Reduced-Dimension Multiuser Detection

Yao Xie, Yonina C. Eldar, Andrea Goldsmith

Abstract

We present a reduced-dimension multiuser detector (RD-MUD) structure that significantly decreases the number of required correlation branches at the receiver front-end, while still achieving performance similar to that of the conventional matched-filter (MF) bank. RD-MUD exploits the fact that the number of active users is typically small relative to the total number of users in the system and relies on ideas of analog compressed sensing to reduce the number of correlators. The correlating signals used by each correlator are chosen as an appropriate linear combination of the users' spreading waveforms, which in turn are chosen from a large class of spreading codes. We derive the probability-of-error when using two methods for recovery of active users and their transmitted symbols: the reduced-dimension decorrelating (RDD) detector, which combines subspace projection and thresholding to determine active users and sign detection for data recovery, and the reduced-dimension decision-feedback (RDDF) detector, which combines decision-feedback orthogonal matching pursuit for active user detection and sign detection for data recovery. We identify conditions such that error is dominated by active user detection. We then show that the number of correlators needed to achieve a small probability-of-error under these conditions is on the order of the logarithm of the number of users in the system for a given projection method based on random discrete Fourier transform (DFT) matrices, which is significantly lower than the number of correlators required by the the conventional MUD using the MF-bank. Our theoretical results take into consideration the effects of correlated signature waveforms as well as near-far issues. The theoretical performance results for both detectors are validated via numerical simulations.

I. Introduction

Multiuser detection (MUD) is a classical problem in multiuser communications and signal processing (see, e.g., [1][2][3] for classic reviews with extensive reference lists and [4][5][6] for some recent development in the field). In multiuser systems, the users communicate simultaneously with a given

Yao Xie (Email: yaoxie@stanford.edu) is with the Department of Electrical Engineering at Stanford University.

Yonina C. Eldar (Email: yonina@ee.technion.ac.il) is with the Department of Electrical Engineering, Technion, Israel Institute of Technology, and was visiting the Department of Electrical Engineering, Stanford University.

Andrea Goldsmith (Email: andrea@wsl.stanford.edu) is with the Department of Electrical Engineering at Stanford University.

This work is partially supported by the Interconnect Focus Center of the Semiconductor Research Corporation, BSF Transformative Science Grant 2010505, and a Stanford General Yao-Wu Wang Graduate Fellowship.

receiver by modulating information symbols onto their unique signature waveforms, which can be characterized by the chip waveform and the signature sequences (also called the spreading code) [7]. The received signal consists of a noisy version of the superposition of the transmitted waveforms. The MUD has to detect the symbols of all users simultaneously. The channel model associated with the MUD may be synchronous or asynchronous. In the synchronous channel model [8], the transmission rate of all users is the same and their symbol epochs are perfectly aligned. This requires closed-loop timing control or synchronization among all transmitters. In the asynchronous channel model [9][10], user time epochs need not be aligned and the transmitted waveforms arrive at the receiver with different time delays. Allowing users to be asynchronous simplifies system design but complicates the system model. The synchronous channel model can be viewed as a special case of the asynchronous channel with delays of all users to be the same. In this paper we will focus on the synchronous channel model. Part of the MUD problem is signature sequence selection, for which there has been a large body of work, both theoretical [11][12][13][14][15][16][17][18][19][20][21] and practical [7][22]. If we require the signature waveforms to be orthogonal, for a system with bandwidth B, the number of orthogonal signature waveforms available (and hence the number of users that the system can support) is approximately $2TB$ [3]. This hard limit on system capacity can be relaxed if we allow nonorthogonal signature waveforms, and instead require the crosscorrelation of the selected signature waveforms to be sufficiently low. In this work, we do not consider optimizing signature waveforms and hence our results will be parameterized by the crosscorrelation properties of the signature waveforms used in our design. When the signature waveforms are nonorthogonal, there is mutual interference among users, which degrades system performance for all users. Hence, an important issue in multiuser systems with nonorthogonal signature waveforms is the *near-far problem* [3]: a strong user with high enough power at the receiver may cause severe performance degradation of a weak user. One of the key challenges in MUD is to design a detector that works well when user signals are received at different power levels. Such detectors are discussed in more details below.

While there has been a large body of work developed for the multiuser detection problem over the last several decades, it is not yet widely implemented in practice, largely due to its complexity and high-precision A/D requirements. The complexity of MUD arises both in the analog circuitry for decorrelation as well as in the digital signal processing for data detection of each user. We characterize the decorrelation complexity using the number of correlators used at the receiver front-end, and measure

the data detection complexity by the complexity-per-bit [8], which is the number of real floating point operations required per bit decision.

The conventional MUD detection structure consists of a matched-filter (MF) bank front-end followed a linear or nonlinear digital multiuser detector. The MF-bank front-end is a set of correlators, each correlating the received signal with the signature waveform of a different user. Hence the conventional MUD requires the number of correlators to be equal to the number of users. The MF-bank front-end obtains a set of sufficient statistics for MUD when the receiver noise is Gaussian.

To recover user data from the MF-bank output, various digital detectors have been developed. Verdu in a landmark paper [9] establishes the optimal MUD detector as the maximum likelihood sequence estimator (MLSE), which minimizes the probability of error for symbol detection. The upper bound on the probability of error for symbol detection, or the bit-error-rate (BER), is due to [9] and the analysis is also given in [23]. Although the MLSE detector can nearly eliminate the degradation in performance due to multiuser interference, it has two main limitations: complexity and required complete channel state information [3]. The complexity-per-bit of the MLSE detector is exponential in the number of users when the signature waveforms are nonorthogonal. To address the complexity issue, other MUD suboptimal detectors have been developed. The MLSE detector is an example of a nonlinear detector that detects symbols of all users jointly. Another example of nonlinear detectors is the decision feedback (DF) detector [24][25], which is based on the idea of interference cancellation and takes various forms. One such form is the successive interference cancellation (SIC) detector [26][27]. This idea of interference cancellation dates back to the information theoretic study of the Gaussian multiple-access channel [28]. The DF detector decodes symbols iteratively and subtracts the detected symbols of strong users first to facilitate detection of weak users. The DF detector also requires complete channel state information but it has less complexity-per-bit than the MLSE detector. The number of correlators required by the DF detector is also equal to the number of users. The DF detector is a good compromise between complexity and performance among all nonlinear and linear techniques (see, e.g., [25]). For this reason we will analyze the DF detector below as an example of a nonlinear detector, but in a reduced dimensional setting.

Linear detection methods, which apply a linear transform to the receiver front-end output and then detect each symbol separately, have lower complexity than nonlinear methods but also worse performance. Linear MUD techniques include the single-user detector, the decorrelating detector and the minimum

mean-square-error (MMSE) detector. The single-user detector is the simplest linear detector, which follows the MF-bank front-end the conventional single-user detection in each branch, which detects symbols by comparing the front-end output with a threshold [8]. When the signature waveforms are orthogonal and synchronous, the single-user detector coincides with the MLSE detector and it minimizes the probability-of-error. In this case, by correlating with each user's signature waveform, the MF-bank maximizes the output signal-to-noise ratio (SNR) for each individual user. When the signature waveforms are nonorthogonal, users interfere with each other, so the probability-of-error of the single-user detector degrades [1]. A linear detector that eliminates user inference is the decorrelating detector [8][10], which, for each user, projects the received signal onto the subspace associated with the signature waveform of that user. This projection amplifies noise when the signature waveforms are nonorthogonal. The decorrelating detector also gives the best joint estimate of symbols and amplitudes in the absence of knowledge of the complete channel state information [8], and it optimizes the near-far resistance among linear detectors [29]. The near-far resistance is a performance measure for the degree of robustness against the near-far problem achieved by a multiuser detector [3]. The MMSE detector [30][31][32] is designed to minimize the mean-square-error (MSE) between symbols and the linear transformed MF-bank front-end output. The MMSE detector takes into account the background noise and interference, and hence to some extent it mitigates the noise amplification effect of the decorrelating detector in the low and medium SNR regimes [25]. A drawback of the MMSE detector is that it requires complete channel state information and it does not achieve the optimal near-far resistance. When the signal-to-noise power ratio (SNR) goes to infinity, the MMSE detector converges to the decorrelating detector [3]. Because of the many advantages of the decorrelating detector, it has received much attention in the MUD literature and is one of the most common linear detectors in MUD [2]. Hence, in this paper, we will focus on the decorrelating detector as an example of a linear detector in the reduced-dimension setting. The decorrelating detector requires inverting the correlation matrix of the signature waveforms, but this computation can be done offline once the signature waveforms are selected. Its complexity-per-bit (other than inverting the correlation matrix) can be shown to be linear in the number of users [8].

Both linear and nonlinear MUDs have sufficiently high complexity to preclude their wide adoption in deployed systems. They both require the number of correlators at the receiver front-end to be equal to the number of users in the system. The complexity-per-bit of the nonlinear detectors are exponential in

the number of users in system for the MLSE detector, and less for the DF detector. The complexity-per-bit of the linear detectors are linear in the number of users, which is much less than that of the nonlinear methods. In a typical communication system, there may be thousands of users. Hence, the complexity of the conventional methods has been a major obstacle for implementing the MF-bank based conventional MUD detectors. Our methods reduce the front-end complexity far below that of the nonlinear and linear techniques in large systems, and their complexity-per-bit are comparable to those of the conventional nonlinear and linear techniques. For example, when the random partial discrete Fourier transform (DFT) matrix is used to construct the correlating signals, the number of correlators used by our methods is proportional to the logarithm of the number of users and proportional to the square of the number of active users, and the complexity-per-bit of our method is comparable to the corresponding conventional MUD detectors.

In this work, we develop a low complexity front-end for MUD along with corresponding digital detectors. We call this structure a reduced-dimension multiuser detector (RD-MUD). The RD-MUD reduces the number of correlators while still achieving performance similar to that of conventional MUDs that are based on the MF-bank front-end. We reduce complexity by exploiting the fact that at any given time the number of active users, K, is typically much smaller than the total number of users, N. This analog signal sparsity allows us to use techniques from analog compressed sensing, which exploits sparsity in analog signals (see, e.g., [33][34][35][36][37][38][39] and [40] for more details on these techniques). Our RD-MUD has a front-end that correlates the received signal with M correlating signals, where M is much smaller than N. The correlating signals are formed as linear combinations of the signature waveforms via a (possibly complex) coefficient matrix $\mathbf{A}$ as is done in the analog compressed sensing literature for sparse signal recovery. Our choice of $\mathbf{A}$ will be shown to be crucial for performance. The output of the RD-MUD front-end can thus be viewed as a projection of the MF-bank output onto a lower dimensional detection subspace. To recover information from this detection subspace, we process the front-end output using algorithms that combine ideas from compressed sensing and MF-bank based conventional MUD. We study two such detectors in detail: the reduced-dimension decorrelating (RDD) detector, a linear detector that combines subspace projection and thresholding to determine active users with a sign detector for data recovery [41][42], and the reduced-dimension decision-feedback (RDDF) detector, a nonlinear detector that combines decision-feedback orthogonal matching pursuit (DF-OMP) [43][44] for active user detection with sign detection for data recovery in

an iterative manner. We present theoretical probability-of-error performance guarantees for these two specific detectors in terms of the coherence of the matrix $\mathbf{A}$ in a non-asymptotic regime with a fixed number of users and active users. Based on these results, we develop a lower bound on the number of correlators M needed to attain a certain probability-of-error performance. For example, if $\mathbf{A}$ is a random partial discrete Fourier transform matrix, the M required by these two specific detectors is on the order of $\log N$. We validate these theoretical performance results via numerical examples.

Previous work on MUD based on active user detection falls into two categories: conventional methods and compressed sensing methods. Specifically, to detect active users in the system, conventional MUD approaches use techniques such as the multiple signal classification algorithm (MUSIC) [45], quickest change detection [46], random set theory, sphere detection and Bayesian filter [47][48][49][50]. In particular, [46] focuses on the on-online detection of the entrance of a new user into the system, and [45][47][50] consider the case when the number of active users is unknown. There has also been work about detecting a subset of active users of interest [51], which belongs to a more general technique called group detection [52]. The group detection technique partitions active users into groups and jointly detects users within a group using the generalized likelihood ratio test. A compressed sensing approach has also been applied for active user detection. However, in contrast to our approach, which processes analog signals, most existing work on exploiting compressed sensing ideas for signal detection based on the original compressed sensing results [53][54] assume discrete signals. In particular, most prior work on MUD exploiting user sparsity applies compressed sensing techniques on discrete signals via matrix multiplication [55][56][57][58][59][60], whereas in our work the compressed sensing techniques are incorporated into the RD-MUD analog front-end. Furthermore, RD-MUD aims at detecting active users as well as their transmitted symbols, whereas prior work [61][62] [55][56][57] aims at detecting only the active users and hence the problem is equivalent to support recovery. These prior works establish conditions on the number of correlators M required to achieve a zero probability-of-error of active user detection when the number of users N tends to infinity. While providing important insights into complexity reduction in large systems, they did not answer questions for practical system design with a finite number of users, such as how many correlators should be used to achieve a target probability-of-error. There is another branch of compressed detection work that focuses on detecting the presence of a discrete signal that is sparse in time [63][59][64]. This work is not relevant to our problem since the MUD signal we consider is sparse in the number of users.

Our RD-MUD consists of two stages: active user detection and data detection of active users. The first stage is closely related to [65]. However, our problem differs in that the probability-of-error must consider errors in both stages. We derive conditions under which the probability-of-error is dominated by errors in the first stage. Also, the decision-feedback (DF) detector we consider is different in its active user detection because it subtracts out detected data symbols whose values are from a finite alphabet (rather than subtracting out the estimated data symbols whose values are real) from the second stage, which makes detection of the remaining active users easier.

The rest of the paper is organized as follows. Section II discusses the system model and reviews conventional detectors using the MF-bank front-end. Section III presents the RD-MUD front-end and detectors. Section IV contains the theoretical performance guarantee of two RD-MUD detectors: the RDD and RDDF detectors. Section V validates the theoretical results through numerical examples, and finally Section VI concludes the paper.

II. System Model

A. Notation

The notation we use is as follows. Vectors and matrices are denoted by boldface lower-case and boldface upper-case letters, respectively. The real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The real part of a scalar x is represented as $\Re[x]$. The conjugate of a complex number x is denoted by x^*. The set of indices of the nonzero entries of a vector $\mathbf{x}$ is called the support of $\mathbf{x}$. Given an index set $\mathcal{I}$, $\mathbf{X}_{\mathcal{I}}$ denotes the submatrix formed by the columns of a matrix $\mathbf{X}$ indexed by $\mathcal{I}$, and $\mathbf{x}_{\mathcal{I}}$ denotes the subvector formed by the entries indexed by $\mathcal{I}$. The entry of a matrix $\mathbf{X}$ at the nth row and mth column is indicated by $[\mathbf{X}]_{nm}$. The identity matrix is denoted by $\mathbf{I}$. The transpose, conjugate transpose, and inverse of a matrix $\mathbf{X}$ are denoted by $\mathbf{X}^\top$, $\mathbf{X}^H$, and $\mathbf{X}^{-1}$, respectively. The inner product between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N\times 1}$ with entries x_n and y_n, respectively, is defined as: $\mathbf{x}^H\mathbf{y} = \sum_{n=1}^{N} x_n^* y_n$. The l_2 norm of a vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\| = (\mathbf{x}^H\mathbf{x})^{1/2}$. The minimum and maximum eigenvalues of a positive-semidefinite matrix $\mathbf{X}$ are denoted by $\lambda_{\min}(\mathbf{X})$ and $\lambda_{\max}(\mathbf{X})$, respectively. The trace of a square matrix $\mathbf{X}$ is denoted as $\mathbf{tr}(\mathbf{X})$. The spectral norm of $\mathbf{X}$ is denoted by $\rho(\mathbf{X}) = [\lambda_{\max}(\mathbf{X}^H\mathbf{X})]^{1/2}$. The function δ_{nm} is defined such that $\delta_{nm} = 1$ only when $n = m$ and otherwise is equal to 0. The sign function is defined as $\mathbf{sgn}(x) = 1$, if $x > 0$, $\mathbf{sgn}(x) = -1$, if $x < 0$, and otherwise is equal to 0. The expectation of a random variable x is denoted as $\mathbb{E}\{x\}$ and the probability of an event $\mathcal{A}$ is denoted as $P(\mathcal{A})$. The union, intersection, and difference of two sets $\{A\}$ and $\{B\}$ are denoted by $\{A\} \cup \{B\}$,

$\{A\} \cap \{B\}$, and $\{A\}/\{B\}$, respectively. The complement of a set $\{A\}$ is denoted as $\{A\}^c$. The notation $A \subset B$ means that set A is a subset of set B.

B. System Model

Consider a multiuser system with N users. Each user is assigned a unique signature waveform from a set $\mathcal{S} = \{s_n(\cdot) : [0, T] \to \mathbb{R}, 1 \leq n \leq N\}$. The signature waveforms are assumed given and known, and possess certain properties discussed in more detail below. Each user modulates its data signal by its signature waveform to transmit its data symbols. The symbols carry information and are chosen from a constellation depending on the specific modulation scheme. The duration of the data symbol T is referred to as the symbol time, which is also equal to the inverse of the data rate for binary modulation.

Define the *inner product* (or *crosscorrelation*) between two real analog signals $x(t)$ and $y(t)$ in L_2 as

$$\langle x(t), y(t) \rangle = \frac{1}{T} \int_0^T x(t) y(t) dt \tag{1}$$

over the symbol time T. We also define the L_2 norm of a real analog signal $x(t)$ as

$$\|x(t)\| = \langle x(t), x(t) \rangle^{1/2}. \tag{2}$$

Two signals are orthogonal if their crosscorrelation is zero. We assume that the signature waveforms are linearly independent. That is, any linear combinations of different signature waveforms cannot be another signature waveform (or its multiples) in order to avoid cancellation of each other's transmission. The crosscorrelations of the signature waveforms are characterized by the Gram matrix $\mathbf{G}$, defined as

$$[\mathbf{G}]_{nl} \triangleq \langle s_n(t), s_l(t) \rangle, \qquad 1 \leq n \leq N, \qquad 1 \leq l \leq N. \tag{3}$$

For convenience, we assume that $s_n(t)$ has unit energy: $\|s_n(t)\| = 1$ for all n so that $[\mathbf{G}]_{nn} = 1$. Due to our assumption of linear independence of the signature waveforms, $\mathbf{G}$ is invertible. The signature waveforms typically have low crosscorrelations, so we also assume that the magnitudes of the off-diagonal elements of $\mathbf{G}$ are much smaller than 1.

We consider a synchronous MUD model with Binary Phase Shift Keying (BPSK) modulation [3]. There are K active users out of N possible users transmitting to the receiver. The set $\mathcal{I}$ contains indices of all active users, and its complement set $\mathcal{I}^c$ contains indices of all non-active users. The active users modulate their signature waveforms using BPSK modulation with the symbol of the user n denoted by

$b_n \in \{1, -1\}$, for $n \in \mathcal{I}$. The nth user transmits its modulated signal at power P_n through a wireless channel with channel amplitude g_n. We assume that the channel amplitude g_n can be estimated and is known to the receiver (CSIR). Define the gain $r_n \triangleq \sqrt{P_n} g_n$ where we assume r_n is also known at the receiver. For simplicity, we assume that g_n is real and, hence, r_n is real and can be negative. The nonactive user can be viewed as transmitting with power $P_n = 0$, or equivalently transmitting zeros: $b_n = 0$, for $n \in \mathcal{I}^c$.

