Compressive Demodulation of Mutually Interfering Signals

Yuejie Chi, Yao Xie and Robert Calderbank

Abstract

Multi-User Detection is fundamental not only to cellular wireless communication but also to Radio-Frequency Identification (RFID) technology that supports supply chain management. The challenge of Multi-user Detection (MUD) is that of demodulating mutually interfering signals, and the two biggest impediments are the asynchronous character of random access and the lack of channel state information. Given that at any time instant the number of active users is typically small, the promise of Compressive Sensing (CS) is the demodulation of sparse superpositions of signature waveforms from very few measurements. This paper begins by unifying two front-end architectures proposed for MUD by showing that both lead to the same discrete signal model. Algorithms are presented for coherent and noncoherent detection that are based on iterative matching pursuit. Noncoherent detection is all that is needed in the application to RFID technology where it is only the identity of the active users that is required. The coherent detector is also able to recover the transmitted symbols. It is shown that compressive demodulation requires $\mathcal{O}(K \log N(\tau+1))$ samples to recover K active users whereas standard MUD requires $N(\tau+1)$ samples to process N total users with a maximal delay τ. Performance guarantees are derived for both coherent and noncoherent detection that are identical in the way they scale with number of active users. The power profile of the active users is shown to be less important than the SNR of the weakest user. Gabor frames and Kerdock codes are proposed as signature waveforms and numerical

Y. Chi is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA (email: chi@ece.osu.edu).

Y. Xie is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 (email: yao.xie@duke.edu).

R. Calderbank is with the Department of Computer Science, Duke University, Durham, NC 27708, USA (email: robert.calderbank@duke.edu).

The work of Y. Chi and R. Calderbank was supported by ONR under Grant N00014-08-1-1110, by AFOSR under Grant FA 9550-09-1-0643, and by NSF under Grants NSF CCF -0915299 and NSF CCF-1017431. The work of Y. Xie is supported by DARPA under Grant N66001-11-4002, MSEE under Grant FA8650-11-1-7150, and ARO under Grant W911NF-09-1-0262.

This paper was presented in part at the 2012 International Statistical Signal Processing Workshop (SSP) and the 2012 Allerton Conference on Communication, Control, and Computing.

examples demonstrate the superior performance of Kerdock codes - the same probability of error with less than half the samples.

Index Terms

multi-user detection, asynchronous random access, sparse recovery, iterative matching pursuit, Gabor frame, Kerdock code

I. INTRODUCTION

Demodulation of mutually interfering signals, or Multi-User Detection (MUD) is central to multiaccess communications [1]. It includes the special case of the "on-off" Random Access Channel (RAC) [2] that arises in modeling control channels in wireless networks, where active users transmitting their signature waveforms can be modeled as sending 1's to the Base Station (BS), and inactive users can be modeled as sending 0's. It also includes the special case of the Radio-Frequency Identification (RFID) system [3] that arises in supply chain management, where each RFID tag is associated with a unique ID and attached to a physical object. In large scale RFID applications, an RFID reader interrogates the environment and all tags within its operational range can be modeled as sending 1's, and tags outside its operational range can be modeled as sending 0's. It also includes the special case of neighbor discovery in wireless ad-hoc networks [4], [5], where neighbors of a query node transmitting their identity information can be modeled as sending 1's, and nonneighbors can be modeled as sending 0's. In all examples, the received signals are possibly corrupted by noise.

State-of-the-art random access protocols, such as IEEE 802.11 standards, rely on retransmission with random delays at each active user to avoid collisions. This accumulates to significant delays as the size of the networks becomes large, for example the scale of RFID tags can easily grow to millions in practice. Therefore it is of great interest to allow multiple active users transmit simultaneously and still be able to recover the active users albeit collisions. The MUD problem becomes the recovery of the active users, and it may be expanded to demodulation of transmitted symbols from each active user in cellular communications. The two biggest impediments are the asynchronous character of random access and the lack of Channel State Information (CSI) at the receiver. The signature waveforms of different users are obtained by modulating a chip waveform using a digital sequence of length L. The total number of users N is severely constrained if all signature waveforms are orthogonal, giving the relationship $N \leq L$. In this paper we are interested in both coherent detection when CSI is known and noncoherent detection when CSI is unknown, under the conditions that the signature waveforms are nonorthogonal and the

delays of each user are unknown.

A. Main Contributions

Our contributions in this paper are three-fold. Given that at any time instant the number of active users K is typically small, the promise of Compressive Sensing (CS) [6], [7] is the demodulation of sparse superpositions of signature waveforms from very few measurements. A baseline architecture for MUD is correlation of the received signal with a bank of matched filters [1], each with respect to a shift of a signature waveform. The first drawback is the huge number of required filters, thus the required number of samples, when the number of total users N is large, which is $N_\tau = N(\tau+1)$ where τ is the maximum delay. A second drawback is that the noise will be colored and amplified by the cross-correlations of selected signature waveforms. An alternative baseline architecture is sampling the received analog signal directly at the chip rate [8]. This approach does not amplify the noise but it does require a high-rate Analog-to-Digital Converter (ADC).

We first demonstrate two front-end architectures for compressive demodulation which can lead to mathematically equivalent discrete signal models. The first architecture is based on subsampling the received signal uniformly at random, which reduces the required rate of ADC in [8]. The second architecture is based on a bank of generalized matched filters, which is the extension to asynchronous communication of the architecture for synchronous MUD proposed by Xie et. al. [9] based on analog compressed sensing [10]. The novelty is that both architectures are unified under the same discrete signal model, and further reduce the number of acquired samples M to be smaller than the length of the signature waveforms L.

Second we present architectures for coherent and noncoherent detection, designed to recover active users and transmitted (QPSK) symbols when the CSI is known, and to recover active users when the CSI is unknown. Both algorithms are based on iterative matching pursuit [11] and assume a flat-fading channel model so that each active user arrives at the receiver on a single path with an *unknown* delay. We note that the generalization to a small number of arrival paths with a prescribed delay pattern is straightforward. Noncoherent detection is more pertinent to applications like RFID and wireless ad hoc networks, where only identification of active users is of interest. Our main theoretical contribution is relating the probability of error for the proposed MUD algorithms to two geometric metrics associated with the set of subsampled signature waveforms. These metrics, the worst case and average coherence, were introduced by Bajwa et. al. in the context of model selection [12]. We provide explicit performance guarantees in terms of these coherence metrics and the distribution of received signal powers. These

fundamental limits quantify robustness of the compressive MUD algorithms to the "near-far" problem [1] in multiple access communications. It is shown our proposed compressive MUD algorithms require $\mathcal{O}(K \log N_\tau)$ samples to recover K active users for both coherent and noncoherent detection, whereas standard MUD requires N_τ samples. We further show that the minimum signal-to-noise ratio dictated by the weakest active user, rather than the power profile of all active users, plays an important role in the performance of the proposed iterative algorithms; therefore power control is less critical.

Finally, we propose deterministic designs of cyclic-extended signature waveforms that satisfy both the geometric metrics linked to the decoding algorithms and the block-circulant structure due to cyclic extensions from the asynchronous character. Gabor frames and Kerdock codes are considered due to their optimal coherence properties proved in [12], [13], and in this paper we extend this analysis to the uniformly random subsampled Gabor frames and Kerdock codes. Gabor frames are block circulant from its construction as a time-frequency expansion of a seed sequence. The Kerdock code is an extended cyclic code over $\mathbb{Z}_4$ (Section IV, [14]) and can be arranged to exhibit a block-circulant structure. We demonstrate through numerical simulations that the performance of the proposed compressive MUD algorithms using Gabor frames and Kerdock codes. The superior performance of Kerdock code is emphasized for practical interests, which can obtain the same probability of error with less than half the samples.

B. Relationship to Prior Work

Here we describe how this paper differs from previous papers that have also formulated MUD as a compressive sensing problem. The focus of most prior work is on synchronous communication, including [2], [4], [5], [9], [15], [16]. In [2], Fletcher et. al. studied MUD in the context of on-off RACs; in [4], [5], Zhang et. al. studied MUD in the context of neighbor discovery in wireless ad hoc networks; in [9], Xie et. al. studied MUD with simultaneous symbol detection in cellular communications. The synchronous model provides insight into what might be possible but it ignores the difficulty in estimating the delays of individual users and in achieving synchronization.

A more general asynchronous model is considered by Applebaum et. al. in [8]. These authors assume synchronization at the chip or symbol level, different signature waveforms arrive with different discrete delays in some finite window, and the receiver uses convex optimization to recover the constituents of the sparse superposition. Thus users are associated with a Toeplitz block in the measurement matrix populated by allowable shifts in the signature waveform. In this paper we introduce a cyclic prefix in order to create a measurement matrix with a block cyclic structure which makes it easier to design codebooks using Gabor frames and Kerdock codes.

The algorithms presented in this paper are based on iterative matching pursuit and for uniformly random delays the number of samples they require is of the same order, $\mathcal{O}(K \log N_\tau)$, as the number required by the convex optimization algorithm presented in [8]. This scaling is a significant improvement over the Reduced-Dimension Decision Feedback (RDDF) detector described in [9] which requires order $\mathcal{O}(K^2 \log N_\tau)$ samples. The reason that we are able to break the *square-root bottleneck* is that by introducing more sophisticated coherence metrics we are able to treat average case rather than worst case performance. These methods may be of independent interest. Note also that the complexity of our algorithms are significantly less than that of of convex optimization when the set of active users is highly sparse ($K \ll N_\tau$) [17]. Moreover, it is possible to further reduce the complexity by terminating the algorithm early and obtaining partial recovery of active users. When the channel is known at the receiver we also improve upon the transmission rate reported by Xie et. al. [9] by incorporating complex channel gains in our model and moving from BPSK to QPSK signaling.

Our focus on deterministic signature waveforms is different from most previous work [2], [4], [5] which considers random waveforms. The fact that random waveforms can be shown to satisfy the Restricted Isometry Property [6] makes analysis possible but they are not very practical. The same criticism can be leveled at the RDDF detector described in [9] where randomness enters the choice of the coefficients determining the filter bank. Randomness also enters into [5] through the pattern of puncturing of Reed-Muller codewords which serve as deterministic signature waveforms.

C. Organization of this paper and Notations

The rest of the paper is organized as follows. Section II describes the system model, and Section III presents two architectures for the compressive MUD front-end. Section IV proposes the coherent and noncoherent detectors, along with their performance guarantees. Section V proves the main theorems. Section VI presents the design of signature waveforms based on Gabor frames and Kerdock codes. Section VII shows the numerical simulations and Section VIII concludes the paper.

Throughout the paper, we use capital bold letters $\boldsymbol{A}$ to denote matrices, small bold letters $\boldsymbol{a}$ to denote vectors, $\|\boldsymbol{A}\|_p$ and $\|\boldsymbol{a}\|_p$ to denote the p-norm of $\boldsymbol{A}$ and $\boldsymbol{a}$, where p = 2 or ∞. $\boldsymbol{I}_N$ denotes the identity matrix of dimension N, $\dagger$ denotes pseudo-inverse, $\boldsymbol{A}^H$ denotes the Hermitian of $\boldsymbol{A}$, and c^* defines the conjugate of a complex number c.

II. System Model

Consider a multi-user system of N total user where the nth users, $n = 1, \cdots, N$, communicate using spread spectrum waveform of the form

$$x_n(t) = \sqrt{P_n} \sum_{\ell=0}^{L-1} a_{n,\ell} p(t - \ell T_c), \quad t \in [0, T), \tag{1}$$

where $p(t)$ is a unit-energy pulse $\int |p(t)|^2 dt = 1$, $\int p^*(t - \ell T_c) p(t - kT_c) dt = 0$, $(\cdot)^*$ denoting the conjugate operation, for $\ell \neq k$. The chip duration T_c determines the system bandwidth, T is the symbol duration, P_n denotes the transmit power of the nth user, and the spreading codeword

$$\boldsymbol{a}_n = [a_{n,0} \quad \cdots \quad a_{n,L-1}]^\mathsf{T}, \quad n = 1, \cdots, N, \tag{2}$$

is the L-length (real- or complex-valued) codeword of unit energy $\|\boldsymbol{a}_n\|_2 = 1$ assigned to the nth user. Typically $L < N$. The notation $^\mathsf{T}$ denotes transpose of a matrix or vector.

To simplify the model, we consider a one-shot model, where the user sends one symbol at a time rather than sending a sequence of symbols. The signal at the receiver is given by

$$y(t) = \sum_{n=1}^{N} g_n \sqrt{P_n} \delta_{\{n \in \mathcal{I}\}} b_n x_n(t - \tau_n') + w(t), \tag{3}$$

where $g_n \in \mathbb{C}$ and $\tau_n' \in \mathbb{R}_+$ are the channel fading coefficient and the continuous delay associated with the nth user, respectively. Define the power profile of all users as $\boldsymbol{r} = [r_1, \cdots, r_N]^\mathsf{T}$, where

$$r_n \triangleq g_n \sqrt{P_n}. \tag{4}$$

The power profile is determined by the power control at the transmitter and the channel coefficients during transmission, which could take complex values.

