Reduced-Dimension Multiuser Detection

Yao Xie, *Member, IEEE*, Yonina C. Eldar, *Fellow, IEEE*, and Andrea Goldsmith, *Fellow, IEEE*

***Abstract*—We present a reduced-dimension multiuser detector (RD-MUD) structure for synchronous systems 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, in some wireless systems, the number of active users may be small relative to the total number of users in the system. Hence, the ideas of analog compressed sensing may be used 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. We derive the probability of symbol 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 matching pursuit for active user detection and sign detection for data recovery. We derive probability of error bounds for both detectors, and show that the number of correlators needed to achieve a small probability of symbol error is on the order of the logarithm of the number of users in the system. The theoretical performance results are validated via numerical simulations.**

***Index Terms*—Compressed sensing, demodulation, multiuser detection.**

I. Introduction

MULTIUSER detection (MUD) is a classical problem in multiuser communications and signal processing (see, e.g., [1]–[3] and the references therein.) In multiuser systems, the users communicate simultaneously with a given receiver by modulating information symbols onto their unique signature waveforms. The received signal consists of a noisy version of the superposition of the transmitted waveforms. MUD has to detect the symbols of all users simultaneously.

Despite the large body of work on MUD, it is not yet widely implemented in practice, largely due to its complexity and high-precision analog-to-digital (A/D) requirements. The complexity arises both in the A/D as well as in the digital signal processing for data detection of each user. A conventional MUD structure consists of a matched-filter (MF) bank front end followed by a linear or nonlinear digital MUD. The MF-bank is a set of correlators, each correlating the received signal with the signature waveform of a different user. The number of correlators is therefore equal to the number of users. In a typical communication system, there may be thousands of users. We characterize the A/D complexity by the number of correlators at the receiver front end, and measure data detection complexity by the number of real floating point operations required per decision bit [4] from the MF-bank output.

Verdú, in the landmark paper [5], established the maximum likelihood sequence estimator (MLSE) as the MUD detector minimizing the probability of symbol error for data detection. However, the complexity per bit of MLSE is exponential in the number of users when the signature waveforms are nonorthogonal. To address the complexity issue, other low-complexity suboptimal detectors have been developed, including the nonlinear decision feedback (DF) detector [6] and linear detectors. The nonlinear DF detector is based on the idea of interference cancellation, which decodes symbols iteratively by subtracting the detected symbols of strong users first to facilitate detection of weak users. The DF detector is a good compromise between complexity and performance (see, e.g., [6]). We will therefore analyze the DF detector below as an example of a nonlinear digital detector, but in a reduced dimensional setting.

Linear detectors apply a linear transform to the receiver front-end output and then detect each user separately. They have lower complexity than nonlinear methods but also worse performance. There are multiple linear MUD techniques. The single-user detector is the simplest linear detector; however, it suffers from user interference when signature waveforms are nonorthogonal. A linear detector that eliminates user inference is the decorrelating detector, 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 provides the best joint estimate of symbols and amplitudes in the absence of knowledge of the complete channel state information [4]. The minimum mean-square-error (MMSE) detector takes into account both background noise and interference, and hence to some extent mitigates the noise amplification of the decorrelating detector in low and medium SNR [6]. Because of the simplicity and interference elimination capability of the decorrelating detector, we will focus on this

Manuscript received October 13, 2011; revised June 10, 2012; accepted January 09, 2013. Date of publication February 25, 2013; date of current version May 15, 2013. This work was supported in part by the Interconnect Focus Center of the Semiconductor Research Corporation, in part by the Binational Science Foundation Transformative Science under Grant 2010505, in part by the Air Force Office of Scientific Research under Grant FA9550-08-1-0010, and in part by a General Yao-Wu Wang Stanford Graduate Fellowship. This paper was presented in part at the 48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, July 2010, and in part at the 2012 IEEE International Conference on Communications.

Y. Xie was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. She is now with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: yao.c.xie@gmail.com).

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

A. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@wsl.stanford.edu).

Communicated by D. P. Palomar, Associate Editor for Detection and Estimation.

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

Digital Object Identifier 10.1109/TIT.2013.2248873

0018-9448/$31.00 © 2013 IEEE

detector as an example of a linear detector in the reduced-dimensional setting.

In many applications, the number of active users K can be much smaller than the total number of users N[7], [8]. This analog signal sparsity allows the use of techniques from analog compressed sensing [9], [10] in order to reduce the number of correlators. While such sparsity has been exploited in various detection settings, there is still a gap in applying these ideas to the multiuser setting we consider here. Most existing work on exploiting compressed sensing [11], [12] for signal detection assumes discrete signals and then applies compressed sensing via matrix multiplication [8], [13]–[16]. In contrast, in MUD, the received signal is continuous. Another complicating factor relative to previous work is that here noise is added in the analog domain prior to A/D conversion at the front end, which corresponds to the measurement stage in compressed sensing. Therefore, A/D conversion will affect both signal and noise. Due to the MFs at the front end, the output noise vector is in general colored. Furthermore, it cannot be whitened without modifying the MF coefficient matrix, which corresponds to the measurement matrix in compressed sensing. In the discrete compressed sensing literature, it is usually assumed that white noise is added after measurement. An exception is the work of [17]. Finally, typically in compressed sensing, the goal is to *reconstruct* a sparse signal from its samples, whereas in MUD, the goal is to *detect* both active users and their symbols. To meet the goal of MUD, we therefore adapt algorithms from compressed sensing for detection and develop results on the probability of symbol error rather than on the mean-squared error (MSE).