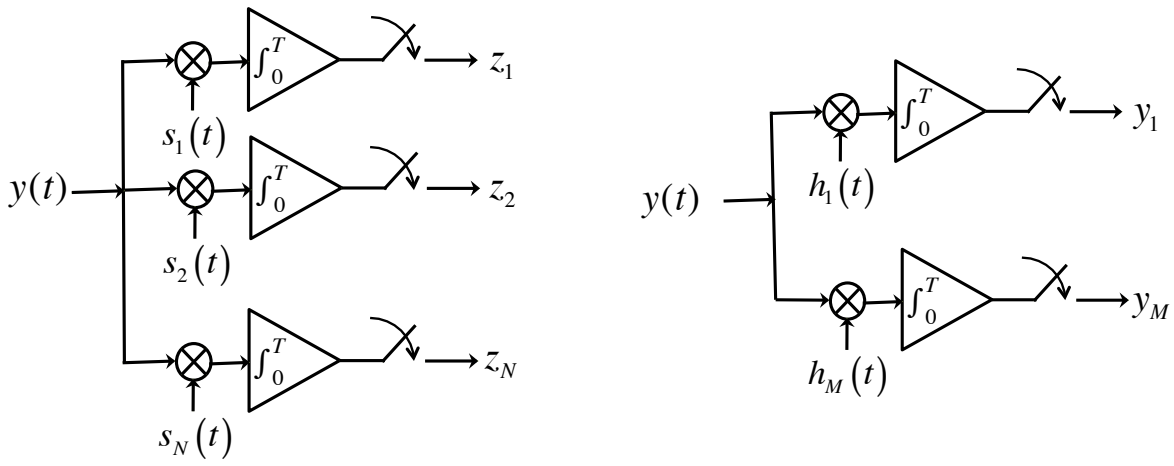
The received signal $y(t)$ is a superposition of the transmitted signals from the active users, plus white Gaussian noise $w(t)$ with zero-mean and variance σ^2:

$$y(t) = \sum_{n=1}^{N} r_n b_n s_n(t) + w(t), \qquad t \in [0, T], \tag{4}$$

with $b_n \in \{1, -1\}$, $n \in \mathcal{I}$, and $b_n = 0$, $n \in \mathcal{I}^c$. In the presence of delays of the asynchronous channel model, the ideas can be combined with the methods developed in [39] for time-delay recovery from low-rate samples.

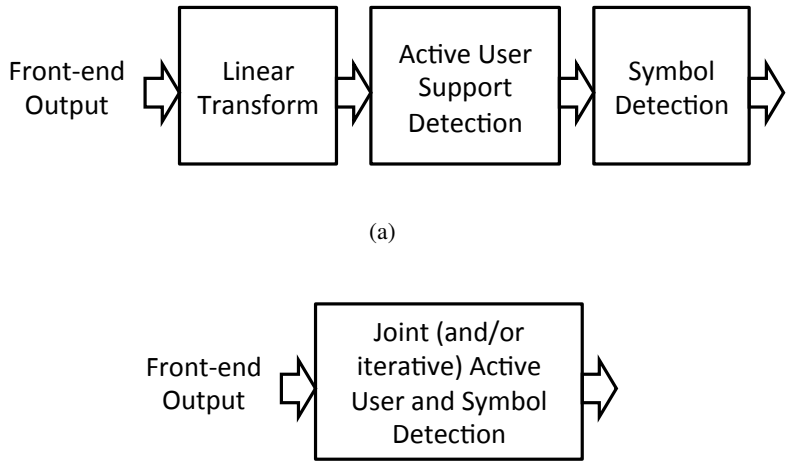
The goal of multiuser detection (MUD) is to detect the set of active users, i.e. users with indices in $\mathcal{I}$, and their transmitted symbols $\{b_n : n \in \mathcal{I}\}$. In practice the number of active users K is typically much smaller than the total number of users N, which is a form of user sparsity. As we will show, this user sparsity enables us to reduce the number of correlators at the front-end and still be able to achieve performance similar to that of a conventional MUD using a bank of MFs. To simplify the detection algorithm, we assume that K is known. To consider joint estimation of K as well as the active users and their symbols greatly complicates analysis for the probability-of-error, since in that case the error will come from three sources: estimating the wrong number of active users, estimating the wrong set of active users, and estimating the wrong symbols. By assuming K is known, we will consider the latter two sources of errors in our analysis. The problem of estimating K can be treated using techniques such as these in [45] and [47].

In the following subsection, we present the structure of conventional MUD detectors using the MF-bank front-end. In Section III we introduce the new reduced-dimension MUD (RD-MUD) front-end and detectors.



(a) (b)

Fig. 1: Front-end of (a) conventional MUD using MF-bank, and (b) RD-MUD.



(a)

(b)

Fig. 2: The diagram of (a) linear detector, and (b) nonlinear detector.

C. Conventional MUD

A conventional MUD detector has a front-end consisting of a bank of matched filters followed by a digital detector. In the following we review the MF-bank front-end and various digital detectors.

1) MF-bank front-end: For single-user systems, the matched filter (MF) passes the signal $s(t)$ through a filter with impulse response $s^*(T-t)$ or, equivalently, multiplies the received signal $s(t)$ with itself and integrates over a symbol time: this maximizes the output SNR of the decision statistic among the class of all linear detectors [3]. The MF-bank is an extension of the MF for multiple users, and it has N

MFs in parallel: the nth branch correlates the received signal with the corresponding signature waveform $s_n(t)$, as illustrated in Fig. 1a. As stated earlier, the output of the MF-bank is a set of sufficient statistics for MUD when the gains r_n are known [3]. Hence, no information is left in the MF-bank front end of the MUD receiver.

Using the model for the received signal (4), the output of the nth correlator of the MF-bank can be written as

$$z_n = \langle y(t), s_n(t) \rangle = r_n b_n + \sum_{l \neq n} [\mathbf{G}]_{nl} r_l b_l + u_n, \qquad 1 \leq n \leq N. \tag{5}$$

The output noise $u_n = \langle w(t), s_n(t) \rangle$ is a Gaussian random variable with zero mean and covariance $\mathbb{E}\{u_n(t)u_m(t)\} = \sigma^2[\mathbf{G}]_{nm}$ (for derivation see Section 2.9 in [3]). Let $\mathbf{z} = [z_1, \cdots, z_N]^\top$, $\mathbf{R} \in \mathbb{R}^{N \times N}$ be a diagonal matrix with $[\mathbf{R}]_{nn} = r_n$, $\mathbf{b} = [b_1, \cdots, b_N]^\top$ and $\mathbf{u} = [u_1, \cdots, u_N]^\top$. We can express the output of the MF-bank (5) in a vector form as:

$$\mathbf{z} = \mathbf{GRb} + \mathbf{u}, \tag{6}$$

where $\mathbf{u}$ is a Gaussian random vector with zero mean and covariance $\mathbb{E}\{\mathbf{u}\mathbf{u}^H\} = \sigma^2\mathbf{G}$.

2) MF-bank detection: Conventional MUD detectors based on using the MF-bank output can be classified into two categories: linear and nonlinear detectors. The diagrams of these two detectors are illustrated in Fig. 2. In the literature, typically the basic synchronous MUD model assumes all users are active, i.e. $b_n \in \{1, -1\}$, and hence the goal of the MUD detectors is to detect all user symbols.

The linear detector applies a linear transform to the MF-bank output (illustrated in Fig. 2a):

$$\mathbf{Tz} = \mathbf{TGRb} + \mathbf{Tu}, \tag{7}$$

and detects symbol for each user separately using a sign detector:

$$\hat{b}_n = \mathbf{sgn}(r_n[\mathbf{Tz}]_n), \qquad 1 \leq n \leq N. \tag{8}$$

Several commonly used linear detectors are the single-user detector, the decorrelating detector and the minimum-mean-square-error (MMSE) detector. The single-user detector [8] is equivalent to having $\mathbf{T} = \mathbf{I}$ in (7), and detecting symbols as $\hat{b}_n = \mathbf{sgn}(r_n z_n)$. The decorrelating detector [8] is motivated by the fact that when the signature waveforms are nonorthogonal, even in the absence of noise, the single-user detector (8) will not lead to good detection, when there is non-negligible interference between

users (captured mathematically in (9) by the second term in parenthesis), since generally

$$\hat{b}_n = \mathbf{sgn}(r_n z_n) = \mathbf{sgn}\left(r_n^2 b_n + \sum_{l \neq n, l=1}^{N} [\mathbf{G}]_{nl} r_n r_l b_l \right) \neq \mathbf{sgn}(r_n^2 b_n) = b_n. \tag{9}$$

The decorrelating detector addresses this problem by applying a linear transform $\mathbf{T} = \mathbf{G}^{-1}$ in (7). The decorrelating detector can recover symbols perfectly in the absence of noise; however, it also amplifies noise when $\mathbf{G} \neq \mathbf{I}$ and requires $\mathbf{G}$ to be invertible. The minimum mean square error (MMSE) detector is based on the criterion to minimize the mean square error (MSE) between the linear transform of the MF-bank output and symbols. Based on this principle, MMSE uses a linear transform $\mathbf{T} = (\mathbf{G} + \sigma^2 \mathbf{R}^{-2})^{-1}$ in (7) [3].

The nonlinear detectors, on the other hand, detect symbols jointly and (or) iteratively as illustrated in Fig. 2b. The nonlinear detectors include the maximum likelihood sequence estimator (MLSE) and the successive interference cancellation (SIC) detector [3]. The MLSE achieves the minimum probability-of-error by solving the following optimization problem to detect symbols jointly:

$$\max_{b_n \in \{1,-1\}} 2\mathbf{y}^H \mathbf{R}\mathbf{b} - \mathbf{b}^H \mathbf{R}\mathbf{G}\mathbf{R}\mathbf{b}. \tag{10}$$

If the signature waveforms are orthogonal, $\mathbf{G}$ is a diagonal matrix and the optimization problem of (10) decouples into N single user problems, and the optimal solution is the sign detector (8) (with $\mathbf{T} = \mathbf{I}$). However, when the signature waveforms are nonorthogonal this optimization problem (10) is exponentially complex in the number of users [66]. It can be solved by exhaustive search which is computationally very expensive. The SIC detector first finds the active user with the largest gain, detects its symbol, subtracts its effect from the received signal, and iterates the above process using the residual signal. After K iterations, the SIC detector determines all active users. As we will show later, the ideas of SIC and the orthogonal matching pursuit (OMP) in compressed sensing are similar.

III. Reduced-Dimension MUD (RD-MUD)

The RD-MUD front-end, illustrated in Fig. 1b, correlates the received signal $y(t)$ with a set of correlating signals $h_m(t)$, $m = 1, \cdots M$, where M is typically much smaller than N. This is in contrast to the conventional matched filter (MF) bank, which correlates the received signal with the full set of N signature waveforms [3]. The front-end output is processed by either a linear or nonlinear detector to detect active users and their symbols, as shown in Fig. 2 for both linear and nonlinear detectors.

A. RD-MUD: Front-End

The design of the correlating signals $h_m(t)$ is the key for RD-MUD to reduce the number of correlators. To construct these signals, we rely on biorthogonal waveforms. The related notion of the biorthogonal function has been used in analog compressed sensing to construct multichannel filters to sample the analog signal [35]. The biorthogonal signals with respect to $\{s_n(t)\}$ are defined as a linear combination of all signature waveforms using a weighting coefficient matrix $\mathbf{G}^{-1}$:

$$\hat{s}_n(t) = \sum_{l=1}^{N} [\mathbf{G}^{-1}]_{nl} s_l(t), \qquad 1 \le n \le N, \tag{11}$$

and have the property that $\langle s_n(t), \hat{s}_m(t) \rangle = \delta_{nm}$, for all n, m, i.e. the set of signals $\{\hat{s}_m(t)\}$ are biorthogonal to the signature waveforms $\{s_n(t)\}$. This is because

$$\langle s_n(t), \hat{s}_m(t) \rangle = \left\langle s_n(t), \sum_{l=1}^{N} [\mathbf{G}^{-1}]_{ml} s_l(t) \right\rangle = \sum_{l=1}^{N} [\mathbf{G}^{-1}]_{ml} \langle s_n(t), s_l(t) \rangle = \sum_{l=1}^{N} [\mathbf{G}^{-1}]_{ml} [\mathbf{G}]_{ln} = \delta_{nm}. \tag{12}$$

Since we assume that the signature waveforms are linearly independent in Section II-B, $\mathbf{G}$ is invertible. Note that when $\{s_n(t)\}$ are orthogonal, $\mathbf{G} = \mathbf{I}$ and $\hat{s}_n(t) = s_n(t)$.

The RD-MUD front-end uses as its correlating signals $\{h_m(t)\}$ which are linear combinations of the biorthogonal waveforms. The linear combination uses (possibly complex) weighting coefficients a_{mn} that we choose:

$$h_m(t) = \sum_{n=1}^{N} a_{mn} \hat{s}_n(t), \qquad 1 \le m \le M. \tag{13}$$

The performance of RD-MUD depends on these coefficients $\{a_{mn}\}$. Define a coefficient matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ with $[\mathbf{A}]_{mn} \triangleq a_{mn}$ and denote the nth column of $\mathbf{A}$ as $\mathbf{a}_n \triangleq [a_{1n}, \cdots, a_{Mn}]^\top$, $n = 1, \cdots, N$. We normalize the columns of $\mathbf{A}$ so that $\|\mathbf{a}_n\| = 1$. The design of the correlating signals is equivalent to the design of the coefficient matrix $\mathbf{A}$ for a given $\{s_n(t)\}$. We will use coherence as a measure of the quality of $\mathbf{A}$, which is defined as [67][65]:

$$\mu \triangleq \max_{n \neq l} \left| \mathbf{a}_n^H \mathbf{a}_l \right|. \tag{14}$$

As we will show later, it is desirable that the columns of $\mathbf{A}$ have small correlation such that μ is small. This requirement for small μ also reflects a tradeoff in choosing how many correlators M to use in the RD-MUD front-end. We will show later that with more correlators, the coherence of $\mathbf{A}$ can be lower and the performance of RD-MUD can be better.

We now derive the output of the RD-MUD front-end. From the choice of the correlating signals (13) and the receive signal model (4), the output of the mth correlator is given by:

$$y_m = \langle h_m(t), y(t)\rangle = \left\langle \sum_{n=1}^{N} a_{mn}\hat{s}_n(t), \sum_{l=1}^{N} r_l b_l s_l(t) \right\rangle + \left\langle \sum_{n=1}^{N} a_{mn}\hat{s}_n(t), w(t) \right\rangle \tag{15}$$

$$= \sum_{l=1}^{N} r_l b_l \sum_{n=1}^{N} a_{mn}\langle \hat{s}_n(t), s_l(t)\rangle + w_m \tag{16}$$

$$= \sum_{l=1}^{N} a_{ml} r_l b_l + w_m, \tag{17}$$

where we have defined $w_m \triangleq \left\langle \sum_{n=1}^{N} a_{mn}\hat{s}_n(t), w(t) \right\rangle$, and used the property that $\langle s_n(t), \hat{s}_m(t)\rangle = \delta_{nm}$. We define the output noise

$$w_m = \sum_{n=1}^{N} a_{mn}\langle \hat{s}_n(t), w(t)\rangle, \qquad 1 \le m \le M. \tag{18}$$

This is a Gaussian random variable with zero-mean, variance $\sigma_m^2 \triangleq \sigma^2[\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H]_{mm}$, and covariance $\rho_{nm} \triangleq \mathbb{E}\{w_n w_m\} = \sigma^2[\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H]_{nm}$ (for derivation see Appendix A). Denoting $\mathbf{y} = [y_1, \cdots, y_M]^\top$ and $\mathbf{w} = [w_1, \cdots, w_M]^\top$, we can express the RD-MUD output (17) in vector form as

$$\mathbf{y} = \mathbf{A}\mathbf{R}\mathbf{b} + \mathbf{w}, \tag{19}$$

where $\mathbf{w}$ is a Gaussian random vector with zero mean and covariance $\sigma^2\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H$. The vector $\mathbf{y}$ can be viewed as a linear projection of the MF-bank front-end output onto a lower dimensional subspace which we call the *detection subspace*. Since there are at most K active users, $\mathbf{b}$ has at most K non-zero entries. The idea of RD-MUD is that when the original signal vector $\mathbf{b}$ is sparse, with proper choice of the matrix $\mathbf{A}$, the detection performance for $\mathbf{b}$ based on $\mathbf{y}$ of (19) in the detection subspace can be similar to the performance based on $\mathbf{z}$ (6), the output of the MF-bank front-end.

B. RD-MUD: Detectors

We now discuss how to recover $\mathbf{b}$ from the RD-MUD front-end output $\mathbf{y}$ of (19) using digital detectors. The model for the output (19) of the RD-MUD front-end has a similar form to the observation model in the compressed sensing literature [56][65], except that the noise in the RD-MUD front-end output is colored due to matched filtering at the front-end. Hence, to recover $\mathbf{b}$, we can adopt the ideas developed in the context of compressed sensing, and combine them with techniques in MF-bank detection.

The linear detector for RD-MUD first estimates active users $\hat{\mathcal{I}}$ using support recovery techniques from compressed sensing [68][56][69]. These support recovery techniques include thresholding [42], which we will describe in more detail later and use for active user detection in the RDD detector and other linear detectors, and orthogonal matching pursuit (OMP) [43][44], which we extend to the decision-feedback OMP and use for joint active user and symbol detection in the RDDF detector.

Once the active users are estimated, their symbols $\mathbf{b}_{\hat{\mathcal{I}}}$ can be detected. Based on the estimated index set of active users $\hat{\mathcal{I}}$ (which may be different from $\mathcal{I}$), we can write the RD-MUD front-end output model (19) as

$$\mathbf{y} = \mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}\mathbf{b}_{\hat{\mathcal{I}}} + \mathbf{w}. \tag{20}$$

The symbols $\mathbf{b}_{\hat{\mathcal{I}}}$ can be detected from (20) by applying a linear transform to the front-end output and detecting symbols separately. The nonlinear detector for RD-MUD detects active users and their symbols jointly (and/or iteratively).

We will focus on recovery based on two algorithms: (1) the reduced-dimension decorrelating (RDD) detector, a linear detector that uses subspace projection along with thresholding [42] to determine active users and sign detection for data recovery; (2) the reduced-dimension decision feedback (RDDF) detector, a nonlinear detector that combines decision-feedback orthogonal matching pursuit (DF-OMP) for active user detection and sign detection for data recovery. DF-OMP differs from the conventional OMP [43][44] in that in each iteration, the binary-valued detected symbols, rather than the real-valued estimates, are subtracted from the received signal to form the residual used by the next iteration. The residual consists of the remaining undetected active users. By subtracting interference from the strongest active user we make it easier to detect the remaining active users. DF-OMP can also be viewed as a successive interference cancellation (SIC) detector [26][27] in the detection subspace: DF-OMP detects each user by computing the inner product of the received signal vector and the signature waveform vector in the projection subspace and subtracts the effect of each user using the data model (19), whereas the SIC detector does this using a difference data model (6). These two algorithms are summarized in Table I.

