, and MI are denoted by $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^*$, $\mathrm{Re}\{\cdot\}$, $\mathrm{Im}\{\cdot\}$, $\mathcal{E}\{\cdot\}$, and $I(\cdot\,;\cdot)$, respectively. $\mathrm{Tr}(\cdot)$ denotes the trace operator, $\|\cdot\|$ is the Euclidean norm when applied to vectors and the Frobenius norm when applied to matrices, $\otimes$ denotes the Kronecker product, $\delta_{k,l}$ is the Kronecker delta function, i.e., $\delta_{k,l}=1$ when $k=l$ and $\delta_{k,l}=0$ otherwise, and $a^+ \triangleq \max(0,a)$. For an $n \times 1$ vector $\mathbf{x}$, $\mathrm{diag}(\mathbf{x})$ is the $n \times n$ diagonal matrix whose diagonal entries are the elements of $\mathbf{x}$, i.e., $(\mathrm{diag}(\mathbf{x}))_{i,i} = (\mathbf{x})_i$. The sets of real and of complex numbers are denoted by $\mathcal{R}$ and $\mathcal{C}$, respectively. Finally, for an $n \times n$ matrix $\mathbb{X}$, $\mathbf{x} = \mathrm{vec}(\mathbb{X})$ is the $n^2 \times 1$ column vector obtained by stacking the columns of $\mathbb{X}$ one below the other. The $n \times n$ matrix $\mathbb{X}$ is recovered from $\mathbf{x}$ via $\mathbb{X} = \mathrm{vec}_n^{-1}(\mathbf{x})$.

B. The Phase Retrieval Setup

We consider the recovery of a random SOI $\mathbf{U} \in \mathcal{C}^n$, from an observation vector $\mathbf{Y} \in \mathcal{R}^m$. Let $\mathbb{A} \in \mathcal{C}^{m \times n}$ be the measurement matrix and $\mathbf{W} \in \mathcal{R}^m$ be the additive noise, modeled as a zero-mean real-valued Gaussian vector with covariance matrix

$\sigma_W^2 \mathbb{I}_m$, $\sigma_W^2 > 0$. As in [12, Eq. (1.5)], [14, Eq. (1)], and [17, Eq. (1.1)], the relationship between $\mathbf{U}$ and $\mathbf{Y}$ is given by:

$$\mathbf{Y} = |\mathbb{A}\mathbf{U}|^2 + \mathbf{W}, \tag{1}$$

where $|\mathbb{A}\mathbf{U}|^2$ denotes *the element-wise squared magnitude*. Since for every $\theta \in \mathcal{R}$, the vectors $\mathbf{U}$ and $\mathbf{U}e^{j\theta}$ result in the same $\mathbf{Y}$, the vector $\mathbf{U}$ can be recovered only up to a global phase.

In this work we study the design of $\mathbb{A}$ aimed at maximizing the MI between the SOI and the observations. Letting $f(\mathbf{u}, \mathbf{y})$ be the joint probability density function (PDF) of $\mathbf{U}$ and $\mathbf{Y}$, $f(\mathbf{u})$ the PDF of $\mathbf{U}$, and $f(\mathbf{y})$ the PDF of $\mathbf{Y}$, the MI between the SOI $\mathbf{U}$ and the observations $\mathbf{Y}$ is given by [24, Ch. 8.5]

$$I(\mathbf{U}; \mathbf{Y}) \triangleq \mathcal{E}_{\mathbf{U},\mathbf{Y}} \left\{ \log \frac{f(\mathbf{U}, \mathbf{Y})}{f(\mathbf{U}) f(\mathbf{Y})} \right\}. \tag{2}$$

Specifically, we study the measurement matrix $\mathbb{A}^{\mathrm{MI}}$ which maximizes[1] the MI for a fixed arbitrary distribution of $\mathbf{U}$, subject to a Frobenious norm constraint $P > 0$, namely,

$$\mathbb{A}^{\mathrm{MI}} = \underset{\mathbb{A} \in \mathcal{C}^{m \times n} : \mathrm{Tr}\left(\mathbb{A}\mathbb{A}^H\right) \leq P}{\arg\max} I(\mathbf{U}; \mathbf{Y}), \tag{3}$$

where $\mathbf{U}$ and $\mathbf{Y}$ are related via (1). In the noiseless non-Bayesian phase retrieval setup, it has been shown that a necessary and sufficient condition for the existence of a bijective mapping from $\mathbf{U}$ to $\mathbf{Y}$ is that the number of observations, m, is linearly related to the dimensions of the SOI[2], n, see [32], [33]. Therefore, we focus on values of m satisfying $n \leq m \leq n^2$.

As discussed in the introduction, in practical scenarios, the structure of the measurement matrix is often constrained. One type of structural constraint commonly encountered in practice is the masked Fourier structure, which arises, for example, when the measurement matrix represents an optical setup consisting of lenses and masks [13], [20]. In this case, $\mathbf{Y}$ is obtained by projecting $\mathbf{U}$ via b optical masks, each modeled as an $n \times n$ diagonal matrix $\mathbb{G}_l$, $l \in \{1, 2, \ldots, b\} \triangleq \mathcal{B}$, followed by an optical lens, modeled as a DFT matrix of size n, denoted $\mathbb{F}_n$ [20, Sec. 3]. Consequently, $m = b \cdot n$ and $\mathbb{A}$ is obtained as

$$\mathbb{A} = \begin{bmatrix} \mathbb{F}_n \mathbb{G}_1 \\ \mathbb{F}_n \mathbb{G}_2 \\ \vdots \\ \mathbb{F}_n \mathbb{G}_b \end{bmatrix} = (\mathbb{I}_b \otimes \mathbb{F}_n) \begin{bmatrix} \mathbb{G}_1 \\ \mathbb{G}_2 \\ \vdots \\ \mathbb{G}_b \end{bmatrix}. \tag{4}$$

Since $n \leq m \leq n^2$, we focus on $1 \leq b \leq n$. In the following sections we study the optimal design of general (unconstrained) measurement matrices, and propose a practical algorithm for designing both general measurement matrices as well as masked Fourier measurement matrices.

III. Optimal Measurement Matrix

In this section we first show that the relationship (1) can be equivalently represented (in the sense of having the same MI) as a MIMO channel with PC Gaussian noise. Then, we use the equivalent representation to study the design of measurement matrices for two cases: The first considers an arbitrary SOI distribution, for which we characterize a necessary condition on the optimal measurement matrix. The second case treats an SOI distribution satisfying a symmetry property (exhibited by, e.g., zero-mean PC Gaussian distributions) focusing on the low SNR regime, for which we obtain the optimal measurement matrix in closed-form.

A. Gaussian MIMO Channel Interpretation

In order to characterize the solution of (3), we first consider the relationship (1): Note that for every $p \in \{1, 2, \ldots, m\} \triangleq \mathcal{M}$, the p-th entry of $|\mathbb{A}\mathbf{U}|^2$ can be written as

$$\left(|\mathbb{A}\mathbf{U}|^2\right)_p = \sum_{k=1}^{n} \sum_{l=1}^{n} (\mathbb{A})_{p,k} (\mathbb{A})_{p,l}^* (\mathbf{U})_k (\mathbf{U})_l^*. \tag{5}$$

Next, define $\mathcal{N} \triangleq \{1, 2, \ldots, n\}$, and the $m \times n^2$ matrix $\tilde{\mathbb{A}}$ such that

$$\left(\tilde{\mathbb{A}}\right)_{p,(k-1)n+l} \triangleq (\mathbb{A})_{p,k} (\mathbb{A})_{p,l}^*, \quad p \in \mathcal{M}, k, l \in \mathcal{N}. \tag{6}$$

Letting $\tilde{\mathbf{U}} \triangleq \mathbf{U} \otimes \mathbf{U}^*$, from (5) we obtain that $|\mathbb{A}\mathbf{U}|^2 = \tilde{\mathbb{A}}(\mathbf{U} \otimes \mathbf{U}^*)$. Thus (3) can be written as

$$\mathbf{Y} = \tilde{\mathbb{A}} (\mathbf{U} \otimes \mathbf{U}^*) + \mathbf{W} \equiv \tilde{\mathbb{A}} \tilde{\mathbf{U}} + \mathbf{W}. \tag{7}$$

We note that the transformation from $\mathbf{U}$ to $\tilde{\mathbf{U}} = \mathbf{U} \otimes \mathbf{U}^*$ is bijective[3], since $\mathbf{U}$ can be obtained from the singular value decomposition (SVD) of the rank one matrix $\mathbf{U}\mathbf{U}^H = \mathrm{vec}_n^{-1}(\mathbf{U} \otimes \mathbf{U}^*)^T$ [34, Ch. 2.4]. We also note that $\tilde{\mathbb{A}}$ corresponds to the *row-wise KRP* of $\mathbb{A}$ and $\mathbb{A}^*$ [34, Ch. 12.3], namely, the rows of $\tilde{\mathbb{A}}$ are obtained as the Kronecker product of the corresponding rows of $\mathbb{A}$ and $\mathbb{A}^*$. Defining $\mathbb{S}_m$ to be the $m \times m^2$ selection matrix such that $(\mathbb{S}_m)_{k,l} = \delta_{l,(k-1)m+k}$, we can write $\tilde{\mathbb{A}}$ as [30, Sec. 2.2]

$$\tilde{\mathbb{A}} = \mathbb{S}_m \cdot (\mathbb{A} \otimes \mathbb{A}^*). \tag{8}$$

The relationship (7) formulates the phase retrieval setup as *a MIMO channel* with complex channel input $\tilde{\mathbf{U}}$, complex channel matrix $\tilde{\mathbb{A}}$, real additive Gaussian noise $\mathbf{W}$, and real channel output $\mathbf{Y}$. We note that $\tilde{\mathbf{U}} = \mathbf{U} \otimes \mathbf{U}^*$ is non-Gaussian for any distribution of $\mathbf{U}$, since, e.g., $\left(\tilde{\mathbf{U}}\right)_1 = |(\mathbf{U})_1|^2$ is non-negative. In order to identify the measurement matrix which maximizes the MI, we wish to apply the gradient of the MI with respect to the measurement matrix, stated in [31, Th. 1]. To facilitate this application, we next formulate the phase retrieval setup as *a complex MIMO channel* with *additive PC Gaussian noise*. To that aim, let $\mathbf{W}_I \in \mathcal{R}^m$ be a random vector, distributed identically to $\mathbf{W}$ and independent of both $\mathbf{W}$ and $\mathbf{U}$, and also let $\mathbf{Y}_C \triangleq \mathbf{Y} + j\mathbf{W}_I$. The relationship between $\mathbf{Y}_C$ and $\tilde{\mathbf{U}}$ corresponds to a complex MIMO channel with additive zero-mean

[1] The optimal matrix $\mathbb{A}^{\mathrm{MI}}$ is not unique since, for example, for any real ϕ, the matrices $\mathbb{A}$ and $\mathbb{A}e^{j\phi}$ result in the same MI $I(\mathbf{U}; \mathbf{Y})$.

[2] Specifically, $m = 4n - 4$ was shown to be sufficient and $m = 4n - \mathcal{O}(n)$ was shown to be necessary.

[3] The transformation from $\mathbf{U}$ to $\tilde{\mathbf{U}}$ is bijective up to a global phase. However, the global phase can be set to an arbitrary value, as (1) is not affected by this global phase. Therefore, bijection up to a global phase is sufficient for establishing equivalence of the two representations in the present setup.

PC Gaussian noise, $\mathbf{W}_C \triangleq \mathbf{W} + j\mathbf{W}_I$, with covariance matrix $2\sigma_W^2 \mathbb{I}_m$:

$$\mathbf{Y}_C = \tilde{\mathbb{A}}\tilde{\mathbf{U}} + \mathbf{W}_C. \tag{9}$$

As the mapping from $\mathbf{U}$ to $\tilde{\mathbf{U}}$ is bijective, it follows from [24, Corollary after Eq. (2.121)] that

$$I\big(\mathbf{U};\mathbf{Y}\big) = I\big(\tilde{\mathbf{U}};\mathbf{Y}\big) \overset{(a)}{=} I\big(\tilde{\mathbf{U}};\mathbf{Y}_C\big), \tag{10}$$

where (a) follows from the MI chain rule [24, Sec. 2.5], since $\mathbf{Y} = \mathrm{Re}\{\mathbf{Y}_C\}$, $\mathbf{W}_I = \mathrm{Im}\{\mathbf{Y}_C\}$, and $\mathbf{W}_I$ is independent of $\mathbf{Y}$ and $\mathbf{U}$. Thus, (3) can be solved by finding $\mathbb{A}$ which maximizes the input-output MI of the MIMO channel representation.

The MIMO channel interpretation represents the non-linear phase retrieval setup (1) as a linear problem (9) *without modifying the MI*. This presents an advantage of using MI as a design criterion over the MMSE, as, unlike MI, MMSE is not invariant to the linear representation, i.e., the error covariance matrices of the MMSE estimator of $\mathbf{U}$ from $\mathbf{Y}$ and of the MMSE estimator of $\tilde{\mathbf{U}}$ from $\mathbf{Y}_C$ are in general not the same.

B. Conditions on $\mathbb{A}^{\mathrm{MI}}$ for Arbitrary SOI Distribution

Let $\mathbb{E}(\mathbb{A})$ be the error covariance matrix of the MMSE estimator of $\tilde{\mathbf{U}}$ from $\mathbf{Y}$ (referred to henceforth as the *MMSE matrix*) for a fixed measurement matrix $\mathbb{A}$, i.e.,

$$\mathbb{E}\left(\mathbb{A}\right) \triangleq \mathcal{E}\Big\{\Big(\tilde{\mathbf{U}} - \mathcal{E}\big\{\tilde{\mathbf{U}}|\mathbf{Y}\big\}\Big)\Big(\tilde{\mathbf{U}} - \mathcal{E}\big\{\tilde{\mathbf{U}}|\mathbf{Y}\big\}\Big)^H\Big\}. \tag{11}$$

Based on the observation that (9) corresponds to a MIMO channel with additive Gaussian noise, we obtain the following necessary condition on $\mathbb{A}^{\mathrm{MI}}$ which solves (3):

Theorem 1 (Necessary condition): Let $\mathbf{a}_k^{\mathrm{MI}}$ be the k-th column of $(\mathbb{A}^{\mathrm{MI}})^T$, $k \in \mathcal{M}$, and define the $n \times n$ matrix

$$\begin{aligned}\mathbb{H}_k\left(\mathbb{A}^{\mathrm{MI}}\right) \triangleq{}& \Big(\mathbb{I}_n \otimes \big(\mathbf{a}_k^{\mathrm{MI}}\big)^T\Big)\Big(\mathbb{E}\left(\mathbb{A}^{\mathrm{MI}}\right)\Big)^T\Big(\mathbb{I}_n \otimes \big(\mathbf{a}_k^{\mathrm{MI}}\big)^*\Big)\\ &+ \Big(\big(\mathbf{a}_k^{\mathrm{MI}}\big)^T \otimes \mathbb{I}_n\Big)\mathbb{E}\left(\mathbb{A}^{\mathrm{MI}}\right)\Big(\big(\mathbf{a}_k^{\mathrm{MI}}\big)^* \otimes \mathbb{I}_n\Big).\end{aligned}$$

Then, $\mathbb{A}^{\mathrm{MI}}$ that solves (3) satisfies:

$$\lambda \mathbf{a}_k^{\mathrm{MI}} = \mathbb{H}_k\left(\mathbb{A}^{\mathrm{MI}}\right)\mathbf{a}_k^{\mathrm{MI}}, \qquad \forall k \in \mathcal{M}, \tag{12}$$

where $\lambda \geq 0$ is selected such that $\mathrm{Tr}(\mathbb{A}^{\mathrm{MI}}(\mathbb{A}^{\mathrm{MI}})^H) = P$.

Proof: See Appendix A.