In this study, we develop a low-complexity MUD structure which we call a reduced-dimension multiuser detector (RD-MUD) exploiting analog signal sparsity, assuming symbol-rate synchronization. The RD-MUD reduces the front-end receiver complexity by decreasing the number of correlators without increasing the complexity of digital signal processing, while still achieving performance similar to that of conventional MUDs that are based on the MF-bank front end. The RD-MUD converts the received analog signal into a discrete output by correlating it with $M \ll N$ correlating signals. We incorporate analog compressed sensing techniques [9] by forming the correlating signals as linear combinations of the signature waveforms via a (possibly complex) coefficient matrix $\boldsymbol{A}$. The RD-MUD output can thus be viewed as a projection of the MF-bank output onto a lower dimensional detection subspace. We then develop several digital detectors to detect both active users and their transmitted symbols, by combining ideas from compressed sensing and 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 [18], [19], and the reduced-dimension decision-feedback (RDDF) detector, a nonlinear detector that combines decision-feedback matching pursuit (DF-MP) [20], [21] for active user detection with sign detection for data recovery in an iterative manner. DF-MP differs from conventional MP [20], [21] 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.

We provide probability-of-symbol-error performance bounds for these detection algorithms, using the coherence of the matrix $\boldsymbol{A}$ in a nonasymptotic regime with a fixed number of total 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-symbol-error performance. When $\boldsymbol{A}$ is a random partial discrete Fourier transform (DFT) matrix, the M required by these two specific detectors is on the order of $\log N$ as compared to N correlators required for conventional MUD. We validate these theoretical results via numerical examples. Our analysis is closely related to [22]. However, Ben-Haim *et al.* [22] considers estimation in white noise, which differs from our problem in the aforementioned aspects. Our work also differs from prior results on compressed sensing for MUD, such as Applebaum *et al.* [7] and Fletcher *et al.* [8], [23], where a so-called on-off random access channel is considered. In these references, the goal is to detect which users are active, and there is no need to detect the transmitted symbols as we consider here. Neither of these works consider front-end complexity.

In this paper, we focus on a synchronous MUD channel model [4], where the transmission rate of all users is the same and their symbol epochs are perfectly aligned. This user synchronization can be achieved using GPS as well as distributed or centralized synchronization schemes (see, e.g., [24] and [25]). Such methods are commonly used in cellular systems, ad hoc networks, and sensor networks to achieve synchronization. Part of the MUD problem is signature sequence selection, for which there has also been a large body of work (see, e.g., [26]). Here, we do not consider optimizing signature waveforms so that our results are parameterized by the cross-correlation properties of the signature waveforms used in our design.

The rest of this 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.

II. Background

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 represented by $\mathbb{R}$ and $\mathbb{C}$, respectively. The real part of a scalar x is denoted as $\Re[x]$, and x^* is its conjugate. The set of indices of the nonzero entries of a vector $\boldsymbol{x}$ is called the support of $\boldsymbol{x}$. Given an index set $\mathcal{I}$, $\boldsymbol{X}_{\mathcal{I}}$ denotes the submatrix formed by the columns of a matrix $\boldsymbol{X}$ indexed by $\mathcal{I}$, and $\boldsymbol{x}_{\mathcal{I}}$ represents the subvector formed by the entries indexed by $\mathcal{I}$. The identity matrix is denoted by $\boldsymbol{I}$. The transpose, conjugate transpose, and inverse of a matrix $\boldsymbol{X}$ are represented by $\boldsymbol{X}^\top$, $\boldsymbol{X}^H$, and $\boldsymbol{X}^{-1}$, respectively, and $\boldsymbol{X}_{nm}$ denotes its nmth value. The ℓ_2 norm is denoted by $\|\boldsymbol{x}\| = (\boldsymbol{x}^H\boldsymbol{x})^{1/2}$, and the ℓ_∞ norm of a vector $\boldsymbol{x}$ is given by

$\|\boldsymbol{x}\|_\infty = \max_{n=1}^N |x_n|$. The minimum and maximum eigenvalues of a positive-semidefinite matrix $\boldsymbol{X}$ are represented by $\lambda_{\min}(\boldsymbol{X})$ and $\lambda_{\max}(\boldsymbol{X})$, respectively. The trace of a square matrix $\boldsymbol{X}$ is denoted as $\mathrm{tr}(\boldsymbol{X})$. The notation $\mathrm{diag}\{x_1, \ldots, x_n\}$ denotes a diagonal matrix with $x_1, \ldots, x_n$ on its diagonal. We use $\boldsymbol{I}$ to denote the identity matrix and $\mathbf{1}$ to denote an all-one vector.

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 represented as P. 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 represented as $\{A\}^c$. The notation $A \subset B$ means that set A is a subset of B. The *inner product* (or *cross correlation*) between two real analog signals $x(t)$ and $y(t)$ in L_2 is written as $\langle x(t), y(t)\rangle = \frac{1}{T}\int_0^T x(t)y(t)dt$, over the symbol time T. The L_2 norm of $x(t)$ is $\|x(t)\| = \langle x(t), x(t)\rangle^{1/2}$. Two signals are orthogonal if their cross correlation is zero.

B. System Model

Consider a synchronous multiuser system [1] with N users. Each user is assigned a known unique signature waveform from a set $\mathcal{S} = \{s_n(\cdot) : [0, T] \to \mathbb{R},\ 1 \le n \le N\}$. Users modulate their data by their signature waveforms. There are K active users out of N possible users transmitting to the receiver. In our setting, we assume that active users modulate their signature waveforms using binary phase shift keying modulation with the symbol of user n denoted by $b_n \in \{1, -1\}$, for $n \in \mathcal{I}$, where $\mathcal{I}$ contains indices of all active users. The amplitude of the nth user's signal at the receiver is given by r_n, which is determined by the transmit power and the wireless channel gain. For simplicity, we assume r_n's are real (but they can be negative), and known at the receiver. 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

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

with $b_n \in \{1, -1\}$, $n \in \mathcal{I}$, and $b_n = 0$, $n \in \mathcal{I}^c$. 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.