1) Reduced-dimension decorrelating (RDD) detector: The RDD detector works as follows. As per (19), the front-end of the RD-MUD projects the received signal $y(t)$ onto the detection subspace as a vector $\mathbf{y}$. By considering the RD-MUD front-end output when the input signal is $s_n(t)$, we can show

that the column $\mathbf{a}_n$ of $\mathbf{A}$ corresponds to the nth signature waveform vector in the detection subspace:

$$\langle h_m(t), s_n(t)\rangle = \sum_{l=1}^{N} a_{ml}\langle \hat{s}_l(t), s_n(t)\rangle = a_{mn}, \qquad 1 \le m \le M. \tag{21}$$

Considering the detection method of the conventional MUD, a natural strategy for RD-MUD is to match the received signal vector $\mathbf{y}$ and the nth signature waveform vector in the detection subspace by computing their inner product, which is given by $\mathbf{a}^H\mathbf{y}$, $n = 1, \cdots, N$. To detect active users, we can rank the magnitudes of these inner products and detect the index of the K largest as active users:

$$\hat{\mathcal{I}} = \{n: \quad \text{if } |\Re[\mathbf{a}_n^H\mathbf{y}]| \text{ is among the } K \text{ largest of } |\Re[\mathbf{a}_n^H\mathbf{y}]|,\ n = 1, \cdots, N\}. \tag{22}$$

This method (22) has also been used for sparsity pattern recovery in the compressed sensing literature (e.g. [56]). To detect their symbols, we use sign detection:

$$\hat{b}_n = \begin{cases} \mathbf{sgn}\left(r_n \Re[\mathbf{a}_n^H\mathbf{y}]\right), & n \in \hat{\mathcal{I}}; \\ 0, & n \notin \hat{\mathcal{I}}. \end{cases} \tag{23}$$

In detecting active users (22) and their symbols (23), we take the real parts of the inner products because the imaginary part of $\mathbf{a}_n^H\mathbf{y}$ contains only noise and interference. To see this, expand

$$\mathbf{a}_n^H\mathbf{y} = r_n b_n + \sum_{l\neq n, l\in\mathcal{I}} r_l b_l(\mathbf{a}_n^H\mathbf{a}_l) + \mathbf{a}_n^H\mathbf{w}, \qquad n \in \mathcal{I}, \tag{24}$$

$$\mathbf{a}_n^H\mathbf{y} = \sum_{l\in\mathcal{I}} r_l b_l(\mathbf{a}_n^H\mathbf{a}_l) + \mathbf{a}_n^H\mathbf{w}, \qquad n \notin \mathcal{I}. \tag{25}$$

Recall that symbols b_n and gains r_n are real and only $\mathbf{A}$ can be complex. Hence the term $r_n b_n$ in (24), which contains the transmitted symbol, is real, and the rest of the terms in (24) and (25), which contain noise and interference, are complex. For real $\mathbf{A}$, (22) and (23) are equivalent to decisions based on the magnitudes of the inner products.

The RDD detector computes the inner products of the received signal vector and the signature waveform vector $\mathbf{a}_n^H\mathbf{y}$, $n = 1, \cdots, N$, which is equivalent to computing $\mathbf{A}^H\mathbf{y}$. This requires MN floating point operations when $\mathbf{A}$ is real (or $2MN$ operations when $\mathbf{A}$ is complex) for detection of $N\log_2 3$ bits (since equivalently we are detecting $b_n \in \{-1, 0, 1\}$). Hence the complexity-per-bit of the RDD detector is proportional to M. For other RD-MUD linear detectors, following the inner product computation, linear processing can be done by multiplying an $N \times N$ matrix that incurs N^2 operations. Hence the complexity-per-bit of other RD-MUD linear detectors is proportional to $M + N$. Since $M \le N$ in

TABLE I: RD-MUD Detection

RDD detector	**RDDF detector**
Detect active users: find $\hat{\mathcal{I}}$ that contains indices of the K largest $\lvert\Re[\mathbf{a}_n^H\mathbf{y}]\rvert$. Detect symbols: $\hat{b}_n = \mathbf{sgn}(r_n\Re[\mathbf{a}_n^H\mathbf{y}])$ for $n \in \hat{\mathcal{I}}$, and $\hat{b}_n = 0$ for $n \notin \hat{\mathcal{I}}$.	Initialize: $\hat{\mathcal{I}}$ is empty, $\mathbf{b}^{(0)} = 0$, $\mathbf{v}^{(0)} = \mathbf{y}$. Iterate K times: $k = 1, \cdots, K$. Detect active user: $n_k = \arg\max_n \lvert\Re[\mathbf{a}_n^H\mathbf{v}^{(k-1)}]\rvert$ Detect symbol: $\hat{b}_n^{(k)} = \mathbf{sgn}(r_{n_k}\Re[\mathbf{a}_{n_k}^H\mathbf{v}^{(k-1)}])$, for $n = n_k$, and $\hat{b}_n^{(k)} = \hat{b}_n^{(k-1)}$ for $n \neq n_k$. Update residual: $\hat{\mathcal{I}}^{(k)} = \hat{\mathcal{I}}^{(k-1)} \cup \{n_k\}$, and $\mathbf{v}^{(k)} = \mathbf{y} - \mathbf{ARb}^{(k)}$.

RD-MUD, the complexity for data detection of the RDD detector and other RD-MUD linear detectors is on the same order as that of the conventional linear MUD detector. But the RDD detector and other linear RD-MUD detectors require much lower decorrelation complexity in the analog front-end than the conventional linear detector.

2) Reduced-dimension decision feedback (RDDF) detector: The RDDF detector determines active users and their corresponding symbols iteratively. It starts with an empty set as the initial estimate for the set of active user $\hat{\mathcal{I}}^0$, zeros as the estimated symbol vector $\mathbf{b}^{(0)} = \mathbf{0}$, and the front-end output as the residual vector $\mathbf{v}^{(0)} = \mathbf{y}$. Subsequently, in each iteration $k = 1, \cdots, K$, the algorithm selects the column $\mathbf{a}_n$ that is most highly correlated with the residual $\mathbf{v}^{(k-1)}$ as the detected active user in the kth iteration, with the active user index:

$$n_k = \arg\max_n \left|\Re[\mathbf{a}_n^H\mathbf{v}^{(k-1)}]\right|. \tag{26}$$

This index is then added to the active user set $\hat{\mathcal{I}}^{(k)} = \hat{\mathcal{I}}^{(k-1)} \cup \{n_k\}$. The symbol for user n_k is detected with other detected symbols staying the same:

$$b_n^{(k)} = \begin{cases} \mathbf{sgn}(\Re[r_{n_k}\mathbf{a}_{n_k}^H\mathbf{v}^{(k-1)}]), & n = n_k; \\ b_n^{(k-1)}, & n \neq n_k. \end{cases} \tag{27}$$

Then the residual vector is updated through

$$\mathbf{v}^{(k)} = \mathbf{y} - \mathbf{ARb}^{(k)}. \tag{28}$$

The residual vector represents the part of $\mathbf{b}$ that has yet to be detected by the algorithm along with noise. The iteration repeats K times (as we will show, with high probability DF-OMP never detects the same active user twice), and finally the active user set is given by $\hat{\mathcal{I}} = \hat{\mathcal{I}}^{(K)}$ with the symbol vector $\hat{b}_n = b_n^{(K)}$, $n = 1, \cdots, N$.

The RDDF detector computes the inner product between the received signal vector and the signature waveform vector $\mathbf{a}_n^H \mathbf{v}^{(k)}$, $n = 1, \cdots, N$, for $k = 1, \cdots, K$. This requires KMN floating point operations when $\mathbf{A}$ is real ($2KMN$ operations when $\mathbf{A}$ is complex) for detection of $N \log_2 3$ bits. Hence the complexity-per-bit is proportional to KM. Since $M \leq N$, this implies that the complexity for data detection of the RDDF detector is on the same order as that of the conventional DF detector (the complexity-per-bit of the DF detector is proportional to KN). But the RDDF detector requires much lower decorrelation complexity in the analog front-end than the conventional DF detector.

3) Noise whitening transform: The noise in the RD-MUD output (19) is in general colored due to the matched filtering at the front-end. We can whiten the noise at the front-end output by applying a linear transform before detecting active users and symbols, as illustrated in Fig. 3. The linear transform to whiten noise in the RD-MUD output is given by $(\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1/2}$, and the whitened output is given by:

$$\mathbf{y}_w \triangleq (\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1/2}\mathbf{y} = (\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1/2}\mathbf{A}\mathbf{R}\mathbf{b} + \mathbf{w}_0, \tag{29}$$

where $\mathbf{w}_0$ is a Gaussian random vector with zero mean and covariance matrix $\sigma^2\mathbf{I}$. If we define a new measurement matrix

$$\mathbf{A}_w \triangleq (\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1/2}\mathbf{A}, \tag{30}$$

then the RDD and RDDF detectors work with the whitened output (29) if we replace $\mathbf{A}$ with $\mathbf{A}_w$ and $\mathbf{y}$ with $\mathbf{y}_w$ in (22), (23), (26) and (27). While whitening the noise in RD-MUD front-end output, the noise whitening transform also distorts the signal component. As we will demonstrate via numerical examples in Section V-B2, the benefits of noise whitening exceed the impact of the corresponding signal detection only when the signature waveforms $\{s_n(t)\}$ are highly correlated. Since this is typically not the case in multiuser systems due to the interference between users that coexists [3], our analysis will focus on detectors without noise whitening, and the benefits of omitting the noise whitening demonstrated in our numerical results.

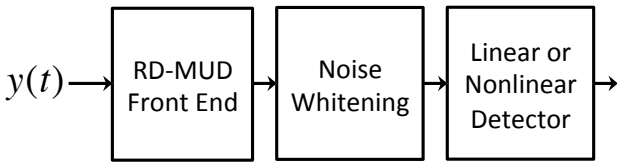


Fig. 3: The diagram of a MUD detector with the noise whitening transform.

4) Other RD-MUD linear detectors: As we discussed in Section I, various linear detectors have been developed for the MF-bank front-end output in the conventional MUD setting. In this section we explore some alternate linear detectors other than the decorrelating detector in the reduced-dimension setting. In particular, we will examine the reduced-dimension MMSE (RD-MMSE) and the reduced-dimension least-squares (RD-LS) detectors.

Reduced-dimension MMSE (RD-MMSE) detector:

Similar to the MMSE detector of the conventional MUD, a linear detector based on the MMSE criterion can be derived for (20) based on the active user set estimate $\hat{\mathcal{I}}$ given by (22). In other words, we want to find a linear transform $\mathbf{M}$ that minimizes $\mathbb{E}\{\|\mathbf{b}_{\hat{\mathcal{I}}} - \mathbf{M}\mathbf{y}\|^2\}$, where the expectation is with respect to the vector of transmitted symbols $\mathbf{b}_{\hat{\mathcal{I}}}$ and the noise vector $\mathbf{w}$. Following the approach for deriving the conventional MMSE detector [3], we assume that $\mathbf{b}_{\hat{\mathcal{I}}}$ has a distribution that is uncorrelated with the noise $\mathbf{w}$ and $\mathbb{E}\{\mathbf{b}_{\hat{\mathcal{I}}}\mathbf{b}_{\hat{\mathcal{I}}}^H\} = \mathbf{I}$. Adapting the techniques for deriving the conventional MMSE detector [3] and taking into account the noise covariance matrix $\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H$ of the reduced-dimension model (20), we obtain the linear transform for the reduced-dimension MMSE (RD-MMSE) detector as:

$$\mathbf{M} = \mathbf{R}_{\hat{\mathcal{I}}}\mathbf{A}_{\hat{\mathcal{I}}}^H(\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}^2\mathbf{A}_{\hat{\mathcal{I}}}^H + \sigma^2\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1}. \tag{31}$$

The derivations are given Appendix B. Accordingly, the RD-MMSE detector determines symbols as:

$$\hat{b}_n = \begin{cases} \mathbf{sgn}([\mathbf{R}_{\hat{\mathcal{I}}}\mathbf{A}_{\hat{\mathcal{I}}}^H(\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}^2\mathbf{A}_{\hat{\mathcal{I}}}^H + \sigma^2\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1}\mathbf{y}]_n), & n \in \hat{\mathcal{I}}; \\ 0, & n \notin \hat{\mathcal{I}}. \end{cases} \tag{32}$$

In summary, the RD-MMSE detector determines active users first through the support recovery method of (22) and then uses (32) to detect symbols.

Reduced-dimension least squares (RD-LS) detector:

In the reduced-dimension model (20), based on the active user set estimate $\hat{\mathcal{I}}$ given by (22), the matrix $\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}$ introduces interference when we detect $\hat{\mathbf{b}}_{\hat{\mathcal{I}}}$. From the view of the system of linear equations, (20)

is an over-determined system with more equations than unknowns since in general we require the number of branches to be greater than the number of active users $M > K$. Hence we cannot directly invert the matrix $\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}$ to remove interference, mimicking the idea of the decorrelating detector of conventional MUD. Alternately, we can alleviate the effect of interference using the method of least-squares (LS) to find an estimate of $\hat{\mathbf{b}}_{\hat{\mathcal{I}}}$: $\hat{\mathbf{b}}_{\hat{\mathcal{I}}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}\mathbf{x}\|^2$, and detecting symbols as the signs of the estimated vector. The solution is given by $\hat{\mathbf{b}}_{\hat{\mathcal{I}}} = \mathbf{R}_{\hat{\mathcal{I}}}^{-1}(\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{A}_{\hat{\mathcal{I}}})^{-1}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}$. This corresponds to the maximum likelihood estimate of $\hat{\mathbf{b}}_{\hat{\mathcal{I}}}$ if (a) we ignore the covariance of the noise $\mathbf{w}$ in (20) and assume it is white; (b) we know the active users, i.e. $\hat{\mathcal{I}} = \mathcal{I}$, and (c) the gains r_n are known. We call this the reduced-dimension least squares (RD-LS) detector. In summary, the RD-LS detector first detects active users by support recovery method of (22). Since $\mathbf{sgn}([\hat{\mathbf{b}}_{\hat{\mathcal{I}}}]_n) = \mathbf{sgn}([\mathbf{R}_{\hat{\mathcal{I}}}^2\hat{\mathbf{b}}_{\hat{\mathcal{I}}}]_n)$, the RD-LS detects symbols by:

$$\hat{b}_n = \begin{cases} \mathbf{sgn}\left(r_n \Re\left[(\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{A}_{\hat{\mathcal{I}}})^{-1}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}\right]_n\right), & n \in \hat{\mathcal{I}}; \\ 0, & n \notin \hat{\mathcal{I}}. \end{cases} \tag{33}$$

We can show that the RD-LS detector (33) and the RDD detector (23) give quite similar results in low noise and with low coherence of $\mathbf{A}$. To see this, write $\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{A}_{\hat{\mathcal{I}}} = \mathbf{I} + \mathbf{E}$, where the symmetric matrix $\mathbf{E}$ has zeros on the diagonal and the off-diagonals are bounded by the coherence μ of $\mathbf{A}$. As discussed in more detail in Section III-C, for the RDD detector to work well, we choose $\mathbf{A}$ with small μ. When $(K-1)\mu < 1$, by Gershgorin's Theorem, we have that the spectral norm of the symmetric matrix $\mathbf{E}$ is bounded by $\rho(\mathbf{E}) \leq (K-1)\mu < 1$. Hence, using Lemma 4 in [70], we can write $(\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{A}_{\hat{\mathcal{I}}})^{-1} = \mathbf{I} + \sum_{n=1}^{\infty}(-\mathbf{E})^n$, and bound the spectral norm of the series by a small number: $\rho\left(\sum_{n=1}^{\infty}(-\mathbf{E})^n\right) \leq (K-1)\mu/[1-(K-1)\mu]$. This means

$$\left\|(\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{A}_{\hat{\mathcal{I}}})^{-1}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y} - \mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}\right\| = \left\|\left[\sum_{n=1}^{\infty}(-\mathbf{E})^n\right]\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}\right\| \leq (K-1)\mu/[1-(K-1)\mu]\left\|\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}\right\|. \tag{34}$$

When the coherence μ of $\mathbf{A}$ is sufficiently small relative to K and $\left\|\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}\right\|$, (34) says that the difference between the two vectors $(\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{A}_{\hat{\mathcal{I}}})^{-1}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}$ and $\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}$ is small. Then when μ is small and with sufficiently small noise, detecting symbols using $(\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{A}_{\hat{\mathcal{I}}})^{-1}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}$ in (33) is similar to detecting symbols using $\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{y}$ in (23). As numerically shown in Section V-B3, the conditional probability of detecting wrong symbols given the correct support of active users, i.e. $P(\hat{\mathbf{b}} \neq \mathbf{b}|\hat{\mathcal{I}} = \mathcal{I})$, for (33) is similar to that for (23).

5) Maximum likelihood detector: The optimal detector that minimizes the probability-of-error for the RD-MUD output is the nonlinear maximum likelihood detector. The maximum likelihood detector

finds the active users and symbols by minimizing the likelihood function, or, equivalently, minimizing the quadratic function $\|(\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1/2}(\mathbf{y}-\mathbf{A}\mathbf{R}\mathbf{b})\|^2$. This is equivalent to solving the following integer optimization problem

$$\max_{b_n\in\{-1,0,1\}} 2\mathbf{y}^H(\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1}\mathbf{A}\mathbf{R}\mathbf{b}-\mathbf{b}^H\mathbf{R}\mathbf{A}^H(\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)^{-1}\mathbf{A}\mathbf{R}\mathbf{b}, \tag{35}$$

where $b_n = 0$ corresponds to the nth user being inactive and this creates an augmented state space with one more state of possible transmitted symbols corresponding to a null symbol "0". Hence (35) is more complex than the conventional maximum likelihood detector for MUD with BPSK modulation (10) since we add the "0" symbol. Similar to the conventional maximum likelihood detector of the conventional MF-bank, the maximization in (35) is a combinatorial optimization problem, which can be solved by exhaustive search with complexity-per-bit exponential in the number of users.

C. *Choice of* $\mathbf{A}$

In Section III-A we have shown that the coefficient matrix $\mathbf{A}$ is our design parameter. In Section III-B1 and Section III-B2 we have shown that both the RDD and RDDF detectors are based on the inner products between the projected received signal vector and the columns of $\mathbf{A}$, which correspond to the signature waveform vectors in the detection subspace. Hence, intuitively, for the RDD and RDDF detectors to work well, the inner products between columns of $\mathbf{A}$, or its coherence defined in (14), should be small, since each column of $\mathbf{A}$ represent a signature waveform vector in the detection subspace. Several commonly used random matrices in compressed sensing that have small coherence with high probability are:

(1) Gaussian random matrices: entries a_{nm} are independent and identically distributed (i.i.d.) with a zero mean and unit variance Gaussian distribution, with columns normalized to have unit norm;

(2) Randomly sampled rows of an unitary matrix that satisfies $\mathbf{X}\mathbf{X}^H = \mathbf{X}^H\mathbf{X} = \mathbf{I}$. For instance, the random partial discrete Fourier transform (DFT) matrix, which is formed by randomly selecting rows of a DFT matrix $\mathbf{F}$: $[\mathbf{F}]_{nm} = e^{i\frac{2\pi}{N}nm}$ and normalizing the columns of the sub-matrix, where $i = \sqrt{-1}$.

We will focus on the random partial DFT matrix for the following reason. If we choose the number of correlators equal to the number of users, i.e. $M = N$, there is no dimension reduction, and the performance of RD-MUD should equal that of the MF-bank. When $M = N$, the random partial DFT

matrix becomes the DFT matrix with the property that $\mathbf{A}^H\mathbf{A} = \mathbf{I}$, i.e, $\mathbf{a}_n^H\mathbf{a}_m = \delta_{nm}$. Consequently, in this case, $\{\mathbf{a}_n^H\mathbf{y}\}$, which is a set of statistics that the RDD and RDDF detectors are based on, has the same distribution as the MF-bank output. To see this, write $\mathbf{a}_n^H\mathbf{y} = \mathbf{a}_n^H\left(\sum_{m=1}^N \mathbf{a}_m r_m b_m\right) + \mathbf{a}_n^H\mathbf{w} = r_n b_n + \mathbf{a}_n^H\mathbf{w}$, where $\mathbf{a}_n^H\mathbf{w}$ is a Gaussian random variable with zero mean and covariance $\sigma^2\mathbf{a}_n^H\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H\mathbf{a}_m = [\mathbf{G}^{-1}]_{nm}$. However, the Gaussian random matrix does not have this property: when $M = N$, $\mathbf{a}_n^H\mathbf{a}_m \neq 0$ for $n \neq m$, and so the performance of RD-MUD using the Gaussian random matrix $\mathbf{A}$ is worse than that using the random partial DFT matrix. This has been validated in our numerical results in Section V-A3 where we will show that when M is relatively large, the Gaussian random matrix performs worse than the random partial DFT matrix.