It follows from (12) that the k-th row of $\mathbb{A}^{\mathrm{MI}}$, $k \in \mathcal{M}$, is an eigenvector of the $n \times n$ Hermitian positive semi-definite matrix $\mathbb{H}_k(\mathbb{A}^{\mathrm{MI}})$, which depends on $\mathbb{A}^{\mathrm{MI}}$. As the optimization problem in (3) is generally non-concave, condition (12) does not uniquely identify the optimal measurement matrix in general. Furthermore, in order to explicitly obtain $\mathbb{A}^{\mathrm{MI}}$ from (12), the MMSE matrix $\mathbb{E}(\mathbb{A}^{\mathrm{MI}})$ must be derived, which is not a simple task. As an example, let the entries of $\mathbf{U}$ be zero-mean i.i.d. PC Gaussian RVs. Then, $\tilde{\mathbf{U}}$ obeys a singular Wishart distribution [35], and $\mathbb{E}(\mathbb{A})$ does not seem to have a tractable analytic expression. Despite this general situation, when the SNR is sufficiently low, we can explicitly characterize $\mathbb{A}^{\mathrm{MI}}$ in certain scenarios, as discussed in the next subsection.

C. Low SNR Regime

We next show that in the low SNR regime, it is possible to obtain an expression for the optimal measurement matrix which does not depend on $\mathbb{E}(\mathbb{A})$. Let $\mathbb{C}_{\mathbf{U}}$ and $\mathbb{C}_{\tilde{\mathbf{U}}}$ denote the covariance matrices of the SOI, $\mathbf{U}$, and of $\tilde{\mathbf{U}} = \mathbf{U} \otimes \mathbf{U}^*$, respectively. In the low SNR regime, i.e., when $\frac{P}{\sigma_W^2} \to 0$, the MI $I\big(\tilde{\mathbf{U}};\mathbf{Y}_C\big)$ satisfies [31, Eq. (41)]:

$$I\left(\tilde{\mathbf{U}};\mathbf{Y}_C\right) \approx \frac{1}{2\sigma_W^2}\mathrm{Tr}\left(\tilde{\mathbb{A}}\mathbb{C}_{\tilde{\mathbf{U}}}\tilde{\mathbb{A}}^H\right). \tag{13}$$

Thus, from (10) and (13), the measurement matrix maximizing the MI in the low SNR regime can be approximated by

$$\mathbb{A}^{\mathrm{MI}} \approx \underset{\mathbb{A}\in\mathcal{C}^{m\times n}:\mathrm{Tr}\left(\mathbb{A}\mathbb{A}^H\right)\leq P}{\arg\max}\ \mathrm{Tr}\left(\tilde{\mathbb{A}}\mathbb{C}_{\tilde{\mathbf{U}}}\tilde{\mathbb{A}}^H\right), \tag{14}$$

where $\tilde{\mathbb{A}}$ is given by (8).

Next, we introduce a new concept we refer to as Kronecker symmetric random vectors:

Definition 1 (Kronecker symmetry): A random vector $\mathbf{X}$ with covariance matrix $\mathbb{C}_{\mathbf{X}}$ is said to be *Kronecker symmetric* if the covariance matrix of $\mathbf{X} \otimes \mathbf{X}^*$ is equal to $\mathbb{C}_{\mathbf{X}} \otimes \mathbb{C}_{\mathbf{X}}^*$.

In particular, zero-mean PC Gaussian distributions satisfy Def. 1, as stated in the following lemma:

Lemma 1: Any $n \times 1$ zero-mean PC Gaussian random vector is Kronecker symmetric.

Proof: See Appendix B.

We now obtain a closed-form solution to (14) when $\mathbf{U}$ is a Kronecker symmetric random vector. The optimal $\mathbb{A}^{\mathrm{MI}}$ for this setup is stated in the following theorem:

Theorem 2: Let $\mathbf{a}_k^{\mathrm{MI}}$ be the k-th column of $(\mathbb{A}^{\mathrm{MI}})^T$, $k \in \mathcal{M}$, and let $\mathbf{v}_{\max}$ be the eigenvector of $\mathbb{C}_{\mathbf{U}}$ corresponding to its maximal eigenvalue. If $\mathbf{U}$ is a Kronecker symmetric random vector with covariance matrix $\mathbb{C}_{\mathbf{U}}$, then, for every $\mathbf{c} \in \mathcal{C}^m$ with $\|\mathbf{c}\|^2 = P$, setting $\mathbf{a}_k^{\mathrm{MI}} = (\mathbf{c})_k \mathbf{v}_{\max}^*$ for all $k \in \mathcal{M}$ solves (14). Thus,

$$\mathbb{A}^{\mathrm{MI}} = \mathbf{c} \cdot \mathbf{v}_{\max}^H. \tag{15}$$

Proof: See Appendix C.

The result of Theorem 2 is quite non-intuitive from an estimation perspective, as it suggests using *a rank-one measurement matrix*. This implies that the optimal measurement matrix projects the multivariate SOI onto a single eigenvector corresponding to the largest spread. Consequently, there are infinitely many realizations of $\mathbf{U}$ which result in the same $|\mathbb{A}\mathbf{U}|^2$. The optimality of rank-one measurements can be explained by noting that the selected scalar projection is, in fact, the least noisy of all possible scalar projections, as it corresponds to the largest eigenvalue of the covariance matrix of the SOI. Hence, when the additive noise is dominant, the optimal strategy is to design the measurement matrix such that it keeps only the least noisy spatial dimension of the signal, and eliminates all other spatial dimensions which are very noisy. From an information

theoretic perspective, this concept is not new, and the strategy of using a single spatial dimension which corresponds to the largest eigenvalue of the channel matrix in memoryless MIMO channels was shown to be optimal in the low SNR regime, e.g., in the design of the optimal precoding matrix for MIMO Gaussian channels [36, Sec. II-B]. However, while in [36, Sec. II-B] the problem was to optimize the input covariance (using the precoding matrix) for a *given channel*, in our case we optimize over the "channel" (represented by the measurement matrix) for a *given SOI covariance matrix*.

Finally, we show that the optimal measurement matrix in Theorem 2 satisfies the necessary condition for optimality in Theorem 1: In the low SNR regime the MMSE matrix (11) satisfies $\mathbb{E}\left(\mathbb{A}\right) \approx \mathbb{C}_{\tilde{\mathbf{U}}}$, see, e.g., [36, Eq. (15)]. The Kronecker symmetry of the SOI implies that $\mathbb{E}\left(\mathbb{A}\right) \approx \mathbb{C}_{\mathbf{U}} \otimes \mathbb{C}_{\mathbf{U}}^{*}$. Plugging this into the definition of $\mathbb{H}_k(\mathbb{A}^{\mathrm{MI}})$ in Theorem 1 results in $\mathbb{H}_k(\mathbb{A}^{\mathrm{MI}}) = 2((\mathbf{a}_k^{\mathrm{MI}})^T \mathbb{C}_{\mathbf{U}} (\mathbf{a}_k^{\mathrm{MI}})^*)\mathbb{C}_{\mathbf{U}}^{*}$. Theorem 1 thus states that for every $k \in \mathcal{M}$, the vector $\mathbf{a}_k^{\mathrm{MI}}$ must be a complex conjugate of an eigenvector of $\mathbb{C}_{\mathbf{U}}$. Consequently, the optimal matrix in Theorem 2 satisfies the necessary condition in Theorem 1.

IV. Practical Design of the Measurement Matrix

As can be concluded from the discussion following Theorem 1, the fact that (12) does not generally have a unique solution combined with the fact that it is often difficult to analytically compute the MMSE matrix, make the characterization of the optimal measurement matrix from condition (12) a very difficult task. Therefore, in this section we propose a practical approach for designing measurement matrices based on Theorem 1, while circumventing the difficulties discussed above by applying appropriate approximations. We note that while the practical design approach proposed in this section assumes that the observations are corrupted by an additive Gaussian noise, the suggested approach can also be used as an ad hoc method for designing measurement matrices for phase retrieval setups with non-Gaussian noise, e.g., Poisson noise [8, Sec. 2.3]. The practical design is performed via the following steps: First, we find the matrix $\tilde{\mathbb{A}}^{\mathrm{MI}}$ which maximizes the MI *without restricting* $\tilde{\mathbb{A}}$ *to satisfy the row-wise KRP structure* (8). Ignoring the structural constraints on $\tilde{\mathbb{A}}$ facilitates characterizing $\tilde{\mathbb{A}}^{\mathrm{MI}}$ via a set of fixed point equations. Then, we obtain a closed-form approximation of $\tilde{\mathbb{A}}^{\mathrm{MI}}$ by using the covariance matrix of the linear MMSE (LMMSE) estimator instead of the actual MMSE matrix. We denote the resulting matrix by $\tilde{\mathbb{A}}'$. Next, noting that the MI is invariant to unitary transformations, we obtain the final measurement matrix by finding $\mathbb{A}$ which minimizes the Frobenious norm between $\mathbb{S}_m\left(\mathbb{A} \otimes (\mathbb{A})^*\right)$ and a given unitary transformation of $\tilde{\mathbb{A}}'$, also designed to minimize the Frobenious norm. Using this procedure we obtain closed-form expressions for general measurement matrices as well as for masked Fourier measurement matrices. In the following we elaborate on these steps.

A. Optimizing Without Structure Constraints

In the first step we replace the maximization of the MI in (3) with respect to the measurement matrix $\mathbb{A}$, with a maximization with respect to $\tilde{\mathbb{A}}$, which denotes the row-wise KRP of $\mathbb{A}$ and $\mathbb{A}^*$. Specifically, we look for the matrix $\tilde{\mathbb{A}}$ which maximizes $I(\tilde{\mathbf{U}}; \mathbf{Y}_C)$, without constraining the structure of $\tilde{\mathbb{A}}$, while satisfying the trace constraint in (3).

We now formulate a constraint on $\tilde{\mathbb{A}}$ which guarantees that the trace constraint in (3) is satisfied. Letting $\mathbf{a}_k$ be the k-th column of $\mathbb{A}^T$, $k \in \mathcal{M}$, we have that

$$\begin{aligned}\left\|\mathbb{A}\right\|^4 &= \sum_{k_1=1}^{m}\sum_{k_2=1}^{m} \|\mathbf{a}_{k_1}\|^2 \|\mathbf{a}_{k_2}\|^2 \\ &\overset{(a)}{\leq} \frac{1}{2}\sum_{k_1=1}^{m}\sum_{k_2=1}^{m}\left(\|\mathbf{a}_{k_1}\|^4 + \|\mathbf{a}_{k_2}\|^4\right) = m\sum_{k=1}^{m}\|\mathbf{a}_k\|^4, \quad (16)\end{aligned}$$

where (a) follows since $a^2 + b^2 \geq 2ab$ for all $a, b \in \mathcal{R}$. Next, it follows from (8) that

$$\begin{aligned}\left\|\tilde{\mathbb{A}}\right\|^2 &= \sum_{k=1}^{m} \|\mathbf{a}_k \otimes \mathbf{a}_k^*\|^2 \\ &\overset{(a)}{=} \sum_{k=1}^{m}\|\mathbf{a}_k\|^4 \overset{(b)}{\geq} \frac{1}{m}\|\mathbb{A}\|^4, \quad (17)\end{aligned}$$

where (a) follows from [34, p. 709] and (b) follows from (16). Therefore, if $\tilde{\mathbb{A}}$ satisfies $\left\|\tilde{\mathbb{A}}\right\| \leq \frac{P}{\sqrt{m}}$, then $\mathrm{Tr}(\mathbb{A}\mathbb{A}^H) = \|\mathbb{A}\|^2 \leq P$, thereby satisfying the constraint in (3). Consequently, we consider the following optimization problem:

$$\tilde{\mathbb{A}}^{\mathrm{MI}} = \underset{\tilde{\mathbb{A}} \in \mathcal{C}^{m \times n^2} : \mathrm{Tr}\left(\tilde{\mathbb{A}}\tilde{\mathbb{A}}^H\right) \leq \frac{P^2}{m}}{\arg\max} I\left(\tilde{\mathbf{U}}; \mathbf{Y}_C\right). \quad (18)$$

Note that without constraining $\tilde{\mathbb{A}}$ to satisfy the structure (8), $\mathbf{Y}$ can be complex, and the MI between the input and the output of the transformed MIMO channel, $I\left(\tilde{\mathbf{U}}; \mathbf{Y}_C\right)$, may not be equal to the MI between the SOI and the observations of the original phase retrieval setup, $I(\mathbf{U}; \mathbf{Y})$.

The solution to (18) is given in the following lemma:

Lemma 2 [26, Th. 4.2], [37, Th. 1], [38, Prop. 2]: Let $\mathbb{E}\left(\tilde{\mathbb{A}}\right)$ be the covariance matrix of the MMSE estimate of $\tilde{\mathbf{U}}$ from $\mathbf{Y}_C$ for a given $\tilde{\mathbb{A}}$, and let $\mathbb{V}_E(\tilde{\mathbb{A}})\mathbb{D}_E(\tilde{\mathbb{A}})(\mathbb{V}_E\left(\tilde{\mathbb{A}}\right))^H$ be the eigenvalue decomposition of $\mathbb{E}\left(\tilde{\mathbb{A}}\right)$, in which $\mathbb{V}_E\left(\tilde{\mathbb{A}}\right)$ is unitary and $\mathbb{D}_E\left(\tilde{\mathbb{A}}\right)$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\mathbb{E}\left(\tilde{\mathbb{A}}\right)$ in descending order. Let $\mathbb{D}_A\left(\tilde{\mathbb{A}}\right)$ be an $m \times n^2$ diagonal matrix whose entries satisfy

$$\left(\mathbb{D}_A\left(\tilde{\mathbb{A}}\right)\right)_{k,k} = 0 \quad \text{if} \quad \left(\mathbb{D}_E\left(\tilde{\mathbb{A}}\right)\right)_{k,k} < \eta \quad (19a)$$

$$\left(\mathbb{D}_A\left(\tilde{\mathbb{A}}\right)\right)_{k,k} > 0 \quad \text{if} \quad \left(\mathbb{D}_E\left(\tilde{\mathbb{A}}\right)\right)_{k,k} = \eta, \quad (19b)$$

where η is selected such that $\sum_{k=1}^{m}(\mathbb{D}_A(\tilde{\mathbb{A}}))_{k,k}^2 = \frac{P^2}{m}$. The matrix $\tilde{\mathbb{A}}^{\mathrm{MI}}$ which solves (18) is given by the solution to

$$\tilde{\mathbb{A}}^{\mathrm{MI}} = \mathbb{D}_A\left(\tilde{\mathbb{A}}^{\mathrm{MI}}\right)\left(\mathbb{V}_E\left(\tilde{\mathbb{A}}^{\mathrm{MI}}\right)\right)^H. \quad (20)$$

Lemma 2 characterizes $\tilde{\mathbb{A}}^{\mathrm{MI}}$ via a set of fixed point equations[4]. Note that the matrix $\mathbb{D}_A(\tilde{\mathbb{A}}^{\mathrm{MI}})$ is constructed such that $\tilde{\mathbb{A}}^{\mathrm{MI}}$ which solves (20) induces a covariance matrix of the MMSE estimate of $\tilde{\mathbf{U}}$ from $\mathbf{Y}_C$, denoted $\mathbb{E}(\tilde{\mathbb{A}}^{\mathrm{MI}})$, whose eigenvalues satisfy (19).

