Block-Sparse Signals: Uncertainty Relations and Efficient Recovery

Yonina C. Eldar, *Senior Member, IEEE*, Patrick Kuppinger, *Student Member, IEEE*, and
Helmut Bölcskei, *Fellow, IEEE*

Abstract—We consider efficient methods for the recovery of block-sparse signals—i.e., sparse signals that have nonzero entries occurring in clusters—from an underdetermined system of linear equations. An uncertainty relation for block-sparse signals is derived, based on a block-coherence measure, which we introduce. We then show that a block-version of the orthogonal matching pursuit algorithm recovers block k-sparse signals in no more than k steps if the block-coherence is sufficiently small. The same condition on block-coherence is shown to guarantee successful recovery through a mixed ℓ_2/ℓ_1-optimization approach. This complements previous recovery results for the block-sparse case which relied on small block-restricted isometry constants. The significance of the results presented in this paper lies in the fact that making explicit use of block-sparsity can provably yield better reconstruction properties than treating the signal as being sparse in the conventional sense, thereby ignoring the additional structure in the problem.

Index Terms—Basis pursuit, block-sparsity, compressed sensing, matching pursuit.

I. INTRODUCTION

THE framework of compressed sensing is concerned with the recovery of an unknown vector from an underdetermined system of linear equations [1], [2]. The key property exploited for recovery of the unknown data is the assumption of sparsity. More concretely, denoting by $\mathbf{x}$ an unknown vector that is observed through a measurement matrix $\mathbf{D}$ according to $\mathbf{y} = \mathbf{D}\mathbf{x}$, it is assumed that $\mathbf{x}$ has only a few nonzero entries. A fundamental observation is that if $\mathbf{D}$ is chosen properly and $\mathbf{x}$ is sufficiently sparse, then $\mathbf{x}$ can be recovered from $\mathbf{y} = \mathbf{D}\mathbf{x}$, irrespectively of the locations of the nonzero entries of $\mathbf{x}$, even if $\mathbf{D}$ has far fewer rows than columns. This result has given rise to a multitude of different recovery algorithms which can be proven to recover a sparse vector $\mathbf{x}$ under a variety of different conditions on $\mathbf{D}$ [1], [3]–[6].

Three widely studied recovery algorithms are the basis pursuit (BP), or ℓ_1-minimization approach [1], [7], the matching pursuit (MP), and the orthogonal matching pursuit (OMP) algorithm [8], [9]. One of the main tools for the characterization of the recovery abilities of BP is the restricted isometry property (RIP) [1], [10]. Specifically, if the measurement matrix $\mathbf{D}$ satisfies the RIP with appropriate restricted isometry constants, then $\mathbf{x}$ can be recovered by BP. Unfortunately, determining the RIP constants of a given matrix is in general an NP-hard problem. A more simple and convenient way to characterize recovery properties of a dictionary is via the coherence measure [5], [11], [12]. It was shown in [5], [13], and [14] that appropriate conditions on the coherence guarantee that BP, MP, and OMP recover the sparse vector $\mathbf{x}$. The coherence also plays an important role in uncertainty relations for sparse signals [11], [12], [15].

In this paper, we consider efficient methods for the recovery of sparse signals that exhibit additional structure in the form of the nonzero coefficients occurring in clusters. Such signals are referred to as block-sparse [16], [17]. Our goal is to explicitly take this block structure into account, both in terms of the recovery algorithms and in terms of the measures that are used to characterize their performance. The significance of the results we obtain lies in the fact that making explicit use of block-sparsity can provably yield better reconstruction properties than treating the signal as being sparse in the conventional sense, thereby ignoring the additional structure in the problem.

Block-sparsity arises naturally, e.g., when dealing with multi-band signals [18]–[21] or in measurements of gene expression levels [22]. Another interesting special case of the block-sparse model appears in the multiple measurement vector (MMV) problem, which deals with the measurement of a set of vectors that share a joint sparsity pattern [16], [23]–[26]. Furthermore, it was shown in [16] and [17] that the block-sparsity model can be used to treat the problem of sampling signals that lie in a union of subspaces [15], [16], [18], [19], [27]–[31].

One approach to exploiting block-sparsity is by suitably extending the BP method, resulting in a mixed ℓ_2/ℓ_1-norm recovery algorithm [16], [32]. It was shown in [16] that if $\mathbf{D}$ has small block-restricted isometry constants, which generalizes the conventional RIP notion, then the mixed norm method is guaranteed to recover any block-sparse signal, irrespectively of the locations of the nonzero blocks. Furthermore, recovery will be robust in the presence of noise and modeling errors (i.e., when the vector is not exactly block-sparse). It was also established in [16] that certain random matrices satisfy the block RIP with overwhelming probability, and that this probability is substantially larger than that of satisfying the standard RIP. In [33] extensions of the CoSaMP algorithm [34] and of iterative hard thresholding [35] to the model-based setting, which includes

block-sparsity as a special case, are proposed and shown to exhibit provable recovery guarantees and robustness properties.

The focus of the present paper is on developing a parallel line of results by generalizing the notion of coherence to the block setting. This can be seen as extending the program laid out in [5] and [13] to the block-sparse case. Specifically, we define two separate notions of coherence: coherence within a block, referred to as sub-coherence and capturing local properties of the dictionary, and block-coherence, describing global dictionary properties. We will show that both coherence notions are necessary to characterize the essence of block-sparsity. We present extensions of the BP, the MP, and the OMP algorithms to the block-sparse case and prove corresponding performance guarantees.

We point out that the term block-coherence was used previously in [36] in the context of quantifying the recovery performance of the MP algorithm in block-incoherent dictionaries. Our definition of block-coherence pertains to block-versions of the MP and the OMP algorithm and is different from that used in [36].

The concept of block-sparsity was previously introduced and studied in the statistics literature in the context of the so-called group Lasso estimator [37]–[42]. The main focus of this paper and the novelty *vis-à-vis* these references are *explicit* recovery guarantees for block-versions of BP and OMP in terms of properties of the dictionary characterized through block-coherence and sub-coherence. In addition, compared to previous work on the group Lasso estimator, our results do not depend on specific assumptions on dictionary properties such as, e.g., intrablock orthogonality, specific block sizes, or the asymptotic regime. Moreover, we establish uncertainty relations for block-sparse signals, which appear to be new. Our work can be seen as extending the program laid out in [5] and [11]–[13], [43] to the block-sparse setting.

We begin, in Section II, by introducing our definitions of block-coherence and sub-coherence. In Section III, we establish an uncertainty relation for block-sparse signals, and we show how the block-coherence measure defined previously occurs naturally in this uncertainty relation. In Section IV, we introduce a block version of the OMP algorithm, termed BOMP, and of the MP algorithm [9], termed BMP. We then derive a sufficient condition on block-coherence to guarantee recovery of block k-sparse signals through BOMP in no more than k steps as well as exponential convergence of BMP. The same condition on block-coherence is shown to guarantee successful recovery through the mixed ℓ_2/ℓ_1-optimization approach. The BOMP algorithm can be viewed as an extension of the subspace OMP algorithm for recovery in the MMV setup [26]. The proofs of our main results are contained in Section V. A discussion on the performance improvements that can be obtained through exploiting block-sparsity is provided in Section VI. Corresponding numerical results are reported in Section VII.

Throughout the paper, we denote vectors by boldface lowercase letters, e.g., $\mathbf{x}$, and matrices by boldface uppercase letters, e.g., $\mathbf{A}$. The identity matrix is written as $\mathbf{I}$ or $\mathbf{I}_d$ when the dimension is not clear from the context. For a given matrix $\mathbf{A}$, $\mathbf{A}^T$, $\mathbf{A}^H$, and $\text{Tr}(\mathbf{A})$ denote its transpose, conjugate transpose, and trace, respectively, $\mathbf{A}^\dagger$ is the pseudo inverse, $\mathcal{R}(\mathbf{A})$ denotes

the range space of $\mathbf{A}$, $\mathbf{A}_{i,j}$ is the element in the ith row and jth column of $\mathbf{A}$, and $\mathbf{a}_\ell$ stands for the ℓth column of $\mathbf{A}$. The ℓth element of a vector $\mathbf{x}$ is denoted by x_ℓ. The Euclidean norm of the vector $\mathbf{x}$ is $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^H \mathbf{x}}$, $\|\mathbf{x}\|_1 = \sum_\ell |x_\ell|$ is the ℓ_1-norm, $\|\mathbf{x}\|_\infty = \max_\ell |x_\ell|$ is the ℓ_∞-norm, and $\|\mathbf{x}\|_0$ designates the number of nonzero entries in $\mathbf{x}$. The Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$ is written as $\mathbf{A} \otimes \mathbf{B}$. The spectral norm of $\mathbf{A}$ is denoted by $\rho(\mathbf{A}) = \lambda_{\max}^{1/2}(\mathbf{A}^H \mathbf{A})$, where $\lambda_{\max}(\mathbf{B})$ is the largest eigenvalue of the positive-semidefinite matrix $\mathbf{B}$.

II. BLOCK-SPARSITY AND BLOCK-COHERENCE

A. Block-Sparsity

We consider the problem of representing a vector $\mathbf{y} \in \mathbb{C}^L$ in a given dictionary $\mathbf{D}$ of size $L \times N$ with $L < N$, so that

$$\mathbf{y} = \mathbf{D}\mathbf{x} \tag{1}$$

for a coefficient vector $\mathbf{x} \in \mathbb{C}^N$. Since the system of equations (1) is underdetermined, there are, in general, infinitely many choices of $\mathbf{x}$ that satisfy (1) for a given $\mathbf{y}$. Therefore, further assumptions on $\mathbf{x}$ are needed to guarantee uniqueness of the representation. Here, we consider the case of sparse vectors $\mathbf{x}$, i.e., $\mathbf{x}$ has only a few nonzero entries relative to its dimension. The standard sparsity model treated in compressed sensing [1], [2] assumes that $\mathbf{x}$ has at most k nonzero elements, which can appear anywhere in the vector. As discussed in [16], [17] there are practical scenarios that involve vectors $\mathbf{x}$ with nonzero entries appearing in blocks (or clusters) rather than being arbitrarily spread throughout the vector. Specific examples include signals that lie in unions of subspaces [16], [27]–[29], [31], and multiband signals [18], [19], [21].

The recovery of block-sparse vectors $\mathbf{x}$ from measurements $\mathbf{y} = \mathbf{D}\mathbf{x}$ is the focus of this paper. To define block-sparsity, we view $\mathbf{x}$ as a concatenation of blocks—assumed throughout the paper to be of length d—with $\mathbf{x}[\ell]$ denoting the ℓth block, i.e.,

$$\mathbf{x} = [\underbrace{x_1 \cdots x_d}_{\mathbf{x}^T[1]} \underbrace{x_{d+1} \cdots x_{2d}}_{\mathbf{x}^T[2]} \cdots \underbrace{x_{N-d+1} \cdots x_N}_{\mathbf{x}^T[M]}]^T \tag{2}$$

where $N = Md$. We furthermore assume that $L = Rd$ with R integer. Similarly to (2), we can represent $\mathbf{D}$ as a concatenation of column-blocks $\mathbf{D}[\ell]$ of size $L \times d$:

$$\mathbf{D} = [\underbrace{\mathbf{d}_1 \cdots \mathbf{d}_d}_{\mathbf{D}[1]} \underbrace{\mathbf{d}_{d+1} \cdots \mathbf{d}_{2d}}_{\mathbf{D}[2]} \cdots \underbrace{\mathbf{d}_{N-d+1} \cdots \mathbf{d}_N}_{\mathbf{D}[M]}]. \tag{3}$$

A vector $\mathbf{x} \in \mathbb{C}^N$ is called block k-sparse if $\mathbf{x}[\ell]$ has nonzero Euclidean norm for at most k indices ℓ. When $d = 1$, block-sparsity reduces to conventional sparsity as defined in [1] and [2]. Denoting

$$\|\mathbf{x}\|_{2,0} = \sum_{\ell=1}^{M} I(\|\mathbf{x}[\ell]\|_2 > 0) \tag{4}$$

with the indicator function $I(\cdot)$, a block k-sparse vector $\mathbf{x}$ is defined as a vector that satisfies $\|\mathbf{x}\|_{2,0} \leq k$. In the remainder of the paper conventional sparsity will be referred to simply as sparsity, in contrast to block-sparsity.