IV. Performance of RD-MUD

In the following, we study the performance of RD-MUD with the RDD and RDDF detectors. We begin by considering the scenario of a single active user without noise, and then move on to analyze the more general scenario with multiple active users and noise.

A. Single Active User

The following discussion shows that, when there is only one active user in the absence of noise, the RDD detector can detect the correct active user and symbol by using only *two* correlators, if the columns of $\mathbf{A}$ are linearly independent. Later we will also show this is a corollary (Corollary 2) from the more general Theorem 1.

Assume there is no noise and only one user with index n_0 is active. In this case $y(t) = r_{n_0}b_{n_0}s_{n_0}(t)$, and by assumption we know only one user is active, i.e. we know $K = 1$. In the conventional MUD, the single-user detector based on the MF-bank detects the active user by finding $\hat{n}_0 = \arg\max_n |\langle y(t), s_n(t)\rangle|$ and the symbol by $\hat{b}_{\hat{n}_0} = \mathbf{sgn}(r_{\hat{n}_0}\langle y(t), s_{\hat{n}_0}(t)\rangle)$. From the Cauchy-Schwarz inequality, for any n,

$$|\langle y(t), s_n(t)\rangle| = |\langle r_{n_0}b_{n_0}s_{n_0}(t), s_n(t)\rangle| \leq |r_{n_0}|\|s_{n_0}(t)\|\|s_n(t)\| = |r_{n_0}|, \tag{36}$$

with equality if and only if $s_n(t) = cs_{n_0}(t)$ for some constant c, and thus $\hat{n}_0 = n_0$. The symbol can also be recovered perfectly, since

$$\hat{b}_{n_0} = \mathbf{sgn}(r_{n_0}\langle y(t), s_{n_0}(t)\rangle) = \mathbf{sgn}(r_{n_0}^2 b_{n_0}\langle s_{n_0}(t), s_{n_0}(t)\rangle) = b_{n_0}. \tag{37}$$

In RD-MUD, with two correlators, the RDD detector determines the active user by finding

$$\hat{n}_0 = \arg \max_{n=1,\cdots,N} |a_{1n}\langle h_1(t), y(t)\rangle + a_{2n}\langle h_2(t), y(t)\rangle|. \tag{38}$$

From the Cauchy-Schwarz inequality,

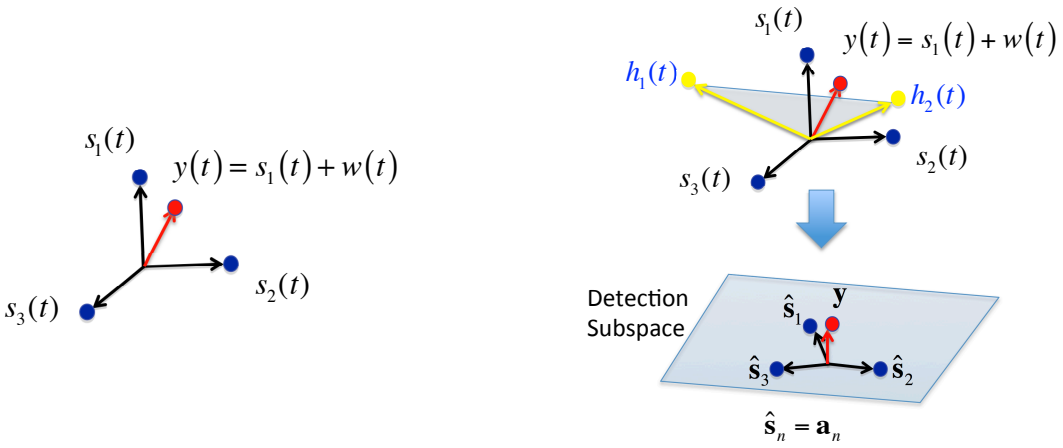
$$|a_{1n}\langle h_1(t), y(t)\rangle + a_{2n}\langle h_2(t), y(t)\rangle|^2 \leq (a_{1n}^2 + a_{2n}^2)\left[\langle h_1(t), y(t)\rangle^2 + \langle h_2(t), y(t)\rangle^2\right], \tag{39}$$

with equality if and only if $a_{mn} = c\langle h_m(t), y(t)\rangle = ca_{mn_0}r_{n_0}b_{n_0} = c(n_0)a_{mn_0}$ for both $m = 1, 2$ with some constant $c(n_0)$. If the columns of $\mathbf{A}$ are linearly independent, we cannot have two indices n such that $a_{mn} = c(n_0)a_{mn_0}$ for $m = 1, 2$. Also recall that the columns of $\mathbf{A}$ are normalized, $a_{1n}^2 + a_{2n}^2 = \|\mathbf{a}_n\|^2 = 1$. Therefore, the maximum is achieved only for $n = n_0$ and $c(n_0) = 1$, which detects the correct active user. The detected symbol is also correct, since

$$\hat{b}_{n_0} = \mathbf{sgn}(r_{n_0}[a_{1n_0}\langle y(t), h_1(t)\rangle + a_{2n_0}\langle y(t), h_2(t)\rangle]) = \mathbf{sgn}(r_{n_0}^2 b_{n_0}[a_{1n_0}^2 + a_{2n_0}^2]) = b_{n_0}. \tag{40}$$

In the presence of noise, detectors in RD-MUD as well as those based on the conventional MF-bank will make detection errors. However, RD-MUD can have a performance similar to the detectors based on the MF-bank, as we now explain using a geometric intuition and later prove formally in Section IV-C. Consider a scenario with three users having orthogonal signature waveforms, where only the first user is active. Suppose that $b_1 = 1$ and $r_1 = 1$, as illustrated in Fig. 4. By correlating with signature waveforms, the MF-bank (Figure 4a) obtains inner products (5) of the received signal $y(t)$ with each of the signature waveforms $\{s_n(t)\}$ and detects based on these inner products. Because of noise, in Fig. 4a the received signal $y(t)$ does not coincide with $s_1(t)$. However, when the noise is sufficiently small, as shown in Fig. 4a, the inner product of the received signal with the first signature waveform is the largest and positive, and hence in this scenario the single-user detector based on the conventional MF-bank detects the correct active user and its symbol. On the other hand, for the same setting, the RDD detector with two correlators (Fig. 4b) projects the received signal onto the detection subspace via (17), and then obtains the decision statistics by computing the inner product between $\mathbf{y}$, the mapping of the received signal onto the subspace, and $\mathbf{a}_n$, the projection of each signature waveform onto the subspace. Because of noise, in Fig. 4b the projected signal vector $\mathbf{y}$ does not coincide with the first signature waveform vector $\mathbf{a}_1$. However, when the noise is sufficiently small, as shown in Fig. 4b, the inner product with the first signature vector is still the largest and positive, and hence in this scenario

the RDD detector also detects the correct active user and symbol. From the above discussion we see that the RDD detector works well when the columns of $\mathbf{A}$ are nearly orthogonal and the noise is sufficiently small. The former requirement is equivalent to requiring the coherence of $\mathbf{A}$ to be as small as possible. The above discussion applies to the RDDF detector as well since its detection is also based on the inner products in the projection space.



(a) MF Bank (b) RD-MUD

Fig. 4: A MUD problem with $N = 3$, $M = 2$, $K = 1$ and the received signal is due to the first user. Illustration of the detection by (a) the single-user detection of the MF-bank, and (b) the RD-MUD. The projection of the signature waveform onto the projection subspace results in signature waveform vectors $\{\hat{\mathbf{s}}_n\}$.

B. Noise Amplification of Subspace Projection

The RDD and RDDF detectors use the set of statistics $\{\mathbf{a}_n^H \mathbf{y}\}$ to detect active users and their symbols, which has noise components $\{\mathbf{a}_n^H \mathbf{w}\}$. We will show that the projection onto the detection subspace amplifies noise. To see this, first consider a special case with orthogonal signature waveforms, i.e. $\mathbf{G} = \mathbf{I}$, and hence the noise amplification is not caused by correlated signature waveforms. Assume the random partial DFT matrix is used as the coefficient matrix $\mathbf{A}$. Using (19), the noise variance of the detection statistic for the nth user is given by $\sigma^2 \mathbf{a}_n^H \mathbf{A}\mathbf{A}^H \mathbf{a}_n = \sigma^2 (N/M)$ by the definition of the random partial DFT matrix in Section III-C. Hence in this special case the noise variance for each user

is amplified by a factor $N/M \geq 1$ due to subspace projection.

In general, from (19) the detection statistic for the nth user has a noise variance $\sigma^2 \mathbf{a}_n^H \mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H \mathbf{a}_n$. We will show this noise variance can be greater than σ^2. First we bound the output noise variance of the RD-MUD for each user as

$$\sigma^2 \lambda_{\min}(\mathbf{G}^{-1}) \|\mathbf{A}^H \mathbf{a}_n\|^2 \leq \sigma^2 \mathbf{a}_n^H \mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H \mathbf{a}_n \leq \sigma^2 \lambda_{\max}(\mathbf{G}^{-1}) \|\mathbf{A}^H \mathbf{a}_n\|^2. \tag{41}$$

We now show that the upper bound in (41) is greater than σ^2. The factor $\lambda_{\max}(\mathbf{G}^{-1})$ captures the largest possible noise amplification due to correlated signature waveforms. We have $\lambda_{\max}(\mathbf{G}^{-1}) \geq 1$ for the following. The sum of all the eigenvalues $\sum_{k=1}^{N} \lambda_k(\mathbf{G}) = \mathbf{tr}(\mathbf{G}) = N$ since all the diagonal entries of $\mathbf{G}$ are one. Since all eigenvalues of $\mathbf{G}$ are nonnegative (a property of $\mathbf{G}$ is that it is positive semi-definite), if $\lambda_{\min}(\mathbf{G}) > 1$, the sum of the eigenvalues will exceed N, and hence $\lambda_{\min}(\mathbf{G}) \leq 1$. Since the eigenvalues of $\mathbf{G}^{-1}$ are the inverse of the eigenvalues of $\mathbf{G}$, $\lambda_{\max}(\mathbf{G}^{-1}) \geq 1$. On the other hand, the factor $\|\mathbf{A}^H \mathbf{a}_n\|^2$ captures the noise amplification due to subspace projection. It is lower-bounded by 1: $\|\mathbf{A}^H \mathbf{a}_n\|^2 = 1 + \sum_{l=1, l\neq n}^{N} (\mathbf{a}_l^H \mathbf{a}_n)^2 \geq 1 + \min_{l\neq n} |\mathbf{a}_l^H \mathbf{a}_n|^2$. When $M < N$, the inner product $\min_{l\neq n} |\mathbf{a}_l^H \mathbf{a}_n|$ is non-zero, and hence the factor $\|\mathbf{A}^H \mathbf{a}_n\|^2$ will be strictly greater than one. As a result, the upper bound on the noise variance in (41) is greater than σ^2.

In the following section, we will capture this noise amplification more precisely by relating the noise variance of the decision statistic to the performance of the RD-MUD detectors.

C. Coherence Based Performance Guarantee

In this section, we present conditions under which the RDD and RDDF detectors can successfully recover active users and their symbols. The conditions depend on $\mathbf{A}$ through its coherence and are parameterized by the crosscorrelations of the signature waveform through the properties of the matrix $\mathbf{G}$. Our performance measure is the probability-of-error, which is defined as the chance of the event that the set of active users is detected incorrectly, *or* any of their symbols are detected incorrectly:

$$P_e = P(\hat{\mathcal{I}} \neq \mathcal{I}) + P(\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} \neq \mathbf{b}\}). \tag{42}$$

We will show in the later section that the second term of (42) is dominated by the first term when (22) and (26) are used for active user detection. The noise plays two roles in the P_e of (42). First, the noise can be sufficiently large relative to the weakest signal such that a nonactive user is determined as active;

second, the noise can be sufficiently large such that the transmitted symbol plus noise is detected in an incorrect decision region and hence decoded in error.

The first error term in (42) is related to the probability-of-error for support recovery (see, e.g. [55] in noise, and the first steps of the greedy algorithms to recover support [65]). There are two major differences in our results on this aspect of RD-MUD performance relative to those previous works. First, although noise in the analog signal model (4) is white, matched filtering at the RD-MUD front-end introduces colored noise in (19). Second, we take into account the second term in (42), which has not been considered in previous work. We find the conditions such that the second term of (42) is dominated by the first term of (42).

Define the largest and smallest channel gains as

$$|r_{\max}| \triangleq \max_{n=1}^{N} |r_n|, \quad |r_{\min}| \triangleq \min_{n=1}^{N} |r_n|. \tag{43}$$

Our main result is the following theorem:

Theorem 1. *Let* $\boldsymbol{b} \in \mathbb{R}^{N\times 1}$ *be an unknown deterministic symbol,* $b_n \in \{-1, 1\}$, $n \in \mathcal{I}$, *and* $b_n = 0$, $n \in \mathcal{I}^c$, $n = 1, \cdots, N$. *Assume that the number of active users* K *is known. Given the RD-MUD front-end output* $\boldsymbol{y} = \boldsymbol{ARb} + \boldsymbol{w}$, *where* $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ *and* $\boldsymbol{G} \in \mathbb{R}^{N\times N}$ *are known, and* $\boldsymbol{w}$ *is a Gaussian random vector with zero mean and covariance* $\sigma^2\boldsymbol{AG}^{-1}\boldsymbol{A}^H$, *if the columns of* $\boldsymbol{A}$ *are linearly independent and the coherence of* $\boldsymbol{A}$ *(14) satisfies the following condition:*

$$|r_{\min}| - (2K-1)\mu|r_{\max}| \geq 2\sigma\sqrt{2(1+\alpha)\log N} \cdot \sqrt{\lambda_{\max}(\boldsymbol{G}^{-1})} \cdot \sqrt{\max_n \left(\boldsymbol{a}_n^H \boldsymbol{A}\boldsymbol{A}^H \boldsymbol{a}_n\right)}, \tag{44}$$

for some constant $\alpha > 0$, *and* $N^{-(1+\alpha)}[\pi(1+\alpha)\log N]^{-1/2} \leq 1$, *then the probability-of-error (42) for the RDD detector is upper bounded as:*

$$P_e \leq N^{-\alpha}[\pi(1+\alpha)\log N]^{-1/2}. \tag{45}$$

If the columns of $\boldsymbol{A}$ *are linearly independent and the coherence of* $\boldsymbol{A}$ *(14) satisfies a weaker condition:*

$$|r_{\min}| - (2K-1)\mu|r_{\min}| \geq 2\sigma\sqrt{2(1+\alpha)\log N} \cdot \sqrt{\lambda_{\max}(\boldsymbol{G}^{-1})} \cdot \sqrt{\max_n \left(\boldsymbol{a}_n^H \boldsymbol{A}\boldsymbol{A}^H \boldsymbol{a}_n\right)}, \tag{46}$$

for some constant $\alpha > 0$, *and* $N^{-(1+\alpha)}[\pi(1+\alpha)\log N]^{-1/2} \leq 1$, *then the probability-of-error (42) for the RDDF detector is upper bounded by the right hand side of (45).*

Proof: See Appendix C. ■

Note in Theorem 1 that the condition of having a small probability-of-error for the RDDF detector is weaker than for the RDD detector. Intuitively, the iterative approach of decision feedback removes the effect of the largest element in **Rb** iteratively, which helps the detection of weaker users. This is an extension of the ideas in SIC for standard MUD, except that now the presence as well as the data of the strongest user is detected, then removed in the detection subspace, which makes it easier to detect the presence and data of the next strongest user, after which the process repeats.

The main idea of the proof is the following. Consider $1 - P_e = P(\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} = \mathbf{b}\})$. First we define an event $\mathcal{G} = \{\max_n |\mathbf{a}_n^H \mathbf{w}| < \tau\}$ for a quantity τ proportional to the right hand side in (44), and prove that $\mathcal{G}$ occurs with high probability. This bounds the probability that the noise projected onto the detection subspace exceeds τ, i.e. it bounds the tail probability of the projected noise. Then we show that under the condition (44), whenever $\mathcal{G}$ occurs, the active users can be correctly detected, which means $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\}$. On the other hand, we show that under a condition weaker than (44), whenever $\mathcal{G}$ occurs, the user data symbols can be correctly detected, which means $\mathcal{G} \subset \{\hat{b}_n = b_n, n \in \mathcal{I}\}$. In other words, condition (44) ensures that whenever $\mathcal{G}$ occurs, both the correct set of active users are detected and that their data are correctly decoded. This means that under condition (44) for the RDD detector, $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} = \mathbf{b}\}$, and thus $P(\mathcal{G}) \leq P(\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} = \mathbf{b}\})$, which concludes the proof. A similar but inductive approach is used to prove the performance guarantee for the RDDF detector.

A special case for Theorem 1 is when $\mathbf{A}\mathbf{A}^H = (N/M)\mathbf{I}$, $\max_n(\mathbf{a}_n^H \mathbf{A}\mathbf{A}^H \mathbf{a}_n) = N/M$ and $\mathbf{G} = \mathbf{I}$, $\lambda_{\max}(\mathbf{G}^{-1}) = 1$. This is true when $\mathbf{A}$ is the random partial DFT matrix and the signature waveforms are orthogonal, and hence the noise in (19) is white. If we scale σ^2 by M/N, the right hand sides of (44) and (46) are then identical to the corresponding quantities in Theorem 4 of [65]. Hence, for the random partial DFT matrix for $\mathbf{A}$, Theorem 1 has the same conditions as those of Theorem 4 in [65]. However, Theorem 4 in [65] guarantees detecting the correct sparsity pattern of $\mathbf{b}$ (equivalently, the correct active users), whereas Theorem 1 guarantees correct detection of not only the active users but their symbols as well. That is because, as mentioned above, correct detection of these transmitted symbols comes "for free" when the conditions to correctly detect the active users are met.

Remarks:

The term $\max_n(\mathbf{a}_n^H \mathbf{A}\mathbf{A}^H \mathbf{a}_n)$ on the right hand side of (44) and (46) is bounded by

$$1 \leq \max_n(\mathbf{a}_n^H \mathbf{A}\mathbf{A}^H \mathbf{a}_n) \leq 1 + (N-1)\mu^2. \tag{47}$$

Equation (47) follows because $\max_n \left(\mathbf{a}_n^H \mathbf{A}\mathbf{A}^H \mathbf{a}_n\right) = \max_n \sum_{l=1}^N (\mathbf{a}_n^H \mathbf{a}_l)^2$, and

$$1 = (\mathbf{a}_n^H \mathbf{a}_n)^2 \leq \max_n \sum_{l=1}^N (\mathbf{a}_n^H \mathbf{a}_l)^2 = 1 + \max_n \sum_{l \neq n} (\mathbf{a}_n^H \mathbf{a}_l)^2 \leq 1 + (N-1)\mu^2. \tag{48}$$

On the other hand, there is a noise phase-transition effect, in the following sense. Conditions (44) and (46) suggest that for the RDD and RDDF detectors to have P_e as small as (45), we need to have

$$|r_{\min}|^2/\sigma^2 > 8 \log N \lambda_{\max}(\mathbf{G}^{-1}), \tag{49}$$

because $\alpha > 0$ and (47) holds. If the minimum SNR, i.e. the SNR associated with the minimum gain $r_{\min}$ is not sufficiently high, these algorithms cannot attain small probability-of-error. We illustrate this effect via numerical examples in Section V-B1 (a similar effect can be observed in standard MUD detectors using linear or DF detection).