B. Replacing the MMSE Matrix with the LMMSE Matrix

In order to obtain $\tilde{\mathbb{A}}^{\mathrm{MI}}$ from Lemma 2, we need the error covariance matrix of the MMSE estimator of $\tilde{\mathbf{U}}$ from $\mathbf{Y}_C$, $\mathbb{E}\left(\tilde{\mathbb{A}}^{\mathrm{MI}}\right)$, which in turn depends on $\tilde{\mathbb{A}}^{\mathrm{MI}}$. As $\mathbb{E}\left(\tilde{\mathbb{A}}\right)$ is difficult to compute, we propose to replace the error covariance matrix of the MMSE estimate with that of the LMMSE estimate[5] of $\tilde{\mathbf{U}}$ from $\mathbf{Y}_C$. The LMMSE matrix is given by [31, Sec. IV-C]

$$\mathbb{E}_L\left(\tilde{\mathbb{A}}\right) = \mathbb{C}_{\tilde{\mathbf{U}}} - \mathbb{C}_{\tilde{\mathbf{U}}}\tilde{\mathbb{A}}^H\left(2\sigma_W^2\,\mathbb{I}_m + \tilde{\mathbb{A}}\mathbb{C}_{\tilde{\mathbf{U}}}\tilde{\mathbb{A}}^H\right)^{-1}\tilde{\mathbb{A}}\mathbb{C}_{\tilde{\mathbf{U}}}.$$

Replacing $\mathbb{E}\left(\tilde{\mathbb{A}}\right)$ with $\mathbb{E}_L\left(\tilde{\mathbb{A}}\right)$ in Lemma 2, we obtain the matrix $\tilde{\mathbb{A}}'$ stated in the following corollary:

Corollary 1: Let $\mathbb{V}_{\tilde{\mathbf{U}}}\mathbb{D}_{\tilde{\mathbf{U}}}\mathbb{V}_{\tilde{\mathbf{U}}}^H$ be the eigenvalue decomposition of $\mathbb{C}_{\tilde{\mathbf{U}}}$, in which $\mathbb{V}_{\tilde{\mathbf{U}}}$ is unitary and $\mathbb{D}_{\tilde{\mathbf{U}}}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\mathbb{C}_{\tilde{\mathbf{U}}}$ arranged in descending order. Let $\tilde{\mathbb{D}}_A$ be an $m \times n^2$ diagonal matrix such that

$$(\tilde{\mathbb{D}}_A)_{k,k}^2 = \left(\tilde{\eta} - \frac{2\sigma_W^2}{(\mathbb{D}_{\tilde{\mathbf{U}}})_{k,k}}\right)^+, \quad \forall k \in \mathcal{M}, \tag{21}$$

where $\tilde{\eta}$ is selected such that $\sum_{k=1}^m (\tilde{\mathbb{D}}_A)_{k,k}^2 = \frac{P^2}{m}$. Finally, let

$$\tilde{\mathbb{A}}' = \tilde{\mathbb{D}}_A\mathbb{V}_{\tilde{\mathbf{U}}}^H. \tag{22}$$

Then, $\tilde{\mathbb{A}}'$ satisfies the conditions in Lemma 2, computed with $\mathbb{E}\left(\tilde{\mathbb{A}}'\right)$ replaced by $\mathbb{E}_L\left(\tilde{\mathbb{A}}'\right)$.

Proof: See Appendix D.

While Lemma 2 corresponds to a generalized mercury waterfilling solution [26, Th. 4.2], Corollary 1 is reminiscent of the conventional waterfilling solution for the optimal $\tilde{\mathbb{A}}$ when $\tilde{\mathbf{U}}$ is Gaussian [26, Th. 4.1]. However, as noted in Section III-A, $\tilde{\mathbf{U}}$ is non-Gaussian for any distribution of $\mathbf{U}$, thus, the resulting $\tilde{\mathbb{A}}'$ has no claim of optimality.

C. Nearest Row-Wise Khatri-Rao Product Representation

The choice of $\tilde{\mathbb{A}}'$ in (22) does not necessarily correspond to a row-wise KRP structure (8). In this case, it is not possible to find a matrix $\mathbb{A}$ such that $|\mathbb{A}\mathbf{U}|^2 = \tilde{\mathbb{A}}'(\mathbf{U}\otimes\mathbf{U}^*)$, which implies that the matrix $\tilde{\mathbb{A}}'$ does not correspond to the model (1). Furthermore, we note that MI is invariant to unitary transformations, and specifically, for any unitary $\mathbb{V} \in \mathcal{C}^{m\times m}$ and for any $\tilde{\mathbb{A}} \in \mathcal{C}^{m\times n^2}$ we have that

$$I\left(\tilde{\mathbf{U}};\tilde{\mathbb{A}}\tilde{\mathbf{U}} + \mathbf{W}_C\right) \overset{(a)}{=} I\left(\tilde{\mathbf{U}};\tilde{\mathbb{A}}\tilde{\mathbf{U}} + \mathbb{V}^H\mathbf{W}_C\right) \overset{(b)}{=} I\left(\tilde{\mathbf{U}};\mathbb{V}\tilde{\mathbb{A}}\tilde{\mathbf{U}} + \mathbf{W}_C\right), \tag{23}$$

where (a) follows from [24, Eq. (8.71)], and (b) since $I\left(\tilde{\mathbf{U}};\mathbf{Y}_C\right) = I\left(\tilde{\mathbf{U}};\mathbb{V}\mathbf{Y}_C\right)$, see [24, p. 35]. Therefore, in order to obtain a measurement matrix, we propose finding an $m \times n$ matrix $\hat{\mathbb{A}}^{\mathrm{O}}$ such that, for a given unitary matrix $\mathbb{V}$,

$$\hat{\mathbb{A}}^{\mathrm{O}} = \underset{\mathbb{A}\in\mathcal{C}^{m\times n}}{\arg\min}\,\|\mathbb{V}\tilde{\mathbb{A}}' - \mathbb{S}_m\left(\mathbb{A}\otimes\mathbb{A}^*\right)\|^2. \tag{24}$$

Note that while the unitary matrix $\mathbb{V}$ does not modify the MI, it can result in reducing the minimal Frobenious norm in (24). We will elaborate on the selection of $\mathbb{V}$ in Section IV-E.

To solve (24), let $\tilde{\mathbf{a}}_k'$ be the $n^2 \times 1$ column vector corresponding to the k-th column of $(\mathbb{V}\tilde{\mathbb{A}}')^T$ and $\tilde{\mathbb{M}}_k^{(H)}$ be the Hermitian part[6] of $\mathrm{vec}_n^{-1}(\tilde{\mathbf{a}}_k')$, $k \in \mathcal{M}$. The solution to (24) can be analytically obtained as stated in the following proposition:

Proposition 1: Let $\hat{\mathbf{a}}_k^{\mathrm{O}}$ be the $n \times 1$ vector corresponding to the k-th column of $\left(\hat{\mathbb{A}}^{\mathrm{O}}\right)^T$, $k \in \mathcal{M}$. Let $\tilde{\mu}_{k,\max}$ be the largest eigenvalue of $\tilde{\mathbb{M}}_k^{(H)}$, and let $\tilde{\mathbf{v}}_{k,\max}$ be the corresponding eigenvector, when the eigenvector matrix is unitary. Then, the columns of $\left(\hat{\mathbb{A}}^{\mathrm{O}}\right)^T$ which solves (24) are given by

$$\hat{\mathbf{a}}_k^{\mathrm{O}} = \sqrt{\max\left(\tilde{\mu}_{k,\max},0\right)}\cdot\tilde{\mathbf{v}}_{k,\max}^*, \qquad k \in \mathcal{M}. \tag{25}$$

Proof: See Appendix E.

The matrix $\hat{\mathbb{A}}^{\mathrm{O}}$ derived in Proposition 1 does not necessarily satisfy the Frobenius norm constraint P. Thus, if the squared norm of $\hat{\mathbb{A}}^{\mathrm{O}}$ is larger than P, then it is scaled down to satisfy the norm constraint. Moreover, since $I\left(\mathbf{U};\gamma|\hat{\mathbb{A}}^{\mathrm{O}}\mathbf{U}|^2 + \mathbf{W}\right)$ is monotonically non-decreasing w.r.t. $\gamma > 0$ [25, Th. 2] for any distribution of $\mathbf{U}$, if the squared norm of $\hat{\mathbb{A}}^{\mathrm{O}}$ is smaller than P, then it is scaled up to the maximal norm to maximize the MI. Consequently, the final measurement matrix is given by $\mathbb{A}^{\mathrm{O}} = \frac{\sqrt{P}}{\|\hat{\mathbb{A}}^{\mathrm{O}}\|}\hat{\mathbb{A}}^{\mathrm{O}}$.

Next, we show that when $\mathbf{U}$ is Kronecker symmetric, then, in the low SNR regime, $\mathbb{A}^{\mathrm{O}}$ coincides with the optimal matrix characterized in Theorem 2, for any unitary transformation matrix $\mathbb{V}$. Let $\mathbf{i}_1$ be an $m \times 1$ vector such that $\left(\mathbf{i}_1\right)_k = \delta_{k,1}$, and let $\mathbb{V}_{\mathbf{U}}\mathbb{D}_{\mathbf{U}}\mathbb{V}_{\mathbf{U}}^H$ be the eigenvalue decomposition of $\mathbb{C}_{\mathbf{U}}$. For a Kronecker symmetric $\mathbf{U}$, we have that $\mathbb{C}_{\tilde{\mathbf{U}}} = \mathbb{C}_{\mathbf{U}}\otimes\mathbb{C}_{\mathbf{U}}^*$, and thus $\mathbb{V}_{\tilde{\mathbf{U}}} = \mathbb{V}_{\mathbf{U}}\otimes\mathbb{V}_{\mathbf{U}}^*$ and $\mathbb{D}_{\tilde{\mathbf{U}}} = \mathbb{D}_{\mathbf{U}}\otimes\mathbb{D}_{\mathbf{U}}^*$ [34, Ch. 12.3.1]. In the low SNR regime, due to the "waterfilling" in (22), the measurement matrix extracts only the least noisy spatial dimension of the SOI, resulting in $\tilde{\mathbb{A}}' = \frac{P}{\sqrt{m}}\mathbf{i}_1\left(\mathbf{v}_{\max}\otimes\mathbf{v}_{\max}^*\right)^H$, where $\mathbf{v}_{\max}$ is the eigenvector corresponding to the maximal eigenvalue of the SOI covariance matrix, $\mathbb{C}_{\mathbf{U}}$. Therefore, letting $\mathbf{v}_1$ denote the leftmost column of $\mathbb{V}$, we have that $\mathbb{V}\tilde{\mathbb{A}}' = \frac{P}{\sqrt{m}}\mathbf{v}_1\left(\mathbf{v}_{\max}\otimes\mathbf{v}_{\max}^*\right)^H$, which results in $\mathrm{vec}_n^{-1}\left(\tilde{\mathbf{a}}_k'\right) = \frac{P}{\sqrt{m}}(\mathbf{v}_1)_k\mathbf{v}_{\max}\mathbf{v}_{\max}^H$ [41, Ch. 9.2] and $\tilde{\mathbb{M}}_k^{(H)} =$

[4]The solution in [26, Th. 4.2] includes a permutation matrix which performs mode alignment. However, for white noise mode alignment is not needed, and the permutation matrix can be set to $\mathbb{I}_{n^2}$ [37, Sec. III].

[5]An inspiration for this approximation stems from the fact that for parallel Gaussian MIMO scenarios, the covariance matrices of the MMSE estimate and of the LMMSE estimate coincide at high SNRs [39].

[6]The Hermitian part of a matrix $\mathbb{Z}$ is given by $\frac{1}{2}(\mathbb{Z} + \mathbb{Z}^H)$.

$\frac{P}{\sqrt{m}}\mathrm{Re}\{(\mathbf{v}_1)_k\}\mathbf{v}_{\max}\mathbf{v}_{\max}^H$. Consequently, $\tilde{\mathbf{v}}_{k,\max}=\mathbf{v}_{\max}$ for every $k\in\mathcal{M}$, and thus $\mathbb{A}^{\mathrm{O}}$ is a rank-one matrix of the form $\mathbb{A}^{\mathrm{O}}=\mathbf{c}\cdot\mathbf{v}_{\max}^H$, which coincides with $\mathbb{A}^{\mathrm{MI}}$ stated in Theorem 2. For example, setting $\mathbb{V}=\mathbb{I}_m$ results in $\mathbf{c}=\sqrt{P}\cdot\mathbf{i}_1$.

D. Masked Fourier Measurement Matrix

As mentioned in Section II-B, in many phase retrieval setups, the measurement matrix represents masked Fourier measurements and is constrained to the structure of (4). In the context of phase retrieval, the design goal is to find the set of masks $\{\mathbb{G}_l\}_{l=1}^b$ in (4) which result in optimal recovery performance. To that aim, define the $n\times 1$ vectors $\mathbf{g}_l$, $l\in\mathcal{B}$, to contain the diagonal elements of $\mathbb{G}_l$, $(\mathbf{g}_l)_k=(\mathbb{G}_l)_{k,k}$, $k\in\mathcal{N}$. With this definition, we can write

$$(\mathbb{A})_{(l-1)n+k,p}=(\mathbf{g}_l)_p\,(\mathbb{F}_n)_{k,p}\,,\quad \forall k,p\in\mathcal{N},l\in\mathcal{B}. \tag{26}$$

Since $\mathbb{A}^{\mathrm{O}}$ does not necessarily represent a masked Fourier structure, based on the rationale detailed in Section IV-C, we suggest to use the masks $\{\mathbf{g}_l^{\mathrm{MF}}\}_{l=1}^b$ that minimize the distance between the resulting measurement matrix and a unitary transformation of $\tilde{\mathbb{A}}'$:

$$\{\mathbf{g}_l^{\mathrm{MF}}\}_{l=1}^b=\underset{\{\mathbf{g}_l\}_{l=1}^b\in\mathcal{C}^n}{\arg\min}\ \|\mathbb{V}\tilde{\mathbb{A}}'-\mathbb{S}_m\,(\mathbb{A}\otimes\mathbb{A}^*)\,\|^2, \tag{27}$$

where $\mathbb{V}$ is a given unitary matrix and $\mathbb{A}$ depends on $\{\mathbf{g}_l^{\mathrm{MF}}\}_{l=1}^b$ via (26). The set of masks which solve (27) is characterized in the following proposition:

Proposition 2: Let $\tilde{\mathbb{F}}_k$ be an $n\times n$ diagonal matrix such that $(\tilde{\mathbb{F}}_k)_{p,p}=(\mathbb{F}_n)_{k,p}$, $k,p\in\mathcal{N}$. For all $l\in\mathcal{B}$, let $\bar{\mu}_{l,\max}$ be the largest eigenvalue of the $n\times n$ Hermitian matrix $\sum_{k=1}^n\tilde{\mathbb{F}}_k\tilde{\mathbb{M}}_{(l-1)n+k}^{(H)}\tilde{\mathbb{F}}_k^*$, where $\tilde{\mathbb{M}}_{(l-1)n+k}^{(H)}$ is the Hermitian part of $\mathrm{vec}_n^{-1}(\tilde{\mathbf{a}}'_{(l-1)n+k})$, and let $\bar{\mathbf{v}}_{l,\max}$ be its corresponding eigenvector, when the eigenvector matrix is unitary. Then, the set of mask coefficients $\{\mathbf{g}_l^{\mathrm{MF}}\}_{l=1}^b$ which solves (27) is obtained as

$$\mathbf{g}_l^{\mathrm{MF}}=\sqrt{n\cdot\max\left(\bar{\mu}_{l,\max},0\right)}\cdot\bar{\mathbf{v}}_{l,\max}^*,\quad l\in\mathcal{B}. \tag{28}$$

Proof: See Appendix F.

The masked Fourier measurement matrix is obtained from the coefficient vectors $\{\mathbf{g}_l^{\mathrm{MF}}\}_{l=1}^b$ via

$$\left(\hat{\mathbb{A}}^{\mathrm{MF}}\right)_{(l-1)\cdot n+k,p}=\left(\mathbf{g}_l^{\mathrm{MF}}\right)_p(\mathbb{F}_n)_{k,p}\,,\quad k,p\in\mathcal{N},l\in\mathcal{B}. \tag{29}$$

Applying the same reasoning used in determining the scaling of $\hat{\mathbb{A}}^{\mathrm{O}}$ in Section IV-C, we conclude that the MI is maximized, subject to the trace constraint, by normalizing $\hat{\mathbb{A}}^{\mathrm{MF}}$ to obtain $\mathbb{A}^{\mathrm{MF}}=\frac{\sqrt{P}}{\|\hat{\mathbb{A}}^{\mathrm{MF}}\|}\hat{\mathbb{A}}^{\mathrm{MF}}$.

Let us again consider a Kronecker symmetric $\mathbf{U}$ in the low SNR regime. For simplicity, we set $\mathbb{V}=\mathbb{I}_m$. As discussed in the previous subsection, for this setting we have that $\tilde{\mathbb{A}}'=\frac{P}{\sqrt{m}}\mathbf{i}_1\left(\mathbf{v}_{\max}\otimes\mathbf{v}_{\max}^*\right)^H$, where $\mathbf{i}_1$ is the $m\times 1$ vector such that $\left(\mathbf{i}_1\right)_k=\delta_{k,1}$, and thus $\tilde{\mathbb{M}}_k^{(H)}$ is non-zero only for $k=1$. Therefore, $\bar{\mu}_{l,\max}$ is zero for all $l\neq 1$, while $\bar{\mu}_{1,\max}$ is the largest eigenvalue of $\tilde{\mathbb{F}}_1^*\tilde{\mathbb{M}}_1^{(H)}\tilde{\mathbb{F}}_1=\mathbb{M}_1^{(H)}=\mathbf{v}_{\max}\mathbf{v}_{\max}^H$, and thus $\bar{\mathbf{v}}_{1,\max}=\mathbf{v}_{\max}$. Consequently, we have that

$$\mathbb{A}^{\mathrm{MF}}=\sqrt{P}\begin{bmatrix}\mathbb{F}_n\,\mathrm{diag}\,(\mathbf{v}_{\max}^*)\\ 0\quad\dots\quad 0\\ \vdots\\ 0\quad\dots\quad 0\end{bmatrix}. \tag{30}$$

Unlike the unconstrained case considered in the previous subsection, the resulting measurement matrix in (30) does not coincide with the optimal matrix given in Theorem 2, due to the masked Fourier structure constraint.