We are interested in providing conditions on the dictionary $\mathbf{D}$ ensuring that the block-sparse vector $\mathbf{x}$ can be recovered from

measurements $\mathbf{y}$ of the form (1) through BMP, BOMP, and a mixed ℓ_2/ℓ_1-optimization program (referred to as L-OPT [16]). Our approach is partly based on [5] and [12]–[14] (and the mathematical techniques used therein) where equivalent results are provided for the sparse case. It was shown in [16] that L-OPT yields perfect recovery if the dictionary $\mathbf{D}$ satisfies appropriate restricted isometry properties. The purpose of this paper is to provide recovery conditions for BMP, BOMP, and L-OPT based on a suitably defined measure of block-coherence. We will see that block-coherence plays a role similar to coherence in the case of conventional sparsity.

Before defining block-coherence, we note that in order to have a unique block k-sparse $\mathbf{x}$ satisfying (1) it is clear that we need $R > k$ and the columns within each block $\mathbf{D}[\ell]$, $\ell = 1, 2, \ldots, M$, need to be linearly independent. More generally, we have the following proposition taken from [16].

Proposition 1: The representation (1) is unique for all block k-sparse vectors $\mathbf{x}$ if and only if $\mathbf{Dg} \neq \mathbf{0}$ for every $\mathbf{g} \neq \mathbf{0}$ that is block $2k$-sparse.

From Proposition 1 it follows that uniqueness of the representation (1) implies linear independence of the columns of $\mathbf{D}[\ell]$ for all ℓ. Throughout the paper, we assume that the dictionaries we consider satisfy the condition of Proposition 1, and, furthermore, $\|\mathbf{d}_r\|_2 = 1$, $r = 1, 2, \ldots, N$.

B. Block-Coherence

The coherence of a dictionary $\mathbf{D}$ measures the similarity between its elements, and is defined by [5], [11], [12]

$$\mu = \max_{\ell, r \neq \ell} |\mathbf{d}_\ell^H \mathbf{d}_r|. \tag{5}$$

This definition was introduced in [9] to heuristically characterize the performance of the MP algorithm, and was later shown to play a fundamental role in quantifying recovery thresholds for the OMP algorithm and for BP [5]. The coherence μ furthermore occurs in ℓ_1-uncertainty relations relevant in the context of decomposing a vector into two orthonormal bases [11], [12]. A definition of coherence for analog signals, along with a corresponding uncertainty relation, is provided in [15].

It is natural to seek a generalization of coherence to the block-sparse setting with the resulting block-coherence measure having the same operational significance as the coherence μ in the sparse case. Below, we propose such a generalization, which is shown—in Sections III and IV—to occur naturally in uncertainty relations and in recovery thresholds for the block-sparse case.

We define the *block-coherence* of $\mathbf{D}$ as

$$\mu_{\mathrm{B}} = \max_{\ell, r \neq \ell} \frac{1}{d} \rho(\mathbf{M}[\ell, r]) \tag{6}$$

with

$$\mathbf{M}[\ell, r] = \mathbf{D}^H[\ell] \mathbf{D}[r]. \tag{7}$$

Note that $\mathbf{M}[\ell, r]$ is the (ℓ, r)th $d \times d$ block of the $N \times N$ matrix $\mathbf{M} = \mathbf{D}^H \mathbf{D}$. When $d = 1$, as expected, $\mu_{\mathrm{B}} = \mu$. While μ_{B}

quantifies global properties of the dictionary $\mathbf{D}$, local properties are characterized by the *sub-coherence* of $\mathbf{D}$, defined as

$$\nu = \max_\ell \max_{i, j \neq i} |\mathbf{d}_i^H \mathbf{d}_j|, \quad \mathbf{d}_i, \mathbf{d}_j \in \mathbf{D}[\ell]. \tag{8}$$

We define $\nu = 0$ for $d = 1$. In addition, if the columns of $\mathbf{D}[\ell]$ are orthonormal for each ℓ, then $\nu = 0$.

Since the columns of $\mathbf{D}$ have unit norm, the coherence μ in (5) satisfies $\mu \in [0, 1]$ and therefore, as a consequence of $\nu \in [0, \mu]$, we have $\nu \in [0, 1]$. The following proposition establishes the same limits for the block-coherence μ_{B}, which explains the choice of normalization by $1/d$ in its definition (6).

In the remainder of the paper, conventional coherence will be referred to simply as coherence, in contrast to block-coherence and sub-coherence.

Proposition 2: The block-coherence μ_{B} satisfies $0 \leq \mu_{\mathrm{B}} \leq \mu$.

Proof: Since the spectral norm is non-negative, clearly $\mu_{\mathrm{B}} \geq 0$. To prove that $\mu_{\mathrm{B}} \leq \mu$, note that the entries of $\mathbf{M}[\ell, r]$ for $\ell \neq r$ have absolute value smaller than or equal to μ. It then follows that

$$\begin{aligned}
\mu_{\mathrm{B}} &= \max_{\ell, r \neq \ell} \frac{1}{d} \rho(\mathbf{M}[\ell, r]) \\
&= \max_{\ell, r \neq \ell} \frac{1}{d} \sqrt{\lambda_{\max}(\mathbf{M}^H[\ell, r] \mathbf{M}[\ell, r])} \\
&\leq \max_{\ell, r \neq \ell} \frac{1}{d} \sqrt{\max_i \sum_{j=1}^d |(\mathbf{M}^H[\ell, r] \mathbf{M}[\ell, r])_{i,j}|} \quad (9) \\
&\leq \max_{\ell, r \neq \ell} \frac{1}{d} \sqrt{\max_i \sum_{j=1}^d d\mu^2} \\
&= \mu \tag{10}
\end{aligned}$$

where (9) is a consequence of Geršgorin's disc theorem [44, Corollary 6.1.5]. ∎

From $\mu \leq 1$, with Proposition 2, it now follows trivially that $\mu_{\mathrm{B}} \leq 1$.

When the columns of $\mathbf{D}[\ell]$ are orthonormal for each ℓ, we can further bound μ_{B}.

Proposition 3: If $\mathbf{D}$ consists of orthonormal blocks, i.e., $\mathbf{D}^H[\ell] \mathbf{D}[\ell] = \mathbf{I}_d$ for all ℓ, then $\mu_{\mathrm{B}} \leq 1/d$.

Proof: Using the submultiplicativity of the spectral norm, we have

$$\begin{aligned}
\mu_{\mathrm{B}} &= \max_{\ell, r \neq \ell} \frac{1}{d} \rho(\mathbf{M}[\ell, r]) \\
&= \max_{\ell, r \neq \ell} \frac{1}{d} \rho(\mathbf{D}^H[\ell] \mathbf{D}[r]) \\
&\leq \max_{\ell, r \neq \ell} \frac{1}{d} \rho(\mathbf{D}^H[\ell]) \rho(\mathbf{D}[r]) \\
&= \frac{1}{d}. \tag{11}
\end{aligned}$$

The equality in (11) follows from $\mathbf{D}^H[\ell] \mathbf{D}[\ell] = \mathbf{I}_d$, for all ℓ, $\lambda_{\max}(\mathbf{D}^H[\ell] \mathbf{D}[\ell]) = \lambda_{\max}(\mathbf{D}[\ell] \mathbf{D}^H[\ell])$, and $\lambda_{\max}(\mathbf{I}_d) = 1$ combined with the definition of the spectral norm. ∎

III. UNCERTAINTY RELATION FOR BLOCK-SPARSE SIGNALS

We next show how the block-coherence μ_{B} defined above naturally appears in an uncertainty relation for block-sparse signals. This relation generalizes the corresponding result for the sparse case derived in [11] and [12].

Uncertainty relations for sparse signals are concerned with representations of a vector $\mathbf{x} \in \mathbb{C}^L$ in two different orthonormal bases for $\mathbb{C}^L$: $\{\boldsymbol{\phi}_\ell, 1 \le \ell \le L\}$ and $\{\boldsymbol{\psi}_\ell, 1 \le \ell \le L\}$ [11], [12]. Any vector $\mathbf{x} \in \mathbb{C}^L$ can be expanded uniquely in terms of each one of these bases according to

$$\mathbf{x} = \sum_{\ell=1}^{L} a_\ell \boldsymbol{\phi}_\ell = \sum_{\ell=1}^{L} b_\ell \boldsymbol{\psi}_\ell. \tag{12}$$

The uncertainty relation sets limits on the sparsity of the decompositions (12) for any $\mathbf{x} \in \mathbb{C}^L$. Specifically, denoting $A = \|\mathbf{a}\|_0$ and $B = \|\mathbf{b}\|_0$, it is shown in [12] that

$$\frac{1}{2}(A + B) \ge \sqrt{AB} \ge \frac{1}{\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})} \tag{13}$$

where $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ is the coherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, defined as

$$\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \max_{\ell,r} |\boldsymbol{\phi}_\ell^H \boldsymbol{\psi}_r|. \tag{14}$$

It is easily seen that for $\mathbf{D}$ consisting of the orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, i.e., $\mathbf{D} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$, we have $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \mu$, where μ is as defined in (5) and associated with $\mathbf{D} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$.

In [11] it is shown that $1/\sqrt{L} \le \mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \le 1$. The upper bound follows from the Cauchy-Schwarz inequality and the fact that the basis elements have norm 1. The lower bound is obtained as follows: The matrix $\mathbf{G} = \boldsymbol{\Phi}^H \boldsymbol{\Psi}$ is unitary so that $\sum_{\ell=1}^{L} \sum_{r=1}^{L} |\boldsymbol{\phi}_\ell^H \boldsymbol{\psi}_r|^2 = \mathrm{Tr}(\mathbf{G}^H \mathbf{G}) = \mathrm{Tr}(\mathbf{I}_L) = L$. Consequently, we have $L^2 \max_{\ell,r} |\boldsymbol{\phi}_\ell^H \boldsymbol{\psi}_r|^2 \ge L$ which implies $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \ge 1/\sqrt{L}$. This lower bound can be achieved, for example, by choosing the two orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ as the spike (identity) and Fourier bases [11]. With this choice, the uncertainty relation (13) becomes

$$A + B \ge 2\sqrt{AB} \ge 2\sqrt{L}. \tag{15}$$

When $\sqrt{L}$ is an integer, the relations in (15) can all be satisfied with equality by choosing $\mathbf{x}$ as a Dirac comb $\boldsymbol{\delta}_{\sqrt{L}}$ with spacing $\sqrt{L}$, resulting in $\sqrt{L}$ nonzero elements. This follows from the fact that the Fourier transform of $\boldsymbol{\delta}_{\sqrt{L}}$ is also $\boldsymbol{\delta}_{\sqrt{L}}$.