D. Bounding probability-of-error of RDD and RDDF

Theorem 1 provides a condition on how small μ has to be to achieve a small probability-of-error. The condition and the achievable small probability-of-error are related by the constant α. We can eliminate this constant and write Theorem 1 in an equivalent form that gives error bounds for the RDD and RDDF detectors explicitly. Define the minimum signal-to-noise ratio (SNR) in the projection subspace as

$$\mathrm{SNR}_{\min} = \frac{|r_{\min}|^2}{\sigma^2 \lambda_{\max}(\mathbf{G}^{-1})}, \tag{50}$$

where the factor $\lambda_{\max}(\mathbf{G}^{-1})$ captures the noise amplification effect in the projection due to nonorthogonal signature waveforms. Also define two factors β_1 and β_2 as

$$\beta_1 \triangleq \frac{[1-(2K-1)\mu|r_{\max}|/|r_{\min}|]^2}{\max_n(\mathbf{a}_n^H \mathbf{A}\mathbf{A}^H \mathbf{a}_n)}, \qquad \beta_2 \triangleq \frac{[1-(2K-1)\mu]^2}{\max_n(\mathbf{a}_n^H \mathbf{A}\mathbf{A}^H \mathbf{a}_n)}. \tag{51}$$

For the RDD detector, we have already implicitly assumed that $1-(2K-1)\mu|r_{\max}|/|r_{\min}| \geq 0$, since the right hand side of (44) in Theorem 1 is non-negative. For the same reason, for the RDDF detector, we have assumed that $1-(2K-1)\mu > 0$. By (47) and (51), $\beta_1 \leq 1$ and $\beta_2 \leq 1$. We have the following corollary from Theorem 1:

Corollary 1. *Under the setting of Theorem 1, with the definitions (50) and (51), the probability-of-error*

for the RDD detector is upper-bounded by

$$P_{e,\mathrm{RDD}} \leq \frac{2N}{\sqrt{\pi}} \left[\frac{\mathrm{SNR}_{\min}}{2} \cdot \beta_1 \right]^{-1/2} e^{-\frac{1}{4}\frac{\mathrm{SNR}_{\min}}{2}\cdot\beta_1}, \tag{52}$$

with $1-(2K-1)\mu|r_{\max}|/|r_{\min}| \geq 0$*, and the probability-of-error for the RDDF detector is upper bounded by*

$$P_{e,\mathrm{RDDF}} \leq \frac{2N}{\sqrt{\pi}} \left[\frac{\mathrm{SNR}_{\min}}{2} \cdot \beta_2 \right]^{-1/2} \cdot e^{-\frac{1}{4}\frac{\mathrm{SNR}_{\min}}{2}\beta_2}, \tag{53}$$

with $1-(2K-1)\mu > 0$.

Proof: We begin by bounding the probability-of-error of the RDD detector from Theorem 1. Under condition (44), by (45) the probability-of-error of the RDD detector is bounded by $N^{-\alpha}[\pi \log(N^{1+\alpha})]^{-1/2}$ for some constant $\alpha > 0$. To make the bound tight, we choose α as large as possible such that it still satisfies (44). With definitions (50) and (51), for the RDD detector, we can rewrite (44) as

$$[\log N^{(1+\alpha)}]^{1/2} \leq \frac{1}{2} \cdot \left[\frac{\mathrm{SNR}_{\min}}{2} \cdot \beta_1 \right]^{1/2}, \tag{54}$$

or equivalently

$$N^{\alpha} \leq N^{-1} e^{\frac{1}{4}\frac{\mathrm{SNR}_{\min}}{2}\cdot\beta_1}. \tag{55}$$

The right hand sides of (54) and (55) are the largest values for $[\log N^{(1+\alpha)}]^{1/2}$ and N^{α} we can obtain under (44), for given $\mathrm{SNR}_{\min}$, $\mathbf{A}$, K and N. Combining (54) and (55) in (45), we have (52). Similarly, by choosing the largest possible α satisfying (46), we derive the bound (53) on the probability-of-error of the RDDF detector. ■

Remarks:

For bounds (52) and (53) to be meaningful, they have to be less than one, and hence $\mathrm{SNR}_{\min}$ should be on the order of $\log N$. Also note that the error bounds (52) and (53) for the RD-MUD detectors are larger when the signature waveforms are correlated, since these error bounds increase in $\mathrm{SNR}_{\min}$, and $\mathrm{SNR}_{\min}$ decreases in $\lambda_{\max}(\mathbf{G}^{-1})$. This implies that the performance of the RD-MUD detectors tend to perform worse when signature waveforms are nonorthogonal (the same effect can also be observed for the conventional MUD detectors based on the MF-bank).

By letting the noise variance σ^2 go to zero in (52) and (53) for the RDD and RDDF detectors, we can derive the following corollary from Theorem 1 (another proof for the RDD detector in this case has been given in Section IV-A).

Corollary 2. *Under the setting of Theorem 1, in the absence of noise, the RDD detector can correctly detect the active users and their symbols if* $\mu < |r_{\min}|/[|r_{\max}|(2K-1)]$*, and the RDDF detector can correctly detect the active users and their symbols if* $\mu < 1/(2K-1)$*. In particular, if* $K = 1$*, with* $M = 2$ *correlators,* $P_e = 0$ *for the RDDF detector, and if furthermore* $|r_{\max}| = |r_{\min}|$*,* $P_e = 0$ *for the RDD detector (which has also been shown in Section IV-A).*

Proof: In Theorem 1, if we let σ^2 go to zero, then $\mathrm{SNR}_{\min}$ goes to infinity, and the right hand sides of both (52) and (53) go to zero, i.e. $P_e = 0$, as long as $\beta_1 > 0$ and $\beta_2 > 0$, or equivalently, $1-(2K-1)|r_{\max}|/|r_{\min}| > 0$ and $1-(2K-1)\mu > 0$. When $K = 1$, the bound on μ for the RDDF detector becomes $1/(2K-1) = 1$, which is satisfied for any μ as long as $M \geq 2$ (since Theorem 1 also requires linear independence of the columns of $\mathbf{A}$ and this rules out the possibility of $M = 1$). ■

E. Comparison with existing bounds

In this section we compare the bound on the probability-of-error in the literature for the decorrelating detector of the conventional MUD with our bound for the RDD and RDDF detectors derived from Theorem 1. The decorrelating detector is the counterpart of the RDD detector in the conventional MUD setting. To see this, note that for the RD-MUD front-end, when $M = N$, we can choose the coefficient matrix $\mathbf{A} = \mathbf{I}$ such that the output data model (19) is equivalent to the MF-bank decorrelating detector (7) (with $\mathbf{T} = \mathbf{G}^{-1}$).

For the decorrelating detector of the conventional MUD, a commonly used performance measure is the probability of error of each user [71][8], which is given by [71] [3]:

$$P(\hat{b}_n \neq b_n) = Q\left(\frac{|r_n|}{\sigma\sqrt{[\mathbf{G}^{-1}]_{nn}}}\right), \tag{56}$$

where $Q(x) = \int_x^\infty (1/\sqrt{2\pi})e^{-z^2/2}dz$ is the Gaussian tail probability. To compare (56) with the P_e bound defined by (42), which consists of both active user detection error and symbol error, we consider the case when all uses are active, i.e. $K = N$, and then P_e is only due to symbol error. In this setting, we have, using the union bound and (56):

$$P_e = P(\hat{\mathbf{b}} \neq \mathbf{b}) \leq \sum_{n=1}^{N} P(\hat{b}_n \neq b_n) \leq NQ\left(\sqrt{\mathrm{SNR}_{\min}}\right) \leq \frac{N}{2\sqrt{\pi}}\left[\frac{\mathrm{SNR}_{\min}}{2}\right]^{-1/2} e^{-\frac{\mathrm{SNR}_{\min}}{2}}, \tag{57}$$

where we have also used the fact that $|r_n|/\left[\sigma\sqrt{[\mathbf{G}^{-1}]_{nn}}\right] \geq \sqrt{\mathrm{SNR}_{\min}}$ and $Q(x)$ is decreasing in x, as

well as the bound on $Q(x)$ [3] on $Q(x)$ given by

$$Q(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}. \tag{58}$$

The bounds on P_e of the RDD and RDDF detectors when $1 \leq K \leq N$ are given in (52) and (53), respectively. Since $\beta_1 \leq 1$ and $\beta_2 \leq 1$, the error bounds (52) for the RDD detector and (53) for the RDDF detector are larger than the bound (57) for conventional MUD. This is because the RDD and RDDF detectors have one extra source of error from detecting the wrong set of active users, and also because the noise can be amplified by the projection onto the detection subspace, as discussed in Section IV-B. The enlargement of the error bound due to subspace projection is captured by factors β_1 and β_2 for the RDD and RDDF detectors, respectively. These factors reduce the effect of $\mathrm{SNR}_{\min}$ in the bounds. Note that β_1 and β_2 increase in μ, and hence we want small μ, which leads to a small error bound for RD-MUD detectors.

A special case is when $K = N$, $\mathbf{A} = \mathbf{I}$. Then $\mu = 0$, $\max_n(\mathbf{a}_n \mathbf{A}\mathbf{A}^H \mathbf{a}_n) = 1$, and by definition (51), $\beta_1 = \beta_2 = 1$. At the beginning of this section, we have shown that this correspond to the decorrelating detector of the conventional MUD. In this case the bounds (53) and (57) become the same expression

$$P_e \leq \frac{2N}{\sqrt{\pi}} \left[\frac{\mathrm{SNR}_{\min}}{2}\right]^{-1/2} e^{-\frac{1}{4}\frac{\mathrm{SNR}_{\min}}{2}}. \tag{59}$$

Compared with the bound (57) for P_e of the conventional decorrelating detector, the bound (59) obtained from our result is larger. This can be explained since (59) is obtained as a special case of RD-MUD which must also detect active users. As we have shown in the proof for Theorem 1, the error from detecting active users dominates the error from detecting symbols.

F. Lower Bound on the Number of Correlators

Theorem 1 is stated for any matrix $\mathbf{A}$. If we substitute the expression for coherence of a given $\mathbf{A}$ in terms of its dimensions M and N into Theorem 1, we can obtain a lower bound on the smallest number of correlators M needed to achieve a certain probability-of-error.

There is a general lower bound on the coherence of any $M \times N$ matrix $\mathbf{A}$ given by [72]:

$$\mu \geq \left[\frac{N-M}{M(N-1)}\right]^{1/2} = M^{-1/2} \left(\frac{N}{N-1}\right)^{1/2} \cdot (1 - M/N)^{1/2} \sim M^{-1/2}, \tag{60}$$

when N is large relative to M and N is much larger than 1. In the absence of noise, the upper bound

on the coherence in Corollary 2 together with the bound (60) imply that, for the RDDF detector to have perfect detection, the number of correlators M should be on the order of $(2K-1)^2$. In the compressed sensing literature, it is known that the bounds obtained using the coherence of the matrix $\mathbf{A}$ may not be as sharp as those obtained using the restricted isometry properties of $\mathbf{A}$ [72]. For example, in compressed sensing, to estimate a sparse vector with κ non-zero entries, the lower bound on the number of measurements obtained from the coherence of $\mathbf{A}$ is proportional to κ^2, while that obtained from the restricted isometry properties is proportional to κ. This effect is referred to as the "quadratic bottleneck" in the compressed sensing literature [72]. Nevertheless, the coherence properties are easy to evaluate, while evaluating the restricted isometry property of a given matrix $\mathbf{A}$ is in general NP-hard [65]. Also as we demonstrate in the proof of Theorem 1, the coherence is a convenient measure of the user interference level in the detection subspace. For this reason, our result is based on the coherence of matrix $\mathbf{A}$.

In the compressed sensing literature, the matrix $\mathbf{A}$ is often chosen to be random, in which case its coherence can be bounded in probability. Consider for example a random partial DFT matrix. We have the following result (which can be proven easily by the complex Hoeffding's inequality [73]):

Lemma 1. *Let $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ be a random partial DFT matrix. Then the coherence of $\boldsymbol{A}$ is bounded by*

$$\mu < [4(2\log N + c)/M]^{1/2}, \tag{61}$$

with probability exceeding $1-2e^{-c}$, for some constant $c>0$.

Using Lemma 1, we have the following corollary to Theorem 1:

Corollary 3. *Consider the setting of Theorem 1, where $\boldsymbol{A}$ is a random partial DFT matrix $\boldsymbol{A}$. Suppose the number of correlators satisfies the following lower bound for the RDD detector*

$$M \geq 4\left[\frac{(2K-1)|r_{\max}|}{|r_{\min}|-2\tau}\right]^2 (2\log N + c), \tag{62}$$

or satisfies the following smaller lower bound for the RDDF detector

$$M \geq 4\left[\frac{(2K-1)|r_{\min}|}{|r_{\min}|-2\tau}\right]^2 (2\log N + c), \tag{63}$$

for some constants $c>0$ and $\alpha>0$, and $|r_{\min}|>2\tau$, for τ defined in (70). Then the probability-of-error

P_e of the RDD detector or the RDDF detector is bounded by

$$1-(1-N^{-\alpha}[\pi(1+\alpha)\log N]^{-1/2})(1-2e^{-c}), \tag{64}$$

for some constant $\alpha > 0$.

This corollary says that to attain a small probability-of-error, the number of correlators needed by the RDD and RDDF detectors is on the order of $\log N$, which is much smaller than that required by the conventional MUD using a MF-bank, which is on the order of N.

V. Numerical Examples

As an illustration of the performance of RD-MUD, we present some numerical examples mainly of both the RDD and RDDF detectors. The results are obtained from 10^5 Monte Carlo trials. For each trial, we generate a Gaussian random noise vector as well as a random partial DFT matrix for $\mathbf{A}$, and form the signal vector according to (19). To simplify, we assume that the gains for all the users are the same: $|r_{\min}| = |r_{\max}| \triangleq r = 1$. First we consider noise-free scenarios with an increasing number of users N for a fixed number of active users K, and then with increasing K for a fixed N. Next we consider two noisy scenarios with orthogonal waveforms $\mathbf{G} = \mathbf{I}$ and nonorthogonal waveforms $\mathbf{G} \neq \mathbf{I}$.

A. Noise-Free Scenario

In the absence of noise, from (7) of the MF-bank, the conventional decorrelating detector has output $\mathbf{z} = \mathbf{Rb}$. The conventional decorrelating detector determines the active users by choosing the K largest of $|z_n|$, which is equivalent to choosing the K largest of $\{|r_n b_n|\}$ in the absence of noise. Recall that the inactive users have $r_n = 0$, which means the conventional decorrelating detector can correctly detect the active users. It then detects symbols by $\hat{b}_n = \mathbf{sgn}(r_n^2 b_n) = b_n$. For the above reasons, the conventional decorrelating detector has $P_e = 0$ in the absence of noise.

1) P_e vs. M, as N increases: Fig. 5 shows the P_e of the RDD detector as a function of M, for fixed $K = 2$, and different values of N. The data points marked on the curves correspond to $M = \log N$, $2\log N$, and $4\log N$. When $M = 8\log N$, $P_e = 0$ for the RDD detector for all values of N. This example clearly demonstrates the $\log N$ scaling factor for the required number of correlators in Corollary 3.

2) P_e vs M, as K increases: Fig. 6a demonstrates the P_e of the RDD detector as a function of M, for a fixed $N = 100$, and different values of K. The points marked on the curves correspond to

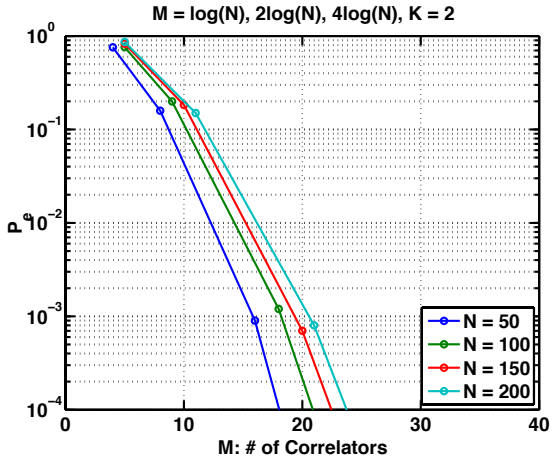
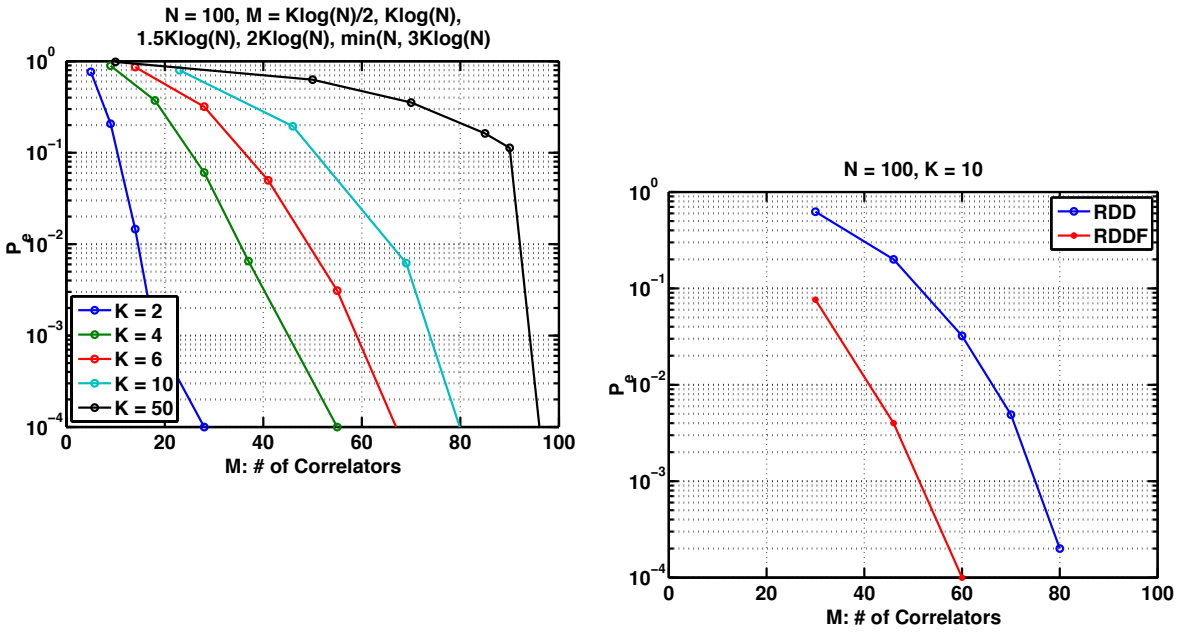


Fig. 5: In the absence of noise, P_e versus M of the RDD detector, for $K = 2$ and different N. The marked points on the curves are equal to $\log N$, $2\log N$, and $4\log N$. When $M = 8\log N$, $P_e = 0$ for the RDD detector.

$M = (K\log N)/2$, $K\log N$, $1.5K\log N$, $2K\log N$ and $\min\{N, 3K\log N\}$. Clearly, the number of correlators needed to obtain $P_e < 10^{-4}$ increases as K increases. When $K = 10$, the RDD detector needs about 80 correlators to obtain $P_e < 10^{-4}$. The number of correlators needed to achieve a small probability-of-error can be improved by using the RDDF detector. As shown in Fig. 6b, when $K = 10$, the RDDF detector uses only 60 correlators to obtain $P_e < 10^{-4}$.

Fig. 6a also demonstrates how Corollary 2 can be used to estimate the number of correlators needed to achieve a small probability-of-error. Corollary 2 says that we need $\mu < 1/(2K-1)$ in the absence of noise to have perfect detection. When $K = 2$, this requires $\mu < 1/3$. We then obtain an estimate for coherence of the random partial DFT matrix with $N = 100$ and various M, by averaging over 10^5 trials, and find that when M is about 30 the coherence is less than 1/3. This is consistent with Fig. 6a, which shows that when $M = 28$, P_e is on the order of 10^{-4}.

3) Random partial DFT vs. Gaussian random matrices: We compare the performance of the RDD detector using the random partial DFT matrix versus using the Gaussian random matrix for $\mathbf{A}$ (defined in Section III-C). In Fig. 7, the probability-of-error of the Gaussian random matrix converges to a value much higher than zero, whereas that of the random partial DFT matrix converges to zero (the value



(a) P_e vs. M for different K, RDD. (b) P_e vs. M for $K = 10$, RDD and RDDF.