We assume that the signature waveforms are linearly independent. The cross correlations of the signature waveforms are characterized by the Gram matrix $\boldsymbol{G}$, defined as

$$[\boldsymbol{G}]_{n\ell} \triangleq \langle s_n(t), s_\ell(t)\rangle, \qquad 1 \le n, \ell \le N. \tag{2}$$

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

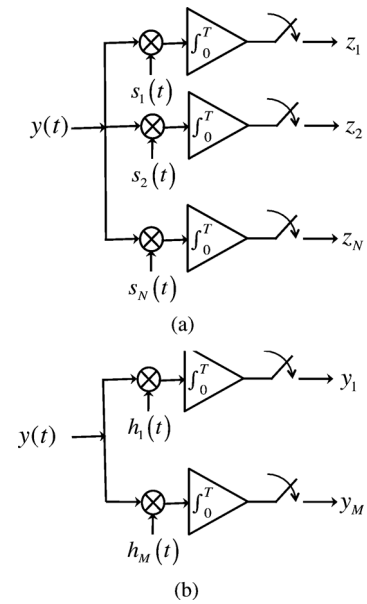


Fig. 1. Front end of (a) the conventional MUD using MF-bank and (b) RD-MUD.

Our goal 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 analog signal sparsity. As we will show, this 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. We will consider two scenarios: the case where K is known, and the case where K is bounded but unknown. The problem of estimating K can be treated using techniques such as those in [27].

C. Conventional MUD

A conventional MUD detector has an MF-bank front-end followed by a digital detector. We now review this architecture.

1) MF-Bank Front-End: For general single-user systems, the MF multiplies the received signal $y(t)$ with the single-user waveform $s(t)$ and integrates over a symbol time. The MF-bank is an extension of the MF for multiple users, and has N MFs in parallel: the nth branch correlates the received signal with the corresponding signature waveform $s_n(t)$, as illustrated in Fig. 1(a). The output of the MF-bank is a set of sufficient statistics for MUD when the amplitudes r_n are known [1]. Using (1), the output of the MF-bank can be written as

$$\boldsymbol{z} = \boldsymbol{G}\boldsymbol{R}\boldsymbol{b} + \boldsymbol{u} \tag{3}$$

where $\boldsymbol{z} = [z_1, \ldots, z_N]^\top$, $z_n = \langle y(t), s_n(t)\rangle$, $\boldsymbol{R} \in \mathbb{R}^{N\times N}$ is a diagonal matrix with $[\boldsymbol{R}]_{nn} = r_n$, $\boldsymbol{b} = [b_1, \ldots, b_N]^\top$ and $\boldsymbol{u} = [u_1, \ldots, u_N]^\top$, where $u_n = \langle w(t), s_n(t)\rangle$. The vector $\boldsymbol{u}$ is Gaussian distributed with zero mean and covariance $\mathbb{E}\{\boldsymbol{u}\boldsymbol{u}^H\} = \sigma^2\boldsymbol{G}$ (for derivation, see [1]).

2) MF-Bank Detection: Conventional MUD detectors based on the MF-bank output can be classified into two categories: linear and nonlinear, as illustrated in Fig. 2. In the literature, the synchronous MUD model typically assumes all users are active,

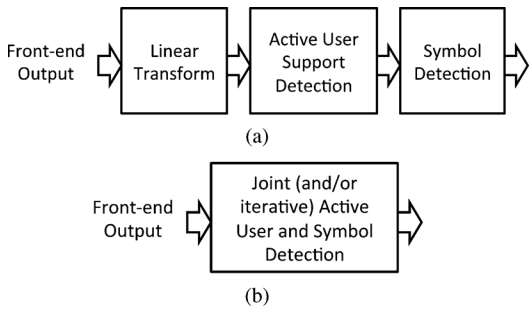


Fig. 2. Diagram of (a) the linear detector and (b) nonlinear detector.

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 $\boldsymbol{T}$ to the MF-bank output [illustrated in Fig. 2(a)]

$$\boldsymbol{Tz} = \boldsymbol{TGRb} + \boldsymbol{Tu} \tag{4}$$

and detects each user's symbol separately using a sign detector

$$\hat{b}_n = \mathbf{sgn}(r_n[\boldsymbol{Tz}]_n), \qquad 1 \le n \le N. \tag{5}$$

Several commonly used linear detectors are the single-user detector, the decorrelating detector, and the MMSE detector. The single-user detector [4] is equivalent to choosing $\boldsymbol{T} = \boldsymbol{I}$ in (4) and (5). By applying a linear transform $\boldsymbol{T} = \boldsymbol{G}^{-1}$ in (4), the decorrelating detector can remove the user interference and recover symbols perfectly in the absence of noise; however, it also amplifies noise when $\boldsymbol{G} \neq \boldsymbol{I}$. The MMSE detector minimizes the MSE between the linear transform of the MF-bank output and symbols, and corresponds to $\boldsymbol{T} = (\boldsymbol{G} + \sigma^2 \boldsymbol{R}^{-2})^{-1}$ in (4)[1].

Nonlinear detectors, on the other hand, detect symbols jointly and (or) iteratively, as illustrated in Fig. 2(b). Examples include MLSE and the successive interference cancellation (SIC) detector [1]. The MLSE solves the following optimization problem:

$$\max_{b_n \in \{1,-1\}} 2\boldsymbol{y}^H \boldsymbol{Rb} - \boldsymbol{b}^H \boldsymbol{RGRb}. \tag{6}$$

However, when the signature waveforms are nonorthogonal, this optimization problem is exponentially complex in the number of users [1]. The SIC detector first finds the active user with the largest amplitude, detects its symbol, subtracts its effect from the received signal, and iterates the aforementioned process using the residual signal. After N iterations, the SIC detector determines all users.