We assume Quadrature Phased Shift Keying (QPSK) modulation, where $b_n \in \{(-1 - j)/\sqrt{2}, (-1 + j)/\sqrt{2}, (1 - j)/\sqrt{2}, (1 + j)/\sqrt{2}\}$ is the transmitted symbol of the nth user, and $w(t)$ is a complex additive white Gaussian noise (AWGN) introduced by the receiver circuitry with zero mean and variance σ_0^2. Denote by $\mathcal{I}$ the set of active users. We assume the support of active users $\mathcal{I}$ is a uniform random K-subset of $[\![N]\!] \triangleq \{1, \ldots, N\}$. The Dirac function $\delta_x = 1$ if x is true and $\delta_x = 0$ otherwise.

Define the individual discrete delays $\tau_n \triangleq \lfloor \tau_n'/T_c \rfloor \in \mathbb{Z}_+$, and the maximum discrete delay $\tau \triangleq \max_n \tau_n \in \mathbb{Z}_+$. While the values of τ_n are unknown, τ is assumed to be known by the transmitters and receivers.

Each vectors $\boldsymbol{a}_n$ is the cyclic prefix of a vector $\tilde{\boldsymbol{a}}_n$ of length $P = L - (\tau + 1)$. As shown in Fig. 1, $\boldsymbol{a}_n$ is obtained by appending the first $\tau + 1$ symbols of $\tilde{\boldsymbol{a}}_n$ to the end of $\tilde{\boldsymbol{a}}_n$, we have $\tilde{a}_{n,\ell} = a_{n,P-\tau-\ell+l}$ for

$\ell = 1, \ldots, \tau + 1$. As a result, any length P sub-sequence of the vectors $\boldsymbol{a}_n$ will be a cyclic shift of $\tilde{\boldsymbol{a}}_n$.

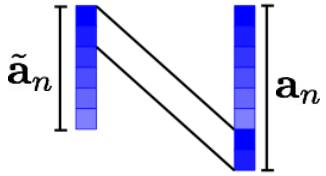


Fig. 1: Illustration of the cyclic prefix in the construction of spreading codewords.

III. Compressive MUD Front-End

In this Section we describe two front-end architectures for compressive MUD. The first is the chip-rate subsampling architecture considered in [18]; and the second is the asynchronous case of a bank of generalized matched filters architecture considered in [9]. We begin by showing that mathematically, the two front-end architectures are equivalent.

A. Chip-rate subsampling architecture

The chip-rate subsampling architecture directly samples the continuous received signal at the chip rate using a high-rate ADC as shown in Fig. 2 (a). The receiver only starts sampling when the waveforms of all active users have arrived. Starting at sample $(\tau + 1)$, it collects M uniformly random samples over a window of length L. These samples, or linear combinations thereof constitute the measurements made by the receiver. We assume the codewords are of a reasonable length relative to the delays such that $L > M$. As a result, the output data vector can be written as

$$\boldsymbol{y} = \bar{\boldsymbol{H}} \boldsymbol{I}_\Omega \boldsymbol{A} \boldsymbol{R} \boldsymbol{b} + \boldsymbol{w}, \tag{5}$$

where $\boldsymbol{y} \in \mathbb{C}^{M \times 1}$, $\boldsymbol{A} \in \mathbb{C}^{P \times N_\tau}$, and the noise $\boldsymbol{w} \in \mathbb{C}^{M \times 1}$ is complex Gaussian distributed with zero mean and variance $\sigma_0^2 \bar{\boldsymbol{H}} \bar{\boldsymbol{H}}^H$. The subsampling matrix is defined as $\boldsymbol{I}_\Omega \in \mathbb{R}^{M \times P}$, where Ω denotes indices of samples, and $\bar{\boldsymbol{H}} \in \mathbb{C}^{M \times M}$ is a matrix that linearly combines the samples. The columns of matrix $\boldsymbol{A}$ have a block structure with each block consisting of circulant shifts of a codeword. Define a circulant matrix $\boldsymbol{A}_n$ as

$$\boldsymbol{A}_n = \begin{bmatrix} \mathcal{T}_0 \tilde{\boldsymbol{a}}_n & \mathcal{T}_1 \tilde{\boldsymbol{a}}_n & \cdots & \mathcal{T}_\tau \tilde{\boldsymbol{a}}_n \end{bmatrix} \in \mathbb{C}^{P \times (\tau+1)}, \tag{6}$$

where the notation $\mathcal{T}_k$ denotes the circulant shift matrix by k, and

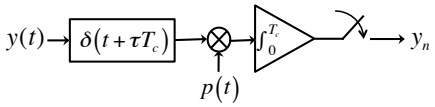
$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \cdots & \boldsymbol{A}_N \end{bmatrix} \in \mathbb{C}^{P \times N_\tau}. \tag{7}$$

The vector $\boldsymbol{b} \in \mathbb{C}^{N_\tau}$ contains the transmitted symbols; it is a concatenation of N vectors $\boldsymbol{b}'_n$ of length $\tau + 1$, each with at most one non-zero entry at the location of τ_n:

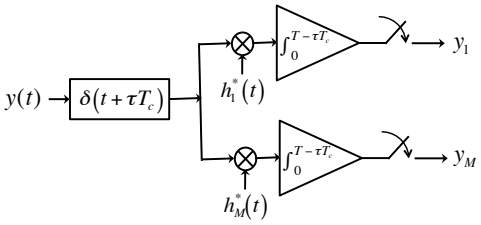
$$b'_{n,m} = b_n \delta_{\{m=\tau_n\}}, \quad m = 0, \cdots, \tau.$$

The entries R_{mm} of the diagonal matrix $\boldsymbol{R} \in \mathbb{C}^{N_\tau}$ are a function of the channel gain, the transmitted power, and the transmitted symbols:

$$R_{mm} = r_n \delta_{\{m=(n-1)(\tau+1)+\tau_n\}}, \quad n = 1, \ldots, N, \quad m = 0, \ldots, N_\tau - 1. \tag{8}$$



(a) chip-rate subsampling



(b) a bank of generalized matched filters

Fig. 2: Illustration of two architectures: (a) the chip-rate subsampling architecture, and (b) a bank of generalized matched filters architecture, where the first block is a linear filter with impulse response $\delta(t + \tau T_c)$.

B. A bank of generalized matched filters

A generalized matched filter for compressive MUD [9] correlates $y(t)$ with a set of signals $\{h_m(t)\}_{m=1}^M$, as shown in Fig. 2 (b). The measurement is taken by multiplying a delayed version of $y(t + \tau T_c)$ with $h_m(t)$, and integrating over a window of length $T - \tau T_c$, where T is the symbol period. The output of

the mth measurement is given by

$$\begin{aligned} y_m &= \int_0^{T-\tau T_c} h_m^*(t) y(t+\tau T_c) dt \\ &\triangleq \langle h_m(t), y(t+\tau T_c) \rangle, \quad m = 1, \cdots, M. \end{aligned}$$

Writing this in a vector notation, we have

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{R}\boldsymbol{b} + \boldsymbol{w},$$

where

$$\boldsymbol{B} = [\boldsymbol{B}_1, \cdots, \boldsymbol{B}_N] \in \mathbb{C}^{M \times N_\tau}, \tag{9}$$

with $\boldsymbol{B}_n \in \mathbb{C}^{M \times (\tau+1)}$, for $n = 1, \cdots, N$. The $(m, \ell+1)$th entry of $\boldsymbol{B}_n$ is given by

$$\begin{aligned} [\boldsymbol{B}_n]_{m,\ell+1} &= \int_0^{T-\tau T_c} h_m^*(t) x_n(t+\tau T_c - \ell T_c), \\ & \ell = 0, \cdots, \tau, \quad m = 1, \cdots, M. \end{aligned}$$

The noise vector $\boldsymbol{w}$ is a M-dimensional complex Gaussian vector with zero mean and covariance matrix

$$[\boldsymbol{\Sigma}]_{mk} = \sigma_0^2 \int_0^{T-\tau T_c} h_m^*(t) h_k(t) dt. \tag{10}$$

We now parameterize for the generalized matched filters $\{h_m(t)\}_{m=1}^M$. In [9], the matched filters are constructed as linear combinations of the bi-orthogonal signals of the user signature waveforms. Here we consider a more general construction that can lead to a discrete model equivalent to that of the chip-rate subsampling (5). Assume the measurement signals are constructed using the chip waveform and chip sequences as

$$h_m(t) = \sum_{\ell=0}^{L} h_{m,\ell} p(t - \ell T_c), \tag{11}$$

where

$$\boldsymbol{h}_m = [h_{m,0} \quad \cdots \quad h_{m,L-1}]^\top, \quad m = 1, \cdots, M, \tag{12}$$

is the L-length (real- or complex-valued) codeword for the mth measurement signal. By this parameter-

ization, for $\ell = 0, \cdots, \tau$,

$$\begin{aligned}
&[\boldsymbol{B}_n]_{m,\ell+1} \\
&= \int_0^{T-\tau T_c} h_m^*(t) x_n(t + (\tau - \ell)T_c) dt \\
&= \sum_{u=0}^{L} \sum_{v=0}^{L} h_{m,u}^* a_{n,v} \int_0^{T-\tau T_c} p^*(t - uT_c) p(t + (\tau - l - v)T_c) dt \\
&= \sum_{u=0}^{L} h_{m,u}^* [\mathcal{T}_l \tilde{\boldsymbol{a}}_n]_u = \boldsymbol{h}_m^H (\mathcal{T}_l \tilde{\boldsymbol{a}}_n),
\end{aligned} \tag{13}$$

where we have used $\int_0^{T-\tau T_c} p^*(t - uT_c) p(t + (\tau - l - v)T_c) dt = \delta_{\{u=\tau-l-v\}}$. Hence, from (9) and (13), we obtain that

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & \cdots & \boldsymbol{B}_N \end{bmatrix} = \boldsymbol{H}\boldsymbol{A}, \tag{14}$$

where

$$\boldsymbol{B}_n = \begin{bmatrix} \boldsymbol{h}_1^H(\mathcal{T}_0 \tilde{\boldsymbol{a}}_n) & \cdots & \boldsymbol{h}_1^H(\mathcal{T}_\tau \tilde{\boldsymbol{a}}_n) \\ \vdots & & \vdots \\ \boldsymbol{h}_M^H(\mathcal{T}_0 \tilde{\boldsymbol{a}}_1) & \cdots & \boldsymbol{h}_M^H(\mathcal{T}_\tau \tilde{\boldsymbol{a}}_1) \end{bmatrix}.$$

The noise in the mth measurement is given by $\langle h_m(t), w(t) \rangle$, which is a complex Gaussian random variable with zero mean and covariance matrix (10) given by $[\boldsymbol{\Sigma}]_{mk} = \sigma_0^2 \boldsymbol{h}_m^H \boldsymbol{h}_k$. Define a matrix

$$\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_M]^H \in \mathbb{C}^{M \times P}.$$

Substituting (14) into (9), we obtain that when the filters $\{h_m(t)\}$ are parameterized by (11), the measurement vector can be written as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{R}\boldsymbol{b} + \boldsymbol{w}, \tag{15}$$

where $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_0^2 \boldsymbol{H}\boldsymbol{H}^H)$ is a complex Gaussian random vector with zero mean and covariance matrix $\sigma_0^2 \boldsymbol{H}\boldsymbol{H}^H$. Given the output (15) of the bank of generalized matched filters, there are two special cases for $\boldsymbol{H}$:

- $\boldsymbol{H} = \bar{\boldsymbol{H}}\boldsymbol{I}_\Omega$, then $\boldsymbol{H}\boldsymbol{H}^H = \bar{\boldsymbol{H}}\bar{\boldsymbol{H}}^H$, which means the output (15) of the second architecture is mathematically equivalent to the chip-rate subsampling architecture (5).
- In the first architecture, if we choose $\bar{\boldsymbol{H}}$ to be an orthogonal matrix, then $\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}^H = \boldsymbol{I}_M$, the output signal power of each measurement is M/N and the noise power is σ_0^2. The signal-to-noise ratio per measurement is $M/(N\sigma_0^2)$.

- In the second architecture, if we choose $\boldsymbol{H}$ to be a tight frame, then $\boldsymbol{H}\boldsymbol{H}^H = (N/M)\boldsymbol{I}_M$. For each measurement, the output signal power is 1, and the noise power is $(N/M)\sigma_0^2$. The signal-to-noise ratio per measurement is $M/(N\sigma_0^2)$.

Table I is a summary of the comparison between these two architectures when $\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}^H = \boldsymbol{I}_M$ and $\boldsymbol{H}\boldsymbol{H}^H = (N/M)\boldsymbol{I}_M$, where $\boldsymbol{I}_M$ is the identity matrix of dimension M. Note that both architectures lead to the same discrete signal model (15). In the following, we will focus on signal recovery and signature waveform designs based on (15).

TABLE I: Comparison of the two architectures.

Architecture	Chip-rate subsampling	Generalized matched filter bank
# of Users	N	N
# of Filters	1	M
# of Samples	M	1
Sampling Rate	$(N/M)T_c$	T $(T \gg T_c)$
Signal Power	M/N	1
Noise Power	σ_0^2	$(N/M)\sigma_0^2$
SNR per measurement	M/N	M/N

IV. Coherent and Noncoherent Detection Algorithms

In the following sections, we choose $\boldsymbol{H}$ as a tight-frame, and hence the noise is white and we assume the noise variance is $\sigma^2 \triangleq N/M\sigma_0^2$. Define

$$\boldsymbol{X} = \boldsymbol{H}\boldsymbol{A} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{N_\tau}] \in \mathbb{C}^{M \times N_\tau}. \tag{16}$$

We further assume that the columns of $\boldsymbol{H}$ and $\boldsymbol{A}$ are scaled so that each column of $\boldsymbol{X}$ is unit-norm: $\|\boldsymbol{x}_n\|_2 = 1$. Hence the model (15) becomes

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{R}\boldsymbol{b} + \boldsymbol{w}, \tag{17}$$

where $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_M)$. Based on this model, we first present a coherent matching pursuit detector based on iterative thresholding to detect active users and their transmitted symbols, when $\boldsymbol{R}$ is assumed known. We also present a noncoherent matching pursuit detector to detect active users when $\boldsymbol{R}$ is assumed unknown, which is adapted from the Orthogonal Matching Pursuit (OMP) algorithm [11].