E. Obtaining the Optimal Unitary Transformation Matrix

In the previous subsections we assumed that the unitary transformation $\mathbb{V}$ applied to $\tilde{\mathbb{A}}'$ is given. In the following we propose an algorithm to jointly identify the optimal transformation $\mathbb{V}$ and the optimal measurement matrix $\mathbb{A}$.

Let $\mathcal{V}$ denote the set of $m\times m$ complex unitary matrices and $\mathcal{A}$ denote the set of $m\times n$ feasible measurement matrices. For example, for unconstrained measurements, $\mathcal{A}=\mathcal{C}^{m\times n}$, and for masked Fourier measurements, $\mathcal{A}$ is the set of all matrices which can be expressed as in (4). The optimal $\mathbb{A}$ and $\mathbb{V}$ are obtained as the solution to the following joint optimization problem:

$$\left(\hat{\mathbb{A}}^{\mathrm{U}},\mathbb{V}^{\mathrm{U}}\right)=\underset{\mathbb{A}\in\mathcal{A},\mathbb{V}\in\mathcal{V}}{\arg\min}\ \|\mathbb{V}\tilde{\mathbb{A}}'-\mathbb{S}_m\,(\mathbb{A}\otimes\mathbb{A}^*)\,\|^2. \tag{31}$$

The solution to (31) *for a fixed* $\mathbb{V}$ is given in Propositions 1 and 2. For a fixed $\mathbb{A}$, the problem in (31) is the unitary Procrustes problem [44, Ch. 7.4]: Letting $\mathbb{V}_{\mathrm{svd}}(\mathbb{A})\mathbb{D}_{\mathrm{svd}}(\mathbb{A})\mathbb{W}_{\mathrm{svd}}^H(\mathbb{A})$ be the SVD of $\mathbb{S}_m(\mathbb{A}\otimes\mathbb{A}^*)\cdot\left(\tilde{\mathbb{A}}'\right)^H$, the solution to (31) *for a fixed* $\mathbb{A}$ is given by

$$\mathbb{V}^{\mathrm{U}}\,(\mathbb{A})=\mathbb{V}_{\mathrm{svd}}\,(\mathbb{A})\,\mathbb{W}_{\mathrm{svd}}^H\,(\mathbb{A})\,. \tag{32}$$

Based on the above, we propose to solve the joint optimization problem (31) in an alternating fashion, i.e., optimize over $\mathcal{A}$ for a fixed $\mathbb{V}$, then optimize over $\mathcal{V}$ for a fixed $\mathbb{A}$, and continue with the alternating optimization process until convergence. The overall matrix design algorithm is summarized in Algorithm 1. As the Frobenious norm objective in (31) is differentiable, convergence of the alternating optimization algorithm is guaranteed [45, Th. 2]. However, since the problem is not necessarily convex[7] w.r.t. both $\mathbb{A}$ and $\mathbb{V}$, the algorithm may converge to a local minima.

Assuming that the computation of $\tilde{\mathbb{A}}'$ in Step 1 of Algorithm 1 is carried out using a computationally efficient waterfilling algorithm, as in, e.g., [46], the complexity of Algorithm 1 is dominated by the computation of the eigenvalue decomposition required in Step 2 and by the matrix product required to compute the SVD in Step 1. Letting $t_{\max}$ denote the maximal number of iterations over Steps 3–4, it follows that the overall

[7]This non-convexity is observed by noting that, for example, for $\phi\in(0,2\pi)$, the right hand side of (31) obtains the same value for $\mathbb{A}$ and for $\mathbb{A}e^{j\phi}$, and a different value for $\frac{1}{2}(1+e^{j\phi})\mathbb{A}$, which is an element of every convex set containing $\mathbb{A}$ and $\mathbb{A}e^{j\phi}$. Consequently, when $\mathbb{A}$ which is not all zero solves (31), the set of all minima is not convex, and the optimization problem is thus not convex [40, Ch. 4.2].

Algorithm 1: Measurement Matrix Design.

1: Initialization: Set $k=0$ and $\mathbb{V}_0=\mathbb{I}_m$.
2: Compute $\tilde{\mathbb{A}}'$ using (22).
3: Obtain $\hat{\mathbb{A}}_{k+1}=\arg\min_{\mathbb{A}\in\mathcal{A}}\|\mathbb{V}_k\tilde{\mathbb{A}}'-\mathbb{S}_m(\mathbb{A}\otimes\mathbb{A}^*)\|^2$ using Proposition 1 (for general measurements) or using Proposition 2 (for masked Fourier measurements).
4: Set $\mathbb{V}_{k+1}=\mathbb{V}_{\rm svd}\left(\hat{\mathbb{A}}_{k+1}\right)\mathbb{W}^H_{\rm svd}\left(\hat{\mathbb{A}}_{k+1}\right)$.
5: If termination criterion is inactive: Set $k:=k+1$ and go to Step 3.
6: $\mathbb{A}^{\rm U}$ is obtained as $\mathbb{A}^{\rm U}=\frac{\sqrt{P}}{\|\hat{\mathbb{A}}_k\|}\hat{\mathbb{A}}_k$.

computational complexity of the algorithm is on the order of $\mathcal{O}(t_{\max}\cdot m^2\cdot n^2+n^6)$ [34, Ch. 1.1, Ch. 8.6].

While in the problem formulation we consider white Gaussian noise, the measurement matrix design in Algorithm 1 can be extended to account for colored Gaussian noise, i.e., for noise $\mathbf{W}$ with covariance matrix $\mathbb{C}_{\mathbf{W}}\neq\sigma_W^2\mathbb{I}_m$, by considering the whitened observations vector $\mathbb{C}_{\mathbf{W}}^{-1/2}\mathbf{Y}$ instead of $\mathbf{Y}$. This is because invertible transformations do not change the MI: $I(\mathbf{U};\mathbf{Y})=I(\mathbf{U};\mathbb{C}_{\mathbf{W}}^{-1/2}\mathbf{Y})$ [24, Corollary after Eq. (2.121)], therefore maximizing the MI for the whitened observations maximizes the MI for the original observations. After applying the whitening transformation, Algorithm 1 can be used on the whitened observations vector $\mathbb{C}_{\mathbf{W}}^{-1/2}\mathbf{Y}$ with noise covariance matrix $\mathbb{I}_m$, with the exception that the objective function in Step 3 is replaced with $\arg\min_{\mathbb{A}\in\mathcal{A}}\|\mathbb{V}_k\mathbb{C}_{\mathbf{W}}^{1/2}\tilde{\mathbb{A}}'-\mathbb{S}_m(\mathbb{A}\otimes\mathbb{A}^*)\|^2$.

V. Simulations Study

In this section we evaluate the performance of phase retrieval with the proposed measurement matrix design in a simulations study. While our design aims at maximizing the statistical dependence between the SOI and the observations via MI maximization, we note that phase retrieval is essentially *an estimation problem*, hence, we evaluate the performance in terms of estimation error. Since the phase retrieval setup inherently has a global phase ambiguity, for an SOI realization $\mathbf{U}=\mathbf{u}$ and its estimate $\hat{\mathbf{U}}=\hat{\mathbf{u}}$, we define the estimation error as

$$\epsilon\left(\mathbf{u},\hat{\mathbf{u}}\right)=\min_{c\in\mathcal{C}:|c|=1}\frac{\|\mathbf{u}-c\cdot\hat{\mathbf{u}}\|}{\|\mathbf{u}\|},\tag{33}$$

namely, the minimum relative distance over all phase rotations, see, e.g., [9, Eq. (19)]. We use both phasecut [9] and TAF (with step-size 1 and truncation threshold 0.9) [11] to estimate the SOI $\mathbf{U}$ from the observations $\mathbf{Y}$. Performance was evaluated for five different measurement matrices:

- $\mathbb{A}^{\rm OK}$ - The optimal measurement matrix for Kronecker symmetric SOI in the low SNR regime, obtained via (15) with $\mathbf{c}$ selected such that $(\mathbf{c})_k=\sqrt{\frac{P}{m}}e^{j2\pi\frac{k-1}{m}}$ for all $k\in\mathcal{M}$.
- $\mathbb{A}^{\rm UC}$ - The unconstrained measurement matrix obtained using Algorithm 1 with $\mathcal{A}=\mathcal{C}^{m\times n}$.

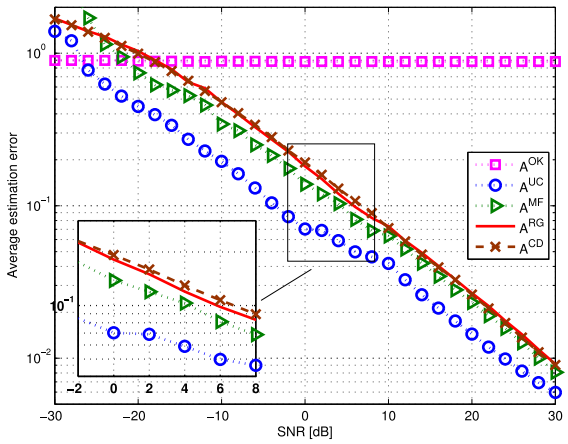

Fig. 1. Average estimation error vs. SNR for $\mathbf{U}_S$ using phasecut, $m=6n$.

- $\mathbb{A}^{\rm MF}$ - The masked Fourier measurement matrix obtained using Algorithm 1 with $\mathcal{A}$ being the set of matrices which can be expressed as in (4).
- $\mathbb{A}^{\rm RG}$ - A random PC Gaussian matrix with i.i.d. entries.
- $\mathbb{A}^{\rm CD}$ - A coded diffraction pattern matrix with random octanary patterns [10, Sec. 4.1], namely, a masked Fourier matrix (4) with i.i.d. random masks, each having i.i.d. entries distributed according to [10, Eq. (4.3)].

For the random matrices, $\mathbb{A}^{\rm RG}$ and $\mathbb{A}^{\rm CD}$, a new realization is generated for each Monte Carlo simulation. The squared Frobenius norm constraint is set to $P=m$, namely, the average row squared norm for all designed matrices is 1. Two different SOI distributions of size $n=10$ were tested:

- $\mathbf{U}_S$ - A sum of complex exponentials (see, e.g., [9, Sec. V]) given by $(\mathbf{U}_S)_k=\sum_{l=1}^{6}M_l e^{j\pi\Phi_l k}$, where $\{M_l\}_{l=1}^6$ are i.i.d. zero-mean unit variance real-valued Gaussian RVs, and $\{\Phi_l\}_{l=1}^6$ are i.i.d. RVs uniformly distributed over $[0,\pi]$, independent of $\{M_l\}_{l=1}^6$.
- $\mathbf{U}_G$ - A zero-mean PC Gaussian vector with covariance matrix $\mathbb{C}_{\mathbf{U}}$ corresponding to an exponentially decaying correlation profile given by $(\mathbb{C}_{\mathbf{U}})_{k,l}=6\cdot e^{-|k-l|+j\frac{2\pi(k-l)}{n}}$, $k,l\in\mathcal{N}$.

Note that all tested SOIs have the same energy, measured as the trace of the covariance matrix. The estimation error is averaged over 1000 Monte Carlo simulations, where a new SOI and noise realization is generated in each simulation.

In Figs. 1–4 we fix the observations dimension to be $m=6\cdot n=60$, and let the SNR, defined as $1/\sigma_W^2$, vary from -30 dB to 30 dB, for $\mathbf{U}_S$ using phasecut, $\mathbf{U}_S$ using TAF, $\mathbf{U}_G$ using phasecut, and $\mathbf{U}_G$ using TAF, respectively. It can be observed from Figs. 1–4 that the *deterministic unconstrained* $\mathbb{A}^{\rm UC}$ achieves the best performance over almost the entire SNR range, for all tested SOI distributions. Notable gains are observed for $\mathbf{U}_S$ in Figs. 1 and 2, where, for example, $\mathbb{A}^{\rm UC}$ attains an average estimation error of $\epsilon=0.1$ for SNRs of -4 dB and -2 dB, for phasecut and for TAF, respectively, while random Gaussian measurements $\mathbb{A}^{\rm RG}$ achieve $\epsilon=0.1$ for SNRs of 4 dB and 8 dB, for phasecut and for TAF, respectively, and random coded

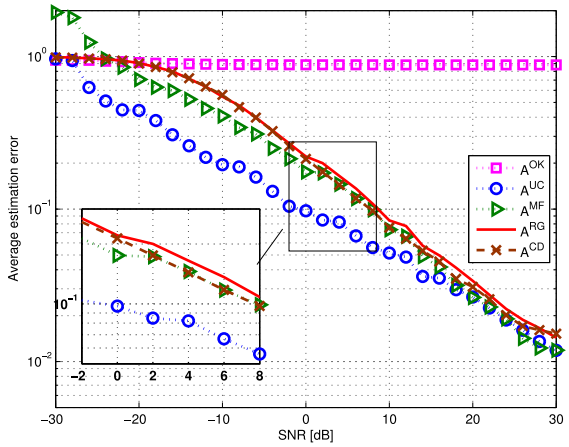

Fig. 2. Average estimation error vs. SNR for $\mathbf{U}_S$ using TAF, $m = 6n$.

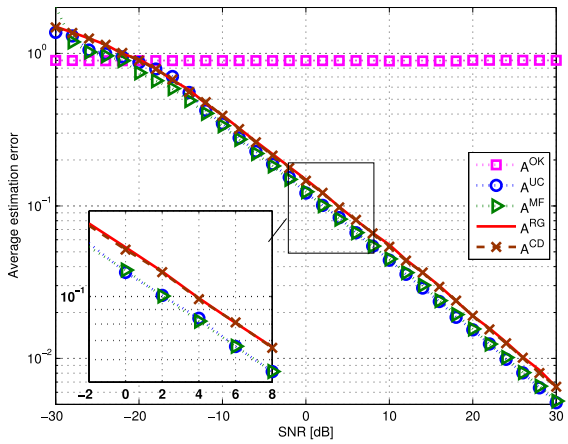

Fig. 3. Average estimation error vs. SNR for $\mathbf{U}_G$ using phasecut, $m = 6n$.

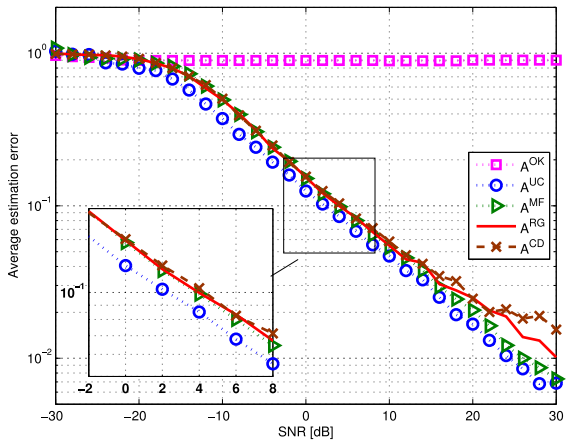

Fig. 4. Average estimation error vs. SNR for $\mathbf{U}_G$ using TAF, $m = 6n$.

diffraction patterns $\mathbb{A}^{\mathrm{CD}}$ achieve $\epsilon = 0.1$ for SNRs of 6 dB and 8 dB, for phasecut and for TAF, respectively. Consequently, for SOI distribution $\mathbf{U}_S$, $\mathbb{A}^{\mathrm{UC}}$ achieves an SNR gain of 8–10 dB at $\epsilon = 0.1$ over Gaussian measurements, and an SNR gain of 10 dB over random coded diffraction patterns. From Figs. 3 and 4 we observe that the corresponding SNR gain at $\epsilon = 0.1$ for the SOI

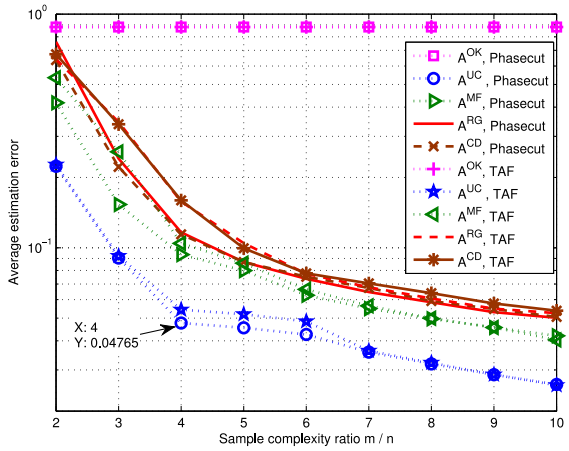

Fig. 5. Average estimation error vs. sample complexity, $\mathbf{U}_S$, SNR $= 10$ dB.