We now develop an uncertainty relation for block-sparse signals. Specifically, we derive a result that is equivalent to (13) with A and B replaced by block-sparsity levels as defined in (4), and $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ replaced by the block-coherence between the orthonormal bases considered, and defined below in (18).

Theorem 1: Let $\boldsymbol{\Phi}, \boldsymbol{\Psi}$ be two unitary $L \times L$ matrices with $L \times d$ blocks $\{\boldsymbol{\Phi}[\ell], \boldsymbol{\Psi}[\ell], 1 \le \ell \le R\}$ and let $\mathbf{x} \in \mathbb{C}^L$ satisfy

$$\mathbf{x} = \sum_{\ell=1}^{R} \boldsymbol{\Phi}[\ell]\mathbf{a}[\ell] = \sum_{\ell=1}^{R} \boldsymbol{\Psi}[\ell]\mathbf{b}[\ell]. \tag{16}$$

Let $A = \|\mathbf{a}\|_{2,0}$ and $B = \|\mathbf{b}\|_{2,0}$. Then

$$\frac{1}{2}(A + B) \ge \sqrt{AB} \ge \frac{1}{d\mu_{\mathrm{B}}(\boldsymbol{\Phi}, \boldsymbol{\Psi})} \tag{17}$$

where

$$\mu_{\mathrm{B}}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \max_{\ell,r} \frac{1}{d} \rho(\boldsymbol{\Phi}^H[\ell]\boldsymbol{\Psi}[r]). \tag{18}$$

Note that for $\mathbf{D}$ consisting of the orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, i.e., $\mathbf{D} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$, we have $\mu_{\mathrm{B}}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \mu_{\mathrm{B}}$, where μ_{B} is as defined in (6) and associated with $\mathbf{D} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$.

Proof: The following proof generalizes the corresponding proof in [12] to block sizes $d \ge 1$. Without loss of generality, we assume that $\|\mathbf{x}\|_2^2 = 1$. Then

$$1 = \|\mathbf{x}\|_2^2 = \sum_{\ell,r=1}^{R} \mathbf{a}^H[\ell]\mathbf{G}[\ell, r]\mathbf{b}[r] \tag{19}$$

$$\le \sum_{\ell,r=1}^{R} |\mathbf{a}^H[\ell]\mathbf{G}[\ell, r]\mathbf{b}[r]| \tag{20}$$

where we set $\mathbf{G}[\ell, r] = \boldsymbol{\Phi}^H[\ell]\boldsymbol{\Psi}[r]$. Now, from the Cauchy-Schwarz inequality, for any $\mathbf{a}, \mathbf{b}$, we have

$$|\mathbf{a}^H \mathbf{G}[\ell, r]\mathbf{b}| \le \|\mathbf{b}\|_2 \|\mathbf{G}^H[\ell, r]\mathbf{a}\|_2$$

$$\le \lambda_{\max}^{1/2}(\mathbf{G}[\ell, r]\mathbf{G}^H[\ell, r])\|\mathbf{b}\|_2\|\mathbf{a}\|_2$$

$$\le d\mu_{\mathrm{B}}\|\mathbf{b}\|_2\|\mathbf{a}\|_2 \tag{21}$$

where, for brevity, we wrote $\mu_{\mathrm{B}} = \mu_{\mathrm{B}}(\boldsymbol{\Phi}, \boldsymbol{\Psi})$. Substituting into (19), we get

$$1 \le d\mu_{\mathrm{B}} \sum_{r=1}^{R} \|\mathbf{b}[r]\|_2 \sum_{\ell=1}^{R} \|\mathbf{a}[\ell]\|_2. \tag{22}$$

Applying the Cauchy-Schwarz inequality yields

$$\sum_{r=1}^{R} \|\mathbf{b}[r]\|_2 \le \sqrt{B}\left(\sum_{r=1}^{R} \|\mathbf{b}[r]\|_2^2\right)^{1/2} = \sqrt{B} \tag{23}$$

where we used the fact that $\sum_{r=1}^{R} \|\mathbf{b}[r]\|_2^2 = \|\mathbf{b}\|_2^2 = 1$ since $\|\mathbf{x}\|_2^2 = 1$ and $\boldsymbol{\Psi}$ is unitary. Similarly, we have that $\sum_{r=1}^{R} \|\mathbf{a}[r]\|_2 \le \sqrt{A}$. Substituting into (22) and using the inequality of arithmetic and geometric means completes the proof. ∎

The bound provided by Theorem 1 can be tighter than that obtained by applying the conventional uncertainty relation (13) to the block-sparse case. This can be seen by using $\|\mathbf{a}\|_0 \le d\|\mathbf{a}\|_{2,0}$ and $\|\mathbf{b}\|_0 \le d\|\mathbf{b}\|_{2,0}$ in (13) to obtain

$$\sqrt{\|\mathbf{a}\|_{2,0}\|\mathbf{b}\|_{2,0}} \ge \frac{1}{d\mu}. \tag{24}$$

Since $\mu_{\mathrm{B}} \le \mu$, this bound may be looser than (17).

A. Block-Incoherent Dictionaries

As already noted, in the sparse case (i.e., $d = 1$) for any two orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, we have $\mu \ge 1/\sqrt{L}$. We next show that a similar inequality holds for the block-coherence.

Proposition 4: The block-coherence (18) satisfies $\mu_{\mathrm{B}} \ge 1/\sqrt{dL}$.

Proof: Let $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ be two orthonormal bases for $\mathbb{C}^L$ and let $\mathbf{G} = \boldsymbol{\Phi}^H \boldsymbol{\Psi}$ with $\mathbf{G}[\ell, r]$ denoting the (ℓ, r)th $d \times d$ block of $\mathbf{G}$. With $R = L/d$, we have

$$R^2 \mu_{\mathrm{B}}^2 \geq \sum_{\ell=1}^R \sum_{r=1}^R \frac{1}{d^2} \lambda_{\max}(\mathbf{G}^H[\ell, r]\mathbf{G}[\ell, r])$$

$$\geq \frac{1}{d^2} \lambda_{\max}\left(\sum_{\ell=1}^R \sum_{r=1}^R \mathbf{G}^H[\ell, r]\mathbf{G}[\ell, r] \right). \quad (25)$$

Now,

$$\sum_{\ell=1}^R \sum_{r=1}^R \mathbf{G}^H[\ell, r]\mathbf{G}[\ell, r] = \sum_{r=1}^R \boldsymbol{\Psi}^H[r] \left(\sum_{\ell=1}^R \boldsymbol{\Phi}[\ell]\boldsymbol{\Phi}^H[\ell] \right) \boldsymbol{\Psi}[r]. \quad (26)$$

Since $\boldsymbol{\Phi}$ is a square matrix consisting of orthonormal columns, we have $\sum_{\ell=1}^R \boldsymbol{\Phi}[\ell]\boldsymbol{\Phi}^H[\ell] = \boldsymbol{\Phi}\,\boldsymbol{\Phi}^H = \mathbf{I}_L$. Furthermore, since $\boldsymbol{\Psi}[r]$ consists of orthonormal columns, for each r, we have $\boldsymbol{\Psi}^H[r]\boldsymbol{\Psi}[r] = \mathbf{I}_d$. Therefore, (25) becomes

$$\mu_{\mathrm{B}}^2 \geq \frac{1}{d^2 R} = \frac{1}{dL} \quad (27)$$

which concludes the proof. ∎

We now construct a pair of bases that achieves the lower bound in (27) and therefore has the smallest possible block-coherence. Let $\mathbf{F}$ be the DFT matrix of size $R = L/d$ with $\mathbf{F}_{\ell, r} = (1/\sqrt{R}) \exp(j 2\pi \ell r / R)$. Define $\boldsymbol{\Phi} = \mathbf{I}_L$ and

$$\boldsymbol{\Psi} = \mathbf{F} \otimes \mathbf{U}_d \quad (28)$$

where $\mathbf{U}_d$ is an arbitrary $d \times d$ unitary matrix. For this choice, $\boldsymbol{\Phi}^H[\ell]\boldsymbol{\Psi}[r] = \mathbf{F}_{\ell, r}\mathbf{U}_d$. Since $\rho(\mathbf{U}_d) = 1$ and $|\mathbf{F}_{\ell, r}| = 1/\sqrt{R}$, we get

$$\mu_{\mathrm{B}} = \frac{1}{d\sqrt{R}} = \frac{1}{\sqrt{dL}}. \quad (29)$$

When $d = 1$, this basis pair reduces to the spike-Fourier pair which is well known to be maximally incoherent [11].

When μ_{B} satisfies (29) the uncertainty relation becomes

$$A + B \geq 2\sqrt{AB} \geq 2\sqrt{R}. \quad (30)$$

If $\sqrt{R}$ is integer, the inequalities in (30) are met with equality for the signal $\mathbf{x} = \boldsymbol{\delta}_{\sqrt{R}} \otimes \mathbf{c}$ where $\mathbf{c}$ is an arbitrary nonzero length-d vector. Indeed, in this case, the representation of $\mathbf{x}$ in the spike basis requires $\sqrt{R}$ blocks (of size d), so that $\|\mathbf{a}\|_{2,0} = \sqrt{R}$. The representation of $\mathbf{x}$ in the basis $\boldsymbol{\Psi}$ in (28) is obtained as

$$\mathbf{b} = (\mathbf{F}^H \otimes \mathbf{U}_d^H)(\boldsymbol{\delta}_{\sqrt{R}} \otimes \mathbf{c}) = \boldsymbol{\delta}_{\sqrt{R}} \otimes \mathbf{U}_d^H \mathbf{c} \quad (31)$$

where we used the fact that the Fourier transform of $\boldsymbol{\delta}_{\sqrt{R}}$ is also $\boldsymbol{\delta}_{\sqrt{R}}$. Therefore, $\mathbf{b}$ has $\sqrt{R}$ nonzero blocks so that $\|\mathbf{b}\|_{2,0} = \sqrt{R}$ and hence $A = B = \sqrt{R}$, which implies that all inequalities in (30) are met with equality.

IV. EFFICIENT RECOVERY ALGORITHMS

We now give operational meaning to block-coherence by showing that if it is small enough, then a block-sparse signal $\mathbf{x}$ can be recovered from the measurements $\mathbf{y} = \mathbf{D}\mathbf{x}$ using either

the mixed ℓ_2/ℓ_1-optimization program (L-OPT) proposed in [16]

$$\min_{\mathbf{x}} \sum_{\ell=1}^M \|\mathbf{x}[\ell]\|_2 \quad \text{s.t. } \mathbf{y} = \mathbf{D}\mathbf{x} \quad (32)$$

or an extension of the MP and OMP algorithms [9] to the block-sparse case described below and termed block-MP (BMP) and block-OMP (BOMP), respectively. We then derive thresholds on the block-sparsity level as a function of μ_{B} and ν for both methods to recover the correct block-sparse $\mathbf{x}$. For L-OPT this complements the results in [16] that establish the recovery capabilities of L-OPT under the condition that $\mathbf{D}$ satisfies a block-RIP with a small enough restricted isometry constant. For the special case of the columns of $\mathbf{D}[\ell]$ being orthonormal for each ℓ, we suggest a block-version of the MP algorithm [9], termed block-MP (BMP).