A. Coherent and Noncoherent Matching Pursuit Detector

The coherent matching pursuit detector is described in Algorithm 1. With knowledge of the number of active users K, the algorithm performs K iterations. In each iteration, Algorithm 1 first finds a user with the strongest correlation with its delayed signature waveforms, then subtracts its exact contribution to the received signal and updates the residual. Since we assume a flat-fading channel[1], there is only one nonzero entry in each user block. Therefore, in the next iteration we can restrict our search to the remaining users. To find the transmitted symbols of each active user, we adopt simple quadrant detectors as in (18) and (19). Our algorithm doubles the rate of the modulation scheme in [9] by considering the complex nature of the power profiles.

Algorithm 1 Coherent Matching Pursuit Detector for Asynchronous MUD

Input: matrices $\boldsymbol{X}$ and $\boldsymbol{R}$, signal vector $\boldsymbol{y}$, number of active users K
Output: active user set $\hat{\mathcal{I}}$, transmitted symbols $\hat{\boldsymbol{b}}$
Initialize: $\mathcal{I}_0 :=$ empty set, $\hat{\boldsymbol{b}}_0 := \boldsymbol{0}$, $\boldsymbol{v}_0 := \boldsymbol{y}$, $\mathcal{X}_0 := \{1, \ldots, N_\tau\}$
for $k = 0$ to $K - 1$ **do**
  Compute: $\boldsymbol{f} := \boldsymbol{X}^H \boldsymbol{v}_k$
  Find $i = \arg\max_{n \in \mathcal{X}_k} |f_n|$
  Detect active users: $\mathcal{I}_{k+1} = \mathcal{I}_k \cup \{\lceil i/(\tau+1) \rceil\}$
  Update: $\mathcal{X}_{k+1} = \mathcal{X}_k \backslash \{\lfloor i/(\tau+1) \rfloor (\tau+1) + 1, \cdots, \lceil i/(\tau+1) \rceil (\tau+1)\}$
  Detect symbols:

$$\Re\{[\hat{\boldsymbol{b}}_{k+1}]_i\} = \frac{1}{\sqrt{2}} \mathrm{sgn}(\Re[r_i^* f_i]), \tag{18}$$

$$\Im\{[\hat{\boldsymbol{b}}_{k+1}]_i\} = \frac{1}{\sqrt{2}} \mathrm{sgn}(\Im[r_i^* f_i]), \tag{19}$$

  where sgn is the sign function, and $\Re(x)$ and $\Im(x)$ takes the real part and imaginary part of x respectively.
  Update $\hat{\boldsymbol{b}}$: $[\hat{\boldsymbol{b}}_{k+1}]_n = [\hat{\boldsymbol{b}}_k]_n$ for $n \neq i$.
  Update residual: $\boldsymbol{v}_{k+1} = \boldsymbol{v}_k - \boldsymbol{X}\boldsymbol{R}\boldsymbol{b}_{k+1}$
end for
$\hat{\mathcal{I}} = \mathcal{I}_K$, $\hat{\boldsymbol{b}} = \hat{\boldsymbol{b}}_K$

The noncoherent matching pursuit detector is described in Algorithm 2. We denote $\boldsymbol{X}_{\mathcal{I}}$ the submatrix (subvector) consisting of columns (entries) of $\boldsymbol{X}$ indexed by $\mathcal{I}$. Given one symbol, it is not possible to resolve the ambiguity in channel phase. Algorithm 2 detects whether a user is active or inactive, and does not recover the transmitted symbols. The residual is updated by subtracting the orthogonal projection of $\boldsymbol{y}$ onto the signal space of the detected users. The noncoherent detector is appropriate for the situation

[1]Our results can be easily generalized to a multipath channel model.

where we do not have access to the channel state information and are only interested in detecting the active users. For example, it is more pertinent in applications like RFID where it is only important to register the presence or absence of a tag.

Algorithm 2 Noncoherent Matching Pursuit Detector for Asynchronous MUD

Input: matrix $\boldsymbol{X}$, signal vector $\boldsymbol{y}$, number of active users K
Output: active user set $\hat{\mathcal{I}}$
Initialize: $\mathcal{I}_0 :=$ empty set, $\boldsymbol{v}_0 := \boldsymbol{y}$, $\mathcal{X}_0 := \{1, \ldots, N_\tau\}$
for $k = 0$ to $K-1$ **do**
  Compute: $\boldsymbol{f} := \boldsymbol{X}^H \boldsymbol{v}_k$
  Find $i = \arg\max_{n \in \mathcal{X}_k} |f_n|$
  Detect active users: $\mathcal{I}_{k+1} = \mathcal{I}_k \cup \{\lceil i/(\tau+1) \rceil\}$
  Update: $\mathcal{X}_{k+1} = \mathcal{X}_k \backslash \{\lfloor i/(\tau+1) \rfloor (\tau+1) + 1, \cdots, \lceil i/(\tau+1) \rceil (\tau+1)\}$
  Update residual: $\boldsymbol{v}_{k+1} = \boldsymbol{y} - \boldsymbol{X}_{\mathcal{I}_{k+1}} \boldsymbol{X}^{\dagger}_{\mathcal{I}_{k+1}} \boldsymbol{y}$.
end for
$\hat{\mathcal{I}} = \mathcal{I}_K$

The complexity of the coherent detector is lower than that of the noncoherent detector, since no orthogonalization is necessary to update the residual. In both detectors, it is possible to terminate the algorithm early and obtain partial recovery of active users[2].

B. Performance Guarantees

The performance guarantee for the two algorithms are expressed in terms of two fundamental metrics of coherence of $\boldsymbol{X}$. The first is the worse-case coherence:

$$\mu(\boldsymbol{X}) \triangleq \max_{n \neq m} |\boldsymbol{x}_n^H \boldsymbol{x}_m|, \tag{20}$$

which is widely used in characterizing the performance of sparse recovery algorithms. The second is the average coherence, defined as

$$\nu(\boldsymbol{X}) \triangleq \frac{1}{N_\tau - 1} \max_n \left| \sum_{m \neq n} \boldsymbol{x}_n^H \boldsymbol{x}_m \right|, \tag{21}$$

where $\mathbf{1}$ is an all-one vector.

We say that a matrix $\boldsymbol{X}$ satisfies the *coherence property* if the following two conditions hold:

$$\mu(\boldsymbol{X}) \leq \frac{0.1}{\sqrt{2 \log N_\tau}}, \qquad \nu(\boldsymbol{X}) \leq \frac{\mu(\boldsymbol{X})}{\sqrt{M}}. \tag{22}$$

[2]The performance guarantees can be easily generalized to partial recovery.

In addition, we say that that a matrix $\boldsymbol{X}$ satisfies the *strong coherence property* if the following two conditions hold:

$$\mu(\boldsymbol{X}) \le \frac{1}{240 \log N_\tau}, \qquad \nu(\boldsymbol{X}) \le \frac{\mu(\boldsymbol{X})}{\sqrt{M}}. \tag{23}$$

Note that the condition on average coherence $\nu(\boldsymbol{X}) \le \mu(\boldsymbol{X})/\sqrt{M}$ can be achieved with essentially no cost via "wiggling", i.e. flipping the signs (or phases) of the columns of $\boldsymbol{X}$, which doesn't change the worst-case coherence $\mu(\boldsymbol{X})$ and the spectral norm $\|\boldsymbol{X}\|_2$ [19]. For simplicity we shall write $\mu = \mu(\boldsymbol{X})$ and $\nu = \nu(\boldsymbol{X})$.

We sort the amplitude of the entries of an K-sparse vector $\boldsymbol{r}$, $|r_n|$ from the largest to the smallest for the active users and denote as $|r|_{(1)}, \ldots, |r|_{(K)}$. Let

$$|r|_{\min} = |r|_{(K)}.$$

We define the nth Signal-to-Noise Ratio (SNR_n) and the nth Largest-to-Average Ratio (LAR_n) as

$$\mathsf{SNR}_n = \frac{|r|^2_{(n)}}{\mathbb{E}\{\|\boldsymbol{w}\|_2^2\}/K}, \quad \mathsf{LAR}_n = \frac{|r|^2_{(n)}}{\|\boldsymbol{r}\|_2^2/K}, \quad n = 1, \ldots, K.$$

The Signal-to-Noise Ratio (SNR) and minimum Signal-to-Noise Ratio ($\mathsf{SNR}_{\min}$) are defined respectively as

$$\mathsf{SNR} = \frac{\|\boldsymbol{r}\|^2}{\mathbb{E}\|\boldsymbol{w}\|_2^2}, \quad \mathsf{SNR}_{\min} = \frac{|r|^2_{\min}}{\mathbb{E}\|\boldsymbol{w}\|_2^2/K}.$$

We then have the following performance guarantee for the coherent matching pursuit detector.

Theorem 1. *Suppose that $N_\tau = N(\tau+1) \ge 128$, that the noise $\boldsymbol{w}$ is distributed as $\mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M)$, and that $\boldsymbol{X}$ satisfies the coherence property. If the number of active users satisfies*

$$K \le \min\left\{\frac{M}{2\log N_\tau}, \frac{1}{c^2\mu^2 \log N_\tau}\right\} \tag{24}$$

for $c = 20\sqrt{2}$, and if the power profile of active users satisfies

$$\mathit{LAR}_{(k)} > \frac{8}{(1 - c\mu\sqrt{(K-k+1)\log N_\tau})^2} \cdot \left(\frac{K \log N_\tau}{M\,\mathit{SNR}}\right), \tag{25}$$

for $1 \le k \le K$, then Algorithm 1 satisfies

$$\Pr\{\hat{\boldsymbol{b}} \ne \boldsymbol{b}\} \le (4 + \pi^{-1}) N_\tau^{-1}.$$

Since $\mathsf{LAR}_{(k)} \geq \mathsf{LAR}_{(K)}$ for $1 \leq k \leq K$, (25) can be satisfied if

$$\mathsf{LAR}_{(K)} > \frac{8}{(1 - c\mu\sqrt{\log N_\tau})^2} \cdot \left(\frac{K \log N_\tau}{M \mathsf{SNR}} \right). \tag{26}$$

Let $\theta = c\mu\sqrt{K \log N_\tau} \in (0, 1)$, then (26) implies that the number of active users is bounded by

$$K < \frac{M(1-\theta)^2 \mathsf{SNR}_{\min}}{8 \log N_\tau}.$$

Combining this with (24), we have the following corollary.

Corollary 2. *Suppose that $N_\tau = N(\tau + 1) \geq 128$, that the noise $\boldsymbol{w}$ is distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I}_M)$, and that $\boldsymbol{X}$ satisfies the coherence property. We write $\mu = c_1 M^{-1/\gamma}$ for some $c_1 > 0$ (c_1 may depend on N_τ and $\gamma \in \{0\} \cup [2, \infty)$). Then Algorithm 1 satisfies $\Pr\{\hat{\boldsymbol{b}} \neq \boldsymbol{b}\} \leq (4 + \pi^{-1}) N_\tau^{-1}$ as long as the number of active users K satisfies*

$$K < \max_{0<\theta<1} \min \left\{ \frac{M}{2 \log N_\tau}, \frac{M(1-\theta)^2 \mathsf{SNR}_{\min}}{8 \log N_\tau}, \frac{\theta^2 M^{2/\gamma}}{c_2^2 \log N_\tau} \right\}, \tag{27}$$

where $c_2 = 20\sqrt{2} c_1$.

We have the following performance guarantee for the noncoherent matching pursuit detector.

Theorem 3. *Suppose that $N_\tau = N(\tau + 1) \geq 128$, that the noise $\boldsymbol{w}$ is distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I}_M)$, and that $\boldsymbol{X}$ satisfies the coherence property. If the number of active users satisfies*

$$K \leq \min \left\{ \frac{N_\tau}{c_4^2 \|\boldsymbol{X}\|_2^2 \log N_\tau}, \frac{1}{c_3^2 \mu^2 \log N_\tau} \right\} \tag{28}$$

for $c_3 = 50\sqrt{2}$ and $c_4 = 104\sqrt{2}$, and if the power profile of active users satisfies

$$\mathit{LAR}_{(k)} > \frac{8}{(1 - c_3\mu\sqrt{(K-k+1)\log N_\tau})^2} \cdot \left(\frac{K \log N_\tau}{M \mathit{SNR}} \right), \tag{29}$$

for $1 \leq k \leq K$, then Algorithm 2 satisfies

$$\Pr\{\hat{\mathcal{I}} \neq \mathcal{I}\} \leq (K\pi^{-1} + 6) N_\tau^{-1}.$$

Similarly, using the fact that $\mathsf{LAR}_{(k)} \geq \mathsf{LAR}_{(K)}$ for $1 \leq k \leq K$, (29) can be satisfied if

$$\mathsf{LAR}_{(K)} > \frac{8}{(1 - c_3\mu\sqrt{\log N_\tau})^2} \cdot \left(\frac{K \log N_\tau}{M \mathsf{SNR}} \right), \tag{30}$$

and the following corollary becomes straightforward.