III. RD-MUD

The RD-MUD front end, illustrated in Fig. 1(b), correlates the received signal $y(t)$ with a set of correlating signals $h_m(t)$, $m = 1, \ldots M$, where M is typically much smaller than N. The front-end output is processed by either a linear or nonlinear detector to detect active users and their symbols; the design of these detectors is adapted to take the analog sparsity into account.

A. RD-MUD Front End

Design of the correlating signals $h_m(t)$ is key for RD-MUD to reduce the number of correlators. To construct these signals, we rely on the ideas introduced in [9] to construct multichannel filters that sample structured analog signals at sub-Nyquist rates. Specifically, we use the biorthogonal signals with respect to $\{s_n(t)\}$, which are defined as

$$\hat{s}_n(t) = \sum_{\ell=1}^{N} [\boldsymbol{G}^{-1}]_{n\ell} s_\ell(t), \qquad 1 \le n \le N. \tag{7}$$

These signals have the property that $\langle s_n(t), \hat{s}_m(t) \rangle = \delta_{nm}$, for all n, m. Note that when $\{s_n(t)\}$ are orthogonal, $\boldsymbol{G} = \boldsymbol{I}$ and $\hat{s}_n(t) = s_n(t)$.

The RD-MUD front end uses as its correlating signals the functions

$$h_m(t) = \sum_{n=1}^{N} a_{mn} \hat{s}_n(t), \qquad 1 \le m \le M \tag{8}$$

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

$$\mu \triangleq \max_{n \neq \ell} \left| \boldsymbol{a}_n^H \boldsymbol{a}_\ell \right|. \tag{9}$$

As we will show later in Section IV-A, it is desirable that μ is small to guarantee small probability of symbol error. This requirement also reflects a tradeoff in choosing how many correlators to use in the RD-MUD front end. With more correlators, the coherence of $\boldsymbol{A}$ can be lower and the performance of RD-MUD improves. Choosing the correlating signals (8) and using the receive signal model (1), the output of the mth correlator is given by

$$\begin{aligned} y_m &= \langle h_m(t), y(t) \rangle \\ &= \left\langle \sum_{n=1}^{N} a_{mn} \hat{s}_n(t), \sum_{\ell=1}^{N} r_\ell b_\ell s_\ell(t) \right\rangle \\ &\quad + \left\langle \sum_{n=1}^{N} a_{mn} \hat{s}_n(t), w(t) \right\rangle \\ &= \sum_{\ell=1}^{N} r_\ell b_\ell \sum_{n=1}^{N} a_{mn} \langle \hat{s}_n(t), s_\ell(t) \rangle + w_m \\ &= \sum_{\ell=1}^{N} a_{m\ell} r_\ell b_\ell + w_m \end{aligned} \tag{10}$$

where the output noise is given by $w_m \triangleq \sum_{n=1}^{N} a_{mn} \langle \hat{s}_n(t), w(t) \rangle$. Denoting $\boldsymbol{y} = [y_1, \ldots, y_M]^\top$ and $\boldsymbol{w} = [w_1, \ldots, w_M]^\top$, we can express the RD-MUD output (10) in vector form as

$$\boldsymbol{y} = \boldsymbol{ARb} + \boldsymbol{w} \tag{11}$$

where $\boldsymbol{w}$ is a Gaussian random vector with zero mean and covariance $\sigma^2 \boldsymbol{AG}^{-1} \boldsymbol{A}^H$ (for derivation, see [28] and [29]). The vector $\boldsymbol{y}$ 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, $\boldsymbol{b}$ has at most K nonzero entries. The idea of RD-MUD is that when the original signal vector $\boldsymbol{b}$ is sparse, with proper choice of the matrix $\boldsymbol{A}$, the detection performance for $\boldsymbol{b}$ based on $\boldsymbol{y}$ of (11) in the detection subspace can be similar to the performance based on the output of the MF-bank front end $\boldsymbol{z}$ of (3).

B. RD-MUD Detection

We now discuss how to recover $\boldsymbol{b}$ from the RD-MUD front-end output using digital detectors. The model (11) for RD-MUD has a similar form to the observation model in the compressed sensing literature [13], [22], except that the noise in the RD-MUD front-end output is colored. Hence, to recover $\boldsymbol{b}$, we can combine ideas developed in the context of compressed sensing and MUD. The linear detector for RD-MUD first estimates active users $\hat{\mathcal{I}}$ using support recovery techniques from compressed sensing. Once the active users are estimated, we can write the RD-MUD front-end output model (11) as

$$\boldsymbol{y} = \boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}}\boldsymbol{b}_{\hat{\mathcal{I}}} + \boldsymbol{w} \tag{12}$$

from which we can detect the symbols $\boldsymbol{b}_{\hat{\mathcal{I}}}$ by applying a linear transform to $\boldsymbol{y}$. 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 RDD detector, a linear detector that uses subspace projection along with thresholding [13], [19] to determine active users and sign detection for data recovery; 2) the RDDF detector, a nonlinear detector that combines DF-MP for active user detection and sign detection for data recovery. These two algorithms are summarized in Algorithms 1 and 2.

Algorithm 1 RDD Detector

1:**Input:** An $M \times N$ matrix $\boldsymbol{A}$, a vector $\boldsymbol{y} \in \mathbb{C}^M$ and the number of active users K.