A. Block OMP and Block MP

The BOMP algorithm begins by initializing the residual as $\mathbf{r}_0 = \mathbf{y}$. At the ℓth stage ($\ell \geq 1$) we choose the block that is best matched to $\mathbf{r}_{\ell-1}$ according to

$$i_\ell = \arg\max_i \|\mathbf{D}^H[i]\mathbf{r}_{\ell-1}\|_2. \quad (33)$$

Once the index i_ℓ is chosen, we find $\mathbf{x}_\ell[i]$ as the solution to

$$\arg\min_{\{\tilde{\mathbf{x}}_\ell[i]\}_{i \in \mathcal{I}}} \left\| \mathbf{y} - \sum_{i \in \mathcal{I}} \mathbf{D}[i]\tilde{\mathbf{x}}_\ell[i] \right\|_2 \quad (34)$$

where $\mathcal{I}$ is the set of chosen indices $i_j, 1 \leq j \leq \ell$. The residual is then updated as

$$\mathbf{r}_\ell = \mathbf{y} - \sum_{i \in \mathcal{I}} \mathbf{D}[i]\mathbf{x}_\ell[i]. \quad (35)$$

In the special case of the columns of $\mathbf{D}[\ell]$ being orthonormal for each ℓ (the elements across different blocks do not have to be orthonormal), we consider an extension of the MP algorithm to the block-case. The resulting algorithm, termed BMP, starts by initializing the residual as $\mathbf{r}_0 = \mathbf{y}$ and at the ℓth stage ($\ell \geq 1$) chooses the block that is best matched to $\mathbf{r}_{\ell-1}$ according to (33). Then, however, the algorithm does not perform a least-squares minimization over the blocks that have already been selected, but directly updates the residual according to

$$\mathbf{r}_\ell = \mathbf{r}_{\ell-1} - \mathbf{D}[i_\ell]\mathbf{D}^H[i_\ell]\mathbf{r}_{\ell-1}. \quad (36)$$

B. Recovery Conditions

Our main result, summarized in Theorems 2 and 3 below, is that any block k-sparse vector $\mathbf{x}$ can be recovered from measurements $\mathbf{y} = \mathbf{D}\mathbf{x}$ using either the BOMP algorithm or L-OPT if the block-coherence satisfies $kd < (\mu_{\mathrm{B}}^{-1} + d - (d-1)\nu\mu_{\mathrm{B}}^{-1})/2$. In the special case of the columns of $\mathbf{D}[\ell]$ being orthonormal for each ℓ, we have $\nu = 0$ and therefore the recovery condition becomes $kd < (\mu_{\mathrm{B}}^{-1} + d)/2$. In this setting BMP exhibits exponential convergence rate (see Theorem 4). If the block-sparse vector $\mathbf{x}$ was treated as a (conventional) kd-sparse vector without exploiting knowledge of the block-sparsity structure, a sufficient

condition for perfect recovery using OMP [5] or (32) for $d = 1$ (known as BP) is $kd < (\mu^{-1} + 1)/2$. Comparing with $kd < (\mu_{\mathrm{B}}^{-1} + d)/2$, we can see that, thanks to $\mu_{\mathrm{B}} \leq \mu$, making explicit use of block-sparsity leads to guaranteed recovery for a potentially higher sparsity level. Later, we will establish such a result for the case of general $\nu \neq 0$.

To formally state our main results, suppose that $\mathbf{x}_0$ is a length-N block k-sparse vector, and let $\mathbf{y} = \mathbf{D}\mathbf{x}_0$. Let $\mathbf{D}_0$ denote the $L \times (kd)$ matrix whose blocks correspond to the nonzero blocks of $\mathbf{x}_0$, and let $\overline{\mathbf{D}}_0$ be the matrix of size $L \times (N - kd)$ which contains the $L \times d$ blocks of $\mathbf{D}$ that are not in $\mathbf{D}_0$. We then have the following theorem, which generalizes the exact recovery condition provided in [5] to the case $d \geq 1$. The proof of the theorem appears in Section V.

Theorem 2: Let $\mathbf{x}_0 \in \mathbb{C}^N$ be a block k-sparse vector with blocks of length d, and let $\mathbf{y} = \mathbf{D}\mathbf{x}_0$ for a given $L \times N$ matrix $\mathbf{D}$. A sufficient condition for the BOMP and the L-OPT algorithm to recover $\mathbf{x}_0$ is that

$$\rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0) < 1 \tag{37}$$

where

$$\rho_c(\mathbf{A}) = \max_r \sum_\ell \rho(\mathbf{A}[\ell, r]) \tag{38}$$

and $\mathbf{A}[\ell, r]$ is the (ℓ, r)th $d \times d$ block of $\mathbf{A}$. In this case, BOMP picks up a correct new block in each step, and consequently converges in at most k steps.

Note that

$$\rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0) = \max_r \rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0[r]). \tag{39}$$

Therefore, (37) implies that for all r

$$\rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0[r]) < 1. \tag{40}$$

The sufficient condition (37) depends on $\mathbf{D}_0$ and therefore on the location of the nonzero blocks in $\mathbf{x}_0$, which, of course, is not known in advance. Nonetheless, as the following theorem, proved in Section V, shows, (37) holds universally under certain conditions on μ_{B} and ν associated with the dictionary $\mathbf{D}$.

Theorem 3: Let μ_{B} be the block-coherence and ν the subcoherence of the dictionary $\mathbf{D}$. Then (37) is satisfied if

$$kd < \frac{1}{2}\left(\mu_{\mathrm{B}}^{-1} + d - (d-1)\frac{\nu}{\mu_{\mathrm{B}}}\right). \tag{41}$$

For $d = 1$, and therefore $\nu = 0$, we recover the corresponding condition $k < (\mu^{-1} + 1)/2$ reported in [5] and [13]. In the special case where the columns of $\mathbf{D}[\ell]$ are orthonormal for each ℓ, we have $\nu = 0$ and (41) becomes

$$kd < \frac{1}{2}(\mu_{\mathrm{B}}^{-1} + d). \tag{42}$$

The next theorem shows that under condition (42), BMP exhibits exponential convergence rate in the case where each block $\mathbf{D}[\ell]$ consists of orthonormal columns.

Theorem 4: If $\mathbf{D}^H[\ell]\mathbf{D}[\ell] = \mathbf{I}_d$, for all ℓ, and $kd < (\mu_{\mathrm{B}}^{-1} + d)/2$, then we have:

1) BMP picks up a correct block in each step.
2) The energy of the residual decays exponentially, i.e., $\|\mathbf{r}_\ell\|_2^2 \leq \beta^\ell \|\mathbf{r}_0\|_2^2$ with

$$\beta = 1 - \frac{1 - (k-1)d\mu_{\mathrm{B}}}{k}. \tag{43}$$

V. PROOFS OF THEOREMS 2, 3, AND 4

Before proceeding with the actual proofs, we start with some definitions and basic results that will be used throughout this section.

For $\mathbf{x} \in \mathbb{C}^N$, we define the general mixed ℓ_2/ℓ_p-norm ($p = 1, 2, \infty$ here and in the following):

$$\|\mathbf{x}\|_{2,p} = \|\mathbf{v}\|_p, \quad \text{where } v_\ell = \|\mathbf{x}[\ell]\|_2 \tag{44}$$

and the $\mathbf{x}[\ell]$ are consecutive length-d blocks. For an $L \times N$ matrix $\mathbf{A}$ with $L = Rd$ and $N = Md$, where R and M are integers, we define the mixed matrix norm (with block size d) as

$$\|\mathbf{A}\|_{2,p} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_{2,p}}{\|\mathbf{x}\|_{2,p}}. \tag{45}$$

The following lemma provides bounds on $\|\mathbf{A}\|_{2,p}$, which will be used in the sequel.

Lemma 1: Let $\mathbf{A}$ be an $L \times N$ matrix with $L = Rd$ and $N = Md$. Denote by $\mathbf{A}[\ell, r]$ the (ℓ, r)th $d \times d$ block of $\mathbf{A}$. Then,

$$\|\mathbf{A}\|_{2,\infty} \leq \max_\ell \sum_r \rho(\mathbf{A}[\ell, r]) \triangleq \rho_r(\mathbf{A}) \tag{46}$$

$$\|\mathbf{A}\|_{2,1} \leq \max_r \sum_\ell \rho(\mathbf{A}[\ell, r]) \triangleq \rho_c(\mathbf{A}). \tag{47}$$

In particular, $\rho_r(\mathbf{A}) = \rho_c(\mathbf{A}^H)$.

Proof: See Appendix A. ∎

Lemma 2: $\rho_c(\mathbf{A})$ as defined in (38) is a matrix norm and as such satisfies the following properties:

- Non-negative: $\rho_c(\mathbf{A}) \geq 0$
- Positive: $\rho_c(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$
- Homogeneous: $\rho_c(\alpha\mathbf{A}) = |\alpha|\rho_c(\mathbf{A})$ for all $\alpha \in \mathbb{C}$
- Triangle inequality: $\rho_c(\mathbf{A} + \mathbf{B}) \leq \rho_c(\mathbf{A}) + \rho_c(\mathbf{B})$
- Submultiplicative: $\rho_c(\mathbf{A}\mathbf{B}) \leq \rho_c(\mathbf{A})\rho_c(\mathbf{B})$.

Proof: See Appendix B. ∎

The proofs of Theorems 2 and 3 follow closely the proof program laid out in [5] for the case $d = 1$ with appropriate modifications accounting for the block structure.

A. Proof of Theorem 2 for BOMP

We begin by proving that (37) is sufficient to ensure recovery using the BOMP algorithm. We first show that if $\mathbf{r}_{\ell-1}$ is in $\mathcal{R}(\mathbf{D}_0)$, then the next chosen index i_ℓ will correspond to a block in $\mathbf{D}_0$. Assuming that this is true, it follows immediately that i_1 is correct since clearly $\mathbf{r}_0 = \mathbf{y}$ lies in $\mathcal{R}(\mathbf{D}_0)$. Noting that $\mathbf{r}_\ell$ lies in the space spanned by $\mathbf{y}$ and $\mathbf{D}[i], i \in \mathcal{I}_\ell$, where $\mathcal{I}_\ell$ denotes the indices chosen up to stage ℓ, it follows that if $\mathcal{I}_\ell$ corresponds to correct indices, i.e., $\mathbf{D}[i]$ is a block of $\mathbf{D}_0$ for all

$i \in \mathcal{I}_\ell$, then $\mathbf{r}_\ell$ also lies in $\mathcal{R}(\mathbf{D}_0)$ and the next index will be correct as well. Thus, at every step a correct $L \times d$ block of $\mathbf{D}$ is selected. As we will show below no index will be chosen twice since the new residual is orthogonal to all the previously chosen subspaces spanned by the columns of the blocks $\mathbf{D}[i]$, $i \in \mathcal{I}_\ell$; consequently the correct $\mathbf{x}_0$ will be recovered in k steps.

Suppose that $\mathbf{r}_{\ell-1} \in \mathcal{R}(\mathbf{D}_0)$. Then, under (37), the next chosen index corresponds to a block in $\mathbf{D}_0$. This is equivalent to requiring that

$$z(\mathbf{r}_{\ell-1}) = \frac{\|\overline{\mathbf{D}}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}}{\|\mathbf{D}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}} < 1. \tag{48}$$

From the properties of the pseudo-inverse, it follows that $\mathbf{D}_0 \mathbf{D}_0^\dagger$ is the orthogonal projector onto $\mathcal{R}(\mathbf{D}_0)$. Hence, it holds that $\mathbf{D}_0 \mathbf{D}_0^\dagger \mathbf{r}_{\ell-1} = \mathbf{r}_{\ell-1}$. Since $\mathbf{D}_0 \mathbf{D}_0^\dagger$ is Hermitian, we have

$$(\mathbf{D}_0^\dagger)^H \mathbf{D}_0^H \mathbf{r}_{\ell-1} = \mathbf{r}_{\ell-1}. \tag{49}$$

Substituting (49) into (48) yields

$$\begin{aligned} z(\mathbf{r}_{\ell-1}) &= \frac{\|\overline{\mathbf{D}}_0^H (\mathbf{D}_0^\dagger)^H \mathbf{D}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}}{\|\mathbf{D}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}} \\ &\leq \rho_r(\overline{\mathbf{D}}_0^H (\mathbf{D}_0^\dagger)^H) \\ &= \rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0) \end{aligned} \tag{50}$$

where we used Lemma 1.