Corollary 4. *Suppose that $N_\tau = N(\tau+1) \geq 128$, that the noise $\boldsymbol{w}$ is distributed as $\mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M)$, and that $\boldsymbol{X}$ satisfies the coherence property. We write $\mu = c_1 M^{-1/\gamma}$ for some $c_1 > 0$ (c_1 may depend on N_τ and $\gamma \in \{0\} \cup [2, \infty)$). Then Algorithm 2 satisfies $\Pr\{\hat{\mathcal{I}} \neq \mathcal{I}\} \leq (K\pi^{-1} + 6)N_\tau^{-1}$ as long as the number of active users K satisfies*

$$K < \max_{0<\theta<1} \min \Big\{ \frac{M(1-\theta)^2 \mathsf{SNR}_{\min}}{8 \log N_\tau}, \frac{\theta^2 M^{2/\gamma}}{c_3^2 \log N_\tau}, \frac{N_\tau}{c_4^2 \|\boldsymbol{X}\|_2^2 \log N_\tau} \Big\}, \tag{31}$$

where $c_3 = 50\sqrt{2}c_1$ and $c_4 = 104\sqrt{2}$.

Theorem 1 and Theorem 3 implies that with both coherent and noncoherent detectors, the system can support $K \sim \mathcal{O}(M/\log N_\tau)$ users with M samples. In other words, the system can support K users with $M \sim \mathcal{O}(K \log N_\tau)$ samples. Since both detectors are based on iterative thresholding, the power profile of different active users, defined in (4), enters the analysis only through the quantities $\mathsf{LAR}_{(k)}$, and plays a less important role than $\mathsf{SNR}_{\min}$, the SNR of the weakest active user in determining the performance. Performance of the two algorithms is identical in terms of scaling.

V. Proofs of Main Theorems

Central to the proof is the notion of (K, ϵ, δ)-Statistical Orthogonality Condition (StOC) introduced in [12], which can be related to the worst-case and average coherence of matrix $\boldsymbol{X}$. We prove that the probability of error is vanishingly small if with high probability $\boldsymbol{X}$ satisfies the StOC and the noise $\boldsymbol{w}$ is uniformly bounded.

A. Preparations

We first introduce an alternative way to represent the measurement model. We can write the vector of transmitted symbols together with the power $\boldsymbol{Rb}$ as a concatenation of a random permutation matrix and a deterministic K-sparse vector $\bar{\boldsymbol{z}} \in \mathbb{C}^{N_\tau}$. The form of the K-sparse vector is given by $\bar{\boldsymbol{z}} \triangleq [z_1, \cdots, z_k, 0, \cdots, 0]^\top$. Let $\bar{\Pi} \triangleq (\pi_1, \cdots, \pi_{N_\tau})$ be a random permutation of $[\![N_\tau]\!]$. Let $\boldsymbol{P}_{\bar{\Pi}}$ be a $N_\tau \times N_\tau$ permutation matrix, and $\boldsymbol{P}_{\bar{\Pi}} \triangleq [\boldsymbol{e}_{\pi_1}, \cdots, \boldsymbol{e}_{\pi_{N_\tau}}]^\top$, with $\boldsymbol{e}_n$ being the nth column of the identity matrix $\boldsymbol{I}_{N_\tau}$. Given this notation, the assumption that $\mathcal{I}$ is a random subset of $[\![N_\tau]\!]$ is equivalent to stating that $\bar{\boldsymbol{z}} = \boldsymbol{P}_{\bar{\Pi}} \boldsymbol{Rb}$. Hence the measurement equation (17) can be written as

$$\boldsymbol{y} = \boldsymbol{XRb} + \boldsymbol{w} = \boldsymbol{X}\boldsymbol{P}_{\bar{\Pi}} \bar{\boldsymbol{z}} + \boldsymbol{w} = \boldsymbol{X}_\Pi \boldsymbol{r}_\mathcal{I} + \boldsymbol{w}, \tag{32}$$

where $\Pi \triangleq (\pi_1, \cdots, \pi_k)$ denotes the first k elements of the random permutation $\bar{\Pi}$, and $\boldsymbol{X}_\Pi$ denotes the $M \times K$ sub-matrix obtained by collecting the columns of $\boldsymbol{X}$ corresponding to the indices in Π, and the vector $\boldsymbol{r}_\mathcal{I} \in \mathbb{C}^K$ represents the K nonzero entries of $\boldsymbol{Rb}$. We next define the (K, ϵ, δ)-*Statistical Orthogonality Condition (StOC)*.

Definition 1 (StOC)**.** *Let $\bar{\Pi}$ be a random permutation of $[\![N_\tau]\!]$. Define $\Pi \triangleq (\pi_1, \cdots, \pi_K)$ and $\Pi_c \triangleq (\pi_{K+1}, \cdots, \pi_{N_\tau})$ for any $K \in [1, N_\tau]$. Then, the $M \times N_\tau$ (normalized) matrix $\boldsymbol{X}$ is said to satisfy the (K, ϵ, δ)-statistical orthogonality condition if there exists $\epsilon, \delta \in [0, 1)$ such that the inequalities:*

$$\|(\boldsymbol{X}_\Pi^H \boldsymbol{X}_\Pi - \boldsymbol{I}_K)\boldsymbol{z}\|_\infty \le \epsilon \|\boldsymbol{z}\|_2 \qquad \textit{(StOC-1)} \tag{33}$$

$$\|\boldsymbol{X}_{\Pi_c}^H \boldsymbol{X}_\Pi \boldsymbol{z}\|_\infty \le \epsilon \|\boldsymbol{z}\|_2 \qquad \textit{(StOC-2)} \tag{34}$$

hold for every fixed $\boldsymbol{z} \in \mathbb{C}^K$ with probability exceeding $1 - \delta$ with respect to the random permutation $\bar{\Pi}$.

The StoC property has proved useful in obtaining average case performance guarantees [12], [20]. It is similar in spirit to the Restricted Isometry Property (RIP) [6] which provides worst case guarantees in CS. An important difference between the two is that while we know of no effective algorithm for testing RIP, it is possible to infer StOC from matrix invariants that can be easily computed.

If (33) and (34) hold for a realization of permutation $\bar{\Pi}$, then for $1 \le k < K$, let $\Pi_k = (\pi_1, \ldots, \pi_k)$ and $\Pi_t^c = (\pi_{k+1}, \ldots, \pi_K)$, so that $\Pi_t \cup \Pi_t^c = \Pi$ and $\Pi_t \cap \Pi_t^c = \varnothing$. For every $\boldsymbol{z} \in \mathbb{C}^k$, we have from (33) that

$$\left\| \begin{bmatrix} \boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k} - \boldsymbol{I}_k & \boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k^c} \\ \boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k} & \boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c} - \boldsymbol{I}_{K-k} \end{bmatrix} \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{0} \end{bmatrix} \right\|_\infty \le \epsilon \|\boldsymbol{z}\|_2.$$

Therefore $\|(\boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k} - \boldsymbol{I}_k)\boldsymbol{z}\|_\infty \le \epsilon \|\boldsymbol{z}\|_2$, and $\|\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k} \boldsymbol{z}\|_\infty \le \epsilon \|\boldsymbol{z}\|_2$. Moreover, from (34) we have

$$\|\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k} \boldsymbol{z}\|_\infty = \left\| \begin{bmatrix} \boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k} & \boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c} \end{bmatrix} \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{0} \end{bmatrix} \right\|_\infty \le \epsilon \|\boldsymbol{z}\|_2.$$

We also need the following two lemmas.

Lemma 1. *An $M \times N_\tau$ matrix $\boldsymbol{X}$ satisfies (K, ϵ, δ)-StOC for any $\epsilon \in [0, 1)$ and $a \ge 1$ with*

$$\delta \le 4N_\tau \exp\left(-\frac{(\epsilon - \sqrt{k}\nu)^2}{16(2 + a^{-1})^2 \mu^2} \right), \tag{35}$$

as long as $K \le \min\{\epsilon^2 \nu^{-2}, (1 + a)^{-1} N_\tau\}$.

The proof for this lemma can be found in [12]. A consequence of this lemma is that if we let $K \le$

$M/(2\log N_\tau)$ and fix $\epsilon = 10\mu\sqrt{2\log N_\tau}$, then the matrix $\boldsymbol{X}$ satisfies (K,ϵ,δ)-StOC with $\delta \le 4N_\tau^{-1}$. Define the event $\mathcal{G}_1$ as:

$$\mathcal{G}_1 \triangleq \{\boldsymbol{X} \text{ satisfies StOC-1 and StOC-2}\}. \tag{36}$$

Then $\mathcal{G}_1$ occurs with probability at least $1-4N_\tau^{-1}$ with respect to $\bar{\Pi}$ given the aforementioned choice of parameters.

In order to prove Theorem 3, we need an argument due to Tropp [21] that shows a random submatrix of $\boldsymbol{X}$ is well-conditioned with high probability. We follow the treatment given by Candès and Plan [22] where this argument appears in a slightly different form, given below.

Lemma 2 ([21], [22])**.** *Let $\bar{\Pi} = (\pi_1,\ldots,\pi_{N_\tau})$ be a random permutation of $[\![N_\tau]\!]$, and define $\Pi = (\pi_1,\ldots,\pi_K)$ for any $K \in [1,N_\tau]$. Then for $q = 2\log N_\tau$ and $K \le N_\tau/(4\|\boldsymbol{X}\|_2^2)$, we have*

$$\begin{aligned}&\left(\mathbb{E}\left[\|\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi - \boldsymbol{I}_K\|_2^q\right]\right)^{1/q}\\ &\le 2^{1/q}\left(30\mu\log N_\tau + 13\sqrt{\frac{2K\|\boldsymbol{X}\|_2^2\log N_\tau}{N_\tau}}\right).\end{aligned} \tag{37}$$

with respect to the random permutation $\bar{\Pi}$.

The following lemma [22] states a probabilistic bound on the extreme singular values of a random submatrix of $\boldsymbol{X}$, by applying the Markov inequality to Lemma 2:

$$\Pr\left(\|\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi - \boldsymbol{I}_K\|_2 \ge 1/2\right) \le 2^q\mathbb{E}\left[\|\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi - \boldsymbol{I}_K\|_2^q\right]$$

Lemma 3 ([22])**.** *Let $\bar{\Pi} = (\pi_1,\ldots,\pi_{N_\tau})$ be a random permutation of $[\![N_\tau]\!]$, and define $\Pi = (\pi_1,\ldots,\pi_K)$ for any $K \le N_\tau$. Suppose that $\mu \le 1/(240\log N_\tau)$ and $K \le N_\tau/(c_2^2\|\boldsymbol{X}\|_2^2\log N_\tau)$ for numerical constant $c_2 = 104\sqrt{2}$, then we have*

$$\Pr\left(\|\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi - \boldsymbol{I}_K\|_2 \ge \frac{1}{2}\right) \le 2p^{-2\log 2}.$$

Define the event

$$\mathcal{G}_2 \triangleq \{\|\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi - \boldsymbol{I}_K\|_2 \le 1/2\},$$

which happens at least $1 - 2N_\tau^{-2\log 2} > 1 - 2N_\tau^{-1}$ with respect to $\bar{\Pi}$ from Lemma 3. Notice that all the eigenvalues of $\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi$ are bounded in $[1/2,3/2]$. Under $\mathcal{G}_2$, we have $\|(\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi)^{-1}\|_2 \le 2$ and $\|\boldsymbol{X}_\Pi(\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi)^{-1}\|_2 \le \sqrt{2}$. Moreover, for $1 \le k < K$ and $\Pi_k = (\pi_1,\ldots,\pi_k)$, we have $\|\boldsymbol{X}_{\Pi_k}^H\boldsymbol{X}_{\Pi_k} - \boldsymbol{I}_k\|_2 \le 1/2$, since eigenvalues of $\boldsymbol{X}_{\Pi_k}^H\boldsymbol{X}_{\Pi_k}$ are majorized by eigenvalues of $\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi$ [23].

Finally, we need that the noise is bounded with high probability.

Lemma 4. *Let $\boldsymbol{P} \in \mathbb{C}^{M\times M}$ be a projection matrix such that $\boldsymbol{P}^2 = \boldsymbol{P}$. Let $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, and $\boldsymbol{X}$ be a unit-column matrix. Then for $\tau > 0$ we have*

$$\Pr(\|\boldsymbol{X}^H\boldsymbol{P}\boldsymbol{w}\|_\infty \le \tau) \ge 1 - \frac{N_\tau}{\pi} e^{-\tau^2/\sigma^2},$$

provided the right hand side is greater than zero.

Proof: See Appendix VIII-B. ■

Now let $\tau = \sigma\sqrt{2\log N_\tau}$, and define

$$\mathcal{H}_0 = \{\|\boldsymbol{X}^H\boldsymbol{P}\boldsymbol{w}\|_\infty \le \tau\}.$$

It follows from Lemma 4 that $\mathcal{H}_0$ occurs with probability at least $1 - \pi^{-1}N_\tau^{-1}$.