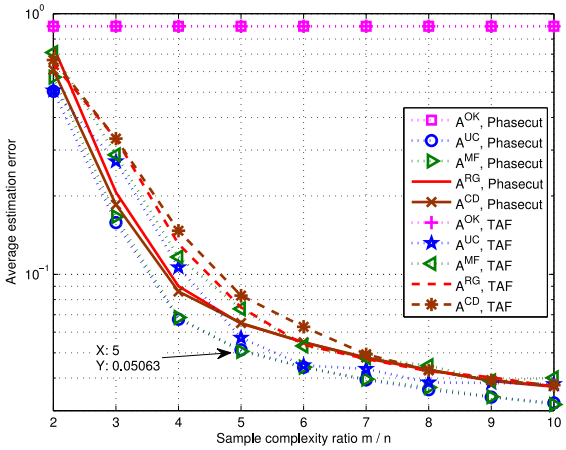

Fig. 6. Average estimation error vs. sample complexity, $\mathbf{U}_G$, SNR $= 10$ dB.

distribution $\mathbf{U}_G$ is 2 dB, compared to both random Gaussian measurements as well as to random coded diffraction patterns. Furthermore, it is observed from Figs. 1–4 that the proposed masked Fourier measurement matrix $\mathbb{A}^{\mathrm{MF}}$, corresponding to *practical deterministic masked Fourier measurements*, achieves an SNR gain of 0–2 dB for both SOI distributions $\mathbf{U}_G$ and $\mathbf{U}_S$, compared to random Gaussian measurements and random coded diffraction patterns. It is also noted in Figs. 1–4 that, as expected, in the low SNR regime, i.e., $1/\sigma_W^2 < -20$ dB, $\mathbb{A}^{\mathrm{OK}}$ obtains the best performance, as it is designed specifically for low SNRs. However, the performance of $\mathbb{A}^{\mathrm{OK}}$ for both recovery algorithms hardly improves with SNR as its rank-one structure does not allow the complete recovery of the SOI at any SNR.

In Figs. 5 and 6 we fix the SNR to be 10 dB, and let the sample complexity ratio $\frac{m}{n}$ [10], [11] vary from 2 to 10, for both $\mathbf{U}_S$ and $\mathbf{U}_G$. From Figs. 5 and 6 we observe that the superiority of the deterministic $\mathbb{A}^{\mathrm{UC}}$ is maintained for different sample complexity values. For example, in Fig. 5 we observe that for $\mathbf{U}_S$ at SNR $1/\sigma_W^2 = 10$ dB, $\mathbb{A}^{\mathrm{UC}}$ obtains an estimation error of less than $\epsilon = 0.05$ for $m = 4n$ and for $m = 6n$, using phasecut and using TAF, respectively, while our masked Fourier design $\mathbb{A}^{\mathrm{MF}}$ requires $m = 8n$ observations, and both random Gaussian measurements and random coded diffraction patterns require

TABLE I
FROBENIUS NORM $\|\mathbb{V}\tilde{\mathbb{A}}' - \mathbb{S}_m(\mathbb{A}\otimes\mathbb{A}^*)\|$ COMPARISON FOR $\mathbf{U}_S$

$1/\sigma_W^2$	$\mathbb{A}^{\mathrm{UC}}$	$\mathbb{A}_{\mathrm{I}}^{\mathrm{UC}}$	$\mathbb{A}^{\mathrm{MF}}$	$\mathbb{A}_{\mathrm{I}}^{\mathrm{MF}}$
-10 dB	2.09	6.93	6.25	7.63
10 dB	2.19	6.99	5.70	7.66
30 dB	2.25	7.08	5.16	7.66

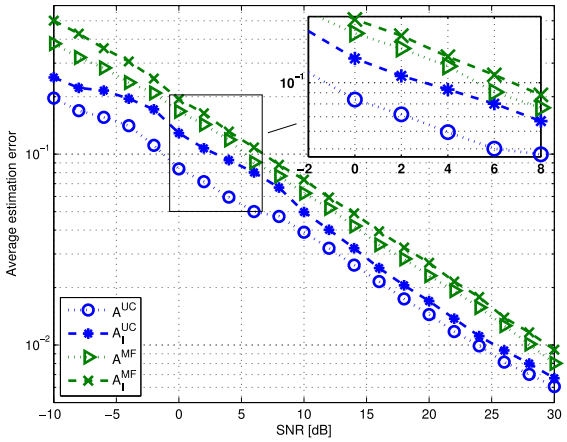

Fig. 7. Average estimation error vs. SNR for $\mathbf{U}_S$, $m = 6n$.

$m = 10n$ observations to achieve a similar estimation error, for both phasecut and TAF. A similar behavior with less notable gains is observed for $\mathbf{U}_G$ in Fig. 6. For example, for $\mathbf{U}_G$ using phasecut, both $\mathbb{A}^{\mathrm{UC}}$ and $\mathbb{A}^{\mathrm{MF}}$ require $m = 5n$ observations to achieve $\epsilon = 0.05$, while both $\mathbb{A}^{\mathrm{RG}}$ and $\mathbb{A}^{\mathrm{CD}}$ require $m = 7n$ observations to achieve similar performance. This implies that our proposed designs *require fewer measurements*, compared to the common random measurement matrices, to achieve the same performance.

Moreover, we observe that the estimation error of both the unconstrained measurements $\mathbb{A}^{\mathrm{UC}}$ and the masked Fourier measurements $\mathbb{A}^{\mathrm{MF}}$ scale w.r.t. SNR (Figs. 1–4) and sample complexity (Figs. 5 and 6) similarly to random measurements $\mathbb{A}^{\mathrm{RG}}$ and $\mathbb{A}^{\mathrm{CD}}$, and that the performance gain compared to random Gaussian measurements and random coded diffraction patterns is maintained for various values of m.

Lastly, we numerically evaluate the performance gain obtained by optimizing over the unitary matrix $\mathbb{V}$, detailed in Section IV-E. To that aim, we set $\mathbb{A}_{\mathrm{I}}^{\mathrm{UC}}$ and $\mathbb{A}_{\mathrm{I}}^{\mathrm{MF}}$ to be the matrices obtained via Propositions 1 and 2, respectively, with the unitary matrix $\mathbb{V}$ fixed to $\mathbb{I}_m$. In Table I we detail the values of Frobenius norm $\|\mathbb{V}\tilde{\mathbb{A}}' - \mathbb{S}_m(\mathbb{A}\otimes\mathbb{A}^*)\|$ computed for $\mathbb{A}_{\mathrm{I}}^{\mathrm{UC}}$ and $\mathbb{A}_{\mathrm{I}}^{\mathrm{MF}}$ with $\mathbb{V} = \mathbb{I}_m$, and for $\mathbb{A}^{\mathrm{UC}}$ and $\mathbb{A}^{\mathrm{MF}}$ with $\mathbb{V}$ obtained via (32), for $m = 6n$, SOI distribution $\mathbf{U}_S$, and $1/\sigma_W^2 = -10, 10, 30$ dB. We note that optimizing over the unitary transformation decreases the Frobenius norm by a factor of approximately 3.3 for $\mathbb{A}^{\mathrm{UC}}$ and 1.4 for $\mathbb{A}^{\mathrm{MF}}$. To illustrate that the Frobenius norm improvement translates into improvement in estimation performance, we depict in Fig. 7 the estimation error obtained with phasecut for the same setup for $1/\sigma_W^2 \in [-10, 30]$ dB. We observe that at $\epsilon = 0.1$ optimizing the unitary matrix yields an SNR gain of 4 dB for $\mathbb{A}^{\mathrm{UC}}$ compared to $\mathbb{A}_{\mathrm{I}}^{\mathrm{UC}}$, and a gain of 2 dB for $\mathbb{A}^{\mathrm{MF}}$ compared to $\mathbb{A}_{\mathrm{I}}^{\mathrm{MF}}$. Fig. 7 demonstrates the benefits of optimizing over $\mathbb{V}$ in Algorithm 1 rather than choosing a fixed $\mathbb{V}$.

The results of the simulation study indicate that significant performance gains can be achieved by the proposed measurement matrix design, for various recovery algorithms, using deterministic and practical measurement setups.

VI. CONCLUSIONS

In this paper we studied the design of measurement matrices for the noisy phase retrieval setup by maximizing the MI between the SOI and the observations. Necessary conditions on the optimal measurement matrix were derived, and the optimal measurement matrix for Kronecker symmetric SOI in the low SNR regime was obtained in closed-form. We also studied the design of practical measurement matrices based on maximizing the MI between the SOI and the observations, by applying a series of approximations. Simulation results demonstrate the benefits of using the proposed approach for various recovery algorithms.

APPENDIX

We first recall the definition of the Kronecker product:

Definition 2 (Kroncker product): For any $n_1 \times n_2$ matrix $\mathbb{N}$ and $m_1 \times m_2$ matrix $\mathbb{M}$, for every $p_1 \in \{1, 2, \ldots, n_1\}$, $p_2 \in \{1, 2, \ldots, n_2\}$, $q_1 \in \{1, 2, \ldots, m_1\}$, $q_2 \in \{1, 2, \ldots, m_2\}$, the entries of $\mathbb{N}\otimes\mathbb{M}$ are given by [34, Ch. 1.3.6]:

$$(\mathbb{N}\otimes\mathbb{M})_{(p_1-1)m_1+q_1,(p_2-1)m_2+q_2} = (\mathbb{N})_{p_1,p_2}(\mathbb{M})_{q_1,q_2}. \quad (34)$$

The following properties of the Kronecker product are repeatedly used in the sequel:

Lemma 3: The Kronecker product satisfies:

P1 For any $n_1^2 \times 1$ vector $\mathbf{x}_1$ and $n_1 \times 1$ vectors $\mathbf{x}_2, \mathbf{x}_3$:

$$\|\mathbf{x}_1 - \mathbf{x}_2\otimes\mathbf{x}_3^*\|^2 = \left\|\mathrm{vec}_{n_1}^{-1}(\mathbf{x}_1) - \mathbf{x}_3^*\mathbf{x}_2^T\right\|^2. \quad (35)$$

P2 For any $n \times 1$ vector $\mathbf{x}$ and $n^2 \times n^2$ matrix $\mathbb{M}$ we have that for every $k \in \mathcal{N}$,

$$\Big(\big(\mathbb{I}_n\otimes\mathbf{x}^T\big)\cdot\mathbb{M}\cdot(\mathbf{x}\otimes\mathbf{x}^*)\Big)_k = \sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\sum_{q_2=1}^{n}(\mathbf{x})_{q_2}(\mathbb{M})_{(k-1)n+q_2,(p_1-1)n+q_1}(\mathbf{x})_{p_1}(\mathbf{x})_{q_1}^*, \quad (36a)$$

and also

$$\Big(\big(\mathbf{x}^T\otimes\mathbb{I}_n\big)\cdot\mathbb{M}^*\cdot(\mathbf{x}^*\otimes\mathbf{x})\Big)_k = \sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\sum_{p_2=1}^{n}(\mathbf{x})_{p_2}(\mathbb{M})^*_{(p_2-1)n+k,(p_1-1)n+q_1}(\mathbf{x})^*_{p_1}(\mathbf{x})_{q_1}. \quad (36b)$$

Proof: Property P1 follows since

$$\left\|\mathbf{x}_1-\mathbf{x}_2\otimes\mathbf{x}_3^*\right\|^2 \overset{(a)}{=} \left\|\mathrm{vec}_{n_1}^{-1}\left(\mathbf{x}_1-\mathbf{x}_2\otimes\mathbf{x}_3^*\right)\right\|^2 \overset{(b)}{=} \left\|\mathrm{vec}_{n_1}^{-1}\left(\mathbf{x}_1\right)-\mathbf{x}_3^*\mathbf{x}_2^T\right\|^2, \quad (37)$$

where (a) follows from the relationship between the Frobenious norm and the Euclidean norm, as for any square matrix $\mathbb{X}$, $\|\mathbb{X}\|^2=\|\mathrm{vec}(\mathbb{X})\|^2$; (b) follows from [34, Ch. 12.3.4].

In the proof of Property P2, we detail only the proof of (36a), as the proof of (36b) follows using similar steps: By explicitly writing the product of the $n\times n^2$ matrix $(\mathbb{I}_n\otimes\mathbf{x}^T)\mathbb{M}$ and the $n^2\times 1$ vector $\mathbf{x}\otimes\mathbf{x}^*$ we have that

$$\begin{aligned}\Big(\left(\mathbb{I}_n\otimes\mathbf{x}^T\right)\cdot\mathbb{M}\cdot\left(\mathbf{x}\otimes\mathbf{x}^*\right)\Big)_k &= \sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\left(\left(\mathbb{I}_n\otimes\mathbf{x}^T\right)\cdot\mathbb{M}\right)_{k,(p_1-1)n+q_1}\left(\mathbf{x}\otimes\mathbf{x}^*\right)_{(p_1-1)n+q_1}\\ &= \sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\sum_{p_2=1}^{n}\sum_{q_2=1}^{n}\left(\mathbb{I}_n\otimes\mathbf{x}^T\right)_{k,(p_2-1)n+q_2}\\ &\quad\cdot\left(\mathbb{M}\right)_{(p_2-1)n+q_2,(p_1-1)n+q_1}\left(\mathbf{x}\otimes\mathbf{x}^*\right)_{(p_1-1)n+q_1}. \quad (38)\end{aligned}$$

Next, from (34) we have that $\left(\mathbb{I}_n\otimes\mathbf{x}^T\right)_{k,(p_2-1)n+q_2}=\left(\mathbb{I}_n\right)_{k,p_2}\cdot\left(\mathbf{x}\right)_{q_2}=\delta_{k,p_2}\left(\mathbf{x}\right)_{q_2}$ and $\left(\mathbf{x}\otimes\mathbf{x}^*\right)_{(p_1-1)n+q_1}=\left(\mathbf{x}\right)_{p_1}\cdot\left(\mathbf{x}\right)_{q_1}^*$. Substituting these computations back into (38) yields

$$\Big(\left(\mathbb{I}_n\otimes\mathbf{x}^T\right)\cdot\mathbb{M}\cdot\left(\mathbf{x}\otimes\mathbf{x}^*\right)\Big)_k = \sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\sum_{q_2=1}^{n}\left(\mathbf{x}\right)_{q_2}\left(\mathbb{M}\right)_{(k-1)n+q_2,(p_1-1)n+q_1}\left(\mathbf{x}\right)_{p_1}\left(\mathbf{x}\right)_{q_1}^*,$$

proving (36a). ■

A. Proof of Theorem 1

Applying the KKT theorem [40, Ch. 5.5.3] to the problem (3), we obtain the following necessary conditions for $\mathbb{A}^{\mathrm{MI}}$:

$$\nabla_{\mathbb{A}}\Big(-I\left(\mathbf{U};\mathbf{Y}\right)-\lambda\left(P-\mathrm{Tr}\left(\mathbb{A}\mathbb{A}^H\right)\right)\Big)\Big|_{\mathbb{A}=\mathbb{A}^{\mathrm{MI}}}=0, \quad (39a)$$

and

$$\lambda\Big(P-\mathrm{Tr}\left(\mathbb{A}^{\mathrm{MI}}\left(\mathbb{A}^{\mathrm{MI}}\right)^H\right)\Big)=0, \quad (39b)$$

where $\lambda\geq 0$. From (39a) it follows that for $\mathbb{A}=\mathbb{A}^{\mathrm{MI}}$

$$\begin{aligned}\nabla_{\mathbb{A}}\Big(I\left(\mathbf{U};\mathbf{Y}\right)\Big)\Big|_{\mathbb{A}=\mathbb{A}^{\mathrm{MI}}} &= \lambda\cdot\nabla_{\mathbb{A}}\Big(\mathrm{Tr}\left(\mathbb{A}\mathbb{A}^H\right)\Big)\Big|_{\mathbb{A}=\mathbb{A}^{\mathrm{MI}}}\\ &= \lambda\cdot\mathbb{A}^{\mathrm{MI}}. \quad (40)\end{aligned}$$