2:Detect active users: find $\hat{\mathcal{I}}$ that contains indices of the K largest $|\Re[\boldsymbol{a}_n^H\boldsymbol{y}]|$.

3:Detect symbols: $\hat{b}_n = \mathbf{sgn}(r_n\Re[\boldsymbol{a}_n^H\boldsymbol{y}])$ for $n \in \hat{\mathcal{I}}$, and $\hat{b}_n = 0$ for $n \notin \hat{\mathcal{I}}$.

Algorithm 2 RDDF Detector

1:**Input:** An $M \times N$ matrix $\boldsymbol{A}$, a vector $\boldsymbol{y} \in \mathbb{C}^M$ and number of active users K.

2:Initialize: $\mathcal{I}^{(0)}$ is empty, $\boldsymbol{b}^{(0)} = 0$, $\boldsymbol{v}^{(0)} = \boldsymbol{y}$.

3:Iterate Steps 4–6 for $k = 1, \ldots, K$:

4:Detect active user: $n_k = \arg\max_n |\Re[\boldsymbol{a}_n^H\boldsymbol{v}^{(k-1)}]|$.

5:Detect symbol: $b_n^{(k)} = \mathbf{sgn}(r_{n_k}\Re[\boldsymbol{a}_{n_k}^H\boldsymbol{v}^{(k-1)}])$, for $n = n_k$, and $b_n^{(k)} = b_n^{(k-1)}$ for $n \neq n_k$.

6:Update: $\mathcal{I}^{(k)} = \mathcal{I}^{(k-1)} \cup \{n_k\}$, and $\boldsymbol{v}^{(k)} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{R}\boldsymbol{b}^{(k)}$.

7:Output: $\hat{\mathcal{I}} = \mathcal{I}^{(K)}$, $\hat{\boldsymbol{b}} = \boldsymbol{b}^{(K)}$.

1) RDD Detector: A natural strategy for detection is to compute the inner product $\boldsymbol{a}^H\boldsymbol{y}$ and detect active users by choosing indices corresponding to the K largest magnitudes of these inner products

$$\begin{aligned}\hat{\mathcal{I}} = \{n: \quad & \text{if } |\Re[\boldsymbol{a}_n^H\ \boldsymbol{y}]| \text{ is among the} \\ & K \text{ largest of } |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]|,\ n = 1, \cdots, N\}.\end{aligned} \tag{13}$$

This corresponds to the thresholding support recovery algorithm in compressed sensing (e.g., [13]). To detect symbols, we use sign detection

$$\hat{b}_n = \begin{cases} \mathbf{sgn}\left(r_n\Re[\boldsymbol{a}_n^H\boldsymbol{y}]\right), & n \in \hat{\mathcal{I}} \\ 0, & n \notin \hat{\mathcal{I}}. \end{cases} \tag{14}$$

In detecting active users (13) and their symbols (14), we take the real parts of the inner products because the imaginary part of $\boldsymbol{a}_n^H\boldsymbol{y}$ contains only noise and interference, since we assume that symbols b_n and amplitudes r_n are real and only $\boldsymbol{A}$ can be complex. When $K = N$ and $M = N$, the RDD detector becomes the decorrelator in conventional MUD.

To compute the complexity per bit of the RDD detector, we note that computing $\boldsymbol{A}^H\boldsymbol{y}$ requires MN floating point operations when $\boldsymbol{A}$ is real (or $2MN$ operations when $\boldsymbol{A}$ 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 RDD is proportional to M. Since $M \leq N$ in RD-MUD, the complexity-per-bit of RDD (and other RD-MUD linear detectors as well) is lower than that of the conventional decorrelating linear MUD detector. Furthermore, RDD and other linear RD-MUD detectors require much lower complexity in the analog front end.

When the number of users is not known, we can replace Step 2 in Algorithm 1 by

$$\hat{\mathcal{I}} = \{n \in \{1, \ldots, N\} : |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]| > \xi\} \tag{15}$$

where $\xi > 0$ is a chosen threshold. We refer to this method as the RDD threshold (RDDt) detector. The RDDt detector is related to the OST algorithm for model selection in [30]. The choice of ξ depends on r_n, σ^2, M, N, μ, and the maximum eigenvalue of $\boldsymbol{G}^{-1}$. Bounds on ξ associated with error probability bounds will be given in Theorem 1. In Section V, we explore numerical optimization of ξ, where we find that to achieve good performance, ξ should increase with N or K, and decrease with M.

2) 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 $\mathcal{I}^{(0)}$, zeros as the estimated symbol vector $\boldsymbol{b}^{(0)} = \boldsymbol{0}$, and the front-end output as the residual vector $\boldsymbol{v}^{(0)} = \boldsymbol{y}$. Subsequently, in each iteration $k = 1, \ldots, K$, the algorithm selects the column $\boldsymbol{a}_n$ that is most highly correlated with the residual $\boldsymbol{v}^{(k-1)}$ as the detected active user in the kth iteration

$$n_k = \arg\max_n \left|\Re[\boldsymbol{a}_n^H\boldsymbol{v}^{(k-1)}]\right| \tag{16}$$

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

$$b_n^{(k)} = \begin{cases} \mathbf{sgn}\left(r_{n_k}\Re[\boldsymbol{a}_{n_k}^H\boldsymbol{v}^{(k-1)}]\right), & n = n_k \\ b_n^{(k-1)}, & n \neq n_k. \end{cases} \tag{17}$$

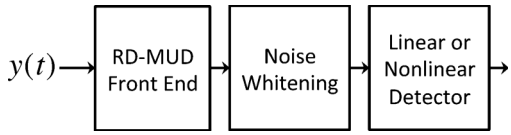


Fig. 3. MUD detector with noise whitening transform.