Fig. 6: In the absence of noise, P_e versus M for $N = 100$, using (a): the RDD detector for different K, and (b): the RDDF detector compared with the RDD detector when $K = 10$.

achieved by the conventional decorrelating detector), when M increases to N.

B. Noisy Scenario

Next, we consider noisy scenarios. For comparison, we also consider the conventional decorrelating detector, which corresponds to the RDD detector with $M = N$ as we explained in Section IV-E.

1) P_e vs. M, as SNR increases, $\boldsymbol{G} = \boldsymbol{I}$: We study P_e versus M for the RDD detector as SNR increases when the signature waveforms are orthogonal and $\mathbf{G} = \mathbf{I}$ and hence the noise in (19) is white. In this case $\mathrm{SNR}_{\min} = |r_{\min}|^2/\sigma^2 = r^2/\sigma^2$, which is denoted as SNR in Fig. 8. Assume $N = 100$ and $K = 2$. In Fig. 8, when SNR increases, the curves converge to the noise-free curve for $K = 2$ shown in Fig. 6a, and to the noise-free curve for $N = 100$ shown in Fig. 5. Note that there is a noise phase-transition effect in Fig. 6a, which is discussed in the Remarks of Section IV-C. The analysis in (49) implies that for $N = 100$ and $\mathbf{G} = \mathbf{I}$, we need SNR to be at least 15.7dB to obtain a small P_e, which is consistent with Fig. 6a.

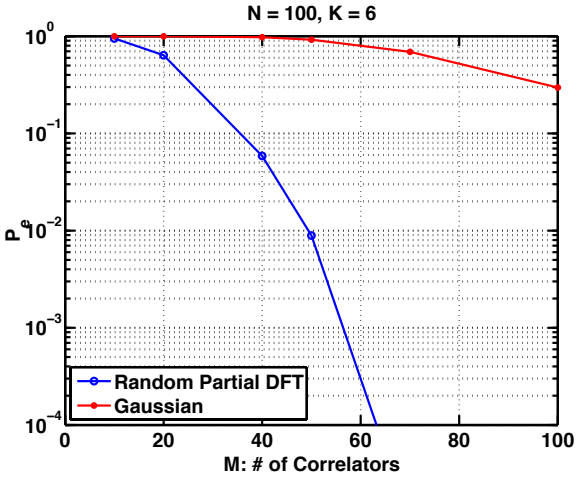


Fig. 7: In the absence of noise, P_e versus M of the RDD detector using the random partial DFT versus using the Gaussian random matrices for $N = 100$ and $K = 6$.

2) P_e vs. M, performance of the noise whitening transform, $\boldsymbol{G} \neq \boldsymbol{I}$: Next we consider a scenario when the signature waveforms are nonorthogonal $\mathbf{G} \neq \mathbf{I}$. We generate an arbitrary symmetric $\mathbf{G}$ with ones on the diagonal and fix it in the Monte Carlo trials. In the first case we consider highly correlated signature waveforms with $\lambda_{\max}(\mathbf{G}^{-1}) = 493.9595$. In the second case we consider nearly orthogonal signature waveforms with $\lambda_{\max}(\mathbf{G}^{-1}) = 4.0771$. Then we compare the P_e of the RDD detector in these two cases without and with the noise whitening transform in Section III-B3. Assume $N = 100$, $K = 2$, $r = 1$ and $\sigma = 0.1$. In Fig. 9a, when the signature waveforms are highly correlated, the noise whitening transform significantly reduces P_e for large M. In this case, the conventional decorrelating detector without the noise whitening transform has a non-negligible probability-of-error, and that with the noise whitening transform has a probability-of-error less than 10^{-4}. In Fig. 9b, when the signature waveforms are nearly orthogonal, the noise whitening transform does not reduce P_e much. In this case, the conventional decorrelating detector without and with the noise whitening transform both have probability-of-error less than 10^{-4}. We also verified that using the noise whitening transform cannot achieve the probability-of-error that is obtained with orthogonal signature waveforms $\mathbf{G} = \mathbf{I}$. This is because the noise whitening transform distorts the signal component.

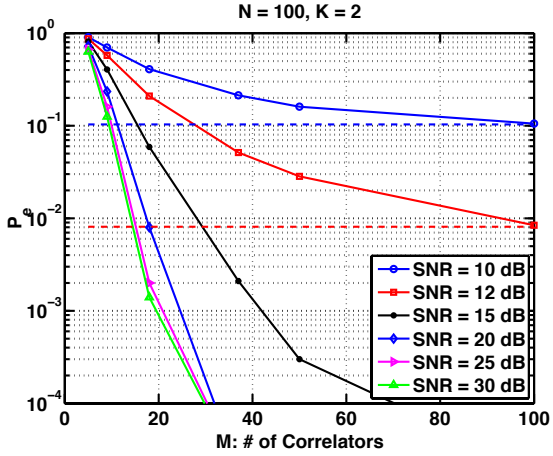
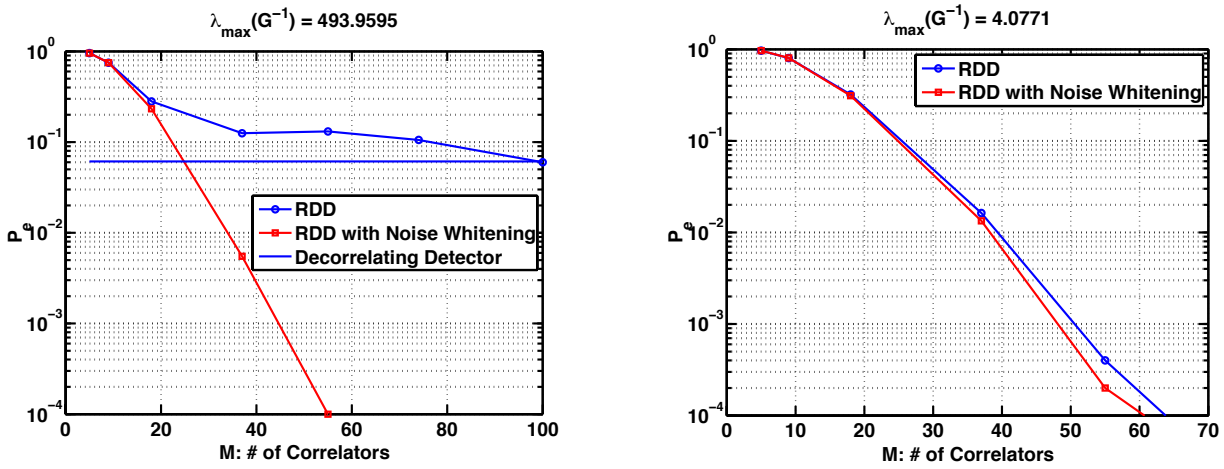


Fig. 8: Performance of the RDD detector, P_e versus M for different SNRs, when the signature waveforms are orthogonal, i.e. $\mathbf{G} = \mathbf{I}$. The dashed lines show P_e for the conventional decorrelating detectors at the corresponding SNR. When SNR is greater than 15 dB (with $N = 100$ correlators), the probability-of-error of the decorrelating detector is less than 10^{-4}.

3) P_e vs. M, RD-MUD linear detectors: In this example, we compare the performance of RD-MUD linear detectors when $\mathbf{G} \neq \mathbf{I}$. In Theorem 1 we have proven that the error is dominated by that from active user detection. So we compare the performance of these RD-MUD linear detectors using their conditional probability of symbol error given the correct detection of active users $P(\hat{\mathbf{b}} \neq \mathbf{b} | \hat{\mathcal{I}} = \mathcal{I})$. Assume $N = 100$, $K = 2$, $r = 1$, $\sigma = 0.1$, and let $\mathbf{G}$ take the same form as those used in Fig. 9. In Fig. 10, the performance of the RDD detector is similar to that of the RD-LS detector (the explanation is given in Section III-B4). In comparison, the RD-MMSE detector has smaller conditional probability of error, especially in Fig. 10a with highly correlated signatures. This improvement is because the linear transform (32) of the RD-MMSE detector alleviates the effect of correlated signature waveforms by including an inversion of $\mathbf{G}$ in the linear transform. The conditional probability-of-error of the conventional decorrelating detector is less than 10^{-4}.



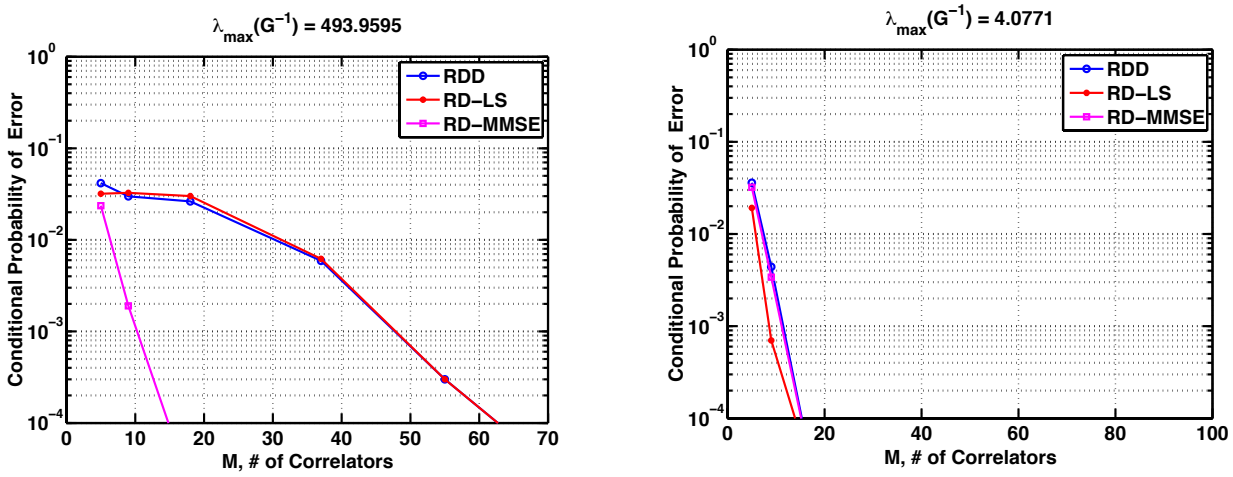
(a) Large $\lambda_{\max}(G^{-1})$: highly correlated signature waveforms

(b) Small $\lambda_{\max}(G^{-1})$: nearly orthogonal signature waveforms

Fig. 9: Comparison of P_e versus M, without and with the noise whitening transform, when $N = 100$, $K = 2$, and for two different cases of $\mathbf{G}$. The conventional decorrelating detector with the noise whitening transform in Fig. 9a, and the conventional decorrelating detectors with and without the noise whitening transform in Fig. 9b have probability-of-error less than 10^{-4}.

VI. Conclusions and Discussions

We have developed a reduced dimension multiuser detection (RD-MUD) structure, which decreases the number of correlators at the front-end of a MUD receiver by exploiting the fact that the number of active users is typically much smaller than the total number of users in the system. Motivated by the idea of analog compressed sensing, the RD-MUD front-end projects the received signal onto a lower dimensional detection subspace by correlating the received signal with a set of correlating signals. The correlating signals are constructed as linear combinations of the signature waveforms using a coefficient matrix $\mathbf{A}$, which determines the performance of RD-MUD and is our key design parameter. Based on the front-end output, RD-MUD detectors recover active users and their symbols in the detection subspace. We have studied in detail two such detectors. The reduced-dimension decorrelating (RDD) detector, which is a linear detector that combines subspace projection along with thresholding for active user detection and sign detection for data recovery. The reduced-dimension decision feedback (RDDF)



(a) Large $\lambda_{\max}(G^{-1})$: highly correlated signature waveforms (b) Small $\lambda_{\max}(G^{-1})$: nearly orthogonal signature waveforms

Fig. 10: Comparison of the conditional probability of error $P(\hat{\mathbf{b}} \neq \mathbf{b}|\hat{\mathcal{I}} = \mathcal{I})$, for RD-MUD linear detectors: RDD, RD-LS, and RD-MMSE. The RD-MMSE detector has the smallest conditional probability of symbol error.

detector is a nonlinear detector that combines decision-feedback orthogonal matching pursuit (DF-OMP) for active user detection with sign detection for data recovery. We have shown that to achieve a desired probability-of-error, the number of correlators used by the RD-MUD can be much smaller than that used by the conventional MUD, and the complexity-per-bit of the RD-MUD detectors are not higher than their counterpart in the conventional MUD setting. In particular, when the random partial DFT matrix is used for the coefficient matrix $\mathbf{A}$ and the RDD and RDDF detectors are used for detection, the RD-MUD front-end requires the number of correlators proportional to the log of the number of users, whereas the conventional MF-bank front-end requires the number of correlators equal to the number of users in the system. We have obtained theoretical performance guarantees for the RDD and RDDF detectors in terms of the coherence of $\mathbf{A}$, which are validated via numerical examples.

Our results are based on binary modulation and can be extended to higher order modulation with symbols taking more possible values. With higher order modulation, however, the conditions to guarantee correct symbol detection may be stronger than the conditions to guarantee correct active user

detection. This general modulation set-up is further related to problems in compressed classification and compressed learning (see, e.g., [74]). We have also assumed that the signature waveforms are given. Better performance of RD-MUD might be obtained through joint optimization of the signature waveforms and the coefficient matrix $\mathbf{A}$.

APPENDIX A
COVARIANCE OF RD-MUD OUTPUT NOISE

Consider the covariance of the output noise at the nth and the mth branches in the RD-MUD front-end:

$$\begin{aligned}\rho_{nm} = \mathbb{E}\{w_n w_m\} &= \mathbb{E}\left\{\sum_{k=1}^{N}\sum_{l=1}^{N} a_{nk}a_{ml}\langle \hat{s}_k(t), w(t)\rangle\langle \hat{s}_l(t), w(t)\rangle\right\} \\ &= \sum_{k=1}^{N}\sum_{l=1}^{N} a_{nk}a_{ml}\mathbb{E}\{\langle \hat{s}_k(t), w(t)\rangle\langle \hat{s}_l(t), w(t)\rangle\}\end{aligned} \tag{65}$$

We want to show that $\rho_{nm} = \sigma^2[\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H]_{nm}$. We have

$$\begin{aligned}&\mathbb{E}\{\langle \hat{s}_k(t), w(t)\rangle\langle \hat{s}_l(t), w(t)\rangle\} = T^{-2}\int_0^T\int_0^T \hat{s}_k(t)\hat{s}_l(u)\mathbb{E}\{w(t)w(u)\}dtdu \\ &= T^{-2}\int_0^T\int_0^T \hat{s}_k(t)\hat{s}_l(u)\sigma^2\delta(t-u)dtdu = \sigma^2 T^{-1}\int_0^T \hat{s}_k(t)\hat{s}_l(t)dt \\ &= \sigma^2\left\langle \sum_n [\mathbf{G}^{-1}]_{nk}s_n(t), \sum_m [\mathbf{G}^{-1}]_{ml}s_m(t)\right\rangle = \sigma^2\sum_n\sum_m [\mathbf{G}^{-1}]_{nk}[\mathbf{G}^{-1}]_{ml}\langle s_n(t), s_m(t)\rangle \\ &= \sigma^2\sum_n\sum_m [\mathbf{G}^{-1}]_{nk}[\mathbf{G}^{-1}]_{ml}[\mathbf{G}]_{nm} = \sigma^2[\mathbf{G}^{-1}]_{lk}\end{aligned} \tag{66}$$

Substituting this back into (65), we have $\rho_{nm} = \mathbb{E}\{w_n w_m\} = \sigma^2\sum_{k=1}^{N}\sum_{l=1}^{N} a_{nk}a_{ml}[\mathbf{G}^{-1}]_{lk} = \sigma^2[\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H]_{nm}$. Setting $n = m$, we have $\sigma_m = \mathbb{E}\{w_m^2\} = \sigma^2[\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H]_{mm}$. This completes the proof.

APPENDIX B
DERIVATION OF RD-MUD MMSE

Proof: Given the active user index set $\hat{\mathcal{I}}$ obtained from (22), we define $\mathbf{W} = \mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}^2\mathbf{A}_{\hat{\mathcal{I}}}^H + \sigma^2\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H$, and $\bar{\mathbf{M}} = \mathbf{R}_{\hat{\mathcal{I}}}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{W}^{-1}$. We want to show that $\bar{\mathbf{M}} = \arg\min_{\mathbf{M}}\mathbb{E}\{\|\mathbf{b}_{\hat{\mathcal{I}}} - \mathbf{M}\mathbf{y}\|^2\}$. Using the same method for deriving the conventional MMSE detector of the MF-bank [3], we assume that $\mathbf{b}_{\hat{\mathcal{I}}}$ has a distribution that is uncorrelated with the noise $\mathbf{w}$ and that $\mathbb{E}\{\mathbf{b}_{\hat{\mathcal{I}}}\mathbf{b}_{\hat{\mathcal{I}}}^H\} = \mathbf{I}$. Based on $\hat{\mathcal{I}}$, we refer to the model (20). Since $\|\mathbf{x}\|^2 = \mathbf{tr}(\mathbf{x}\mathbf{x}^H)$, we can write the MSE as $\mathbb{E}\{\|\mathbf{b}_{\hat{\mathcal{I}}} - \mathbf{M}\mathbf{y}\|^2\} = \mathbf{tr}(\mathbb{E}\{(\mathbf{b}_{\hat{\mathcal{I}}} - \mathbf{M}\mathbf{y})(\mathbf{b}_{\hat{\mathcal{I}}} - \mathbf{M}\mathbf{y})^H\})$.

Now we expand

$$\begin{aligned}
&\mathbb{E}\{(\mathbf{b}_{\hat{\mathcal{I}}}-\mathbf{M}\mathbf{y})(\mathbf{b}_{\hat{\mathcal{I}}}-\mathbf{M}\mathbf{y})^H\}\\
&=\mathbb{E}\{\mathbf{b}_{\hat{\mathcal{I}}}\mathbf{b}_{\hat{\mathcal{I}}}^H\}-\mathbb{E}\{\mathbf{b}_{\hat{\mathcal{I}}}\mathbf{y}^H\}\mathbf{M}^H-\mathbf{M}\mathbb{E}\{\mathbf{y}\mathbf{b}_{\hat{\mathcal{I}}}^H\}+\mathbf{M}\mathbb{E}\{\mathbf{y}\mathbf{y}^H\}\mathbf{M}^H\\
&=\mathbf{I}+\mathbf{M}(\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}^2\mathbf{A}_{\hat{\mathcal{I}}}^H+\sigma^2\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H)\mathbf{M}^H-\mathbf{R}_{\hat{\mathcal{I}}}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{M}^H-\mathbf{M}\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}.
\end{aligned}\tag{67}$$

It can be verified that $\mathbf{M}\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}=\mathbf{M}\mathbf{W}\bar{\mathbf{M}}^H$. Hence from (67), we have

$$\begin{aligned}
\mathbb{E}\{(\mathbf{b}_{\hat{\mathcal{I}}}-\mathbf{M}\mathbf{y})(\mathbf{b}_{\hat{\mathcal{I}}}-\mathbf{M}\mathbf{y})^H\}&=\mathbf{I}+\mathbf{M}\mathbf{W}\mathbf{M}^H-\bar{\mathbf{M}}\mathbf{W}\mathbf{M}^H-\mathbf{M}\mathbf{W}\bar{\mathbf{M}}^H\\
&=\mathbf{I}-\bar{\mathbf{M}}\mathbf{W}\bar{\mathbf{M}}^H+(\mathbf{M}-\bar{\mathbf{M}})\mathbf{W}(\mathbf{M}-\bar{\mathbf{M}})^H\\
&=\mathbf{I}-\mathbf{R}_{\hat{\mathcal{I}}}\mathbf{A}_{\hat{\mathcal{I}}}^H\mathbf{W}^{-1}\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{R}_{\hat{\mathcal{I}}}+(\mathbf{M}-\bar{\mathbf{M}})\mathbf{W}(\mathbf{M}-\bar{\mathbf{M}})^H.
\end{aligned}\tag{68}$$

Since $\mathbf{W}$ is a positive semidefinite matrix, the trace of the second term in (68) is always nonnegative. So we conclude that the matrix $\mathbf{M}$ that minimizes the MSE $\mathbf{tr}(\mathbb{E}\{(\mathbf{b}_{\hat{\mathcal{I}}}-\mathbf{M}\mathbf{y})(\mathbf{b}_{\hat{\mathcal{I}}}-\mathbf{M}\mathbf{y})^H\})$ is $\bar{\mathbf{M}}$, or equivalently, $\mathbb{E}\{\|\mathbf{b}_{\hat{\mathcal{I}}}-\mathbf{M}\mathbf{y}\|^2\}$, as required. ■

APPENDIX C
PROOF OF THEOREM 1

The proof of Theorem 1 for both the RDD and RDDF detectors are closely related. To exploit this similarity, we first prove several lemmas that are useful for both results. First, we will demonstrate that the random event

$$\mathcal{G}\triangleq\left\{\max_{1\le n\le N}|\mathbf{a}_n^H\mathbf{w}|<\tau\right\}\tag{69}$$

occurs with high probability, where

$$\tau\triangleq\sigma\sqrt{2(1+\alpha)\log N}\cdot\sqrt{\lambda_{\max}(\mathbf{G}^{-1})}\cdot\sqrt{\max_n(\mathbf{a}_n^H\mathbf{A}\mathbf{A}^H\mathbf{a}_n)},\tag{70}$$

and $\alpha>0$. Then we show that when $\mathcal{G}$ occurs, both algorithms can detect the active users and their symbols. The proofs follow the arguments in [65] with modifications to account for the fact that $\mathbf{w}$ is colored noise, and the error can also be caused by incorrect symbol detection. However, as we will show, the error probability of active user detection dominates the latter case.