B. Proof of Theorem 1

When applied to Algorithm 1, the next lemma shows that under appropriate conditions, ranking the inner products between $\boldsymbol{x}_n$ and $\boldsymbol{y}$ is an effective method of detecting the set of active users.

Lemma 5. *Let $\boldsymbol{b}$ be a vector with support $\mathcal{I}$ corresponding to K active users, and let $\boldsymbol{y}$ be a noisy measurement as in* (17)*. Suppose that*

$$|r|_{(1)} - 2\epsilon\|\boldsymbol{r}_{\mathcal{I}}\|_2 > 2\tau. \tag{38}$$

Then, if the event $\mathcal{G}_1 \cap \mathcal{H}_0$ occurs, we have

$$\max_{n\in\mathcal{I}} |\boldsymbol{x}_n^H\boldsymbol{y}| > \max_{n\notin\mathcal{I}} |\boldsymbol{x}_n^H\boldsymbol{y}|. \tag{39}$$

and $sgn(\Re[r_{n_1}^*\boldsymbol{x}_{n_1}^H\boldsymbol{y}]) = \sqrt{2}\Re[b_{n_1}]$, $sgn(\Im[r_{n_1}^*\boldsymbol{x}_{n_1}^H\boldsymbol{y}]) = \sqrt{2}\Im[b_{n_1}]$, for

$$n_1 = \arg\max_n |\boldsymbol{x}_n^H\boldsymbol{y}|. \tag{40}$$

Proof: See Appendix VIII-C. ■

We now prove the performance guarantee for the coherent detector in Algorithm 1. First we show that under the event $\mathcal{G}_1 \cap \mathcal{H}_0$, which happens with probability at least $1 - (\pi^{-1} + 4)N_\tau^{-1}$, Algorithm 1 correctly detects all active users and symbols in the first K iterations. Define a subset Π_k which contains the k variables that are selected until the kth iteration, $0 \le k \le K$.

We want to prove $\Pi_k \subset \Pi$ by induction. First at $k = 0$, $\Pi_0 = \varnothing \subset \Pi$. Suppose we are currently at the kth iteration of Algorithm 1, $0 \le k \le K-1$, and assume that $\Pi_k \subset \Pi$. The kth step is to detect the user with the largest $|\boldsymbol{x}_n^H \boldsymbol{v}_k|$. We have

$$\boldsymbol{v}_k = \boldsymbol{X}(\boldsymbol{b} - \boldsymbol{b}_k) + \boldsymbol{w} = \boldsymbol{X}\boldsymbol{\eta}_k + \boldsymbol{w}, \tag{41}$$

where $\boldsymbol{\eta}_k \triangleq \boldsymbol{b} - \boldsymbol{b}_k$. This vector has support $\Pi_k^c = \Pi \backslash \Pi_k$ and has at most $(K-k)$ non-zero elements, since $\boldsymbol{b}_{k-1}$ contains correct symbols at the correct locations for k active users, i.e. $[\boldsymbol{b}_k]_n = [\boldsymbol{b}]_n$, for $n \in \Pi_k$. This $\boldsymbol{v}_k$ is a noisy measurement of the vector $\boldsymbol{X}\boldsymbol{\eta}_k$. The signal model in (41) for the kth iteration is identical to the signal model in the first iteration with $\boldsymbol{b}$ replaced by $\boldsymbol{\eta}_k$ (with a smaller sparsity $K-k$ rather than K), Π replaced by Π_k^c, and $\boldsymbol{y}$ replaced by $\boldsymbol{v}_k$. Hence, from Lemma 5 we have that under the condition

$$\|\boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty - 2\epsilon \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2 > 2\tau, \tag{42}$$

we have

$$\max_{n \in \mathcal{I} \backslash \mathcal{I}_{(k-1)}} |\boldsymbol{x}_n^H \boldsymbol{v}_k| > \max_{n \notin \mathcal{I} \backslash \mathcal{I}_{(k-1)}} |\boldsymbol{x}_n^H \boldsymbol{v}_k|. \tag{43}$$

i.e. Algorithm 1 can detect an active user correctly, and no index of an active user that has been detected before will be chosen again. Note that $\|\boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty \ge |r|_{(k+1)}$, $\|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2 \le \sqrt{K-k}|r|_{(k+1)}$, (42) is satisfied by

$$|r|_{(k+1)} > 2\epsilon\sqrt{K-k}|r|_{(k+1)} + 2\tau.$$

Since $K < 1/(c^2\mu^2 \log N_\tau)$ and $\epsilon = 10\mu\sqrt{2\log N_\tau}$, this is equivalent to the condition in (25) for $0 \le k \le K-1$, therefore a correct user is selected at the kth iteration, so that $\Pi_{k+1} \subset \Pi$. On the other hand, since condition (38) is true, the symbol can be detected correctly as well. Then we have that under the event $\mathcal{G}_1 \bigcap \mathcal{H}_0$, $\mathrm{sgn}(\Re[r_{n_1}^* \boldsymbol{x}_{n_1}^H \boldsymbol{y}]) = \sqrt{2}\Re[b_{n_1}]$, $\mathrm{sgn}(\Im[r_{n_1}^* \boldsymbol{x}_{n_1}^H \boldsymbol{y}]) = \sqrt{2}\Im[b_{n_1}]$, that is $\mathcal{G}_1 \bigcap \mathcal{H}_0 \subset \{b_{n_k}^{(k)} = b_{n_k}\}$. By induction, since no active users will be detected twice, it follows that the first K steps of Algorithm 1 can detect all active users.

C. Proof of Theorem 3

We note that in Algorithm 2, the residual $\boldsymbol{v}_k$, $k = 0, \cdots, K-1$ is orthogonal to the selected columns in previous iterations, so in each iteration a new column will be selected. Define a subset Π_k which contains the k variables that are selected until the kth iteration. Then $\boldsymbol{P}_k = \boldsymbol{X}_{\Pi_k}(\boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k})^{-1}\boldsymbol{X}_{\Pi_k}^H$ is the projection matrix onto the linear subspace spanned by the columns of $\boldsymbol{X}_{\Pi_k}$, and we assume $\boldsymbol{P}_0 = \boldsymbol{0}$.

Again we want to prove $\Pi_k \subset \Pi$ by induction. First at $k = 0$, $\Pi_0 = \varnothing \subset \Pi$. Assume at the kth iteration, $\Pi_k \subset \Pi$, $0 \le k \le K-1$, then the residual $\boldsymbol{v}_k$ can be written as

$$\begin{aligned}\boldsymbol{v}_k &= (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{y} \\ &= (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{X}_\Pi \boldsymbol{r}_\mathcal{I} + (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{w} \triangleq \boldsymbol{s}_k + \boldsymbol{n}_k,\end{aligned}$$

where $\boldsymbol{s}_k = (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{X}_\Pi \boldsymbol{r}_\mathcal{I}$ and $\boldsymbol{n}_k = (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{w}$ are the signal and noise components respectively at the kth iteration.

Let $M_\Pi^k = \|\boldsymbol{X}_\Pi^H \boldsymbol{s}_k\|_\infty$, $M_{\Pi^c}^k = \|\boldsymbol{X}_{\Pi^c}^H \boldsymbol{s}_k\|_\infty$ and $N^k = \|\boldsymbol{X}^H \boldsymbol{n}_k\|_\infty$, then a sufficient condition for $\Pi_{k+1} \subset \Pi$, i.e. for Algorithm 2 to select a correct active user at the next iteration is that

$$M_\Pi^k - M_{\Pi^c}^k > 2N^k \tag{44}$$

since under (44) we have

$$\|\boldsymbol{X}_\Pi^H \boldsymbol{v}_k\|_\infty \ge M_\Pi^k - N^K > M_{\Pi^c}^k + N^K \ge \|\boldsymbol{X}_{\Pi^c}^H \boldsymbol{v}_k\|_\infty.$$

Let the event

$$\mathcal{G} = \mathcal{G}_1 \cap \mathcal{G}_2.$$

From Lemma 1 and Lemma 3 the event $\mathcal{G}$ holds with probability at least $1 - 4N_\tau^{-1} - 2N_\tau^{-2\log 2}$ with respect to $\bar{\Pi}$.

Now we bound M_Π^k and $M_{\Pi^c}^k$ under the event $\mathcal{G}$. Let $\Pi_k^c = \Pi \backslash \Pi_k$ be the index set of yet to be selected active users, and $\boldsymbol{r}_{\Pi_k^c} = \boldsymbol{r}_{\mathcal{I}_k^c}$ be the corresponding coefficients. We can find a vector $\boldsymbol{z}$ of dimension $(K-k)$ such that $\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c} \boldsymbol{z} = \boldsymbol{X}_{\Pi_k^c}^H (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{X}_\Pi \boldsymbol{r}_\mathcal{I}$, where the vector $\boldsymbol{z}$ can be written as

$$\begin{aligned}\boldsymbol{z} &= (\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c})^{-1} \boldsymbol{X}_{\Pi_k^c}^H (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c} \\ &= \boldsymbol{r}_{\mathcal{I}_k^c} - (\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c})^{-1} \boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{P}_k \boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c}.\end{aligned}$$

Since we have

$$\|\boldsymbol{z}\|_2 \le \|(\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c})^{-1}\|_2 \|\boldsymbol{X}_{\Pi_k^c}^H (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c}\|_2 \tag{45}$$

$$\le 2\|\boldsymbol{X}_\Pi^H \boldsymbol{X}_\Pi\|_2 \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2 \le 3\|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2, \tag{46}$$

where (45) follows from

$$\|\boldsymbol{X}_{\Pi_k^c}^H (\boldsymbol{I} - \boldsymbol{P}_k)\boldsymbol{X}_{\Pi_k^c})\|_2 \le \|\boldsymbol{X}_\Pi^H \boldsymbol{X}_\Pi\|_2,$$

whose proof can be found in [24], and (46) follows from Lemma 3. Also,

$$\begin{aligned}
&\|\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{P}_k \boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty \\
&= \|\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k} (\boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k})^{-1} \boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty \\
&\le \epsilon \|(\boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k})^{-1} \boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c}\|_2 \\
&\le \epsilon \|(\boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k})^{-1}\|_2 \|\boldsymbol{X}_{\Pi_k}^H \boldsymbol{X}_{\Pi_k^c}\|_2 \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2 \\
&\le \epsilon \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2,
\end{aligned}$$

therefore M_Π^k can be bounded as

$$\begin{aligned}
M_\Pi^k &= \|\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c} - \boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{P}_k \boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty \\
&\ge \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty - \|(\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{X}_{\Pi_k^c} - \boldsymbol{I}) \boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty - \|\boldsymbol{X}_{\Pi_k^c}^H \boldsymbol{P}_k \boldsymbol{X}_{\Pi_k^c} \boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty \\
&\ge \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty - 2\epsilon \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2. \qquad (47)
\end{aligned}$$

where (47) follows from (33). Next, $M_{\Pi^c}^k$ can be bounded as

$$\begin{aligned}
M_{\Pi^c}^k &= \|\boldsymbol{X}_{\Pi^c}^H (\boldsymbol{I} - \boldsymbol{P}_k) \boldsymbol{X}_\Pi \boldsymbol{r}_\mathcal{I}\|_\infty \\
&= \|\boldsymbol{X}_{\Pi^c}^H \boldsymbol{X}_{\Pi_k^c} \boldsymbol{z}\|_\infty \\
&\le \epsilon \|\boldsymbol{z}\|_2 \le 3\epsilon \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2. \qquad (48)
\end{aligned}$$

where (48) follows from (34).

Conditioned on the event $\mathcal{G}$, for each $\boldsymbol{P}_k$, since $\boldsymbol{I} - \boldsymbol{P}_k$ is also a projection matrix, define the event

$$\mathcal{H}_k = \{N^k \le \tau\}, \quad k = 0, \cdots, K-1. \qquad (49)$$

Then from Lemma 4, $\mathcal{H}_k$ happens with probability at least $1 - \pi^{-1} N_\tau^{-1}$ with respect to $\boldsymbol{w}$. We further define the event $\mathcal{H} = \cap_{k=0}^{K-1} \mathcal{H}_k$, then from the union bound $\Pr(\mathcal{H}|\mathcal{G}) = \Pr(\mathcal{H}) \ge 1 - K\pi^{-1} N_\tau^{-1}$.

Under the event $\mathcal{G} \cap \mathcal{H}$, from the above discussions which happens with probability $\Pr(\mathcal{G} \cap \mathcal{H}) \ge 1 - K\pi^{-1} N_\tau^{-1} - 2N_\tau^{-2\log 2} - 4N_\tau^{-1} \ge 1 - (K\pi^{-1} + 6) N_\tau^{-1}$. Now we are ready to analyze the performance of Algorithm 2 under the event $\mathcal{G} \cap \mathcal{H}$. Substituting the bounds (47), (48) and (49) into (44), it is sufficient that at the kth iteration

$$\|\boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty > 5\epsilon \|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2 + 2\tau. \qquad (50)$$

Note that $\|\boldsymbol{r}_{\mathcal{I}_k^c}\|_\infty \geq |r|_{(k+1)}$, $\|\boldsymbol{r}_{\mathcal{I}_k^c}\|_2 \leq \sqrt{K-k}|r|_{(k+1)}$, (50) is satisfied by

$$|r|_{(k+1)} > 5\epsilon\sqrt{K-k}|r|_{(k+1)} + 2\tau.$$

Since $K < 1/(c_1^2\mu^2 \log N_\tau)$ and $\epsilon = 10\mu\sqrt{2\log N_\tau}$, this is equivalent to the condition in (29) for $0 \leq k \leq K-1$, therefore a correct user is selected at the kth iteration, so that $\Pi_{k+1} \subset \Pi$. Since the number of active users is K, Algorithm 2 successfully finds Π in K iterations under the event $\mathcal{G} \cap \mathcal{H}$, and we have proved Theorem 3.