The residual vector is then updated through $\boldsymbol{v}^{(k)} = \boldsymbol{y} - \boldsymbol{ARb}^{(k)}$. The residual vector represents the part of $\boldsymbol{b}$ 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-MP never detects the same active user twice), and finally, the active user set is given by $\hat{\mathcal{I}} = \mathcal{I}^{(K)}$ with the symbol vector $\hat{b}_n = b_n^{(K)}$, $n = 1, \ldots, N$. When $K = N$ and $M = N$, the RDDF detector becomes the successive interference cancellation technique in the conventional MUD.

The complexity per bit of RDDF is proportional to KM. Since $M \leq N$, this implies that the complexity for data detection of RDDF is lower than that of the conventional DF detector (the complexity per bit of the DF detector is proportional to KN). Note that RDDF is similar to MP in compressed sensing but with symbol detection.

We can modify the RDDF detector to account for an unknown number of users by iterating only when the residual does not contain any significant "component," i.e., when $\|A^H \boldsymbol{v}^{(k)}\|_\infty < \epsilon$ for some threshold $\epsilon > 0$. We refer to this method as the RDDF threshold (RDDFt) detector. The choice of ϵ depends on r_n, σ^2, M, N, μ, and the maximum eigenvalue of $\boldsymbol{G}^{-1}$. As with the threshold ξ, bounds on ϵ to ensure given error probability bounds, and its numerical optimization, are presented in Theorem 1 and Section V, respectively. As in the RDDt, here we also found in numerical optimizations that ϵ should increase with N or K, and decrease with M.

3) Noise Whitening Transform: The noise in the RD-MUD output (11) is in general colored due to the matched filtering at the front end. We can whiten the noise by applying a linear transform $(\boldsymbol{AG}^{-1}\boldsymbol{A}^H)^{-1/2}$ before detecting active users and symbols, as illustrated in Fig. 3. The whitened output is given by

$$\begin{aligned}\boldsymbol{y}_w &\triangleq (\boldsymbol{AG}^{-1}\boldsymbol{A}^H)^{-1/2}\boldsymbol{y} \\ &= (\boldsymbol{AG}^{-1}\boldsymbol{A}^H)^{-1/2}\boldsymbol{ARb} + \boldsymbol{w}_0\end{aligned} \tag{18}$$

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

$$\boldsymbol{A}_w \triangleq (\boldsymbol{AG}^{-1}\boldsymbol{A}^H)^{-1/2}\boldsymbol{A} \tag{19}$$

then the RDD and RDDF detectors can be applied by replacing $\boldsymbol{A}$ with $\boldsymbol{A}_w$ and $\boldsymbol{y}$ with $\boldsymbol{y}_w$ in (13), (14), (16), and (17). While whitening the noise, the whitening transform also distorts the signal component. As we demonstrate via numerical examples in Section V-6, noise whitening is beneficial when the signature waveforms are highly correlated.

4) Other RD-MUD Linear Detectors: By combining ideas developed in the context of compressed sensing and conventional linear MUD detection, we can develop alternative linear detectors in the reduced-dimension setting.

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 (12). The RD-MMSE detector determines active users through the support recovery method of (13), and then uses a linear transform $\boldsymbol{M}$ that minimizes $\mathbb{E}\{\|\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{My}\|^2\}$ to estimate the symbols. Here, the expectation is with respect to the vector of transmitted symbols $\boldsymbol{b}_{\hat{\mathcal{I}}}$ and the noise vector $\boldsymbol{w}$. Following the approach for deriving the conventional MMSE detector [1], assuming that $\boldsymbol{b}_{\hat{\mathcal{I}}}$ is uncorrelated with $\boldsymbol{w}$ and $\mathbb{E}\{\boldsymbol{b}_{\hat{\mathcal{I}}}\boldsymbol{b}_{\hat{\mathcal{I}}}^H\} = \boldsymbol{I}$, we obtain the linear transform for the reduced-dimension MMSE (RD-MMSE) detector as (see Appendix A for details): $\boldsymbol{M} = \boldsymbol{R}_{\hat{\mathcal{I}}}\boldsymbol{A}_{\hat{\mathcal{I}}}^H(\boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}}^2\boldsymbol{A}_{\hat{\mathcal{I}}}^H + \sigma^2\boldsymbol{AG}^{-1}\boldsymbol{A}^H)^{-1}$. The symbols are then determined as

$$\hat{b}_n = \begin{cases} \mathbf{sgn}(r_n\Re\{[\boldsymbol{My}]_n\}), & n \in \hat{\mathcal{I}} \\ 0, & n \notin \hat{\mathcal{I}}. \end{cases} \tag{20}$$

Similarly, we can modify RDDF by replacing symbol detection by (20) on the detected support $\mathcal{I}^{(k)}$ in each iteration.

Reduced-dimension least squares (RD-LS) detector: In the reduced-dimensional model (12), the matrix $\boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}}$ introduces interference when we detect the symbols. Borrowing from the idea of conventional MUD decorrelator, we can alleviate the effect of interference using the method of least squares, and estimate the symbols by solving $\hat{\boldsymbol{b}}_{\hat{\mathcal{I}}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}}\boldsymbol{x}\|^2$. We call this the RD-LS detector. Since $\mathbf{sgn}([\hat{\boldsymbol{b}}_{\hat{\mathcal{I}}}]_n) = \mathbf{sgn}([\boldsymbol{R}_{\hat{\mathcal{I}}}^2\hat{\boldsymbol{b}}_{\hat{\mathcal{I}}}]_n)$, RD-LS detects symbols by

$$\hat{b}_n = \begin{cases} \mathbf{sgn}\left(r_n\Re\left[(\boldsymbol{A}_{\hat{\mathcal{I}}}^H\boldsymbol{A}_{\hat{\mathcal{I}}})^{-1}\boldsymbol{A}_{\hat{\mathcal{I}}}^H\boldsymbol{y}\right]_n\right), & n \in \hat{\mathcal{I}} \\ 0, & n \notin \hat{\mathcal{I}}. \end{cases} \tag{21}$$

Similarly, we can modify RDDF by replacing symbol detection by (21) on detected support $\mathcal{I}^{(k)}$ in each iteration.