Sidak's lemma [75] states the following:

Lemma 2 (Sidak's lemma)**.** *Let $[X_1,\cdots,X_n]^\top$ be a vector of random variables having the n-dimensional normal distribution with zero means, arbitrary variances $\sigma_1^2,\cdots,\sigma_n^2$, and an arbitrary correlation matrix*

$[\boldsymbol{P}]_{mk} = \rho_{mk}$. *Then, for any positive numbers* $c_1, \cdots, c_n$,

$$P(|X_1| \leq c_1, |X_2| \leq c_2, \cdots, |X_n| \leq c_n) \geq P(|X_1| \leq c_1) \cdot P(|X_2| \leq c_2, \cdots, |X_n| \leq c_n). \tag{71}$$

Lemma 3. *Suppose that* $\boldsymbol{w}$ *is a Gaussian random vector with zero mean and covariance* $\sigma^2 \boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H$. *If* $N^{-(1+\alpha)}[\pi(1+\alpha)\log N]^{-1/2} \leq 1$ *for some* $\alpha > 0$, *then the event* $\mathcal{G}$ *of (69) occurs with probability at least one minus (45).*

Proof: The random variables $\{\mathbf{a}_n^H \mathbf{w}\}_{n=1}^N$ are jointly Gaussian, with means equal to zero, variances σ_n^2 equal to $\sigma^2 \mathbf{a}_n^H \mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H \mathbf{a}_n$, and covariances ρ_{nm} between the nth and mth random variables equal to $\sigma^2 \mathbf{a}_n^H \mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H \mathbf{a}_m$. Define

$$\hat{\tau} \triangleq \sigma[2(1+\alpha)\log N]^{1/2} \cdot \left[\max_n (\mathbf{a}_n^H \mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H \mathbf{a}_n)\right]^{1/2}, \tag{72}$$

and an event

$$\hat{\mathcal{G}} \triangleq \left\{ \max_{1 \leq n \leq N} |\mathbf{a}_n^H \mathbf{w}| < \hat{\tau} \right\}. \tag{73}$$

Using Sidak's lemma, we have

$$P\left(\hat{\mathcal{G}}\right) = P\left(|\mathbf{a}_1^H \mathbf{w}| < \hat{\tau}, \cdots, |\mathbf{a}_N^H \mathbf{w}| < \hat{\tau}\right) \geq \prod_{n=1}^N P(|\mathbf{a}_n^H \mathbf{w}| < \hat{\tau}). \tag{74}$$

Since $\mathbf{a}_n^H \mathbf{w}$ is a Gaussian random variable with zero mean and variance σ_n^2, the tail probability of the colored noise can be written as

$$P(|\mathbf{a}_n^H \mathbf{w}| < \hat{\tau}) = 1 - 2Q\left(\frac{\hat{\tau}}{\sigma_n}\right). \tag{75}$$

By the bound (58) on $Q(x)$, (75) can be bounded as

$$P(|\mathbf{a}_n^H \mathbf{w}| < \hat{\tau}) \geq 1 - \eta_n, \tag{76}$$

where $\eta_n \triangleq \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma_n}{\hat{\tau}} e^{-\hat{\tau}^2/(2\sigma_n^2)}$. Define

$$\begin{aligned} \sigma_{\max} &\triangleq \max_n \sigma_n = \sigma \left[\max_n (\mathbf{a}_n^H \mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H \mathbf{a}_n)\right]^{1/2}, \\ \eta_{\max} &\triangleq \sqrt{\frac{2}{\pi}} \frac{\sigma_{\max}}{\hat{\tau}} e^{-\hat{\tau}^2/(2\sigma_{\max}^2)}. \end{aligned} \tag{77}$$

Since $\sigma_{\max}/\hat{\tau} = [2(1+\alpha)\log N]^{-1/2}$ by the definition of $\hat{\tau}$, we have $\eta_{\max} = \sqrt{\frac{2}{\pi}}[2(1+\alpha)\log N]^{-1/2}e^{-(1+\alpha)\log N}$. It is easy to show that η_n increases as σ_n increases. Hence $\eta_n \leq \eta_{\max}$. When $\eta_{\max} \leq 1$, we can use the inequality $(1-x)^N \geq 1 - Nx$ when $x \geq 0$ and substitute the value of $\eta_{\max}$ to write (74) as

$$P(\hat{\mathcal{G}}) \geq \prod_{n=1}^{N}(1-\eta_n) \geq (1-\eta_{\max})^N \geq 1 - N\eta_{\max} = 1 - N^{-\alpha}[\pi(1+\alpha)\log N]^{-1/2}, \tag{78}$$

which holds for any $\eta_{\max} \leq 1$ and $N \geq 1$.

Next we show that $\hat{\tau} \leq \tau$. Note that

$$\mathbf{a}_n^H \mathbf{A}\mathbf{G}^{-1}\mathbf{A}^H\mathbf{a}_n \leq \|\mathbf{A}^H\mathbf{a}_n\|^2 \lambda_{\max}(\mathbf{G}^{-1}) \leq [\max_n \left(\mathbf{a}_n^H \mathbf{A}\mathbf{A}^H\mathbf{a}_n\right)]\lambda_{\max}(\mathbf{G}^{-1}). \tag{79}$$

From inequality (79) and definitions (70) for τ and (72) for $\hat{\tau}$, we obtain $\hat{\tau} \leq \tau$. Hence

$$P(\mathcal{G}) = P(\max_n |\mathbf{a}_n^H\mathbf{w}| < \tau) \geq P(\max_n |\mathbf{a}_n^H\mathbf{w}| < \hat{\tau}) = P(\hat{\mathcal{G}}). \tag{80}$$

Combining (78) and (80), we conclude that $P(\mathcal{G})$ is greater than one minus the expression (45), as required. ■

The next lemma shows that, under appropriate conditions, ranking the inner products between $\mathbf{a}_n$ and $\mathbf{y}$ is an effective method of detecting the set of active users. The proof of this lemma is adapted from Lemma 3 in [65] to account for the fact that the signal vector $\mathbf{y}$ here can be complex as $\mathbf{A}$ can be complex. Since only the real part contains all the useful information, to prove this lemma, we basically follow the proof for Lemma 3 in [65] while using the following inequality whenever needed: $|\Re[\mathbf{a}_n^H\mathbf{a}_m]| \leq |\mathbf{a}_n^H\mathbf{a}_m| \leq \mu$ for $n \neq m$, and $|\Re[\mathbf{a}_n^H\mathbf{w}]| \leq |\mathbf{a}_n^H\mathbf{w}|$.

Lemma 4. *Let $\boldsymbol{b}$ be a vector with support $\mathcal{I}$ which consists of K active users, and let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{R}\boldsymbol{b} + \boldsymbol{w}$ for a Gaussian noise vector $\boldsymbol{w}$ with zero mean and covariance $\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H$. Define $|r_{\max}|$ and $|r_{\min}|$ as in (43), and suppose that*

$$|r_{\min}| - (2K-1)\mu|r_{\max}| \geq 2\tau. \tag{81}$$

Then, if the event $\mathcal{G}$ of (69) occurs, we have

$$\min_{n\in\mathcal{I}} |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]| > \max_{n\notin\mathcal{I}} |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]|. \tag{82}$$

If, rather than (81), a weaker condition holds:

$$|r_{\max}| - (2K-1)\mu|r_{\max}| \geq 2\tau. \tag{83}$$

Then, if the event $\mathcal{G}$ of (69) occurs, we have

$$\max_{n\in\mathcal{I}} |\Re[\boldsymbol{a}_n^H \mathbf{y}]| > \max_{n\notin\mathcal{I}} |\Re[\boldsymbol{a}_n^H \mathbf{y}]|. \tag{84}$$

Proof: We begin by deriving a lower-bound for $\min_{n\in\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]|$ under the event $\mathcal{G}$:

$$\begin{aligned}
\min_{n\in\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]| &= \min_{n\in\mathcal{I}} \left| b_n r_n + \sum_{m\neq n} b_m r_m \Re[\mathbf{a}_n^H \mathbf{a}_m] + \Re[\mathbf{a}_n^H \mathbf{w}] \right| \\
&\geq \min_{n\in\mathcal{I}} \left(|b_n||r_n| - \left| \sum_{m\neq n} b_m r_m \Re[\mathbf{a}_n^H \mathbf{a}_m] \right| - |\Re[\mathbf{a}_n^H \mathbf{w}]| \right) \\
&\geq |r_{\min}| - \max_{n\in\mathcal{I}} \sum_{m\neq n} |b_m||r_m||\mathbf{a}_n^H \mathbf{a}_m| - \max_{n\in\mathcal{I}} |\mathbf{a}_n^H \mathbf{w}| \\
&> |r_{\min}| - (K-1)\mu|r_{\max}| - \tau,
\end{aligned} \tag{85}$$

where we have used the triangle inequality, the fact that $|b_n| = 1$, $|r_{\min}| \leq |r_n| \leq |r_{\max}|$, $|\Re[\mathbf{a}_n^H \mathbf{a}_m]| \leq |\mathbf{a}_n^H \mathbf{a}_m| \leq \mu$ for $n \neq m$, and $|\Re[\mathbf{a}_n^H \mathbf{w}]| \leq |\mathbf{a}_n^H \mathbf{w}|$. On the other hand, we can similarly expand and upper-bound $\max_{n\notin\mathcal{I}} |\mathbb{R}[\mathbf{a}_n^H \mathbf{y}]|$, under the event $\mathcal{G}$, as

$$\begin{aligned}
\max_{n\notin\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]| &= \max_{n\notin\mathcal{I}} \left| \sum_{m\in\mathcal{I}} b_m r_m \Re[\mathbf{a}_n^H \mathbf{a}_m] + \Re[\mathbf{a}_n^H \mathbf{w}] \right| \\
&\leq \max_{n\notin\mathcal{I}} \sum_{m\in\mathcal{I}} |b_m||r_m||\Re[\mathbf{a}_n^H \mathbf{a}_m]| + \max_{n\notin\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{w}]| \\
&< K\mu|r_{\max}| + \tau.
\end{aligned} \tag{86}$$

Combining (85) and (86), we have that under the event $\mathcal{G}$,

$$\min_{n\in\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]| > |r_{\min}| - (2K-1)\mu|r_{\max}| - 2\tau + \max_{n\notin\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]|. \tag{87}$$

So when $\mathcal{G}$ occurs, under the condition (81), we obtain (82).

Similarly, when $\mathcal{G}$ occurs, we expand and lower-bound $\max_{n\in\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]|$. Assume that n_0 is the index

achieving the largest absolute gain: $|r_{n_0}| = |r_{\max}|$. Then under event $\mathcal{G}$:

$$\begin{aligned}\max_{n\in\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]| \geq |\Re[\mathbf{a}_{n_0}^H \mathbf{y}]| &= \left| b_{n_0} r_{n_0} + \sum_{m\neq n_0} b_m r_m \Re[\mathbf{a}_{n_0}^H \mathbf{a}_m] + \Re[\mathbf{a}_{n_0}^H \mathbf{w}] \right| \\ &\geq |r_{\max}| - \sum_{m\neq n_0} |b_m||r_m||\Re[\mathbf{a}_{n_0}^H \mathbf{a}_m]| - |\Re[\mathbf{a}_{n_0}^H \mathbf{w}]| \\ &> |r_{\max}| - (K-1)\mu|r_{\max}| - \tau.\end{aligned} \tag{88}$$

Combining (88) and (86), we have that under the event $\mathcal{G}$,

$$\max_{n\in\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]| > |r_{\max}| - (2K-1)\mu|r_{\max}| - 2\tau + \max_{n\notin\mathcal{I}} |\Re[\mathbf{a}_n^H \mathbf{y}]|. \tag{89}$$

So when $\mathcal{G}$ occurs, under the condition (83), we obtain (84), as required. ■

The following lemma demonstrates that the sign detector can effectively detect transmitted symbols for the RDD and RDDF detectors.

Lemma 5. *Let $\boldsymbol{b}$ be a vector with $b_n \in \{1, -1\}$, for $n \in \mathcal{I}$ and $b_n = 0$ otherwise, and let $\mathbf{y} = \boldsymbol{ARb} + \boldsymbol{w}$ for a Gaussian noise vector $\boldsymbol{w}$ with zero mean and covariance $\sigma^2 \boldsymbol{AG}^{-1}\boldsymbol{A}^H$. Suppose that*

$$|r_{\min}| - (K-1)\mu|r_{\max}| \geq \tau. \tag{90}$$

Then, if the event $\mathcal{G}$ occurs, we have

$$\mathbf{sgn}(r_n \Re[\boldsymbol{a}_n^H \mathbf{y}]) = b_n, \qquad n \in \mathcal{I}. \tag{91}$$

If, instead of (90), a weaker condition

$$|r_{\max}| + |r_{\min}| - 2(K-1)\mu|r_{\max}| \geq 2\tau \tag{92}$$

holds, then under the event $\mathcal{G}$, we have

$$\mathbf{sgn}(r_{n_1} \Re[\boldsymbol{a}_{n_1}^H \mathbf{y}]) = b_{n_1}, \tag{93}$$

for

$$n_1 = \arg\max_n |\Re[\boldsymbol{a}_n^H \mathbf{y}]|. \tag{94}$$

Proof: To detect correctly, for $b_n = 1$, $\Re[r_n \mathbf{a}_n^H \mathbf{y}]$ has to be positive, and for $b_n = -1$, $\Re[r_n \mathbf{a}_n^H \mathbf{y}]$ has to be negative. First assume $b_n = 1$. We expand $\Re[r_n \mathbf{a}_n^H \mathbf{y}]$, find the lower-bound and the condition

such that the lower bound is positive. Substituting in the expression for $\mathbf{y}$, using the inequality that $x+y+z \geq x-|y|-|z|$, under the event $\mathcal{G}$, we obtain

$$\begin{aligned}\Re[r_n \mathbf{a}_n^H \mathbf{y}] &= |r_n|^2 + \sum_{m\neq n} b_m r_n r_m \Re\left[\mathbf{a}_n^H \mathbf{a}_m\right] + r_n \Re\left[\mathbf{a}_n^H \mathbf{w}\right] \\ &\geq |r_n||r_{\min}| - \sum_{m\neq n} |r_n||r_m||\Re\left[\mathbf{a}_n^H \mathbf{a}_m\right]| - |r_n||\Re\left[\mathbf{a}_n^H \mathbf{w}\right]| \\ &> |r_n|\left[|r_{\min}| - (K-1)\mu|r_{\max}| - \tau\right].\end{aligned} \tag{95}$$

From (95), $\Re[r_n \mathbf{a}_n^H \mathbf{y}] > 0$ for $n \in \mathcal{I}$ if (90) holds and $b_n = 1$. Similarly, we can show for $b_n = -1$, under event $\mathcal{G}$, if (90) holds, $\Re[r_n \mathbf{a}_n^H \mathbf{y}] < 0$. Hence if (90) holds we obtain (91).

Recall that n_0 is the index of the largest gain: $|r_{n_0}| = |r_{\max}|$. Due to (94), we have

$$|\Re[\mathbf{a}_{n_1}^H \mathbf{y}]| \geq |\Re[\mathbf{a}_{n_0}^H \mathbf{y}]|. \tag{96}$$

We will show that under the event $\mathcal{G}$, once (92) holds, then $\mathbf{sgn}(r_{n_1}\Re[\mathbf{a}_{n_1}^H \mathbf{y}]) \neq b_{n_1}$ leads to a contradiction to (96). First assume $b_{n_1} = 1$. If $\hat{b}_{n_1} = \mathbf{sgn}(r_{n_1}\Re[\mathbf{a}_{n_1}^H \mathbf{y}]) \neq b_{n_1}$, then

$$\hat{b}_{n_1} = \mathbf{sgn}\left(r_{n_1}^2 + \sum_{m\neq n_1} b_m r_{n_1} r_m \Re\left[\mathbf{a}_{n_1}^H \mathbf{a}_m\right] + r_{n_1}\Re\left[\mathbf{a}_{n_1}^H \mathbf{w}\right]\right) = -1. \tag{97}$$

So the expression inside the **sgn** operator of (97) must be negative. Since $r_{n_1}^2 > 0$, we must have

$$\sum_{m\neq n_1} b_m r_{n_1} r_m \Re\left[\mathbf{a}_{n_1}^H \mathbf{a}_m\right] + r_{n_1}\Re\left[\mathbf{a}_{n_1}^H \mathbf{w}\right] < 0. \tag{98}$$

Multiplying the left-hand-side of (96) by $|r_{n_1}|$, and using the equality $|x|\cdot|y| = |xy|$, we obtain

$$\begin{aligned}|r_{n_1}||\Re[\mathbf{a}_{n_1}^H \mathbf{y}]| &= |r_{n_1}|\left|r_{n_1} + \sum_{m\neq n_1} b_m r_m \Re[\mathbf{a}_{n_1}^H \mathbf{a}_m] + \Re[\mathbf{a}_{n_1}^H \mathbf{w}]\right| \\ &= \left|r_{n_1}^2 + \sum_{m\neq n_1} b_m r_{n_1} r_m \Re[\mathbf{a}_{n_1}^H \mathbf{a}_m] + r_{n_1}\Re[\mathbf{a}_{n_1}^H \mathbf{w}]\right|.\end{aligned} \tag{99}$$

Due to (97), the last line of (99) inside the $|\cdot|$ operator is negative. Using the fact that $r_{n_1}^2 > 0$ and (98), and the identity $|x+y| = -(x+y) = |y| - x$ when $x+y<0$ and $y<0$, under the event $\mathcal{G}$, we

obtain that

$$|r_{n_1}||\Re[\mathbf{a}_{n_1}^H\mathbf{y}]| = \left|\sum_{m\neq n_1} b_m r_{n_1} r_m \Re\left[\mathbf{a}_{n_1}^H\mathbf{a}_m\right] + r_{n_1}\Re\left[\mathbf{a}_{n_1}^H\mathbf{w}\right]\right| - r_{n_1}^2$$
$$< |r_{n_1}|(K-1)\mu|r_{\max}| + |r_{n_1}|\tau - |r_{n_1}||r_{\min}| = |r_{n_1}|[(K-1)\mu|r_{\max}| + \tau - |r_{\min}|]. \tag{100}$$

On the other hand, multiply the right-hand-side of (96) by $|r_{n_1}|$. Similarly, using the equality $|x|\cdot|y| = |xy|$ and triangle inequality, under the event $\mathcal{G}$, we obtain

$$|r_{n_1}||\Re[\mathbf{a}_{n_0}^H\mathbf{y}]| = \left|r_{n_1}r_{n_0}b_{n_0} + \sum_{m\neq n_0} b_m r_{n_1} r_m \Re\left[\mathbf{a}_{n_0}^H\mathbf{a}_m\right] + r_{n_1}\Re\left[\mathbf{a}_{n_0}^H\mathbf{w}\right]\right|$$
$$> |r_{n_1}|[|r_{\max}| - (K-1)\mu|r_{\max}| - \tau]. \tag{101}$$

Combining (100) and (101), we have that once (92) holds, if $b_{n_1} = 1$, then $\mathbf{sgn}(r_{n_1}\Re[\mathbf{a}_{n_1}^H\mathbf{y}]) = -1$ leads to $|\Re[\mathbf{a}_{n_1}^H\mathbf{y}]| < |\Re[\mathbf{a}_{n_0}^H\mathbf{y}]|$, which contradicts (96), and hence $\mathbf{sgn}(r_{n_1}\Re[\mathbf{a}_{n_1}^H\mathbf{y}]) = 1$. A similar argument can be made for $b_{n_1} = -1$, which completes the proof.