It remains to show that BOMP in each step chooses a new block participating in the (unique) representation $\mathbf{y} = \mathbf{Dx}$. We start by defining $\mathbf{D}_\ell = [\mathbf{D}[i_1] \cdots \mathbf{D}[i_\ell]]$ where $i_j \in \mathcal{I}_\ell, 1 \leq j \leq \ell$. It follows that the solution of the minimization problem in (34) is given by

$$\hat{\mathbf{x}} = (\mathbf{D}_\ell^H \mathbf{D}_\ell)^{-1} \mathbf{D}_\ell^H \mathbf{y} \tag{51}$$

which upon inserting into (35) yields

$$\mathbf{r}_\ell = (\mathbf{I} - \mathbf{D}_\ell (\mathbf{D}_\ell^H \mathbf{D}_\ell)^{-1} \mathbf{D}_\ell^H) \mathbf{y}. \tag{52}$$

Now, we note that $\mathbf{D}_\ell (\mathbf{D}_\ell^H \mathbf{D}_\ell)^{-1} \mathbf{D}_\ell^H$ is the orthogonal projector onto the range space of $\mathbf{D}_\ell$. Therefore $\|\mathbf{D}^H[i] \mathbf{r}_\ell\|_2 = 0$ for all blocks $\mathbf{D}[i]$ that lie in the span of the matrix $\mathbf{D}_\ell$. By the assumption in Proposition 1 we are guaranteed that as long as $\ell < k$ there exists at least one block in $\mathbf{D}_0$ which does not lie in the span of $\mathbf{D}_\ell$. Since this block will lead to strictly positive $\|\mathbf{D}^H[i] \mathbf{r}_\ell\|_2$ the result is established, concluding the proof.

B. Proof of Theorem 2 for L-OPT

We next show that (37) is also sufficient to ensure recovery using L-OPT. To this end we rely on the following lemma:

Lemma 3: Suppose that $\mathbf{v} \in \mathbb{C}^{kd}$ with $\|\mathbf{v}[\ell]\|_2 > 0$, for all ℓ, and that $\mathbf{A}$ is a matrix of size $L \times (kd)$, with $L = Rd$ and the $d \times d$ blocks $\mathbf{A}[\ell, r]$. Then, $\|\mathbf{Av}\|_{2,1} \leq \rho_c(\mathbf{A}) \|\mathbf{v}\|_{2,1}$. If in addition the values of $\rho_c(\mathbf{AJ}_\ell)$ are not all equal, then the inequality is strict. Here, $\mathbf{J}_\ell$ is a $(kd) \times d$ matrix that is all zero except for the ℓth $d \times d$ block which equals $\mathbf{I}_d$.

Proof: See Appendix C. ∎

To prove that L-OPT recovers the correct vector $\mathbf{x}_0$, let $\mathbf{x}' \neq \mathbf{x}_0$ be another length-N block k-sparse vector for which $\mathbf{y} =$

$\mathbf{Dx}'$. Denote by $\mathbf{c}_0$ and $\mathbf{c}'$ the length-kd vectors consisting of the nonzero elements of $\mathbf{x}_0$ and $\mathbf{x}'$, respectively. Let $\mathbf{D}_0$ and $\mathbf{D}'$ denote the corresponding columns of $\mathbf{D}$ so that $\mathbf{y} = \mathbf{D}_0 \mathbf{c}_0 = \mathbf{D}' \mathbf{c}'$. From the assumption in Proposition 1, it follows that there cannot be two different representations using the same blocks $\mathbf{D}_0$. Therefore, $\mathbf{D}'$ must contain at least one block, $\mathbf{Z}$, that is not included in $\mathbf{D}_0$. From (40), we get $\rho_c(\mathbf{D}_0^\dagger \mathbf{Z}) < 1$. For any other block $\mathbf{U}$ in $\mathbf{D}$, we must have that

$$\rho_c(\mathbf{D}_0^\dagger \mathbf{U}) \leq 1. \tag{53}$$

Indeed, if $\mathbf{U} \in \mathbf{D}_0$, then $\mathbf{U} = \mathbf{D}_0[\ell] = \mathbf{D}_0 \mathbf{J}_\ell$ for some ℓ where $\mathbf{J}_\ell$ was defined in Lemma 3. In this case, $\mathbf{D}_0^\dagger \mathbf{D}_0[\ell] = \mathbf{J}_\ell$ and, therefore, $\rho_c(\mathbf{D}_0^\dagger \mathbf{U}) = \rho_c(\mathbf{J}_\ell) = 1$. If, on the other hand, $\mathbf{U} = \overline{\mathbf{D}}_0[\ell]$ for some ℓ, then it follows from (40) that $\rho_c(\mathbf{D}_0^\dagger \mathbf{U}) < 1$.

Now, suppose first that the $(kd) \times d$ blocks in $\mathbf{D}_0^\dagger \mathbf{D}'$ do not all have the same[1] ρ_c. Then,

$$\begin{aligned} \|\mathbf{c}_0\|_{2,1} &= \|\mathbf{D}_0^\dagger \mathbf{D}_0 \mathbf{c}_0\|_{2,1} \\ &= \|\mathbf{D}_0^\dagger \mathbf{D}' \mathbf{c}'\|_{2,1} \\ &< \rho_c(\mathbf{D}_0^\dagger \mathbf{D}') \|\mathbf{c}'\|_{2,1} \tag{54} \\ &\leq \|\mathbf{c}'\|_{2,1} \tag{55} \end{aligned}$$

where the first equality is a consequence of the columns of $\mathbf{D}_0$ being linearly independent (a consequence of the assumption in Proposition 1), the first inequality follows from Lemma 3 since $\|\mathbf{c}'[\ell]\|_2 > 0$, for all ℓ, and the last inequality follows from (53). If all the $(kd) \times d$ blocks in $\mathbf{D}_0^\dagger \mathbf{D}'$ have identical ρ_c, then the inequality (54) is no longer strict, but the second inequality (55) becomes strict instead as a consequence of $\rho_c(\mathbf{D}_0^\dagger \mathbf{Z}) < 1$; therefore $\|\mathbf{c}_0\|_{2,1} < \|\mathbf{c}'\|_{2,1}$ still holds.

Since $\|\mathbf{x}_0\|_{2,1} = \|\mathbf{c}_0\|_{2,1}$ and $\|\mathbf{x}'\|_{2,1} = \|\mathbf{c}'\|_{2,1}$, we conclude that under (40), any set of coefficients used to represent the original signal that is not equal to $\mathbf{x}_0$ will result in a larger ℓ_2/ℓ_1-norm.

C. Proof of Theorem 3

We start by deriving an upper bound on $\rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0)$ in terms of μ_{B} and ν. Writing out $\mathbf{D}_0^\dagger$, we have that

$$\rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0) = \rho_c((\mathbf{D}_0^H \mathbf{D}_0)^{-1} \mathbf{D}_0^H \overline{\mathbf{D}}_0). \tag{56}$$

Submultiplicativity of $\rho_c(\mathbf{A})$ (Lemma 2) implies that

$$\begin{aligned} \rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0) &\leq \rho_c((\mathbf{D}_0^H \mathbf{D}_0)^{-1}) \rho_c(\mathbf{D}_0^H \overline{\mathbf{D}}_0) \\ &= \rho_c((\mathbf{D}_0^H \mathbf{D}_0)^{-1}) \max_{j \notin \Lambda_0} \sum_{i \in \Lambda_0} \rho(\mathbf{D}^H[i] \mathbf{D}[j]) \end{aligned} \tag{57}$$

where Λ_0 is the set of indices ℓ for which $\mathbf{D}[\ell]$ is in $\mathbf{D}_0$. Since Λ_0 contains k indices, the last term in (57) is bounded above by $kd\mu_{\mathrm{B}}$, which allows us to conclude that

$$\rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0) \leq \rho_c((\mathbf{D}_0^H \mathbf{D}_0)^{-1}) kd\mu_{\mathrm{B}}. \tag{58}$$

It remains to develop a bound on $\rho_c((\mathbf{D}_0^H \mathbf{D}_0)^{-1})$. To this end, we express $\mathbf{D}_0^H \mathbf{D}_0$ as $\mathbf{D}_0^H \mathbf{D}_0 = \mathbf{I} + \mathbf{A}$, where $\mathbf{A}$ is a $(kd) \times (kd)$ matrix with blocks $\mathbf{A}[\ell, r]$ of size $d \times d$ such that

---

[1]Note that for an $(sd) \times d$ matrix $\mathbf{A}$, $\rho_c(\mathbf{A}) = \sum_\ell \rho(\mathbf{A}[\ell])$, where $\mathbf{A}[\ell]$, $\ell = 1, 2, \ldots, s$, denotes the $d \times d$ block of $\mathbf{A}$ made up of the rows $\{(\ell-1)d + 1, \ldots, \ell d\}$.

$\mathbf{A}_{i,i} = 0$, for all i. This follows from the fact that the columns of $\mathbf{A}$ are normalized. Since $\mathbf{A}[\ell, r] = \mathbf{D}_0^H[\ell]\mathbf{D}_0[r]$, for all $\ell \neq r$, and $\mathbf{A}[r, r] = \mathbf{D}_0^H[r]\mathbf{D}_0[r] - \mathbf{I}_d$, we have

$$\rho_c(\mathbf{A}) = \max_r \sum_\ell \rho(\mathbf{A}[\ell, r])$$

$$\leq \max_r \rho(\mathbf{A}[r, r]) + \max_r \sum_{\ell \neq r} \rho(\mathbf{A}[\ell, r]) \quad (59)$$

$$\leq (d-1)\nu + (k-1)d\mu_{\mathrm{B}} \quad (60)$$

where the first term in (60) is obtained by applying Geršgorin's disc theorem [44, Corollary 6.1.5] together with the definition of ν; the second term in (60) follows from the fact that the summation in the second term of (59) is over $k-1$ elements and $\rho(\mathbf{A}[\ell, r])$, for all $\ell \neq r$, can be upper-bounded by $d\mu_{\mathrm{B}}$. Assumption (41) now implies that $(d-1)\nu + (k-1)d\mu_{\mathrm{B}} < 1$ and therefore, from (60), we have $\rho_c(\mathbf{A}) < 1$.

We next use the following result.

Lemma 4: Suppose that $\rho_c(\mathbf{A}) < 1$. Then $(\mathbf{I} + \mathbf{A})^{-1} = \sum_{k=0}^\infty (-\mathbf{A})^k$.

Proof: Follows immediately by using the fact that $\rho_c(\mathbf{A})$ is a matrix norm (cf. Lemma 2) and applying [44, Corollary 5.6.16]. ∎

Thanks to Lemma 4, we have that

$$\rho_c((\mathbf{D}_0^H\mathbf{D}_0)^{-1}) = \rho_c\left(\sum_{k=0}^\infty (-\mathbf{A})^k\right)$$

$$\leq \sum_{k=0}^\infty (\rho_c(\mathbf{A}))^k \quad (61)$$

$$= \frac{1}{1 - \rho_c(\mathbf{A})}$$

$$\leq \frac{1}{1 - (d-1)\nu - (k-1)d\mu_{\mathrm{B}}}. \quad (62)$$

Here, (61) is a consequence of $\rho_c(\mathbf{A})$ satisfying the triangle inequality and being submultiplicative and (62) follows by using (60).