5) Reduced-Dimension Maximum Likelihood (RD-ML) Detector: The RD-ML detector finds the active users and symbols by solving the following integer optimization problem:

$$\begin{aligned}\max_{b_n \in \{-1,0,1\}} \; & 2\boldsymbol{y}^H(\boldsymbol{AG}^{-1}\boldsymbol{A}^H)^{-1}\boldsymbol{ARb} \\ & - \boldsymbol{b}^H\boldsymbol{RA}^H(\boldsymbol{AG}^{-1}\boldsymbol{A}^H)^{-1}\boldsymbol{ARb},\end{aligned} \tag{22}$$

where $b_n = 0$ corresponds to the nth user being inactive. Similar to the conventional maximum likelihood detector, the complexity per bit of the RD-ML is exponential in the number of users. We therefore do not consider this algorithm further.

C. Choice of $\boldsymbol{A}$

In Sections III-B1 and III-B2, we have shown that both the RDD and RDDF detectors are based on the inner products between $\boldsymbol{y}$ and the columns of $\boldsymbol{A}$. Since $\boldsymbol{y}$ consists of $\boldsymbol{a}_n$ corresponding to the active users plus noise, intuitively, for RDD and RDDF to work well, the inner products between columns of $\boldsymbol{A}$, or its coherence as defined in (9), should be small. Several commonly used random matrices in compressed sensing that have small coherence with high probability are the following.

1) Gaussian random matrices, where 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 a unitary matrix. For instance, the random partial DFT matrix, which is formed by randomly selecting rows of a DFT matrix $\boldsymbol{F}$: $[\boldsymbol{F}]_{nm} = e^{i\frac{2\pi}{N}nm}$ and normalizing the columns of the submatrix.
3) Kerdock codes [31]: these codes have dimension restricted to $M \times M^2$, where $M = 2^{m+1}$ with m an odd integer greater than or equal to 3. They have very good coherence properties, with $\mu = 1/\sqrt{M}$ which meets the Welch lower bound on coherence. The Welch bound imposes a general lower bound on the coherence of any $M \times N$ matrix $\boldsymbol{A}$ [32] leading to $\mu \gtrapprox M^{-1/2}$, when N is large relative to M and N is much larger than 1.

Among these three possible matrix choices, the random partial DFT matrix has some important properties that simplify closed-form analysis in some cases. In practice, if we choose the number of correlators to be equal to the number of users, i.e., $M = N$, then 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 $\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{I}$, i.e., $\boldsymbol{a}_n^H\boldsymbol{a}_m = \delta_{nm}$. In this case, the set of statistics $\{\boldsymbol{a}_n^H\boldsymbol{y}\}$ that RDD and RDDF are based on has the same distribution as the decorrelator output. To see this, write $\boldsymbol{a}_n^H\boldsymbol{y} = \boldsymbol{a}_n^H\left(\sum_{m=1}^N \boldsymbol{a}_m r_m b_m\right) + \boldsymbol{a}_n^H\boldsymbol{w} = r_n b_n + \boldsymbol{a}_n^H\boldsymbol{w}$, where $\boldsymbol{a}_n^H\boldsymbol{w}$ is a Gaussian random variable with zero mean and covariance $\sigma^2\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H\boldsymbol{a}_m = [\boldsymbol{G}^{-1}]_{nm}$. In contrast, the Gaussian random matrix does not have this property. Therefore, in this setting, the performance of RD-MUD using a Gaussian random matrix $\boldsymbol{A}$ is worse than that using the random partial DFT matrix. This is also validated in our numerical results in Section V-3. We will also see in Section V-3 that Kerdock codes outperform both random partial DFT and Gaussian random matrices for a large number of users. This is due to their good coherence properties. However, as discussed previously, Kerdock codes have restricted dimensions and are thus less flexible for system design.

IV. Performance of RD-MUD

We now study the performance of RD-MUD with the RDD and RDDF detectors. We begin by considering the case of a single active user without noise as a motivating example.

A. Single Active User

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 every two columns of $\boldsymbol{A}$ are linearly independent. Later, we will also show this is a corollary (Corollary 2) of 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 $K = 1$. In RD-MUD, with two correlators, the RDD detector determines the active user by finding

$$\hat{n}_0 = \arg\max_{n=1,\cdots,N} |a_{1n}\langle h_1(t), y(t)\rangle + a_{2n}\langle h_2(t), y(t)\rangle|. \quad (23)$$

From the Cauchy–Schwarz inequality

$$\begin{aligned}&|a_{1n}\langle h_1(t), y(t)\rangle + a_{2n}\langle h_2(t), y(t)\rangle|^2\\ &\le (a_{1n}^2 + a_{2n}^2)\left[\langle h_1(t), y(t)\rangle^2 + \langle h_2(t), y(t)\rangle^2\right] \quad (24)\end{aligned}$$