■

We are now ready to prove Theorem 1. The proof for the RDD detector is obtained by combining Lemmas 3, 4 and 5. Lemma 3 ensures that the event $\mathcal{G}$ occurs with probability at least as high as one minus (45). Whenever $\mathcal{G}$ occurs, Lemma 4 guarantees by using (22), that the RDD detector can correctly detect active users under the condition (44), i.e. $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\}$. Finally, whenever $\mathcal{G}$ occurs, Lemma 5 guarantees that, based on the correct support of active users, their transmitted symbols can be detected correctly under the condition (90), i.e. $\mathcal{G} \subset \{\hat{b}_n = b_n, n \in \mathcal{I}\}$. Clearly condition (90) is weaker than (44), since (44) can be written as $|r_{\min}| - (K-1)\mu|r_{\max}| \geq \tau + (\tau + K\mu|r_{\max}|) > \tau$, and hence if (44) holds then (90) also holds. In summary, under condition (44), $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} = \mathbf{b}\}$, and $1 - P_e = P(\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} = \mathbf{b}\}) \geq P(\mathcal{G})$, which is greater than one minus (45), which concludes the proof for the RDD detector.

We now prove the performance guarantee for the RDDF detector adopting the technique used in proving Theorem 4 in [65]. First we show that whenever $\mathcal{G}$ occurs, the RDDF detector correctly detects an active user in the first iteration, which follows from Lemmas 3 and 4. Note that (46) implies (83), and therefore, by Lemma 4, we have that by choosing the largest $|\Re[\mathbf{a}_n^H\mathbf{y}]|$, the RDDF detector can detect a correct user in the set $\mathcal{I}$. Second, we show that whenever $\mathcal{G}$ occurs, the RDDF detector

correctly detects the transmitted symbol of this active user. Note that (46) also implies (92), since (46) can be written as $|r_{\min}| \geq 2\tau/[1-(2K-1)\mu]$, which implies $|r_{\max}| \geq 2\tau/[1-(2K-1)\mu]$, and hence $|r_{\max}|+|r_{\min}|-2(K-1)\mu|r_{\max}| \geq 2\tau[1-2(K-1)\mu]/[1-(2K-1)\mu]+|r_{\min}| > 2\tau$, since $[1-2(K-1)\mu]/[1-(2K-1)\mu] \geq 1$. Therefore, by Lemma 5, using a sign detector, we can detect the symbol correctly. Consequently, the first step of the RDDF detector correctly detect the active user and its symbol, i.e. $\mathcal{G} \subset \{\hat{\mathcal{I}}^{(1)} \subset \mathcal{I}, b_{n_1}^{(1)} = b_{n_1}\}$.

The proof now continues by induction. Suppose we are currently in the kth iteration of the RDDF detector, $1 \leq k \leq K$, and assume that correct users and their symbols have been detected in all the $k-1$ previous steps. The kth step is to detect the user with the largest $|\Re[\mathbf{a}_n^H \mathbf{v}^{(k)}]|$. Using the same notations as those in Section III-B2 and by definition of $\mathbf{v}^{(k)}$, we have

$$\mathbf{v}^{(k)} = \mathbf{AR}(\mathbf{b}-\mathbf{b}^{(k-1)}) + \mathbf{w} = \mathbf{ARx}^{(k-1)} + \mathbf{w}, \tag{102}$$

where $\mathbf{x}^{(k-1)} \triangleq \mathbf{b}-\mathbf{b}^{(k-1)}$. This vector has support $\mathcal{I}/\mathcal{I}^{(k-1)}$ and has at most $K-k+1$ non-zero elements, since $\mathbf{b}^{(k-1)}$ contains correct symbols at the correct locations for $(k-1)$ active users, i.e. $b_n^{(k-1)} = b_n$, for $n \in \mathcal{I}^{(k-1)}$. This $\mathbf{v}^{(k)}$ is a noisy measurement of the vector $\mathbf{ARx}^{(k-1)}$. The data model in (102) for the kth iteration is identical to the data model in the first iteration with $\mathbf{b}$ replaced by $\mathbf{x}^{(k-1)}$ (with a smaller sparsity $K-k+1$ rather than K), $\mathcal{I}$ replaced by $\mathcal{I}/\mathcal{I}^{(k-1)}$, and $\mathbf{y}$ replaced by $\mathbf{v}^{(k)}$. Since

$$|r_{\max}^{(k)}| \triangleq \max_{n \in \mathcal{I}/\mathcal{I}^{(k-1)}} |r_n| \geq |r_{\min}|, \tag{103}$$

we have that under condition (46) this model (102) also satisfies the requirement (83). Consequently, by Lemma 4, we have that under the event $\mathcal{G}$,

$$\max_{n \in \mathcal{I}/\mathcal{I}^{(k-1)}} |\Re[\mathbf{a}_n^H \mathbf{v}^{(k-1)}]| > \max_{n \in (\mathcal{I}/\mathcal{I}^{(k-1)})^c} |\Re[\mathbf{a}_n^H \mathbf{v}^{(k-1)}]|. \tag{104}$$

Therefore, in the kth iteration, the RDDF detector can detect an active user correctly, i.e. $\mathcal{G} \subset \{\hat{\mathcal{I}}^{(k)} \subset \mathcal{I}\}$, and no index of active users that has been detected before will be chosen again. On the other hand, since (46) can be written as $|r_{\min}| \geq 2\tau/[1-(2K-1)\mu]$, from (103) this implies $|r_{\max}^{(k)}| \geq 2\tau/[1-(2K-1)\mu]$, and hence $|r_{\max}^{(k)}| - (2K-1)\mu|r_{\max}^{(k)}| \geq 2\tau$, and consequently $|r_{\max}^{(k)}| - (2K-2)\mu|r_{\max}^{(k)}| + |r_{\min}| \geq 2\tau$. Consequently, condition (92) is true for (102). Then by Lemma 5, we have that under the event $\mathcal{G}$,

$$\mathbf{sgn}(\Re[r_{n_k} \mathbf{a}_{n_k}^H \mathbf{v}^{(k-1)}]) = b_{n_k}, \tag{105}$$

i.e. $\mathcal{G} \subset \{b_{n_k}^{(k)} = b_{n_k}\}$. By induction, since no active users will be detected twice, it follows that the first K steps of the RDDF detector can detect all active users and their symbols, i.e.

$$\mathcal{G} \subset \cup_{k=1}^{K} \{\hat{\mathcal{I}}^{(k)} \subset \mathcal{I}, b_{n_k}^{(k)} = b_{n_k}\} = \{\hat{\mathcal{I}}^{(K)} = \mathcal{I}, b_n^{(K)} = b_n, n \in \mathcal{I}^{(K)}\}. \tag{106}$$

Note that condition (83) is weaker than (46), since (46) can be written as $|r_{\min}|[1-(2K-1)\mu] \geq 2\tau$, which implies $|r_{\max}|[1-(2K-1)\mu] \geq 2\tau$. This further implies $|r_{\max}|[1-2(K-1)\mu] + |r_{\min}| \geq 2\tau$, since $1-2(K-1)\mu \geq 1-(2K-1)\mu$ and $|r_{\min}| \geq 0$. Consequently, under condition (46), from (106), $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} = \mathbf{b}\}$, and $1 - P_e = P(\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\mathbf{b}} = \mathbf{b}\}) \geq P(\mathcal{G})$ which is greater than one minus (45), which concludes the proof for the RDDF detector. This completes the proof of Theorem 1.

REFERENCES

[1] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Personal Communications*, pp. 46–58, April 1995.

[2] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Comm. Mag.*, vol. 34, pp. 124 – 136, Oct. 1996.

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

[4] J. G. Andrews, "Interference cancellation for cellular systems: A contemporary overview," *IEEE Wireless Comm. Mag.: on advanced receiver for CDMA*, 2005.

[5] C. Schlegal and A. Grant, *Coordinated multiuser communications*. Springer, May 2006.

[6] M. L. Honig, ed., *Advances in multiuser detection*. Wiley Series in Telecommunications and Signal Processing, Wiley-IEEE Press, Aug. 2009.

[7] A. J. Viterbi, *CDMA: principles of spread spectrum communication*. Addison-Wesley, Apr. 1995.

[8] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channel," *IEEE Trans. Info. Theory*, vol. 35, pp. 123 – 136, Jan 1989.

[9] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Info. Theory*, vol. 32, pp. 85 – 96, Jan. 1986.

[10] R. Lupas and S. Verdu, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Comm.*, vol. 38, pp. 496 – 508, April 1990.

[11] S. Verdu and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Info. Theory*, vol. 45, pp. 622 – 640, March 1999.

[12] M. Rupf and J. L. Massey, "Optimal sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Info. Theory*, vol. 40, pp. 1261 – 1266, July 1994.

[13] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Trans. Info. Theory*, vol. 45, pp. 1984 – 1991, Sept. 1999.

[14] P. Viswanath and V. Anantharam, "Optimal sequences for CDMA under colored noise: a Schur-saddle function property," *IEEE Trans. Info. Theory*, vol. 48, pp. 1295 – 1318, June 2002.

[15] S. Ulukus and R. D. Yates, "Iterative construction of optimum signature sequence sets in synchronous CDMA systems," *IEEE Trans. Info. Theory*, vol. 47, pp. 1989 – 1998, July 2001.

[16] J. A. Tropp, I. S. Dhillon, and R. W. Heath, "Finite-step algorithms for constructing optimal CDMA signature sequences," *IEEE Trans. Info. Theory*, vol. 50, pp. 2916 – 2921, Nov. 2004.

[17] R. W. Heath, T. Strohmer, and A. J. Paulraj, "On quasi-orthogonal signatures for CDMA systems," *IEEE Trans. Info. Theory*, vol. 52, pp. 1217 – 1226, Mar. 2006.

[18] J. Luo, S. Ulukus, and A. Ephremides, "Optimal sequence and sum capacity of symbol asynchronous CDMA systems," *IEEE Trans. Info. Theory*, vol. 51, pp. 2760 – 2769, Aug. 2005.

[19] T. Guess, "CDMA with power control and sequence design: the capacity region with and without multidimensional signaling," *IEEE Trans. Info. Theory*, vol. 50, pp. 2604 – 2619, Nov. 2004.

[20] A. M. Tulino and S. Verdu, *Random matrix theory and wireless communications*, ch. 3.1. Now Publishers Inc, June 2004.

[21] S. Buzzi, H. V. Poor, and A. Zappone, "Transmitter waveform and widely linear receiver design: Noncooperative games for wireless multiple-access networks," *IEEE Trans. Info. Theory*, vol. 56, pp. 4874 – 4892, Oct. 2010.

[22] M. Simon, J. Omura, R. Scholtz, B. Levitt, J. K. Omura, R. A. Scholtz, and M. K. Simon, *Spread spectrum communications handbook*. McGraw-Hill Professional, Sept. 2001.

[23] M. K. Varanasi and B. Aazhang, "Near-optimal detection in synchronous code-division multiple-access systems," *IEEE Trans. Comm.*, vol. 39, pp. 725 – 736, May 1991.

[24] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code division multiple access communications," *IEEE Trans. Comm.*, vol. 38, pp. 509 – 519, April 1990.

[25] M. K. Varanasi, "Decision feedback multiuser detection: A systematic approach," *IEEE Trans. Info. Theory*, vol. 45, pp. 219 – 240, Jan. 1999.

[26] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE Journal On Selected Areas in Comm.*, vol. 8, pp. 641 – 649, May 1990.

[27] A. Duel-Hallen, "Decorrelating decision feedback multiuser detector for synchronous code-division multiple-access channel," *IEEE Trans. Comm.*, vol. 41, pp. 285 – 290, Feb. 1993.

[28] T. Cover, "Some advances in broadcoast channels," in *Advances in Communication Systems*, New York, NY, USA: Academic, 1975.

[29] S. Verdu, "Near-far resistant receivers for DS/SSMA communications," tech. rep., U.S. Army Research Proposal, Contract DAAL03-87-K-0062, Princeton University, 1986.

[30] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimal detectors for coherent multiuser communications," *IEEE Journal On Selected Areas in Comm.*, vol. 8, pp. 683 – 690, May 1990.

[31] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Comm.*, vol. 42, pp. 3221 – 3231, Dec. 1994.

[32] H. V. Poor and S. Verdu, "Probability of error in MMSE multiuser detector," *IEEE Trans. Info. Theory*, vol. 43, pp. 858 – 871, May 1997.

[33] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse wideband analog signals," *IEEE Sel. Topics Signal Process.*, vol. 56, pp. 4692 – 4702, Oct. 2008.

[34] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Info. Theory*, vol. 55, pp. 5302 – 5316, Nov. 2009.

[35] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, pp. 2986–2997, August 2009.

[36] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, pp. 993 – 1009, Mar. 2009.

[37] Y. C. Eldar, "Uncertainty relations for shift-invariant analog signals," *IEEE Trans. Info. Theory*, vol. 55, pp. 5742 – 5757, Dec. 2009.

[38] 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.

[39] K. Gedalyahu and Y. C. Eldar, "Time-delay estimation from low-rate samples: A union of subspaces approach," *IEEE Trans. on Signal Process.*, vol. 58, pp. 3017 – 3031, June 2010.

[40] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: from theory to applications," *submitted to IEEE Trans. Signal Process. and arXived*, 2011.

[41] R. Gribonval, B. Mailhe, H. Rauhut, K. Schnass, and P. Vandergheynst, "Average case analysis of multichannel thresholding," in *Proc. IEEE ICASSP07*, (Honolulu), 2007.

[42] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, pp. 265 – 274, 2009.

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

[44] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231 – 2242, Oct. 2004.

[45] W.-C. Wu and K.-C. Chen, "Identification of active users in synchronous CDMA multiuser detection," *IEEE Journal On Selected Areas in Comm.*, vol. 16, pp. 1723 – 1735, Dec. 1998.

[46] T. Oskiper and H. V. Poor, "Online activity detection in a multiuser environment using a matrix CUSUM algorithm," *IEEE Trans. Info. Theory*, vol. 48, pp. 477 – 493, Feb. 2002.

[47] E. Biglieri and M. Lops, "Multiuser detection in dynamic environment - Part I: user identification and data detection," *IEEE Trans. Info. Theory*, vol. 53, pp. 3158 – 3170, Sept. 2007.

[48] D. Angelosante, E. Biglieri, and M. Lops, "Multiuser detection in dynamic environment - Part II: joint user identification and parameter estimation," *IEEE Trans. Info. Theory*, vol. 55, pp. 2365 – 2374, May 2009.

[49] D. Angelosante, E. Biglieri, and M. Lops, "Low-complexity receivers for multiuser detection with an unknown number of active users," *Signal Processing*, vol. 90, pp. 1486 – 1495, May 2010.

[50] A. T. Campo, A. G. Fabregas, and E. Biglieri, "Large-system analysis of multiuser detection with an unknown number of users: A high-SNR approach," *IEEE Trans. Info. Theory*, vol. 57, pp. 3416 – 3428, June 2011.

[51] G. Ricci and M. K. Varanasi, "Blind multiuser detection of L out of K active users of an N user synchronous CDMA system," in *IEEE 49th Vehicular Tech. Conf.*, pp. 1445 – 1449, May 1999.

[52] M. K. Varanasi, "Group detection for synchronous Gaussian code-division multiple-access channels," *IEEE Trans. Info. Theory*, vol. 41, pp. 1083 – 1096, July 1995.

[53] 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.

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

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

[56] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," *IEEE Trans. Info. Theory*, vol. 55, pp. 5758 – 5772, Jan. 2009.

[57] Y. Jin, Y.-H. Kim, and B. D. Rao, "Support recovery of sparse signals," *submitted to IEEE Trans. Info. Theory and arXived*, March 2010.

[58] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Info. Theory*, vol. 52, pp. 4036 – 4080, Sept. 2006.

[59] J. Haupt and R. Nowak, "Compressive sampling for signal detection," *IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 3, pp. 1509–1512, 2007.

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

[61] L. Applebaum, W. Bajwa, M. F. Duarte, and R. Calderbank, "Multiuser detection in asynchronous on-off random access channels using lasso," in *Proc. 48th Annu. Allerton Conf. Comm., Control, and Computing*, (Monticello, IL), Sept. 2010.

[62] L. Applebaum, W. Bajwa, M. Duarte, and R. Calderbank, "Asynchronous code-division random access using convex optimization," *Submitted for journal publication and arXived*, Jan. 2011.

[63] J. Haupt and R. Nowak, "A generalized restricted isometry property," tech. rep., Univ. of Wisconsin, Department of ECE, 2007.

[64] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE Journal On Selected Areas in Comm.*, vol. 26, pp. 421 – 431, April 2008.

[65] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, pp. 5030 – 5043, Oct. 2010.

[66] S. Verdu, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, pp. 303 – 312, 1989.

[67] T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," Tech. Rep. [Online]. Available: http://www-stat.wharton.upenn.edu/ tcai/paper/Stable-Recovery-MIP.pdf, U. Penn., 2009.

[68] G. Reeves and M. Gastpar, "A note on optimal support recovery in compressed sensing," in *Forty-Third Asilomar Conf. on Sig., Sys, and Computers*, pp. 1576 – 1580, Nov. 2009.

[69] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Info. Theory*, vol. 56, pp. 1383 – 1399, Mar. 2010.

[70] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. on Signal Process.*, vol. 58, pp. 3042 – 3054, June 2010.

[71] R. Lupas and S. Verdu, "Asymptotic efficiency of linear multiuser detectors," in *Proc. of 25th Conf. on Decision and Control (CDC)*, pp. 2094 – 2100, Dec. 1986.

[72] M. Fornasier and H. Rauhut, "Compressive sensing," in *Handbook of Mathematical Methods in Imaging* (O. Scherzer, ed.), ch. 2, Springer, 2011.

[73] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. of Amer. Stat. Asso.*, vol. 58, pp. 13 – 30, March 1963.

[74] R. Calderbank, S. Jafarpour, and J. Kent, *Finding needles in compressed haystacks*, ch. 10. Compressed Sensing: Theory and Applications, Cambridge University Press, Apr. 2012.

[75] Z. Sidak, "Rectangular confidence regions for the means of multivariate normal distributions," *J. of Amer. Stat. Asso.*, vol. 62, pp. 626 – 633, Jun. 1967.