Combining (62) with (58), we get

$$\rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0) \leq \frac{kd\mu_{\mathrm{B}}}{1 - (d-1)\nu - (k-1)d\mu_{\mathrm{B}}} < 1 \quad (63)$$

where the last inequality is a consequence of (41).

D. Proof of Theorem 4

The proof of the first part of Theorem 4 follows from the arguments in the proofs of Theorems 2 and 3 for $\nu = 0$. As a consequence of the first statement of Theorem 4, we get that the residual $\mathbf{r}_\ell$ in each step of the algorithm will be in $\mathcal{R}(\mathbf{D}_0)$. For the proof of the second statement in Theorem 4, we mimic the corresponding proof in [14]. We first need the following lemma, which is an extension of [45, Lemma 3.5] to the block-sparse case. This lemma will provide us with a lower bound on the amount of energy that can be removed from the residual $\mathbf{r}_\ell$ in one step of the BMP algorithm.

Lemma 5: Let $\mathbf{D}_0$ denote the $L \times (kd)$ matrix whose blocks correspond to the nonzero blocks of $\mathbf{x}_0$. Then, we have

$$\max_i \|\mathbf{D}_0^H[i]\mathbf{r}_\ell\|_2 \geq \frac{\|\mathbf{r}_\ell\|_2^2}{\|\mathbf{c}_\ell\|_{2,1}} \quad (64)$$

where $\mathbf{c}_\ell$ is the coefficient vector corresponding to $\mathbf{r}_\ell \neq \mathbf{0}$, i.e., $\mathbf{r}_\ell = \mathbf{D}_0\mathbf{c}_\ell$.

Proof: We start by noting that $\mathbf{r}_\ell = \mathbf{D}_0\mathbf{c}_\ell = \sum_{i=1}^k \mathbf{D}_0[i]\mathbf{c}_\ell[i]$, where $\mathbf{c}_\ell[i] \neq \mathbf{0}$ for at least one index $i \in \{1, 2, \ldots, k\}$. Therefore,

$$\|\mathbf{r}_\ell\|_2^2 = \sum_{i=1}^k \mathbf{c}_\ell^H[i]\mathbf{D}_0^H[i]\mathbf{r}_\ell$$

$$\leq \sum_{i=1}^k |\mathbf{c}_\ell^H[i]\mathbf{D}_0^H[i]\mathbf{r}_\ell|$$

$$\leq \sum_{i=1}^k \|\mathbf{c}_\ell[i]\|_2 \|\mathbf{D}_0^H[i]\mathbf{r}_\ell\|_2$$

$$\leq \left(\max_i \|\mathbf{D}_0^H[i]\mathbf{r}_\ell\|_2\right) \sum_{i=1}^k \|\mathbf{c}_\ell[i]\|_2. \quad (65)$$

The result then follows by noting that $\sum_{i=1}^k \|\mathbf{c}_\ell[i]\|_2 = \|\mathbf{c}_\ell\|_{2,1}$. ∎

Next, we compute an upper bound on $\|\mathbf{c}_\ell\|_{2,1}$. With $\mathbf{M}[i, j] = \mathbf{D}_0^H[i]\mathbf{D}_0[j]$, where $i, j \in \{1, \ldots, k\}$, we get

$$\|\mathbf{r}_\ell\|_2^2 = \mathbf{c}_\ell^H\mathbf{D}_0^H\mathbf{D}_0\mathbf{c}_\ell$$

$$= \sum_{i=1}^k \sum_{j=1}^k \mathbf{c}_\ell^H[i]\mathbf{M}[i, j]\mathbf{c}_\ell[j]$$

$$= \sum_{i=1}^k \mathbf{c}_\ell^H[i]\mathbf{M}[i, i]\mathbf{c}_\ell[i] + \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k \mathbf{c}_\ell^H[i]\mathbf{M}[i, j]\mathbf{c}_\ell[j]$$

$$\geq \sum_{i=1}^k \|\mathbf{c}_\ell[i]\|_2^2 - \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k |\mathbf{c}_\ell^H[i]\mathbf{M}[i, j]\mathbf{c}_\ell[j]| \quad (66)$$

where we used the fact that $\mathbf{M}[i, i] = \mathbf{I}_d$, for all i, as a consequence of each of the blocks of $\mathbf{D}_0$ consisting of orthonormal vectors. Applying the Cauchy-Schwarz inequality to the second term in (66), we have

$$\|\mathbf{r}_\ell\|_2^2 \geq \sum_{i=1}^k \|\mathbf{c}_\ell[i]\|_2^2 - \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k \|\mathbf{c}_\ell[i]\|_2 \|\mathbf{M}[i, j]\mathbf{c}_\ell[j]\|_2$$

$$\geq \|\mathbf{c}_\ell\|_{2,2}^2 - \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k \|\mathbf{c}_\ell[i]\|_2 \|\mathbf{c}_\ell[j]\|_2 d\mu_{\mathrm{B}} \quad (67)$$

$$= \|\mathbf{c}_\ell\|_{2,2}^2 - d\mu_{\mathrm{B}} \sum_{s=1}^{k-1} \sum_{i=1}^k \|\mathbf{c}_\ell[i]\|_2 \|\mathbf{c}_\ell[(i+s)_k]\|_2 \quad (68)$$

where $(i + s)_k$ stands for $(i + s)$ modulo k, (67) follows from $\|\mathbf{M}[i, j]\mathbf{c}_\ell[j]\|_2 \leq d\mu_{\mathrm{B}}\|\mathbf{c}_\ell[j]\|_2$, and (68) is obtained

by merely rearranging terms in the summation in (67). Applying the Cauchy-Schwarz inequality to the inner product $\sum_{i=1}^{k} \|\mathbf{c}_\ell[i]\|_2 \|\mathbf{c}_\ell[(i+s)_k]\|_2$, we obtain

$$
\begin{aligned}
\|\mathbf{r}_\ell\|_2^2 &\geq \|\mathbf{c}_\ell\|_{2,2}^2 - d\mu_{\mathrm{B}} \sum_{s=1}^{k-1} \|\mathbf{c}_\ell\|_{2,2}^2 \\
&= (1-(k-1)d\mu_{\mathrm{B}})\|\mathbf{c}_\ell\|_{2,2}^2 \\
&\geq \frac{(1-(k-1)d\mu_{\mathrm{B}})}{k} \|\mathbf{c}_\ell\|_{2,1}^2
\end{aligned}
\tag{69}
$$

where (69) follows by the same argument as used in (23). Combining (65) with (69) leads to

$$
\max_i \|\mathbf{D}_0^H[i]\mathbf{r}_\ell\|_2 \geq \sqrt{\frac{(1-(k-1)d\mu_{\mathrm{B}})}{k}} \|\mathbf{r}_\ell\|_2.
\tag{70}
$$

The proof is then completed by noting that by the first statement in Theorem 4, BMP picks a block in $\mathbf{D}_0$ in each step. Therefore, we can bound the energy of the residual in the $(\ell+1)$st step as

$$
\begin{aligned}
\|\mathbf{r}_{\ell+1}\|_2^2 &= \|\mathbf{r}_\ell\|_2^2 - \|\mathbf{D}^H[i_{\ell+1}]\mathbf{r}_\ell\|_2^2 \\
&= \|\mathbf{r}_\ell\|_2^2 - \max_i \|\mathbf{D}_0^H[i]\mathbf{r}_\ell\|_2^2
\end{aligned}
\tag{71}
$$

$$
\leq \left(1 - \frac{(1-(k-1)d\mu_{\mathrm{B}})}{k} \right) \|\mathbf{r}_\ell\|_2^2
\tag{72}
$$

where in (71) we used the fact that $\mathbf{r}_{\ell+1}$ is orthogonal to $\mathbf{D}[i_{\ell+1}]\mathbf{D}^H[i_{\ell+1}]\mathbf{r}_\ell$.

VI. DISCUSSION

Theorem 3 indicates under which conditions exploiting block-sparsity leads to higher recovery thresholds than treating the block-sparse signal as a conventionally sparse signal. For dictionaries $\mathbf{D}$ where the individual blocks $\mathbf{D}[\ell]$ consist of orthonormal columns for each ℓ, we have $\nu = 0$. Therefore, thanks to $\mu_{\mathrm{B}} \leq \mu$, recovery through exploiting block-sparsity is guaranteed for a potentially higher sparsity level. If the individual blocks $\mathbf{D}[\ell]$ are, however, not orthonormal, then $\nu > 0$, and (41) shows that ν has to be small for block-sparse recovery to result in higher recovery thresholds than sparse recovery.

It is natural to consider the case where one starts with a general dictionary $\mathbf{D}$ and orthogonalizes the individual blocks $\mathbf{D}[\ell]$ so that $\nu = 0$. The comparison that is meaningful here is between the recovery threshold of the original dictionary $\mathbf{D}$ without exploiting block-sparsity and the recovery threshold of the orthogonalized dictionary taking block-sparsity into account. To this end, we start by noting that the assumption in Proposition 1 implies that the columns of $\mathbf{D}[\ell]$ are linearly independent, for each ℓ. We can therefore write $\mathbf{D}[\ell] = \mathbf{A}[\ell]\mathbf{W}_\ell$ where $\mathbf{A}[\ell]$ consists of orthonormal columns that span $\mathcal{R}(\mathbf{D}[\ell])$ and $\mathbf{W}_\ell$ is invertible. The orthogonalized dictionary is given by the $L \times N$ matrix $\mathbf{A}$ with blocks $\mathbf{A}[\ell]$. Since $\mathbf{D} = \mathbf{A}\mathbf{W}$ with the $N \times N$ block-diagonal matrix $\mathbf{W}$ with blocks $\mathbf{W}_\ell$, we conclude that $\mathbf{c} = \mathbf{W}\mathbf{x}$ is block-sparse and—thanks to the invertibility of the $\mathbf{W}_\ell$—of the same block-sparsity level as $\mathbf{x}$, i.e., orthogonalization preserves the block-sparsity level. It is easy to see that the definition of block-coherence in (6) is invariant to the choice of orthonormal basis $\mathbf{A}[\ell]$ for $\mathcal{R}(\mathbf{A}[\ell])$.

This is because any other basis has the form $\mathbf{A}[\ell]\mathbf{U}_\ell$ for some unitary matrix $\mathbf{U}_\ell$, and from the properties of the spectral norm

$$
\rho(\mathbf{M}[\ell,r]) = \rho(\mathbf{U}_\ell^H \mathbf{M}[\ell,r]\mathbf{U}_r)
\tag{73}
$$

for any unitary matrices $\mathbf{U}_\ell, \mathbf{U}_r$.

Unfortunately, it seems difficult to derive general results on the relation between μ before and μ_{B} after orthogonalization. Nevertheless, we can establish a minimum block size d above which orthogonalization followed by block-sparse recovery leads to a guaranteed improvement in the recovery thresholds. We first note that the coherence μ of a dictionary consisting of $N = Md$ elements in a vector space of dimension $L = Rd$ can be lower-bounded as [46]

$$
\mu \geq \sqrt{\frac{M-R}{R(Md-1)}} \overset{Md \gg 1}{\approx} \sqrt{\frac{M-R}{RMd}}.
\tag{74}
$$

Using this lower bound together with Proposition 3 and the fact that after orthogonalization we have $\nu = 0$, it can be shown that if $d > RM/(M-R)$, then the recovery threshold obtained from taking block-sparsity into account in the orthogonalized dictionary is higher than the recovery threshold corresponding to conventional sparsity in the original dictionary. This is true irrespectively of the dictionary we start from as long as it satisfies the conditions of Proposition 1.