VI. Deterministic Signature Waveforms

A. Gabor Frames

In the following we will construct the signature sequences $\tilde{\boldsymbol{a}}_n$ from Gabor frames. Let $\boldsymbol{g} \in \mathbb{C}^P$ be a seed vector with each entry $|g_n|^2 = 1/M$ and let $\boldsymbol{T}(\boldsymbol{g}) \in \mathbb{C}^{P\times P}$ be the circulant matrix generated from $\boldsymbol{g}$ as $\boldsymbol{T}(\boldsymbol{g}) = [\mathcal{T}_0\boldsymbol{g} \quad \cdots \quad \mathcal{T}_\tau\boldsymbol{g}]$. Its eigen-decomposition can be written as

$$\boldsymbol{T}(\boldsymbol{g}) = \boldsymbol{F}\mathrm{diag}(\boldsymbol{F}^H\boldsymbol{g})\boldsymbol{F}^H \triangleq \boldsymbol{F}\mathrm{diag}(\hat{\boldsymbol{g}})\boldsymbol{F}^H,$$

where $\boldsymbol{F} = \frac{1}{\sqrt{P}}[\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_{P-1}]$ is the DFT matrix with columns

$$\boldsymbol{\omega}_m = \left[e^{j2\pi\frac{m}{P}\cdot 0}, e^{j2\pi\frac{m}{P}\cdot 1}, \ldots, e^{j2\pi\frac{m}{P}\cdot(P-1)}\right]^\top.$$

Define corresponding diagonal matrices $\boldsymbol{W}_m = \mathrm{diag}[\boldsymbol{\omega}_m]$, for $m = 0, 1, \ldots, P-1$. Then the Gabor frame $\boldsymbol{\Phi} = [\phi_m]$ generated from $\boldsymbol{g}$ is an $P \times P^2$ block matrix of the form

$$\boldsymbol{\Phi} = [\boldsymbol{W}_0\boldsymbol{T}(\boldsymbol{g}),\ \boldsymbol{W}_1\boldsymbol{T}(\boldsymbol{g}),\ \ldots,\ \boldsymbol{W}_{P-1}\boldsymbol{T}(\boldsymbol{g})]. \tag{51}$$

where each column has norm $\sqrt{P/M}$. When we apply the DFT to the Gabor frame $\boldsymbol{\Phi}$, and obtain $\hat{\boldsymbol{\Phi}} = \boldsymbol{F}^H\boldsymbol{\Phi}$, the order of time-shift and frequency modulation is reversed, and therefore $\hat{\boldsymbol{\Phi}}$ is composed of circulant matrices after appropriate ordering of columns. In fact, if we index each column m from P^2 to $P \times P$ by $m = Pq + \ell$, the matrix $\boldsymbol{\Phi}_\ell$ is obtained by keeping all columns with $r = \ell \pmod P$. So $\boldsymbol{\Phi}_\ell$ can be written as

$$\boldsymbol{\Phi}_\ell = \sqrt{P}\cdot\mathrm{diag}(\boldsymbol{S}^\ell\boldsymbol{g})\boldsymbol{F},$$

where $\boldsymbol{S}$ is the right-shift matrix by one, and its DFT transform

$$\hat{\boldsymbol{\Phi}}_\ell = \boldsymbol{F}\boldsymbol{\Phi}_\ell = \sqrt{P}\boldsymbol{T}(\boldsymbol{W}_\ell\hat{\boldsymbol{g}})$$

is a circulant matrix. We use $[\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_{P-1}]$ as the matrix $\boldsymbol{A}$.

At the receiver, a random partial DFT is applied to the received symbol, so $\boldsymbol{H} = \boldsymbol{F}_\Omega$ is a partial DFT matrix, and the resulted matrix $\boldsymbol{X} = \boldsymbol{\Phi}_\Omega$ is a subsampled Gabor frame defined in (51), with unit-norm columns. The maximum discrete delay τ which this Gabor frame construction can support is $P-1$, where $\boldsymbol{W}_\ell \hat{\boldsymbol{g}}$ can be assigned as signature sequences to a user, so the maximum number of total users should satisfy $N \leq P$. In general, if $\tau < P-1$, we can split $\boldsymbol{\Phi}_\ell$ into blocks to support multiple users, and send $\mathcal{T}_{d(\tau+1)} \boldsymbol{W}_\ell \hat{\boldsymbol{g}}$ as signature sequences for $d = 0, \cdots, \lfloor P/(\tau+1) \rfloor$ and $\ell = 1, \cdots, P$, so the maximum number of total user satisfies $N \leq P\lfloor P/(\tau+1) \rfloor$ in general.

Now we consider the coherence properties of $\boldsymbol{X}$. We have the following proposition.

Proposition 1. *Let $\boldsymbol{X}$ be a unit-column matrix with M rows subsampled uniformly at random from a Gabor frame $\boldsymbol{\Phi}$ that satisfies the strong coherence property. If $M \geq \gamma \log^3 P$ for some constant γ, then with probability at least $1 - 2P^{-1}$, we have $\mu(\boldsymbol{X}) \leq \gamma_2 / \log P$ for some constant γ_2, and $\nu(\boldsymbol{X}) \leq 2/M$ deterministically.*

Proof: See Appendix VIII-D. ■

It is established in [12] that Gabor frames satisfy the strong coherence property when the seed sequence is the Alltop sequence, or with high probability when the seed sequence is randomly generated. Proposition 1 implies that we can find an M such that the subsampled Gabor frame satisfies the (strong) coherence property as long as M is not too small.

B. Kerdock Codes

The set of Kerdock codewords is given as columns of the matrix $\tilde{\boldsymbol{\Psi}} \in \{\pm 1, \pm j\}^{P \times P^2}$, where $P = 2^m$. Since the Kerdock code is a cyclic extended code over $\mathbb{Z}^4$, we can find a map of the columns of Kerdock code into P blocks (Theorem 10, [14]), such that the $P-1$ columns within each block are cyclic. Then we can assign adjacent codewords in the same cyclic block to one user, and set the first code as user's transmitted codeword. We denote the final code book as $\boldsymbol{\Psi} \in \{\pm 1, \pm j\}^{P \times (P^2 - P)}$, and we denote the discarded P columns by the set $\boldsymbol{\Psi}_c$.

The coherence property of the subsampled Kerdock code set is summarized in the Proposition below.

Proposition 2. *Let $\boldsymbol{X}$ be a unit-column matrix with M rows subsampled uniformly at random from a Kerdock code $\boldsymbol{\Psi}$. If $M \geq \gamma \log^3 P$ for some constant γ, then with probability at least $1 - 2P^{-1}$, we have $\mu(\boldsymbol{X}) \leq \gamma_2 / \log P$ for some constant γ_2, and $\nu(\boldsymbol{X}) \leq 2/M$ deterministically.*

Proof: See Appendix VIII-E. ■

The worst-case coherence of $\mathbf{\Psi}$ meets the Welch bound $\mu(\mathbf{\Psi}) = 1/\sqrt{P}$ and the average coherence of $\mathbf{\Psi}$ is $\nu(\mathbf{\Psi}) = 1/P$. Proposition 1 implies that we can find an M such that the subsampled Kerdock code set satisfies the (strong) coherence property as long as M is not too small.

VII. Numerical Examples

A. Gabor Signature Waveforms

We first consider when each circulant matrix in the Gabor frame supports only one user. This corresponds to the maximum delay the algorithm can work in the asynchronous case. Let the seed vector $\boldsymbol{g}$ for the Gabor frame be either an Alltop sequence of length $P = 127$, given as

$$\boldsymbol{g} = \frac{1}{\sqrt{P}}[e^{j2\pi\frac{1^3}{P}}, e^{j2\pi\frac{2^3}{P}}, \ldots, e^{j2\pi\frac{P^3}{P}}];$$

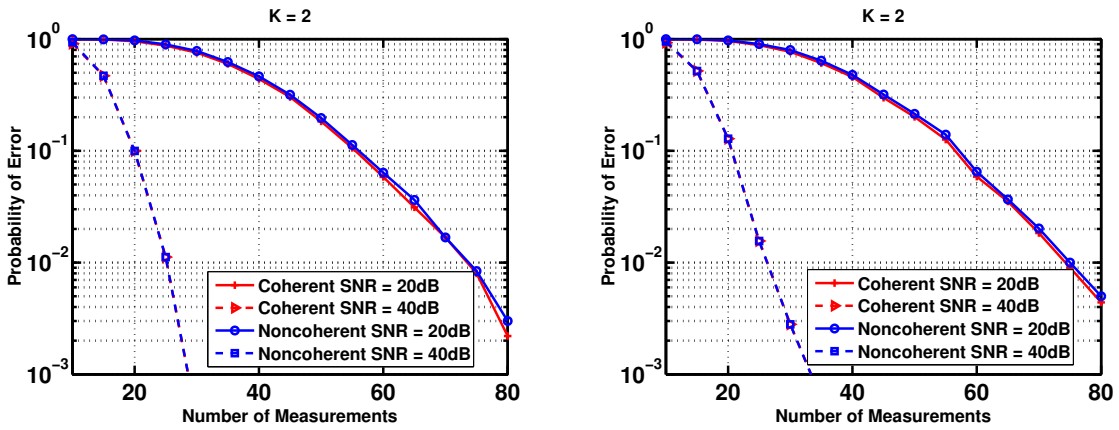
or a unit vector with random uniform phase of length $P = 128$, given as

$$\boldsymbol{g} = \frac{1}{\sqrt{P}}[e^{j2\pi\theta_1}, e^{j2\pi\theta_1}, \ldots, e^{j2\pi\theta_P}], \tag{52}$$

where θ_i is uniformly distributed on $[0, 1]$, $1 \le i \le P$. The power profile is assumed known as $r_n = 1$ for all $n = 1, \cdots, N$ in the coherent case, and are assume unknown in the noncoherent case.

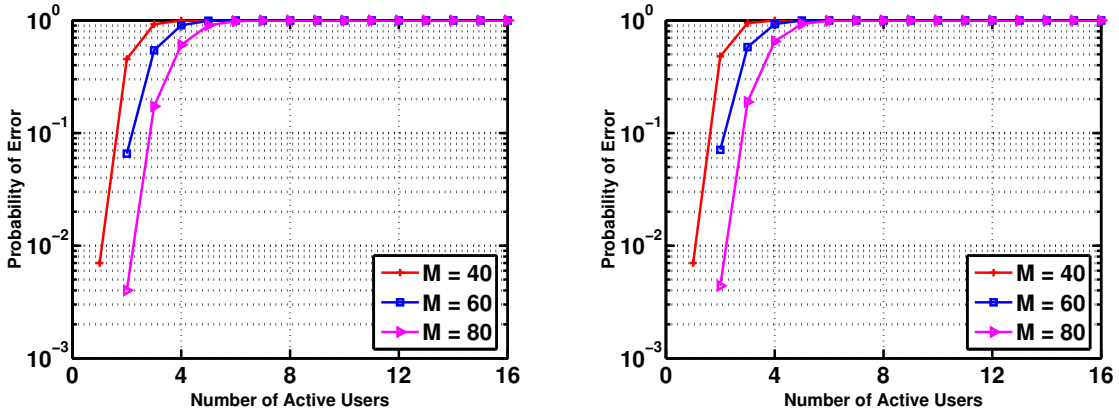
The active users are selected first by uniformly choosing a number at random from 1 to P, and then, for each active user, the delay is chosen uniformly at random. First, we fix the number of active users, namely $K = 2$, and apply the coherent detector described in Algorithm 1 and noncoherent detector described in Algorithm 2 for $\mathsf{SNR} = 20$dB and $\mathsf{SNR} = 40$dB. The partial DFT matrix is applied with randomly selected rows and the number of Monto Carlo runs is $5,000$. Fig. 3 shows the probability of error for multi-user detector with respect to the number of measurements. The performance of the Alltop Gabor frame is better than that of the random Gabor frame due to its optimal coherence. It is also worth noting that the performance of the noncoherent detector is almost the same as that of the coherent detector, albeit it does not perform symbol detection. This may suggest that channel state information and power control are less important in sparse recovery of active users.

Finally, we consider when the maximum delay is relatively small, for example $\tau = 15$ when $P = 128$ for a random Gabor frame. We transmit the first sequence within the block of the circulant matrix, resulting in a total number of $P^2/(\tau + 1) = 1024$ users, and Fig. 4 and Fig. 5 show the probability of error for multi-user detection with respect to the number of active users K for different number of random measurements $M = 40, 60, 80$ when $\mathsf{SNR} = 20$dB and $\mathsf{SNR} = 40$dB respectively.