To determine the derivative of the left-hand side of (40), we use the chain rule for complex gradients [41, Ch. 4.1.1], from which we have that for every $k_1\in\mathcal{M}, k_2\in\mathcal{N}$,

$$\begin{aligned}\Big(\nabla_{\mathbb{A}}\Big(I\left(\mathbf{U};\mathbf{Y}\right)\Big)\Big)_{k_1,k_2} &= \mathrm{Tr}\left(\Big(\nabla_{\tilde{\mathbb{A}}}\Big(I\left(\mathbf{U};\mathbf{Y}\right)\Big)\Big)^T\frac{\partial\tilde{\mathbb{A}}^*}{\partial\left(\mathbb{A}\right)_{k_1,k_2}^*}\right)\\ &\quad+\mathrm{Tr}\left(\Big(\nabla_{\tilde{\mathbb{A}}^*}\Big(I\left(\mathbf{U};\mathbf{Y}\right)\Big)\Big)^T\frac{\partial\tilde{\mathbb{A}}}{\partial\left(\mathbb{A}\right)_{k_1,k_2}^*}\right). \quad (41)\end{aligned}$$

Next, we let $\mathbb{E}_C\left(\mathbb{A}\right)$ denote the MMSE matrix for estimating $\tilde{\mathbf{U}}$ from $\mathbf{Y}_C$, and note that (10) implies that

$$\begin{aligned}\nabla_{\tilde{\mathbb{A}}}\Big(I\left(\mathbf{U};\mathbf{Y}\right)\Big) &= \nabla_{\tilde{\mathbb{A}}}\Big(I\left(\tilde{\mathbf{U}};\mathbf{Y}_C\right)\Big)\\ &\overset{(a)}{=}\tilde{\mathbb{A}}\cdot\mathbb{E}_C\left(\mathbb{A}\right)\overset{(b)}{=}\tilde{\mathbb{A}}\cdot\mathbb{E}\left(\mathbb{A}\right), \quad (42)\end{aligned}$$

where (a) follows from [31, Eq. (4)], since the relationship between $\mathbf{Y}_C$ and $\tilde{\mathbf{U}}$ corresponds to a PC Gaussian MIMO channel with input $\tilde{\mathbf{U}}$ and output $\mathbf{Y}_C$; (b) follows since $\mathbf{W}_I=\mathrm{Im}\{\mathbf{Y}_C\}$ is independent of $\mathbf{Y}=\mathrm{Re}\{\mathbf{Y}_C\}$ and of $\tilde{\mathbf{U}}$, thus the MMSE matrix for estimating $\tilde{\mathbf{U}}$ from $\mathbf{Y}_C$, $\mathbb{E}_C\left(\mathbb{A}\right)$, is equal to the MMSE matrix for estimating $\tilde{\mathbf{U}}$ from $\mathbf{Y}$, $\mathbb{E}(\mathbb{A})$. As MI is real-valued, it follows from (42) and from the definition of the generalized complex derivative [41, Ch. 4.1.1] that

$$\nabla_{\tilde{\mathbb{A}}^*}\Big(I\left(\mathbf{U};\mathbf{Y}\right)\Big)=\Big(\tilde{\mathbb{A}}\cdot\mathbb{E}\left(\mathbb{A}\right)\Big)^*. \quad (43)$$

Plugging (42) and (43) into (41) results in

$$\begin{aligned}\Big(\nabla_{\mathbb{A}}\Big(I\left(\mathbf{U};\mathbf{Y}\right)\Big)\Big)_{k_1,k_2} &= \sum_{l_1=1}^{m}\sum_{l_2=1}^{n^2}\Big(\tilde{\mathbb{A}}\cdot\mathbb{E}\left(\mathbb{A}\right)\Big)_{l_1,l_2}\\ &\quad\frac{\partial(\tilde{\mathbb{A}})_{l_1,l_2}^*}{\partial\left(\mathbb{A}\right)_{k_1,k_2}^*}+\sum_{l_1=1}^{m}\sum_{l_2=1}^{n^2}\Big(\tilde{\mathbb{A}}\cdot\mathbb{E}\left(\mathbb{A}\right)\Big)_{l_1,l_2}^*\frac{\partial(\tilde{\mathbb{A}})_{l_1,l_2}}{\partial\left(\mathbb{A}\right)_{k_1,k_2}^*}. \quad (44)\end{aligned}$$

By writing the index l_2 as $l_2=(p_2-1)n+q_2$, where $p_2,q_2\in\mathcal{N}$, it follows from the definition of $\tilde{\mathbb{A}}$ in (6) that

$$\frac{\partial\big(\tilde{\mathbb{A}}\big)_{l_1,(p_2-1)n+q_2}^*}{\partial\left(\mathbb{A}\right)_{k_1,k_2}^*}=\left(\mathbb{A}\right)_{k_1,q_2}\delta_{l_1,k_1}\delta_{p_2,k_2}, \quad (45a)$$

and

$$\frac{\partial\big(\tilde{\mathbb{A}}\big)_{l_1,(p_2-1)n+q_2}}{\partial\left(\mathbb{A}\right)_{k_1,k_2}^*}=\left(\mathbb{A}\right)_{k_1,p_2}\delta_{l_1,k_1}\delta_{q_2,k_2}. \quad (45b)$$

Thus, (44) yields

$$\begin{aligned}\Big(\nabla_{\mathbb{A}}\Big(I(\mathbf{U};\mathbf{Y})\Big)\Big)_{k_1,k_2} &= \sum_{q_2=1}^{n}\Big(\tilde{\mathbb{A}}\cdot\mathbb{E}\left(\mathbb{A}\right)\Big)_{k_1,(k_2-1)n+q_2}\left(\mathbb{A}\right)_{k_1,q_2}\\ &\quad+\sum_{p_2=1}^{n}\Big(\tilde{\mathbb{A}}\cdot\mathbb{E}\left(\mathbb{A}\right)\Big)_{k_1,(p_2-1)n+k_2}^*\left(\mathbb{A}\right)_{k_1,p_2}. \quad (46)\end{aligned}$$

Next, we note that

$$\begin{aligned}&\left(\tilde{\mathbb{A}}\cdot\mathbb{E}\left(\mathbb{A}\right)\right)_{k_1,(p_2-1)n+q_2}\\&=\sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\left(\tilde{\mathbb{A}}\right)_{k_1,(p_1-1)n+q_1}\left(\mathbb{E}\left(\mathbb{A}\right)\right)_{(p_1-1)n+q_1,(p_2-1)n+q_2}\\&\overset{(a)}{=}\sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\left(\mathbb{A}\right)_{k_1,p_1}\left(\mathbb{A}\right)^{*}_{k_1,q_1}\left(\mathbb{E}\left(\mathbb{A}\right)\right)_{(p_1-1)n+q_1,(p_2-1)n+q_2},\end{aligned}\tag{47}$$

where (a) follows from the definition of $\tilde{\mathbb{A}}$ in (6). Plugging (46) and (47) into (40), we conclude that the entries of the optimal measurement matrix $\mathbb{A}^{\mathrm{MI}}$ satisfy

$$\begin{aligned}\lambda\cdot\left(\mathbb{A}^{\mathrm{MI}}\right)_{k_1,k_2}&=\sum_{q_2=1}^{n}\sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\left(\mathbb{A}^{\mathrm{MI}}\right)_{k_1,p_1}\left(\mathbb{A}^{\mathrm{MI}}\right)^{*}_{k_1,q_1}\\&\quad\left(\mathbb{A}^{\mathrm{MI}}\right)_{k_1,q_2}\cdot\left(\mathbb{E}\left(\mathbb{A}^{\mathrm{MI}}\right)\right)_{(p_1-1)n+q_1,(k_2-1)n+q_2}\\&\quad+\sum_{p_2=1}^{n}\sum_{p_1=1}^{n}\sum_{q_1=1}^{n}\left(\mathbb{A}^{\mathrm{MI}}\right)^{*}_{k_1,p_1}\left(\mathbb{A}^{\mathrm{MI}}\right)_{k_1,q_1}\\&\quad\left(\mathbb{A}^{\mathrm{MI}}\right)_{k_1,p_2}\cdot\left(\mathbb{E}\left(\mathbb{A}^{\mathrm{MI}}\right)\right)^{*}_{(p_1-1)n+q_1,(p_2-1)n+k_2},\end{aligned}\tag{48}$$

where λ is set to satisfy the power constraint.

We now use Property P2 of Lemma 3 to express (48) in vector form. Letting $\mathbf{a}_k^{\mathrm{MI}}$ denote the k-th column of $(\mathbb{A}^{\mathrm{MI}})^T$, we note that the first and second summands in the right hand side of (48) correspond to (36a) and (36b), respectively, with $\mathbf{x}=\mathbf{a}_{k_1}^{\mathrm{MI}}$ and $\mathbb{M}=\mathbb{E}^T(\mathbb{A}^{\mathrm{MI}})$. Thus, (48) can be written as

$$\begin{aligned}&\lambda\cdot\left(\mathbb{A}^{\mathrm{MI}}\right)_{k_1,k_2}\\&=\left(\left(\mathbb{I}_n\otimes\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^T\right)\cdot\mathbb{E}^T\left(\mathbb{A}^{\mathrm{MI}}\right)\cdot\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\otimes\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^*\right)\right)_{k_2}\\&+\left(\left(\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^T\otimes\mathbb{I}_n\right)\cdot\mathbb{E}^H\left(\mathbb{A}^{\mathrm{MI}}\right)\cdot\left(\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^*\otimes\mathbf{a}_{k_1}^{\mathrm{MI}}\right)\right)_{k_2}.\end{aligned}\tag{49}$$

Consequently, as the MMSE matrix is Hermitian, we have

$$\begin{aligned}\lambda\cdot\mathbf{a}_{k_1}^{\mathrm{MI}}&=\left(\mathbb{I}_n\otimes\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^T\right)\cdot\mathbb{E}^T\left(\mathbb{A}^{\mathrm{MI}}\right)\cdot\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\otimes\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^*\right)\\&\quad+\left(\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^T\otimes\mathbb{I}_n\right)\cdot\mathbb{E}\left(\mathbb{A}^{\mathrm{MI}}\right)\cdot\left(\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^*\otimes\mathbf{a}_{k_1}^{\mathrm{MI}}\right)\\&=\left(\left(\mathbb{I}_n\otimes\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^T\right)\cdot\mathbb{E}^T\left(\mathbb{A}^{\mathrm{MI}}\right)\cdot\left(\mathbb{I}_n\otimes\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^*\right)\right.\\&\quad\left.+\left(\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^T\otimes\mathbb{I}_n\right)\cdot\mathbb{E}\left(\mathbb{A}^{\mathrm{MI}}\right)\cdot\left(\left(\mathbf{a}_{k_1}^{\mathrm{MI}}\right)^*\otimes\mathbb{I}_n\right)\right)\mathbf{a}_{k_1}^{\mathrm{MI}}\\&=\mathbb{H}_{k_1}\left(\mathbb{A}^{\mathrm{MI}}\right)\cdot\mathbf{a}_{k_1}^{\mathrm{MI}},\qquad k_1\in\mathcal{M},\end{aligned}\tag{50}$$

proving the theorem. ■

B. Proof of Lemma 1

We first write the indexes $k_1,k_2\in\{1,2,\ldots,n^2\}$ as $k_1=(p_1-1)n+q_1$ and $k_2=(p_2-1)n+q_2$, where $p_1,p_2,q_1,q_2\in\mathcal{N}$. Using (34), the entries of the covariance matrix of $\mathbf{X}\otimes\mathbf{X}^*$, denoted $\mathbb{C}_{\mathbf{X}\otimes\mathbf{X}^*}$, can then be written as

$$\begin{aligned}&\left(\mathbb{C}_{\mathbf{X}\otimes\mathbf{X}^*}\right)_{(p_1-1)n+q_1,(p_2-1)n+q_2}\\&\quad=\mathcal{E}\left\{\left(\mathbf{X}\right)_{p_1}\left(\mathbf{X}\right)^*_{q_1}\left(\mathbf{X}\right)^*_{p_2}\left(\mathbf{X}\right)_{q_2}\right\}\\&\qquad-\mathcal{E}\left\{\left(\mathbf{X}\right)_{p_1}\left(\mathbf{X}\right)^*_{q_1}\right\}\mathcal{E}\left\{\left(\mathbf{X}\right)^*_{p_2}\left(\mathbf{X}\right)_{q_2}\right\}\\&\quad\overset{(a)}{=}\mathcal{E}\left\{\left(\mathbf{X}\right)_{p_1}\left(\mathbf{X}\right)^*_{q_1}\right\}\mathcal{E}\left\{\left(\mathbf{X}\right)^*_{p_2}\left(\mathbf{X}\right)_{q_2}\right\}\\&\qquad+\mathcal{E}\left\{\left(\mathbf{X}\right)_{p_1}\left(\mathbf{X}\right)^*_{p_2}\right\}\mathcal{E}\left\{\left(\mathbf{X}\right)^*_{q_1}\left(\mathbf{X}\right)_{q_2}\right\}\\&\qquad+\mathcal{E}\left\{\left(\mathbf{X}\right)_{p_1}\left(\mathbf{X}\right)_{q_2}\right\}\mathcal{E}\left\{\left(\mathbf{X}\right)^*_{p_2}\left(\mathbf{X}\right)^*_{p_1}\right\}\\&\qquad-\mathcal{E}\left\{\left(\mathbf{X}\right)_{p_1}\left(\mathbf{X}\right)^*_{q_1}\right\}\mathcal{E}\left\{\left(\mathbf{X}\right)^*_{p_2}\left(\mathbf{X}\right)_{q_2}\right\}\\&\quad\overset{(b)}{=}\mathcal{E}\left\{\left(\mathbf{X}\right)_{p_1}\left(\mathbf{X}\right)^*_{p_2}\right\}\mathcal{E}\left\{\left(\mathbf{X}\right)^*_{q_1}\left(\mathbf{X}\right)_{q_2}\right\}\\&\quad=\left(\mathbb{C}_{\mathbf{X}}\right)_{p_1,p_2}\left(\mathbb{C}_{\mathbf{X}}\right)^*_{q_1,q_2}\\&\quad\overset{(c)}{=}\left(\mathbb{C}_{\mathbf{X}}\otimes\mathbb{C}^*_{\mathbf{X}}\right)_{(p_1-1)n+q_1,(p_2-1)n+q_2},\end{aligned}\tag{51}$$

where (a) follows from Isserlis theorem for complex Gaussian random vectors [42, Ch. 1.4]; (b) follows from the proper complexity of $\mathbf{X}$, which implies that $\mathcal{E}\{(\mathbf{X})_{p_1}(\mathbf{X})_{q_2}\}\mathcal{E}\{(\mathbf{X})^*_{p_2}(\mathbf{X})^*_{p_1}\}=0$; and (c) follows from (34). Eq. (51) proves the lemma. ■

C. Proof of Theorem 2

To solve the optimization problem (14), we employ the following auxiliary lemma:

Lemma 4: Let $\mathbf{a}_k$ be the k-th column of $\mathbb{A}^T$, $k\in\mathcal{M}$. If $\mathbf{U}$ is Kronecker symmetric with covariance matrix $\mathbb{C}_{\mathbf{U}}$, then

$$\mathrm{Tr}\left(\tilde{\mathbb{A}}\mathbb{C}_{\tilde{\mathbf{U}}}\tilde{\mathbb{A}}^H\right)=\sum_{k=1}^{m}\left(\mathbf{a}_k^H\mathbb{C}^*_{\mathbf{U}}\mathbf{a}_k\right)^2.\tag{52}$$