Finally, we note that finding dictionaries that lead to significant improvements in the theoretical recovery thresholds when exploiting block-sparsity seems to be a difficult design problem, especially in the most general case where the blocks of the dictionary do not consist of orthonormal columns. For example, partitioning the realizations of i.i.d. Gaussian matrices into blocks will, in general, not lead to satisfactory results in the sense that the theoretical recovery thresholds taking block-sparsity into account are not significantly larger than the conventional sparsity thresholds. However, we note that the theoretical thresholds are extremely pessimistic and simulation results presented below suggest that the performance of recovery algorithms are far better than the corresponding theoretical guarantees. Nevertheless, there do exist dictionaries where significant improvements also in the theoretical recovery thresholds are possible.

Consider, for example, the pair of bases $\boldsymbol{\Phi} = \mathbf{I}_L$ and $\boldsymbol{\Psi} = \mathbf{F} \otimes \mathbf{U}_d$ shown in Section III-A to achieve the lower bound in (27). For the corresponding dictionary $\mathbf{D} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$, we have $M = 2R$, $\mu_{\mathrm{B}} = 1/(d\sqrt{R})$, with the recovery threshold, assuming that block-sparsity is exploited, given by $kd < d(\sqrt{R}+1)/2$. The coherence of the dictionary is $\mu = \|\mathrm{vec}(\mathbf{U}_d)\|_\infty/\sqrt{R}$. Fig. 1, obtained by averaging over randomly chosen unitary matrices $\mathbf{U}_d$, shows that the recovery thresholds for dictionaries of the form $\mathbf{D} = [\mathbf{I}_L\ (\mathbf{F} \otimes \mathbf{U}_d)]$ obtained by taking block-sparsity into account can be significantly higher than those for conventional sparsity. Here, $\mathbf{U}_d$ is chosen randomly according to $\mathbf{U}_d = \exp(j\mathbf{A}\mathbf{A}^H)$ where $\mathbf{A}$ is a $d \times d$ i.i.d. Gaussian matrix. In particular, for $\mathbf{U}_d = \mathbf{I}_d$, we obtain the conventional recovery threshold as $k = kd < (\sqrt{R}+1)/2$, which allows us to conclude that exploiting block-sparsity can result in guaranteed recovery for a sparsity level that is d times higher than what would be obtained in the (conventional) sparse case.
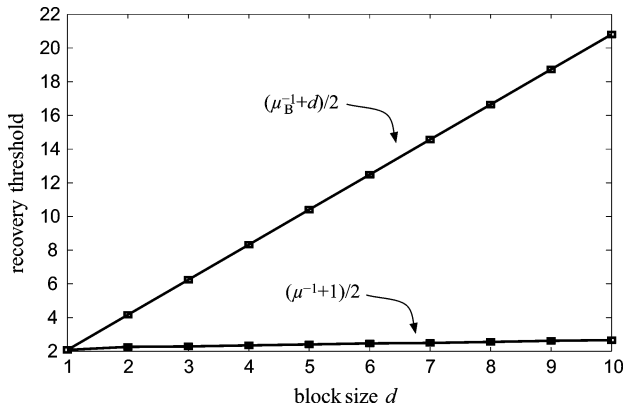
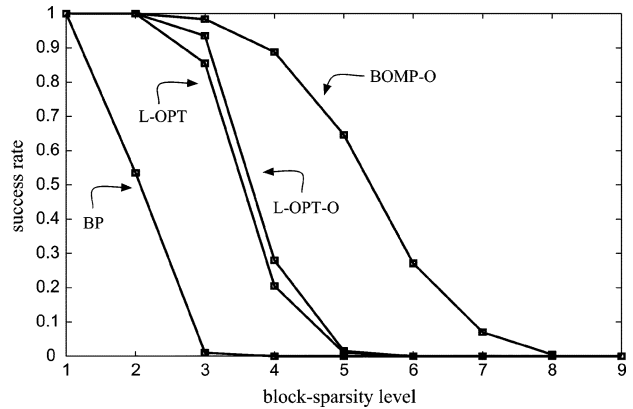Fig. 1. Recovery thresholds for both block-sparsity and conventional sparsity for $R = 10$ as a function of d.

Fig. 2. Performance of OMP, BOMP, and BOMP-O for a dictionary with $L = 40$, $N = 400$, and $d = 4$.
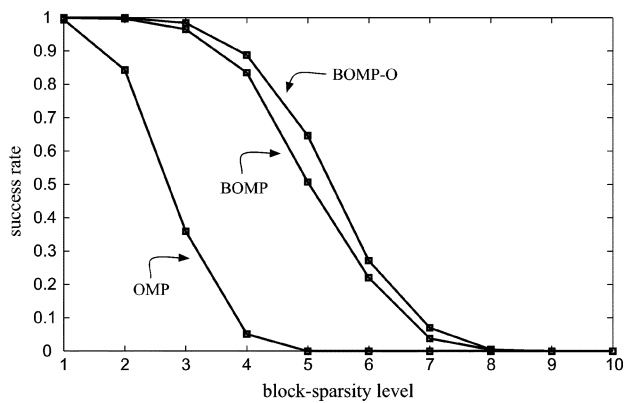

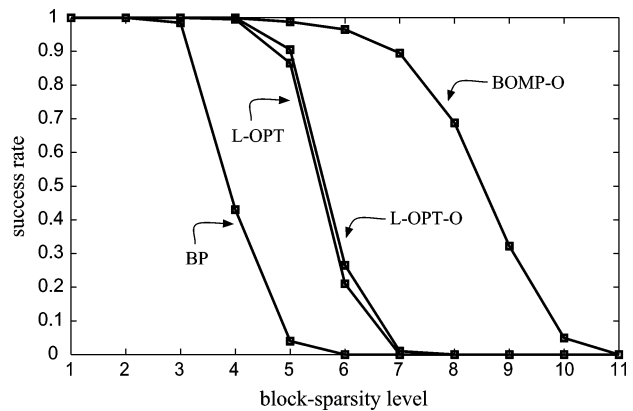Fig. 3. Performance of OMP, BOMP, and BOMP-O for a dictionary with $L = 80$, $N = 160$, and $d = 8$.

Fig. 4. Performance of BP, L-OPT, L-OPT-O, and BOMP-O for a dictionary with $L = 40$, $N = 400$, and $d = 4$.

Fig. 5. Performance of BP, L-OPT, L-OPT-O, and BOMP-O for a dictionary with $L = 80$, $N = 160$, and $d = 8$.
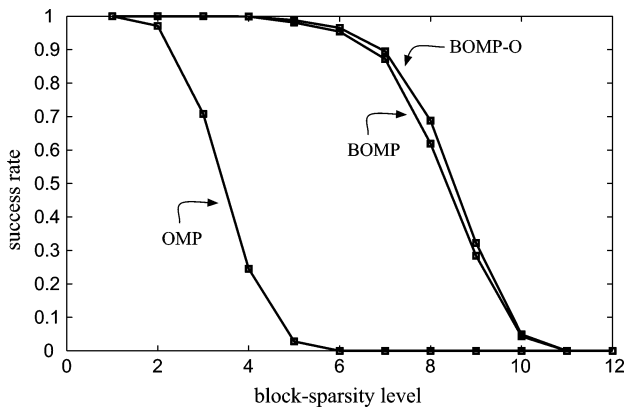
VII. NUMERICAL RESULTS

The aim of this section is to quantify the improvement in the recovery properties of OMP and BP obtained by taking block-sparsity explicitly into account and performing recovery using BOMP and L-OPT, respectively. In all simulation examples below, we randomly generate dictionaries by drawing from i.i.d. Gaussian matrices and normalizing the resulting columns to have norm 1. The dictionary is divided into consecutive blocks of length d. The sparse vector to be recovered has i.i.d. Gaussian entries on the randomly chosen support set according to a uniform prior, i.e., the locations of the nonzero blocks of the block-k sparse vector are chosen uniformly at random among all possible $\binom{M}{k}$ locations.

In Figs. 2 and 3, we plot the recovery success rate[2] as a function of the block-sparsity level of the signal to be recovered. For each block-sparsity level we average over 1000 pairs of realizations of the dictionary and the block-sparse signal. We can see that BOMP outperforms OMP significantly and BOMP with orthogonalized blocks, denoted as BOMP-O, yields slightly better performance than BOMP. We also evaluate the performance of L-OPT compared to BP, as well as L-OPT run on orthogonalized blocks, termed L-OPT-O. For each block-sparsity level we average over 200 pairs of realizations of the dictionary and the block-sparse signal. The corresponding results, depicted in Figs. 4 and 5, show that L-OPT outperforms BP, and L-OPT-O slightly outperforms L-OPT. Furthermore, we can see that BOMP-O significantly outperforms L-OPT-O.

Next, we compare the numerical results depicted in Figs. 3 and 5 for $L = 80$, $N = 160$, and $d = 8$ to the analytical thresholds (41) and (42) and to the thresholds for conventional sparsity reported in [5], [13], and [43]. Specifically, for each realization

---

[2]Success is declared if the recovered vector is within a certain small Euclidean distance of the original vector.

of the random dictionary, we compute the corresponding analytical thresholds and find the following average analytical thresholds (over all realizations of the dictionary):

- BOMP and L-OPT: $kd < 3.2 \Longrightarrow k = 0$
- BOMP-O and L-OPT-O: $kd < 17.7 \Longrightarrow k \leq 2$
- OMP and BP: $kd < 3 \Longrightarrow k = 0$.

Evidently, the analytical thresholds are more pessimistic than the numerical thresholds in Figs. 3 and 5. The latter indicate success rates close to 100% for BOMP, BOMP-O, L-OPT, and L-OPT-O up to a block-sparsity level of $k = 4$ and for OMP and BP up to $k = 2$.

VIII. CONCLUSION

This paper extends the concepts of uncertainty relations, coherence, and recovery thresholds for matching pursuit and basis pursuit to the case of sparse signals that have block-sparsity structure. The extension is made possible by an appropriate definition of block-coherence.

The motivation for considering block-sparse signals is twofold. First, in many applications the nonzero elements of sparse vectors tend to cluster in blocks; several examples are given in [16]. Second, it is shown in [16] that sampling problems over unions of subspaces can be converted into block-sparse recovery problems. Specifically, this is true when the union has a direct-sum decomposition, which is the case in many applications including multiband signals [18]–[20], [29]. As a consequence we obtain the first general class of concrete recovery methods for union of subspace problems. This was the main contribution of [16] together with equivalence and robustness proofs for L-OPT based on a suitably modified definition of the restricted isometry property. Here, we complement this contribution by developing similar results using the concept of block-coherence.

Interesting avenues of further research on block-sparsity include investigating the stability of BOMP in the presence of noise (as done in [47] for the $d = 1$ case) and finding recovery guarantees for mixed ℓ_p/ℓ_q-norm algorithms (see [48] and references therein).