(a) Alltop Gabor Frame (b) Random Gabor Frame

Fig. 3: Probability of error for multi-user detection with respect to the number of measurements from coherent and noncoherent detectors using (a) an Alltop Gabor frame with length $P = 127$, and (b) a random Gabor frame with length $P = 128$, for $K = 2$ active users and SNR = 20dB and 40dB, where the maximum chip delay is $\tau = 126$.

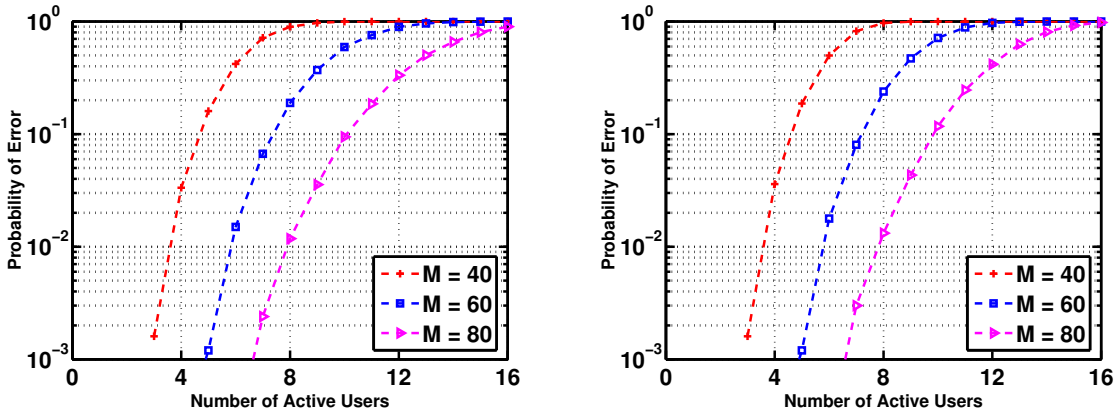


(a) Coherent detector (b) Noncoherent detector

Fig. 4: Probability of error for coherent and noncoherent multi-user detection with respect to the number of active users using a random Gabor frame with $P = 128$ for $M = 40, 60, 80$ when SNR = 20dB, where the maximum chip delay is $\tau = 15$. The total number of users is $N = 1024$.

B. Kerdock Signature Waveforms

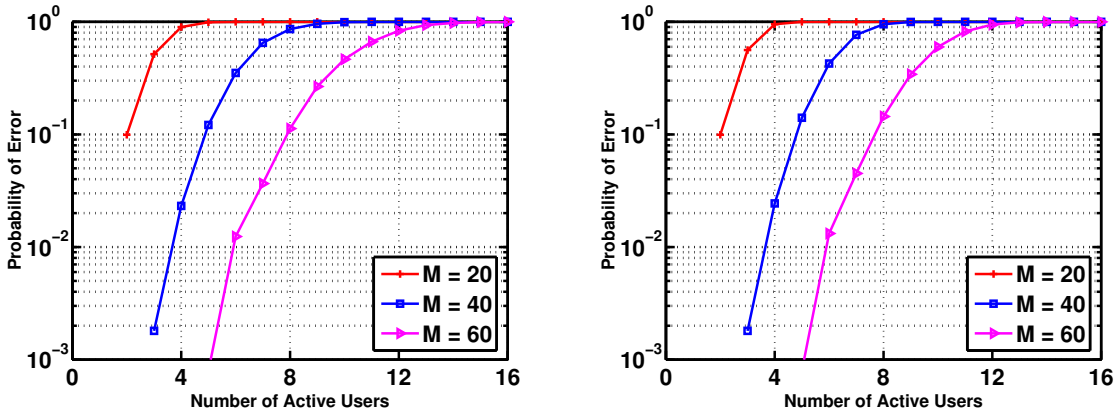
We first generate a Kerdock code set $\mathbf{\Psi}$ of length $P = 128$ with P^2 codewords. By removing the all-one row in $\mathbf{\Psi}$, and removing two column in each block of size P, we obtain a block-circulant matrix of size $(P-1) \times P(P-2)$, where there are P circulant blocks of size $(P-1) \times (P-2)$. As earlier, we assume the maximal delay is $\tau = 15$, the total number of users is given as $\lfloor (P-2)/(\tau+1) \rfloor \cdot P = 896$.



(a) Coherent detector (b) Noncoherent detector

Fig. 5: Probability of error for coherent and noncoherent multi-user detection with respect to the number of active users using a random Gabor frame with $P = 128$ for $M = 40, 60, 80$ when SNR = 40dB, where the maximum chip delay is $\tau = 15$. The total number of users is $N = 1024$.

Fig. 6 show the probability of error for multi-user detection with respect to the number of active users K for different number of random measurements $M = 20, 40, 60$ when SNR = 20dB.



(a) Coherent detector (b) Noncoherent detector

Fig. 6: Probability of error for coherent and noncoherent multi-user detection with respect to the number of active users K using a Kerdock code set with $P = 127$ for $M = 20, 40, 60$ when SNR = 20dB, where the maximum chip delay is $\tau = 15$. The total number of user is $N = 896$.

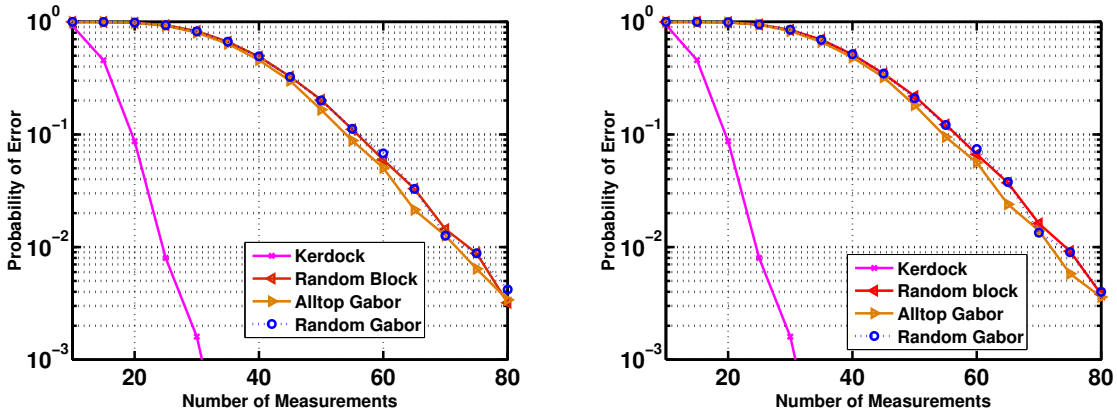
C. Comparison of Signature Waveforms

In this section we compare the performance of different signatures for multi-user detector when SNR = 20dB and $K = 2$. We use the above considered Kerdock code, Alltop Gabor frame and random Gabor

Signatures	# of total users
Kerdock	896
Random Block	1024
Alltop Gabor	889
Random Gabor	1024

TABLE II: Total number of users for different signatures.

frame when $P = 128$. We also consider the cyclic extensions of random matrix whose columns are generated from (52). Table II summarizes the total number of users for different signature waveforms, notice that both Kerdock and Alltop suffer from the floor operation in calculating the number of total users. As shown in Fig. 7, the performance of Kerdock code is significantly better than other choices. The performance of cyclic extensions of random matrices and Gabor frames are similar, since the subsampling degenerates the optimal coherence properties of the unsampled Gabor frame. The Alltop Gabor frame is slightly better than its random counterparts.



(a) Coherent detector (b) Noncoherent detector

Fig. 7: Comparison of performance with respect to the number of measurements for multi-user detection when $K = 2$ and $\mathsf{SNR} = 20\text{dB}$, where the maximum chip delay is $\tau = 15$.

VIII. Conclusions

This paper describes two MUD front-end architectures that lead to mathematically equivalent discrete signal models. Both coherent and noncoherent detectors based on iterative matching pursuit are presented to recover active users, and their transmitted symbols are also detected in the coherent case. It is shown that compressive demodulation requires $\mathcal{O}(K \log N_\tau)$ samples to recover K active users. Gabor frames

and Kerdock codes are proposed as signature waveforms and numerical examples are provided where the superior performance of Kerdock code is emphasized. The resilience of iterative matching pursuit to variability in relative strength of the entries of the signal might be an advantage in multi-user detection in wireless communications because it makes power control less critical. We make the final remark that the noncoherent detectors can be extended to detect transmitted symbols by assigning two different signature waveforms to the BPSK signaling.

APPENDIX

A. Sidak's lemma

Lemma 6 (Sidak's lemma)**.** *[25] Let* $[X_1, \cdots, X_n]$ *be a vector of random multivariate normal variables with zero means, arbitrary variances* σ_1^2*,* $\cdots$*,* σ_n^2 *and and an arbitrary correlation matrix. Then, for any positive numbers* $c_1, \cdots, c_n$*, we have*

$$\Pr(|X_1| \le c_1, \cdots, |X_n| \le c_n) \ge \prod_{i=1}^{n} \Pr(|X_i| \le c_i).$$

B. Proof of Lemma 4

Since $\boldsymbol{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I}_M)$, $\boldsymbol{X}^H \boldsymbol{P} \boldsymbol{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{X}^H \boldsymbol{P} \boldsymbol{X})$ but it is a colored Gaussian noise. We want to bound $\Pr(\|\boldsymbol{X}^H \boldsymbol{P} \boldsymbol{w}\|_\infty \ge \tau)$ for some $\tau > 0$. Note that each $\boldsymbol{x}_n^H \boldsymbol{P} \boldsymbol{w} \sim \mathcal{CN}(0, \sigma_n^2)$, where $\sigma_n^2 = \sigma^2 \boldsymbol{x}_n^H \boldsymbol{P} \boldsymbol{x}_n \le \sigma^2$ (recall that $\|\boldsymbol{x}_n\|_2 = 1$). Then

$$\Pr(|\boldsymbol{x}_n^H \boldsymbol{P} \boldsymbol{w}| \le \tau) = 1 - \frac{1}{\pi} e^{-\tau^2/\sigma_n^2} \ge 1 - \frac{1}{\pi} e^{-\tau^2/\sigma^2}.$$

Following Lemma 6, for $\tau > 0$ we have

$$\begin{aligned} \Pr(\|\boldsymbol{X}^H \boldsymbol{P} \boldsymbol{w}\|_\infty \le \tau) &\ge \prod_{n=1}^{N_\tau} \Pr(|\boldsymbol{x}_n^H \boldsymbol{P} \boldsymbol{w}| \le \tau) \\ &\ge (1 - \frac{1}{\pi} e^{-\tau^2/\sigma^2})^{N_\tau} \ge 1 - \frac{N_\tau}{\pi} e^{-\tau^2/\sigma^2}, \end{aligned}$$

provided the right hand side is greater than zero.

C. Proof of Lemma 5

We begin by deriving a lower-bound for $\max_{n\in\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{y}|$ when $\mathcal{G}_1\cap\mathcal{H}_0$ occurs. Assume that n_0 is the index achieving the largest absolute gain: $|r_{n_0}| = |r|_{(1)}$. Then under the event $\mathcal{G}_1\cap\mathcal{H}_0$:

$$\begin{aligned}\max_{n\in\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{y}| \geq |\boldsymbol{x}_{n_0}^H\boldsymbol{y}| &= \left|b_{n_0}r_{n_0} + \sum_{m\neq n_0} b_m r_m \boldsymbol{x}_{n_0}^H\boldsymbol{x}_m + \boldsymbol{x}_{n_0}^H\boldsymbol{w}\right| \\ &\geq |r|_{(1)} - \left|\sum_{m\neq n_0} b_m r_m \boldsymbol{x}_{n_0}^H\boldsymbol{x}_m\right| - |\boldsymbol{x}_{n_0}^H\boldsymbol{w}| \\ &= |r|_{(1)} - \|(\boldsymbol{X}_\Pi^H\boldsymbol{X}_\Pi - \boldsymbol{I})\boldsymbol{r}_\mathcal{I}\| - |\boldsymbol{x}_{n_0}^H\boldsymbol{w}| \\ &> |r|_{(1)}| - \epsilon\|\boldsymbol{r}_\mathcal{I}\|_2 - \tau. \end{aligned} \tag{53}$$

On the other hand, we can similarly expand and upper-bound $\max_{n\notin\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{y}|$, under the event $\mathcal{G}\bigcup\mathcal{G}_\mathcal{I}$, as

$$\begin{aligned}\max_{n\notin\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{y}| &= \max_{n\notin\mathcal{I}}\left|\sum_{m\in\mathcal{I}} b_m r_m \boldsymbol{x}_n^H\boldsymbol{x}_m + \boldsymbol{x}_n^H\boldsymbol{w}\right| \\ &\leq \max_{n\notin\mathcal{I}}\left|\sum_{m\in\mathcal{I}} b_m r_m \boldsymbol{x}_n^H\boldsymbol{x}_m\right| + \max_{n\notin\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{w}| \\ &= \|\boldsymbol{X}_{\Pi^c}^H\boldsymbol{X}_\Pi\boldsymbol{z}\|_\infty + \max_{n\notin\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{w}| \\ &< \epsilon\|\boldsymbol{r}_\mathcal{I}\|_2 + \tau. \end{aligned} \tag{54}$$

Combining (53) and (54), we have that under the event $\mathcal{G}_1\cap\mathcal{H}_0$,

$$\max_{n\in\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{y}| > |r|_{(1)} - 2\epsilon\|\boldsymbol{r}_\mathcal{I}\|_2 - 2\tau + \max_{n\notin\mathcal{I}}|\boldsymbol{x}_n^H\boldsymbol{y}|. \tag{55}$$

So when $\mathcal{G}$ occurs, under the condition (38), we obtain (39), as required.