Proof: Using Def. 1 and the representation (8) it follows that

$$\begin{aligned}&\mathrm{Tr}\left(\tilde{\mathbb{A}}\mathbb{C}_{\tilde{\mathbf{U}}}\tilde{\mathbb{A}}^H\right)\\&\quad=\mathrm{Tr}\left(\mathbb{S}_m\left(\mathbb{A}\otimes\mathbb{A}^*\right)\cdot\left(\mathbb{C}_{\mathbf{U}}\otimes\mathbb{C}^*_{\mathbf{U}}\right)\cdot\left(\mathbb{A}^H\otimes\mathbb{A}^T\right)\mathbb{S}_m^H\right)\\&\quad\overset{(a)}{=}\mathrm{Tr}\left(\mathbb{S}_m^H\mathbb{S}_m\left(\left(\mathbb{A}\mathbb{C}_{\mathbf{U}}\mathbb{A}^H\right)\otimes\left(\mathbb{A}\mathbb{C}_{\mathbf{U}}\mathbb{A}^H\right)^*\right)\right),\end{aligned}\tag{53}$$

where (a) follows from the properties of the trace operator [41, Ch. 1.1] and the Kronecker product [41, Ch. 10.2]. Note that $\mathbb{S}_m^H\mathbb{S}_m$ is an $m^2\times m^2$ diagonal matrix which satisfies $(\mathbb{S}_m^H\mathbb{S}_m)_{l,l}=1$ if $l=(k-1)m+k$ for some $k\in\mathcal{M}$ and

$(\mathbb{S}_m^H \mathbb{S}_m)_{l,l} = 0$ otherwise. Therefore, (53) can be written as

$$\begin{aligned}\mathrm{Tr}\left(\tilde{\mathbb{A}}\mathbb{C}_{\tilde{\mathbf{U}}}\tilde{\mathbb{A}}^H\right) &= \sum_{k=1}^{m}\Big(\left(\mathbb{A}\mathbb{C}_{\mathbf{U}}\mathbb{A}^H\right) \\ &\qquad \otimes \left(\mathbb{A}\mathbb{C}_{\mathbf{U}}\mathbb{A}^H\right)^*\Big)_{(k-1)m+k,(k-1)m+k} \\ &\overset{(a)}{=} \sum_{k=1}^{m}\left|\mathbf{a}_k^T\mathbb{C}_{\mathbf{U}}\mathbf{a}_k^*\right|^2 \overset{(b)}{=} \sum_{k=1}^{m}\left(\mathbf{a}_k^H\mathbb{C}_{\mathbf{U}}^*\mathbf{a}_k\right)^2, \quad (54)\end{aligned}$$

where (a) follows from (34) and from the definition of $\mathbf{a}_k$ as the k-th column of $\mathbb{A}^T$, and (b) follows since $\mathbb{C}_{\mathbf{U}}$ is Hermitian and positive semi-definite. ■

Using Lemma 4, (14) can be written as

$$\begin{aligned}\mathbb{A}^{\mathrm{MI}} &= \left[\mathbf{a}_1^{\mathrm{MI}}, \mathbf{a}_2^{\mathrm{MI}}, \ldots, \mathbf{a}_m^{\mathrm{MI}}\right]^T \\ &= \underset{\{\mathbf{a}_k\}_{k=1}^m : \sum_{k=1}^m \|\mathbf{a}_k\|^2 \le P}{\arg\max} \sum_{k=1}^{m}\left(\mathbf{a}_k^H\mathbb{C}_{\mathbf{U}}^*\mathbf{a}_k\right)^2 \\ &= \underset{\{\mathbf{a}_k\}_{k=1}^m : \sum_{k=1}^m \|\mathbf{a}_k\|^2 \le P}{\arg\max} \sum_{k=1}^{m}\left(\frac{\mathbf{a}_k^H\mathbb{C}_{\mathbf{U}}^*\mathbf{a}_k}{\|\mathbf{a}_k\|}\right)^2\|\mathbf{a}_k\|^2. \quad (55)\end{aligned}$$

The maximal value of the ratio $\frac{\mathbf{a}_k^H\mathbb{C}_{\mathbf{U}}^*\mathbf{a}_k}{\|\mathbf{a}_k\|}$ is the largest eigenvalue of $\mathbb{C}_{\mathbf{U}}^*$, denoted $\mu_{\max}$. This maximum is obtained by setting $\frac{\mathbf{a}_k}{\|\mathbf{a}_k\|} = e^{j2\pi\phi_k}\mathbf{v}_{\max}^*$, where $\mathbf{v}_{\max}^*$ is the eigenvector of $\mathbb{C}_{\mathbf{U}}^*$ corresponding to $\mu_{\max}$, for any real ϕ_k [43, p. 550]. Thus,

$$\sum_{k=1}^{m}\left(\frac{\mathbf{a}_k^H\mathbb{C}_{\mathbf{U}}^*\mathbf{a}_k}{\|\mathbf{a}_k\|}\right)^2\|\mathbf{a}_k\|^2 \le \mu_{\max}^2\sum_{k=1}^{m}\|\mathbf{a}_k\|^2 \le \mu_{\max}^2 P. \quad (56)$$

It follows from (56) that any selection of $\{\mathbf{a}_k\}_{k=1}^m$ such that $\mathbf{a}_k = (\mathbf{c})_k\mathbf{v}_{\max}^*$ and $\sum_{k=1}^{m}|(\mathbf{c})_k|^2 = P$ solves (55). As $\mathbb{C}_{\mathbf{U}}^*$ is Hermitian positive semi-definite, it follows that $\mu_{\max}$ is also the largest eigenvalue of $\mathbb{C}_{\mathbf{U}}^*$, and that its corresponding eigenvector is $\mathbf{v}_{\max}$, thus proving the theorem. ■

D. Proof of Corollary 1

In order to prove the corollary we show that if the MMSE matrix is replaced by the LMMSE matrix $\mathbb{E}_L\left(\tilde{\mathbb{A}}\right)$, then $\tilde{\mathbb{A}}'$ in (22) satisfies the conditions of Lemma 2, namely, $\mathbb{V}_{\tilde{\mathbf{U}}}$ diagonalizes $\mathbb{E}_L\left(\tilde{\mathbb{A}}'\right)$ and $\tilde{\mathbb{D}}_A$ satisfies (19).

Using (22) it follows that $\mathbb{E}_L\left(\tilde{\mathbb{A}}'\right)$ is given by

$$\begin{aligned}\mathbb{E}_L\big(\tilde{\mathbb{A}}'\big) &= \mathbb{C}_{\tilde{\mathbf{U}}} - \mathbb{C}_{\tilde{\mathbf{U}}}\mathbb{V}_{\tilde{\mathbf{U}}}\tilde{\mathbb{D}}_A^T\big(2\sigma_W^2\mathbb{I}_m + \tilde{\mathbb{D}}_A\mathbb{V}_{\tilde{\mathbf{U}}}^H\mathbb{C}_{\tilde{\mathbf{U}}}\mathbb{V}_{\tilde{\mathbf{U}}}\tilde{\mathbb{D}}_A^T\big)^{-1} \\ &\qquad \cdot \tilde{\mathbb{D}}_A\mathbb{V}_{\tilde{\mathbf{U}}}^H\mathbb{C}_{\tilde{\mathbf{U}}} \\ &= \mathbb{C}_{\tilde{\mathbf{U}}} - \mathbb{C}_{\tilde{\mathbf{U}}}\mathbb{V}_{\tilde{\mathbf{U}}}\tilde{\mathbb{D}}_A^T\left(2\sigma_W^2\mathbb{I}_m + \tilde{\mathbb{D}}_A\mathbb{D}_{\tilde{\mathbf{U}}}\tilde{\mathbb{D}}_A^T\right)^{-1} \\ &\qquad \cdot \tilde{\mathbb{D}}_A\mathbb{V}_{\tilde{\mathbf{U}}}^H\mathbb{C}_{\tilde{\mathbf{U}}}. \quad (57)\end{aligned}$$

From (57) it follows that $\mathbb{E}_L\left(\tilde{\mathbb{A}}'\right)$ is diagonalized by $\mathbb{V}_{\tilde{\mathbf{U}}}$, and the eigenvalue matrix is the diagonal matrix given by

$$\begin{aligned}&\mathbb{V}_{\tilde{\mathbf{U}}}^H\mathbb{E}_L\left(\tilde{\mathbb{A}}'\right)\mathbb{V}_{\tilde{\mathbf{U}}} \\ &= \mathbb{D}_{\tilde{\mathbf{U}}} - \mathbb{D}_{\tilde{\mathbf{U}}}\tilde{\mathbb{D}}_A^T\left(2\sigma_W^2\mathbb{I}_m + \tilde{\mathbb{D}}_A\mathbb{D}_{\tilde{\mathbf{U}}}\tilde{\mathbb{D}}_A^T\right)^{-1}\tilde{\mathbb{D}}_A\mathbb{D}_{\tilde{\mathbf{U}}}. \quad (58)\end{aligned}$$

In order to satisfy (19), for all $k \in \mathcal{M}$, $\left(\tilde{\mathbb{D}}_A\right)_{k,k}$ must be non-negative, and if $\left(\tilde{\mathbb{D}}_A\right)_{k,k} > 0$, then from (58):

$$\eta = \left(\mathbb{D}_{\tilde{\mathbf{U}}}\right)_{k,k} - \frac{\left(\mathbb{D}_{\tilde{\mathbf{U}}}\right)_{k,k}^2\left(\tilde{\mathbb{D}}_A\right)_{k,k}^2}{2\sigma_W^2 + \left(\tilde{\mathbb{D}}_A\right)_{k,k}^2\left(\mathbb{D}_{\tilde{\mathbf{U}}}\right)_{k,k}}. \quad (59)$$

Extracting $\left(\tilde{\mathbb{D}}_A\right)_{k,k}^2$ from (59) and setting $\tilde{\eta} \triangleq \frac{2\sigma_W^2}{\eta}$ yields (21), and concludes the proof. ■

E. Proof of Proposition 1

Letting $\mathbf{a}_k$ be the k-th column of $\mathbb{A}^T$, $k \in \mathcal{M}$, we note that

$$\|\mathbb{V}\tilde{\mathbb{A}}' - \mathbb{S}_m\left(\mathbb{A}\otimes\mathbb{A}^*\right)\|^2 = \sum_{k=1}^{m}\|\tilde{\mathbf{a}}_k' - \mathbf{a}_k\otimes\mathbf{a}_k^*\|^2. \quad (60)$$

Therefore, the solution to the nearest row-wise KRP problem (24) is given by the solutions to the m nearest Kronecker product problems, i.e., for any $k \in \mathcal{M}$,

$$\begin{aligned}\hat{\mathbf{a}}_k^{\mathrm{O}} &= \underset{\mathbf{a}_k\in\mathcal{C}^n}{\arg\min}\|\tilde{\mathbf{a}}_k' - \mathbf{a}_k\otimes\mathbf{a}_k^*\|^2 \\ &\overset{(a)}{=} \underset{\mathbf{a}_k\in\mathcal{C}^n}{\arg\min}\left\|\mathrm{vec}_n^{-1}\left(\tilde{\mathbf{a}}_k'\right) - \mathbf{a}_k^*\mathbf{a}_k^T\right\|^2, \quad (61)\end{aligned}$$

where (a) follows from (35). Solving (61) is facilitated by the following Lemma:

Lemma 5: For an $n\times n$ matrix $\mathbb{X}$ with Hermitian part $\mathbb{M}_X$, it holds that

$$\underset{\mathbf{v}\in\mathcal{C}^n}{\mathrm{argmin}}\left\|\mathbb{X} - \mathbf{v}^*\mathbf{v}^T\right\|^2 = \underset{\mathbf{v}\in\mathcal{C}^n}{\mathrm{argmin}}\left\|\mathbb{M}_X - \mathbf{v}^*\mathbf{v}^T\right\|^2. \quad (62)$$

Proof: We note that since $\|\mathbb{B}\|^2 = \mathrm{Tr}(\mathbb{B}\mathbb{B}^H)$, then

$$\begin{aligned}\left\|\mathbb{X} - \mathbf{v}^*\mathbf{v}^T\right\|^2 &= \|\mathbb{X}\|^2 + \left\|\mathbf{v}^*\mathbf{v}^T\right\|^2 - \mathbf{v}^T\left(\mathbb{X}+\mathbb{X}^H\right)\mathbf{v}^* \\ &\overset{(a)}{=} \|\mathbb{X}\|^2 + \left\|\mathbf{v}^*\mathbf{v}^T\right\|^2 - 2\mathbf{v}^T\mathbb{M}_X\mathbf{v}^*, \quad (63)\end{aligned}$$

where (a) follows since $\mathbb{M}_X = \frac{1}{2}\left(\mathbb{X}+\mathbb{X}^H\right)$. Applying the argmin operation to (63) proves the lemma. ■

From Lemma 5 it follows that (61) is equivalent to

$$\begin{aligned}\hat{\mathbf{a}}_k^{\mathrm{O}} &= \underset{\mathbf{a}_k\in\mathcal{C}^n}{\arg\min}\|\tilde{\mathbb{M}}_k^{(H)} - \mathbf{a}_k^*\mathbf{a}_k^T\|^2, \quad (64) \\ &= \underset{\mathbf{a}_k\in\mathcal{C}^n}{\arg\min}\left(\|\tilde{\mathbb{M}}_k^{(H)}\|^2 + \|\mathbf{a}_k^*\mathbf{a}_k^T\|^2 - 2\mathbf{a}_k^T\tilde{\mathbb{M}}_k^{(H)}\mathbf{a}_k^*\right) \\ &\overset{(a)}{=} \underset{\mathbf{a}_k\in\mathcal{C}^n}{\arg\min}\left(\|\mathbf{a}_k^*\mathbf{a}_k^T\|^2 - 2\mathbf{a}_k^H\left(\tilde{\mathbb{M}}_k^{(H)}\right)^*\mathbf{a}_k\right), \quad (65)\end{aligned}$$

where (a) follows since $\tilde{\mathbb{M}}_k^{(H)}$ does not depend on $\mathbf{a}_k$, and since $\mathbf{a}_k^T\tilde{\mathbb{M}}_k^{(H)}\mathbf{a}_k^*$ is real valued [43, p. 549]. Since the rank one

Hermitian matrix $\mathbf{a}_k^* \mathbf{a}_k^T$ is positive semi-definite, the Eckart-Young theorem [34, Th. 2.4.8] cannot be used to solve (64). Consequently, we compute the gradient of the right hand side of (65) w.r.t. $\mathbf{a}_k$ and set it to zero. This results in

$$2 \left\| \mathbf{a}_k \right\|^2 \mathbf{a}_k - 2 \left(\tilde{\mathbb{M}}_k^{(H)} \right)^* \mathbf{a}_k = 0. \tag{66}$$

In order to satisfy (66), $\hat{\mathbf{a}}_k^{\mathrm{O}}$ must be either the zero vector or an eigenvector of the Hermitian matrix $\left(\tilde{\mathbb{M}}_k^{(H)} \right)^*$ with a non-negative eigenvalue. Specifically, for any non-negative eigenvalue $\tilde{\mu}_k^p$ of $\tilde{\mathbb{M}}_k^{(H)}$ and its corresponding unit-norm eigenvector $\tilde{\mathbf{v}}_k^p$, we have that $\left(\tilde{\mathbf{v}}_k^p \right)^*$ is an eigenvector of $\left(\tilde{\mathbb{M}}_k^{(H)} \right)^*$ with eigenvalue $\tilde{\mu}_k^p$, and thus (66) is satisfied by $\mathbf{a}_k^p = \sqrt{\tilde{\mu}_k^p} \cdot \left(\tilde{\mathbf{v}}_k^p \right)^*$, $p \in \mathcal{N}$. In order to select the eigenvalue-eigenvector pair which minimizes the Frobenious norm, we plug $\mathbf{a}_k^p$ into the right hand side of (65), which results in

$$\begin{aligned} &\left\| \mathbf{a}_k^p \right\|^4 - 2 \left(\mathbf{a}_k^p \right)^H \left(\tilde{\mathbb{M}}_k^{(H)} \right)^* \mathbf{a}_k^p \\ &\quad = \left(\tilde{\mu}_k^p \right)^2 - 2 \left(\tilde{\mu}_k^p \right)^2 = - \left(\tilde{\mu}_k^p \right)^2 . \end{aligned} \tag{67}$$