APPENDIX A
PROOF OF LEMMA 1

We first prove (46):

$$
\begin{aligned}
\|\mathbf{A}\mathbf{x}\|_{2,\infty} &= \max_j \left\| \sum_i \mathbf{A}[j,i]\mathbf{x}[i] \right\|_2 \\
&\leq \max_j \sum_i \|\mathbf{A}[j,i]\mathbf{x}[i]\|_2 \\
&\leq \max_j \sum_i \|\mathbf{x}[i]\|_2 \, \rho(\mathbf{A}[j,i]) \\
&\leq \|\mathbf{x}\|_{2,\infty} \max_j \sum_i \rho(\mathbf{A}[j,i]).
\end{aligned} \tag{75}
$$

Therefore, for any $\mathbf{x} \in \mathbb{C}^N$ with $\mathbf{x} \neq \mathbf{0}$, we have

$$
\frac{\|\mathbf{A}\mathbf{x}\|_{2,\infty}}{\|\mathbf{x}\|_{2,\infty}} \leq \rho_r(\mathbf{A}) \tag{76}
$$

which establishes (46). The proof of (47) is similar:

$$
\begin{aligned}
\|\mathbf{A}\mathbf{x}\|_{2,1} &= \sum_j \left\| \sum_i \mathbf{A}[j,i]\mathbf{x}[i] \right\|_2 \\
&\leq \sum_j \sum_i \|\mathbf{A}[j,i]\mathbf{x}[i]\|_2 \\
&\leq \sum_i \|\mathbf{x}[i]\|_2 \sum_j \rho(\mathbf{A}[j,i]) \\
&\leq \rho_c(\mathbf{A})\|\mathbf{x}\|_{2,1} \tag{77}
\end{aligned}
$$

from which the result follows. Finally, we have $\rho_c(\mathbf{A}^H) = \max_r \sum_\ell \rho(\mathbf{A}^H[\ell,r]) = \max_r \sum_\ell \rho(\mathbf{A}[r,\ell]) = \rho_r(\mathbf{A})$.

APPENDIX B
PROOF OF LEMMA 2

Nonnegativity and positivity follow immediately from the fact that the spectral norm is a matrix norm [44, p. 295]. Homogeneity follows by noting that

$$
\begin{aligned}
\rho_c(\alpha\mathbf{A}) &= \max_r \sum_\ell \rho(\alpha\mathbf{A}[\ell,r]) \\
&= \max_r \sum_\ell |\alpha|\rho(\mathbf{A}[\ell,r]) \\
&= |\alpha|\rho_c(\mathbf{A}). \tag{78}
\end{aligned}
$$

The triangle inequality is obtained as follows:

$$
\begin{aligned}
\rho_c(\mathbf{A} + \mathbf{B}) &= \max_r \sum_\ell \rho(\mathbf{A}[\ell,r] + \mathbf{B}[\ell,r]) \\
&\leq \max_r \left(\sum_\ell \rho(\mathbf{A}[\ell,r]) + \sum_\ell \rho(\mathbf{B}[\ell,r]) \right) \tag{79} \\
&\leq \max_r \sum_\ell \rho(\mathbf{A}[\ell,r]) + \max_r \sum_\ell \rho(\mathbf{B}[\ell,r]) \\
&= \rho_c(\mathbf{A}) + \rho_c(\mathbf{B}) \tag{80}
\end{aligned}
$$

where (79) is a consequence of the spectral norm satisfying the triangle inequality.

Finally, to verify submultiplicativity, we use the fact that

$$
\rho_c(\mathbf{A}\mathbf{B}) = \max_\ell \rho_c(\mathbf{A}\mathbf{B}[\ell]). \tag{81}
$$

Therefore, if we prove the inequality

$$
\rho_c(\mathbf{A}\mathbf{B}[\ell]) \leq \rho_c(\mathbf{A})\rho_c(\mathbf{B}[\ell]) \tag{82}
$$

then the result follows from (81) and the fact that $\max_\ell \rho_c(\mathbf{B}[\ell]) = \rho_c(\mathbf{B})$. To prove (82), note that

$$
\begin{aligned}
\rho_c(\mathbf{A}\mathbf{B}[\ell]) &= \sum_i \rho\left(\sum_j \mathbf{A}[i,j]\mathbf{B}[j,\ell] \right) \\
&\leq \sum_i \sum_j \rho(\mathbf{A}[i,j]\mathbf{B}[j,\ell]) \\
&\leq \sum_i \sum_j \rho(\mathbf{A}[i,j])\rho(\mathbf{B}[j,\ell]) \tag{83}
\end{aligned}
$$

where we used the triangle inequality for, and the submultiplicativity of, the spectral norm. Now, we have

$$\sum_i \rho(\mathbf{A}[i,j]) \le \max_\ell \sum_i \rho(\mathbf{A}[i,\ell]) = \rho_c(\mathbf{A}). \tag{84}$$

Substituting into (83) yields

$$\rho_c(\mathbf{AB}[\ell]) \le \rho_c(\mathbf{A}) \sum_j \rho(\mathbf{B}[j,\ell]) = \rho_c(\mathbf{A})\rho_c(\mathbf{B}[\ell]) \tag{85}$$

which completes the proof.

APPENDIX C
PROOF OF LEMMA 3

The proof of the statement $\|\mathbf{Av}\|_{2,1} \le \rho_c(\mathbf{A})\|\mathbf{v}\|_{2,1}$ follows directly from (77) by replacing $\mathbf{A}$ by an $L \times (kd)$ matrix and $\mathbf{x} \in \mathbb{C}^N$ by $\mathbf{v} \in \mathbb{C}^{kd}$ with $\|\mathbf{v}[l]\|_2 > 0$, for all ℓ. If the $a_i = \sum_j \rho(\mathbf{A}[j,i])$ are not all equal, then the last inequality in (77) is strict. Since $a_i = \rho_c(\mathbf{AJ}_i)$ the result follows.

REFERENCES

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constr. Approx.*, vol. 13, no. 1, pp. 57–98, 1997.

[4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[5] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[8] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Systems, Comput.*, Nov. 1993, pp. 40–44.

[9] S. G. Mallat and Z. Zhang, "Matching pursuits and time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[10] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.

[11] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[12] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[13] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[14] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 255–261, Jan. 2006.

[15] Y. C. Eldar, "Uncertainty relations for shift-invariant analog signals," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5742–5757, Dec. 2009.

[16] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[17] Y. C. Eldar and M. Mishali, "Block-sparsity and sampling over a union of subspaces," in *Proc. 16th Int. Conf. Digital Signal Processing*, Jul. 2009, pp. 1–8.

[18] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[19] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2009.

[20] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, no. 1, pp. 37–52, 1967.

[21] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," Electr. Eng. Dept., Technion—Israel Institute of Technology, Haifa, CCIT Report #751 Dec-09, EE Pub no. 1708; arXiv 0912.2495, 2009.

[22] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275–285, Jun. 2008.

[23] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[24] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[25] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.

[26] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[27] Y. M. Lu and M. N. Do, "Sampling signals from a union of subspaces," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 41–47, Mar. 2008.

[28] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, Apr. 2009.

[29] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, Aug. 2009.

[30] P. G. Casazza and G. Kutyniok, "Frames of subspaces," in *Wavelets, Frames, and Operator Theory*, ser. Contemp. Math. Providence, RI: Amer. Math. Soc., 2004, vol. 345, pp. 87–113.

[31] K. Gedalyahu and Y. C. Eldar, "Time delay estimation from low rate samples: A union of subspaces approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, Jun. 2010.

[32] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, May 2010.

[33] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[34] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[35] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[36] L. Peotta and P. Vandergheynst, "Matching pursuit with block incoherent dictionaries," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4549–4557, Sep. 2007.

[37] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[38] C. Chesneau and M. Hebiri, "Some theoretical results on the Grouped Variables Lasso," *Math. Methods Statist.*, vol. 17, no. 4, pp. 317–326, Dec. 2008.

[39] Y. Nardi and A. Rinaldo, "On the asymptotic properties of the group lasso estimator for linear models," *Electron. J. Statist.*, vol. 2, pp. 605–633, 2008.

[40] F. R. Bach, "Consistency of the group Lasso and multiple kernel learning," *J. Mach. Learn. Res.*, vol. 9, pp. 1179–1225, 2008.

[41] L. Meier, S. v. d. Geer, and P. Bühlmann, "The group lasso for logistic regression," *J. R. Statist. Soc. B*, vol. 70, no. 1, pp. 53–77, 2008.

[42] J. Huang and T. Zhang, "The benefit of group sparsity," Feb. 2009, ArXiv:0901.2962v2 [stat.ML].

[43] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[44] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge Press, 1985.

[45] R. A. DeVore and V. N. Temlyakov, "Some remarks on greedy algorithms," *Adv. Comp. Math.*, vol. 5, pp. 173–187, 1996.

[46] T. Strohmer and R. W. Heath, Jr, "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, May 2003.

[47] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[48] M. Kowalski, "Sparse regression using mixed norms," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 303–324, Nov. 2009.

Patrick Kuppinger (S'07) was born in Freiburg, Germany, on May 5, 1983. He received the B.Sc. degree in electrical engineering and information technology in 2005 from ETH Zurich, and the M.Sc. degree with distinction in electrical engineering in 2006 from Imperial College London, London, U.K.

Since 2006, he has been working toward the Dr.sc. degree at ETH Zurich. His research interests are in signal processing and communication theory.

Mr. Kuppinger held a scholarship from the Klaus Murmann Fellowship Program (Stiftung der Deutschen Wirtschaft) (2004–2006) and was the recipient of an IDEA League scholarship during his stay at Imperial College London.

Yonina C. Eldar (S'98–M'02–SM'07) 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 2001 from the Massachusetts Institute of Technology (MIT), Cambridge.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group, MIT. She is currently a Professor with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa. She is also a Research Affiliate with the Research Laboratory of Electronics, MIT, and a visiting Professor at Stanford University, Stanford, CA. Her research interests are in the general areas of sampling theory, statistical signal processing, optimization methods, and their applications to biology and optics.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel and David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, and in 2009 the Technion's Award for Excellence in Teaching. She is a member of the IEEE Signal Processing Theory and Methods technical committee and the Bio Imaging Signal Processing Technical Committee, 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*, and on the Editorial Board of *Foundations and Trends in Signal Processing*.

Helmut Bölcskei (M'98–SM'02–F'09) was born in Mödling, Cisleithania, Austria-Hungary, on May 29, 1970, and received the Dipl.-Ing. and Dr.techn. degrees in electrical engineering from Vienna University of Technology, Vienna, Austria, in 1994 and 1997, respectively.

From 1999 to 2001, he was a Postdoctoral Researcher with the Department of Electrical Engineering and the Department of Statistics, Stanford University, Stanford, CA. He was in the founding team of Iospan Wireless Inc., a Silicon Valley-based startup company (acquired by Intel Corporation in 2002) specialized in multiple-input multiple-output (MIMO) wireless systems for high-speed Internet access. From 2001 to 2002, he was an Assistant Professor of Electrical Engineering with the University of Illinois at Urbana-Champaign. He has been with ETH Zurich, since 2002, where he is Professor of Electrical Engineering. He was a visiting researcher with Philips Research Laboratories Eindhoven, The Netherlands, ENST Paris, France, and the Heinrich Hertz Institute Berlin, Germany. His research interests are in information and communication theory, signal processing, and harmonic analysis.

Dr. Bölcskei received the 2001 IEEE Signal Processing Society Young Author Best Paper Award, the 2006 IEEE Communications Society Leonard G. Abraham Best Paper Award, the ETH "Golden Owl" Teaching Award, and was an Erwin Schrödinger Fellow (1999–2001) of the Austrian National Science Foundation (FWF). He was a plenary speaker at several IEEE conferences and served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the EURASIP *Journal on Applied Signal Processing*. He is currently on the editorial board of *Foundations and Trends in Networking*, serves as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, was TPC Co-Chair of the 2008 IEEE International Symposium on Information Theory, and serves on the Board of Governors of the IEEE Information Theory Society.