Furthermore, to detect correctly, for $\Re[b_n] = 1/\sqrt{2}$, $\Re[r_n^H\boldsymbol{x}_n^H\boldsymbol{y}]$ has to be positive, and for $\Re[b_n] = -1/\sqrt{2}$, $\Re[r_n^*\boldsymbol{x}_n^H\boldsymbol{y}]$ has to be negative. Similarly we can detect $\Im[b_n]$. First assume $\Re[b_n] = 1/\sqrt{2}$, then

$$\begin{aligned}&\Re[r_{n_1}^*\boldsymbol{x}_{n_1}^H\boldsymbol{y}] \\ &= |r_{n_1}|^2 + \sum_{m\neq n_1}\Re[b_m]\Re\left[r_{n_1}^* r_m \boldsymbol{x}_{n_1}^H\boldsymbol{x}_m\right] + \Re\left[r_{n_1}^*\boldsymbol{x}_{n_1}^H\boldsymbol{w}\right]\end{aligned}$$

must be positive. Suppose this does not hold, and $\Re[r_{n_1}^*\boldsymbol{x}_{n_1}^H\boldsymbol{y}] < 0$. Recall that n_0 is the index of the largest gain: $|r|_{n_0} = |r|_{(1)}$. From (40), we have

$$|r_{n_1}^*\boldsymbol{x}_{n_1}^H\boldsymbol{y}| \geq |r_{n_1}^*\boldsymbol{x}_{n_0}^H\boldsymbol{y}|. \tag{56}$$

Since

$$\begin{aligned} |r_{n_1}^* \boldsymbol{x}_{n_1}^H \boldsymbol{y}| &= \left| |r_{n_1}|^2 + \sum_{m \neq n_1} b_m r_{n_1}^* r_m \boldsymbol{x}_{n_1}^H \boldsymbol{x}_m + r_{n_1}^* \boldsymbol{x}_{n_1}^H \boldsymbol{w} \right| \\ &\leq \left| \sum_{m \neq n_1} b_m r_{n_1}^* r_m \boldsymbol{x}_{n_1}^H \boldsymbol{x}_m + r_{n_1}^* \boldsymbol{x}_{n_1}^H \boldsymbol{w} \right| \qquad (57) \\ &\leq |r_{n_1}|(\epsilon \|\boldsymbol{r}_{\mathcal{I}}\|_2 + \tau), \end{aligned}$$

where (57) follows from $\Re[r_{n_1}^* \boldsymbol{x}_{n_1}^H \boldsymbol{y}] < 0$. Similarly to earlier derivations, we have

$$|r_{n_1}^* \boldsymbol{x}_{n_0}^H \boldsymbol{y}| > |r_{n_1}|(|r|_{(1)} - \epsilon \|\boldsymbol{r}_{\mathcal{I}}\|_2 - \tau), \qquad (58)$$

we have that once (38) holds, $|r_{n_1}^* \boldsymbol{x}_{n_0}^H \boldsymbol{y}| > |r_{n_1}^H \boldsymbol{x}_{n_1}^H \boldsymbol{y}|$, which contradicts (56), then $\operatorname{sgn}(\Re[r_{n_1}^* \boldsymbol{a}_{n_1}^H \boldsymbol{y}]) = 1$. A similar argument can be made for $\Re[b_{n_1}] = -1/\sqrt{2}$ and the cases associated with $\Im[b_{n_1}]$, which completes the proof.

D. Proof of Proposition 1

Denote the index set of subsampled rows of the Gabor frame as Λ. Let $\phi_m(i)$ be the ith entry of $\boldsymbol{\phi}_m$, the coherence between two distinct columns of $\boldsymbol{X}$ is given as $m \neq m'$,

$$\langle \boldsymbol{x}_m, \boldsymbol{x}_{m'} \rangle = \sum_{i \in \Lambda} \phi_m^*(i) \phi_{m'}(i),$$

with the expectation $\mathbb{E}\langle \boldsymbol{x}_m, \boldsymbol{x}_{m'} \rangle = \langle \boldsymbol{\phi}_m, \boldsymbol{\phi}_{m'} \rangle$, whose absolute value is upper bounded by $\mu(\boldsymbol{\Phi})$, the worst case coherence of $\boldsymbol{\Phi}$. Applying the triangle inequality and the Hoeffding's inequality [26] we have for $\gamma > 0$,

$$\Pr\{|\langle \boldsymbol{x}_m, \boldsymbol{x}_{m'} \rangle| - \mu(\boldsymbol{\Phi}) \geq \gamma\} \leq 4 \exp\left(-\frac{\gamma^2 M}{4}\right),$$

Now we consider all pairs of different inner products and apply the union bound,

$$\begin{aligned} \Pr\{\mu(\boldsymbol{X}) - \mu(\boldsymbol{\Phi}) \geq \gamma\} &\leq 2P^2(P^2-1) \exp\left(-\frac{\gamma^2 M}{4}\right) \\ &< 2P^4 \exp\left(-\frac{\gamma^2 M}{4}\right). \end{aligned}$$

Let $\gamma = \sqrt{\frac{20 \log P}{M}}$, then with probability at least $1 - 2P^{-1}$, we have

$$\mu(\boldsymbol{X}) \leq \mu(\boldsymbol{\Phi}) + \sqrt{\frac{20 \log P}{M}}. \qquad (59)$$

If the Gabor frame satisfies the coherence property such that $\mu(\mathbf{\Phi}) \le \gamma_1/\log P$ for some constant γ_1, then by choosing $M \ge \gamma \log^3 P$, we have $\mu(\boldsymbol{X}) < \gamma_2/\log P$ for some constant γ_2 with probability at least $1 - 2P^{-1}$.

We next consider the average coherence of $\boldsymbol{X}$. Let $m = Pq + r$, $m' = Pq' + r'$, we have

$$\sum_{m' \neq m} \langle \boldsymbol{x}_m, \boldsymbol{x}_{m'} \rangle = \sum_{i \in \Lambda} \sum_{m' \neq m} \phi_m^*(i) \phi_{m'}(i).$$

Since each column in a Gabor frame can be written as,

$$\boldsymbol{\phi}_m = \left[g_{(1-r)_P} e^{j2\pi \frac{q}{P} \cdot 0}, \ldots, \, g_{(P-r)_P} e^{j2\pi \frac{q}{P} \cdot (P-1)} \right]^\top,$$

where $(1-r)_P = \mathrm{mod}(1-r, P)$. If $r \neq r'$, we have

$$\sum_{q'=0}^{P-1} \sum_{r \neq r'} \phi_m^H(i) \phi_{m'}(i) = P g^*_{(1-r)_P} \sum_{r \neq r'} g_{(1-r')_P} \cdot \delta_{\{i=1\}}$$

If $r = r'$, $q \neq q'$, we have

$$\sum_{q' \neq q} \phi_m^H(i) \phi_{m'}(i) = \frac{1}{M} \left[(P-1) \cdot \delta_{\{i=1\}} - \delta_{\{i \neq 1\}} \right],$$

where we use the fact $|g_i|^2 = 1/M$. To sum up, we have

$$\begin{aligned} &\sum_{m' \neq m} \phi_m^H(i) \phi_{m'}(i) \\ &= \begin{cases} P g^*_{(1-r)_P} \sum_{r \neq r'} g_{(1-r')_P} + (P-1)/M & i = 1 \\ -1/M & i \neq 1 \end{cases} \end{aligned}$$

Then $|\sum_{m' \neq m} \langle \boldsymbol{x}_m, \boldsymbol{x}_{m'} \rangle| \le \frac{P}{\sqrt{M}} \|\boldsymbol{g}\|_1 + 1 = \frac{P^2}{M} + 1$, and the average coherence $\nu(\boldsymbol{X})$ can be bounded deterministically as

$$\nu(\boldsymbol{X}) = \frac{P^2 + M}{P^2 - 1} \cdot \frac{1}{M} \le \frac{2}{M}.$$

E. Proof of Proposition 2

The analysis of the worst-case coherence is exactly as in the proof of Proposition 1 hence is not repeated. Regarding the average coherence, the columns in the Kerdock set $\tilde{\mathbf{\Psi}}$ form an abelian group $\mathcal{G}$ under point-wise multiplication. By the fundamental group property, if every row contains some entry not equal to 1, then the column group $\mathcal{G}$ satisfies $\sum_{g \in \mathcal{G}} g = 0$. Let $x_m(i)$ and $\psi_m(i)$ be the ith entry

respectively of $\boldsymbol{x}_m$ and $\boldsymbol{\psi}_m$. When $i \neq 0$, the subsampled Kerdock set $\boldsymbol{X}$ then satisfies

$$\begin{aligned}
&\left| \sum_{m' \neq m} \langle x_m(i), x_{m'}(i) \rangle \right| \\
&= \left| \sum_{m' \neq m} \langle \psi_m(i), \psi_{m'}(i) \rangle - \sum_{m \in \boldsymbol{\Psi}_c} \langle \psi_m(i), \psi_{m'}(i) \rangle \right| \\
&\leq \left| \sum_{m' \neq m} \langle \psi_m(i), \psi_{m'}(i) \rangle \right| + \sum_{m \in \boldsymbol{\Psi}_c} |\langle \psi_m(i), \psi_{m'}(i) \rangle| \\
&\leq 1 - \frac{1}{M} + \frac{P}{M},
\end{aligned}$$

and

$$\sum_{m' \neq m} \langle x_m(0), x_{m'}(0) \rangle = \frac{P^2 - P - 1}{M}.$$

Then the average coherence is bounded as

$$\nu(\boldsymbol{X}) = \frac{1}{P^2 - 1} \left| \sum_{m' \neq m} \langle \boldsymbol{x}_{m'}, \boldsymbol{x}_m \rangle \right| \leq \frac{P^2 + M - 2}{(P^2 - 1)M} \leq \frac{2}{M}.$$

REFERENCES

[1] S. Verdu, *Multiuser Detection*. Cambridge University Press, 1998.

[2] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," *submitted to IEEE Trans. Information Theory and arXived.*, March 2010.

[3] K. Finkenzeller, *RFID handbook: fundamentals and applications in contactless smart cards, radio frequency identification and near-field communication*. Wiley, 2010.

[4] L. Zhang and D. Guo, "Wireless peer-to-peer mutual broadcast via sparse recovery," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 1901–1905, IEEE, 2011.

[5] L. Zhang, J. Luo, and D. Guo, "Neighbor discovery for wireless networks via compressed sensing," *Performance Evaluation*, 2012.

[6] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Info. Theory*, vol. 52, pp. 5406 – 5424, Dec. 2006.

[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, pp. 1289 – 1306, April 2006.

[8] L. Applebaum, W. U. Bajwa, M. F. Duarte, and R. Calderbank, "Asynchronous code-division random access using convex optimization," *Physical Communication*, vol. In Press, 2011.

[9] Y. Xie, Y. Eldar, and A. Goldsmith, "Reduced-dimension multiuser detection," *accepted, IEEE Trans. Information Theory*, Jan. 2013.

[10] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics in Signal Process.*, vol. 4, pp. 375 – 391, April 2010.

[11] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[12] W. U. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? two fundamental measures of coherence and their role in model selection," *J. of Comm. and Networks*, vol. 12, Aug. 2010.

[13] R. Calderbank and S. Jafarpour, "Reed muller sensing matrices and the lasso," in *Proceedings of the 6th international conference on Sequences and their applications*, SETA'10, (Berlin, Heidelberg), pp. 442–463, Springer-Verlag, 2010.

[14] A. R. Hammons. Jr., P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, and P. Sole, "The Z4-linearity of Kerdock, Preparata, Goethals and related codes," *IEEE Trans. Information Theory*, vol. 40, pp. 301–319, April 1994.

[15] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. on Comm.*, vol. 59, pp. 454 – 465, Feb. 2011.

[16] Y. Jin, Y.-H. Kim, and B. D. Rao, "Limits on support recovery of sparse signals via multiple access communication techniques," *IEEE Trans. Info. Theory*, vol. 57, pp. 7877–7892, December 2011.

[17] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[18] Y. Xie, Y. Chi, L. Applebaum, and R. Calderbank, "Compressive demodulation of mutually interfering signals," in *2012 Statistical Signal Processing Workshop*, (Ann Arbor, MI), Aug. 2012.

[19] W. Bajwa, R. Calderbank, and D. Mixon, "Two are better than one: Fundamental parameters of frame coherence," *Appl. Comput. Harmon. Anal.*, p. to appear, 2012.

[20] Y. Chi and R. Calderbank, "Coherence-based performance guarantees of orthogonal matching pursuit," in *Proc. 50th Allerton Conference on Communication, Control and Computing*, (Monticello, Illinois), Sep. 2012.

[21] J. A. Tropp, "On the conditioning of random subdictionaries," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 1–24, 2008.

[22] E. J. Candés and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Annals of Statistics*, vol. 37, no. 5A, pp. 2145–2177, 2009.

[23] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*. Springer, 2010.

[24] T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 1–26, 2011.

[25] Z. Sidak, "Rectangular confidence regions for the means of multivariate normal distributions," *J. of Amer. Stat. Asso.*, vol. 12, pp. 626 –633, Jun. 1967.

[26] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. pp. 13–30, 1963.