Note that (67) is minimized by the largest eigenvalue. Thus, when some eigenvalues are non-negative then the expression (65) is minimized by taking the largest non-negative eigenvalue. When all the eigenvalues are negative, $\left(\tilde{\mathbb{M}}_k^{(H)} \right)^*$ is negative definite. In this case, the expression in (65) is strictly non-negative, hence its minimal value is obtained by setting $\mathbf{a}_k$ to be the all-zero vector. Consequently, $\hat{\mathbf{a}}_k^{\mathrm{O}} = \sqrt{\max(\tilde{\mu}_{k,\max}, 0)} \cdot \tilde{\mathbf{v}}_{k,\max}^*$. ■

F. Proof of Proposition 2

Let $\mathbf{a}_q$ be the q-th column of $\mathbb{A}^T$, and recall that $m = b \cdot n$. When $\mathbb{A}$ corresponds to a masked Fourier measurement matrix (4) we have that the right hand side of (65), which results in

$$\begin{aligned} &\left\| \mathbb{V} \tilde{\mathbb{A}}' - \mathbb{S}_m \left(\mathbb{A} \otimes \mathbb{A}^* \right) \right\|^2 \\ &= \sum_{l=1}^{b} \sum_{k=1}^{n} \left\| \tilde{\mathbf{a}}'_{(l-1)n+k} - \mathbf{a}_{(k-1)n+p} \otimes \mathbf{a}^*_{(l-1)n+k} \right\|^2 \\ &\overset{(a)}{=} \sum_{l=1}^{b} \sum_{k=1}^{n} \left\| \mathrm{vec}_n^{-1} \left(\tilde{\mathbf{a}}'_{(l-1)n+k} \right) - \mathbf{a}^*_{(k-1)n+p} \mathbf{a}^T_{(l-1)n+k} \right\|^2 \\ &\overset{(b)}{=} \sum_{l=1}^{b} \sum_{k=1}^{n} \left\| \mathrm{vec}_n^{-1} \left(\tilde{\mathbf{a}}'_{(l-1)n+k} \right) - \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \right\|^2 , \end{aligned} \tag{68}$$

where (a) follows from (35); (b) follows from (4) since $\mathbf{a}_{(l-1)n+k} = \tilde{\mathbb{F}}_k \mathbf{g}_l$. From (68), in order to minimize the Frobenious norm, the mask vectors $\mathbf{g}_l^{\mathrm{MF}}$ should satisfy

$$\mathbf{g}_l^{\mathrm{MF}} = \underset{\mathbf{g}_l \in \mathcal{C}^n}{\mathrm{argmin}} \sum_{k=1}^{n} \left\| \mathrm{vec}_n^{-1} \left(\tilde{\mathbf{a}}'_{(l-1)n+k} \right) - \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \right\|^2 . \tag{69}$$

As $\tilde{\mathbb{M}}_{(l-1)n+k}^{(H)}$ is the Hermitian part of $\mathrm{vec}_n^{-1} \left(\tilde{\mathbf{a}}'_{(l-1)n+k} \right)$, it follows from Lemma 5 and (69) that $\mathbf{g}_l^{\mathrm{MF}}$ can be obtained from

$$\begin{aligned} \mathbf{g}_l^{\mathrm{MF}} &= \underset{\mathbf{g}_l \in \mathcal{C}^n}{\mathrm{argmin}} \sum_{k=1}^{n} \left\| \tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} - \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \right\|^2 \\ &\overset{(a)}{=} \underset{\mathbf{g}_l \in \mathcal{C}^n}{\mathrm{argmin}} \sum_{k=1}^{n} \left\| \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \right\|^2 \\ &\qquad - 2 \mathbf{g}_l^H \tilde{\mathbb{F}}_k^* \left(\tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} \right)^* \tilde{\mathbb{F}}_k \mathbf{g}_l , \end{aligned} \tag{70}$$

where (a) follows from the same arguments as those leading to (65). Next, we recall that the diagonal elements of $\tilde{\mathbb{F}}_k$ are in fact the k-th row of $\mathbb{F}_n$, hence $\tilde{\mathbb{F}}_k \tilde{\mathbb{F}}_k^* = \frac{1}{n} \mathbb{I}_n$. Therefore,

$$\begin{aligned} \left\| \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \right\|^2 &= \mathrm{Tr} \left(\tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \right) \\ &= \mathrm{Tr} \left(\mathbf{g}_l^T \tilde{\mathbb{F}}_k \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \mathbf{g}_l^T \tilde{\mathbb{F}}_k \tilde{\mathbb{F}}_k^* \mathbf{g}_l^* \right) = \frac{1}{n^2} \left\| \mathbf{g}_l \right\|^4 . \end{aligned}$$

Plugging this into (70) yields

$$\begin{aligned} \mathbf{g}_l^{\mathrm{MF}} &= \underset{\mathbf{g}_l \in \mathcal{C}^n}{\mathrm{argmin}} \sum_{k=1}^{n} \frac{1}{n^2} \left\| \mathbf{g}_l \right\|^4 - 2 \sum_{k=1}^{n} \mathbf{g}_l^H \tilde{\mathbb{F}}_k^* \left(\tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} \right)^* \tilde{\mathbb{F}}_k \mathbf{g}_l \\ &= \underset{\mathbf{g}_l \in \mathcal{C}^n}{\mathrm{argmin}} \frac{\left\| \mathbf{g}_l \right\|^4}{n} - 2 \mathbf{g}_l^H \left(\sum_{k=1}^{n} \tilde{\mathbb{F}}_k \tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} \tilde{\mathbb{F}}_k^* \right)^* \mathbf{g}_l . \end{aligned} \tag{71}$$

In order to find the minimizing vector, we compute the gradient of the right hand side of (71) with respect to $\mathbf{g}_l$ and equate it to zero, which results in

$$\frac{2}{n} \left\| \mathbf{g}_l \right\|^2 \mathbf{g}_l - 2 \left(\sum_{k=1}^{n} \tilde{\mathbb{F}}_k \tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} \tilde{\mathbb{F}}_k^* \right)^* \mathbf{g}_l = 0. \tag{72}$$

In order to satisfy (72), $\mathbf{g}_l^{\mathrm{MF}}$ must be an eigenvector of the $n \times n$ Hermitian matrix $\left(\sum_{k=1}^{n} \tilde{\mathbb{F}}_k \tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} \tilde{\mathbb{F}}_k^* \right)^*$ with a non-negative eigenvalue, and specifically, for any non-negative eigenvalue $\bar{\mu}_l^p$ of $\sum_{k=1}^{n} \tilde{\mathbb{F}}_k \tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} \tilde{\mathbb{F}}_k^*$ and its corresponding unit-norm eigenvector $\bar{\mathbf{v}}_l^p$, (72) is satisfied by $\mathbf{g}_l^p = \sqrt{n \bar{\mu}_l^p} \cdot \left(\bar{\mathbf{v}}_l^p \right)^*$, $p \in \mathcal{N}$. In order to characterize the vector $\mathbf{g}_l$ which minimizes the Frobenious norm, we plug $\mathbf{g}_l^p$ into the right hand side of (71), which results in

$$\begin{aligned} &\frac{1}{n} \left\| \mathbf{g}_l^p \right\|^4 - 2 \left(\mathbf{g}_l^p \right)^H \left(\sum_{k=1}^{n} \tilde{\mathbb{F}}_k \tilde{\mathbb{M}}_{(l-1)n+k}^{(H)} \tilde{\mathbb{F}}_k^* \right)^* \mathbf{g}_l^p \\ &\qquad = \frac{1}{n} \left(n \bar{\mu}_l^p \right)^2 - 2n \left(\bar{\mu}_l^p \right)^2 = -n \left(\bar{\mu}_l^p \right)^2 . \end{aligned} \tag{73}$$

Note that (73) is minimized by the largest eigenvalue. Thus, when some eigenvalues are non-negative then the expression (71) is minimized by taking the largest non-negative eigenvalue. When all the eigenvalues are negative, it follows that

$\left(\sum_{k=1}^{n} \tilde{\mathbb{F}}_k \tilde{\mathbb{M}}^{(H)}_{(l-1)n+k} \tilde{\mathbb{F}}^*_k\right)^*$ is negative definite. In this case, the expression in (71) is strictly non-negative, hence its minimal value is obtained by setting $\mathbf{g}_l$ to be the all-zero vector. Consequently, $\mathbf{g}_l^{\mathrm{MF}} = \sqrt{n \cdot \max(\bar{\mu}_{l,\max}, 0)} \cdot \bar{\mathbf{v}}^*_{l,\max}$. ■

REFERENCES

[1] R. W. Harrison, "Phase problem in crystallography," *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1046–1055, May 1993.

[2] O. Bunk *et al.*, "Diffractive imaging for periodic samples: Retrieving one-dimensional concentration profiles across microfluidic channels," *Acta Crystallogr. Section A, Found. Crystallogr.*, vol. 63, no. 4, pp. 306–314, 2007.

[3] J. C. Dainty and J. R. Fienup, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery: Theory and Application*, H. Stark, Ed. New York, NY, USA: Academic, 1987, pp. 231–275.

[4] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, "Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes," *Annu. Rev. Phys. Chem.*, vol. 59, pp. 387–410, May 2008.

[5] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 56, no. 2, pp. 87–109, May 2015.

[6] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[7] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, Aug. 1982, pp. 2758–2769.

[8] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Rev.*, vol. 57, no. 2, pp. 225–251, May 2015.

[9] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Math. Programm.*, vol. 149, no. 1, pp. 47–81, Feb. 2015.

[10] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.

[11] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Trans. Inf. Theory*, to be published.

[12] E. J. Candes and X. Li, "Solving quadratic equations via phaselift when there are about as many equations as unknowns," *Found. Comput. Math.*, vol. 14, no. 5, pp. 1017–1026, Oct. 2012.

[13] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Appl. Comput. Harmon. Anal.*, vol. 39, no. 2, pp. 277–299, Sep. 2015.

[14] K. Huang, Y. C. Eldar, and N. D. Sidiropoulos, "Phase retrieval from 1D Fourier measurements: Convexity, uniqueness, and algorithms," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6105–6117, Dec. 2016.

[15] B. Rajaei, E. W. Tramel, S. Gigan, F. Krzakala, and L. Daudet, "Intensity-only optical compressive imaging using a multiply scattering material and a double phase retrieval approach," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Shanghai, China, Mar. 2016, pp. 4054–4058.

[16] A. Dremeau *et al.*, "Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques," *Opt. Expr.*, vol. 23, no. 9, pp. 11898–11911, May 2015.

[17] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," *Appl. Comput. Harmon. Anal.*, vol. 36, no. 3, pp. 473–494, May 2014.

[18] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "STFT phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 770–781, Jun. 2016.

[19] T. Bendory, Y. C. Eldar, and N. Boumal, "Non-convex phase retrieval from STFT measurements," *IEEE Trans. Inf. Theory*, preprint, doi: 10.1109/TIT.2017.2745623.

[20] R. Pedarsani, D. Yin, K. Lee, and K. Ramchandran, "PhaseCode: Fast and efficient compressive phase retrieval based on sparse-graph codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3663–3691, Jun. 2017.

[21] M. A. Iwen, A. Viswanathan, and Y. Wang., "Fast phase retrieval from local correlation measurements," *SIAM J. Imag. Sci.*, vol. 9, no. 4, pp. 1655–1688, Oct. 2016.

[22] B. G. Bodmann and N. Hammen, "Algorithms and error bounds for noisy phase retrieval with low-redundancy frames," *Appl. Comput. Harmon. Anal.*, vol. 43, no. 3, pp. 482–503, Nov. 2017.

[23] D. Guo,Y. Wu, S. Shamai, and S. Verdu, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.

[25] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[26] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM J. Imag. Sci.*, vol. 5, no. 4, pp. 1185–1212, Oct. 2012.

[27] M. R. Bell, "Information theory and radar waveform design," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1578–1597, Sep. 1993.

[28] Y. Yang and R. S. Blum, "MIMO radar waveform design based on mutual information and minimum mean-square error estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 330–343, Jan. 2007.

[29] E. Riegler and G. Taubock, "Almost lossless analog compression without phase information," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 999–1003.

[30] S. Liu and G. Trenkler, "Hadamard, Khatri-Rao, Kronecker and other matrix products," *Int. J. Inf. Syst. Sci.*, vol. 4, no. 1, pp. 160–177, 2008.

[31] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[32] T. Heinosaari, L. Mazzarella, and M. M. Wolf, "Quantum tomography under prior information," *Commun. Math. Phys.*, vol. 318, no. 2, pp. 355–374, Feb. 2013.

[33] B. G. Bodmann and N. Hammen, "Stable phase retrieval with low-redundancy frames," *Adv. Comput. Math.*, vol. 41, no. 2, pp. 317–331, Apr. 2015.

[34] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2013.

[35] T. Ratnarajah and R. Vaillancourt, "Quadratic forms on complex random matrices and multiple-antenna systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2976–2984, Aug. 2005.

[36] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdu, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1085, Mar. 2010.

[37] M. Lamarca, "Linear precoding for mutual information maximization in MIMO systems," in *Proc. Int. Symp. Wireless Commun. Syst.*, Sienna, Italy, Sep. 2009, pp. 26–30.

[38] M. Payaro and D. P. Palomar, "On optimal precoding in linear vector Gaussian channels with arbitrary input distribution," *IEEE Int. Symp. Inf. Theory*, Seoul, South Korea, Jun. 2009, pp. 1085–1089.

[39] R. Bustin, M. Payaro, D. P. Palomar, and S. Shamai, "On MMSE crossing properties and implications in parallel vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 818–844, Feb. 2013.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[41] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Kongens Lyngby, Denmark: Technical Univ. of Denmark, Nov. 2012.

[42] L. H. Koopmans, *The Spectral Analysis of Time Series*. New York, NY, USA: Academic, 1995.

[43] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.

[44] R. A. Horn and C. A. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[45] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural Parall. Sci. Comput.*, vol. 11, no. 4, pp. 351–368, Dec. 2003.

[46] D. P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686–695, Feb. 2005.

Nir Shlezinger (M'17) received the B.Sc., M.Sc., and Ph.D. degrees in 2011, 2013, and 2017, respectively, from Ben-Gurion University, Be'er Sheva, Israel, all in electrical and computer engineering. He is currently a postdoctoral researcher in the Signal Acquisition Modeling and Processing Lab in the Technion, Israel Institute of Technology, Haifa, Israel. From 2009 to 2013 he worked as an engineer at Yitran Communications. His research interests include information theory and signal processing for communications.

Ron Dabora (M'07–SM'14) received the B.Sc. and M.Sc. degrees from Tel-Aviv University, Tel Aviv, Israel, in 1994 and 2000, respectively, and the Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2007, all in electrical engineering. From 1994 to 2000, he was with the Ministry of Defense of Israel. From 2000 to 2003 he was with the algorithms group, Millimetrix Broadband Networks, Israel, and from 2007 to 2009, he was a postdoctoral researcher with the Department of Electrical Engineering, Stanford University. Since 2009, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, Ben-Gurion University, Be'er Sheva, Israel. His research interests include network information theory, wireless communications, and power line communications. He served as a TPC Member in a number of international conferences, including WCNC, PIMRC, and ICC. From 2012 to 2014, he served as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He currently serves as a Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics in 1995 and the B.Sc. degree in electrical engineering in 1996 both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in electrical engineering and computer science in 2002 from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

She is currently a Professor in the Department of Electrical Engineering at the Technion—Israel Institute of Technology, Haifa, Israel, where she holds the Edwards Chair in Engineering. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT, an Adjunct Professor at Duke University, and was a Visiting Professor at Stanford University, Stanford, CA, USA. She is a member of the Israel Academy of Sciences and Humanities (elected 2017), an IEEE Fellow and a EURASIP Fellow. Her research interests include the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics.

Dr. Eldar has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), the IEEE/AESS Fred Nathanson Memorial Radar Award (2014), and the IEEE Kiyo Tomiyasu Award (2016). She was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technions Award for Excellence in Teaching (two times). She received several best paper awards and best demo awards together with her research students and colleagues including the SIAM outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award and the IET Circuits, Devices and Systems Premium Award, and was selected as one of the 50 most influential women in Israel.

She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor in Chief of *Foundations and Trends in Signal Processing*, a member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*. She was Cochair and Technical Cochair of several international conferences and workshops.

She is author of the book "*Sampling Theory: Beyond Bandlimited Systems*" and coauthor of the books "*Compressed Sensing*" and "*Convex Optimization Methods in Signal Processing and Communications*," all published by Cambridge University Press.