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 every two columns of $\boldsymbol{A}$ 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 $\boldsymbol{A}$ are normalized, $a_{1n}^2 + a_{2n}^2 = \|\boldsymbol{a}_n\|^2 = 1$. 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

$$\begin{aligned}\hat{b}_{n_0} &= \mathbf{sgn}(r_{n_0}[a_{1n_0}\langle y(t), h_1(t)\rangle + a_{2n_0}\langle y(t), h_2(t)\rangle])\\ &= \mathbf{sgn}(r_{n_0}^2 b_{n_0}[a_{1n_0}^2 + a_{2n_0}^2]) = b_{n_0}. \quad (25)\end{aligned}$$

B. Noise Amplification of Subspace Projection

The projection onto the detection subspace amplifies the variance of noise. When the RDD and RDDF detectors detect the nth user, they are affected by a noise component $\boldsymbol{a}_n^H\boldsymbol{w}$. Consider the special case with orthogonal signature waveforms, i.e., $\boldsymbol{G} = \boldsymbol{I}$, and $\boldsymbol{A}$ chosen as the random partial DFT matrix. In this case, the noise variance is given by $\sigma^2\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{a}_n = \sigma^2(N/M)$ so that it is amplified by a factor $N/M \ge 1$. In general, with subspace projection, the noise variance is amplified by a factor of $\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H\boldsymbol{a}_n$ [17]. In the following, we capture this noise amplification more precisely by relating the noise variance 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 $\boldsymbol{A}$ through its coherence and are parameterized by the cross correlations of the signature waveforms through the properties of the matrix $\boldsymbol{G}$. Our performance measure is the probability of symbol error, which is defined as the probability that the set of active users is detected incorrectly, *or* any of their symbols are detected incorrectly

$$P_e = \mathbb{P}\{\hat{\mathcal{I}} \ne \mathcal{I}\} + \mathbb{P}\{\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\boldsymbol{b}} \ne \boldsymbol{b}\}\}. \quad (26)$$

We will show in the proof of Theorem 1 that the second term of (26) is dominated by the first term when (13) and (16) are used for active user detection. Define the largest and smallest channel gains as

$$|r_{\max}| \triangleq \max_n |r_n|, \quad |r_{\min}| \triangleq \min_n |r_n|. \quad (27)$$

Also define the kth largest channel gain as $|r^{(k)}|$. Hence, $|r_{\max}| = |r^{(1)}|$ and $|r_{\min}| = |r^{(K)}|$. Our main result is the following theorem.

Theorem 1: Let $\boldsymbol{b} \in \mathbb{R}^{N\times 1}$ be an unknown deterministic symbol vector, $b_n \in \{-1, 1\}$, $n \in \mathcal{I}$, and $b_n = 0$, $n \in \mathcal{I}^c$, $n = 1, \ldots, N$. Denote the RD-MUD front-end output by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{R}\boldsymbol{b} + \boldsymbol{w}$, where $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{G} \in \mathbb{R}^{N\times N}$ are known, $\boldsymbol{w}$ is a Gaussian random vector with zero mean and covariance $\sigma^2\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H$ and $\boldsymbol{R} = \mathrm{diag}\{r_1, \ldots, r_N\}$. Let

$$\tau \triangleq \sigma\sqrt{2(1+\alpha)\log N} \cdot \sqrt{\lambda_{\max}(\boldsymbol{G}^{-1})} \cdot \sqrt{\max_n(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{a}_n)} \quad (28)$$

for a given constant $\alpha > 0$.

1) Assume that the number of active users K is known. If the coherence (9) of $\boldsymbol{A}$ satisfies the following condition:

$$|r_{\min}| - (2K-1)\mu|r_{\max}| \geq 2\tau \tag{29}$$

for some constant $\alpha > 0$, then the probability of symbol error (26) for the RDD detector is upper bounded by

$$P_e \leq N^{-\alpha}[\pi(1+\alpha)\log N]^{-1/2}. \tag{30}$$

2) Assume K_0 is an upper bound for the number of active users. If the coherence (9) of $\boldsymbol{A}$ satisfies (29) for $K = K_0$, and we choose a threshold $\xi > 0$ that satisfies

$$K_0\mu|r_{\max}| + \tau < \xi < |r_{\min}| - (K_0-1)\mu|r_{\max}| - \tau \tag{31}$$

then the probability of symbol error (26) for the RDDt detector is upper bounded by the right-hand side of (30).

3) Assume that the number of active users K is known. If the coherence (9) of $\boldsymbol{A}$ satisfies the following condition:

$$|r_{\min}| - (2K-1)\mu|r_{\min}| \geq 2\tau \tag{32}$$

for some constant $\alpha > 0$, then the probability of symbol error (26) for the RDDF detector is upper bounded by the right-hand side of (30).

4) If the coherence (9) of $\boldsymbol{A}$ satisfies (32) and we choose a threshold $\epsilon > 0$ such that

$$\tau < \epsilon < \min_{k=1}^{K}\{|r^{(k)}|[1-(K-k)\mu] - \tau\} \tag{33}$$

then the probability of symbol error (26) for the RDDFt detector is upper bounded by the right-hand side of (30).

Proof: See Appendix B. ■

The main idea of the proof is the following. We define an event

$$\mathcal{G} = \left\{\max_{n=1}^{N} |\boldsymbol{a}_n^H \boldsymbol{w}| < \tau\right\} \tag{34}$$

for τ defined in (28), and prove that $\mathcal{G}$ occurs with high probability. This bounds the two-sided tail probability of the noise. Then, we show that under (29), whenever $\mathcal{G}$ occurs, the active users can be correctly detected. On the other hand, we show that under a condition weaker than (29), whenever $\mathcal{G}$ occurs, the user data symbols can be correctly detected. A similar but inductive approach is used to prove the performance guarantee for the RDDF detector.

A special case of Theorem 1 is when $\boldsymbol{A}\boldsymbol{A}^H = (N/M)\boldsymbol{I}$, $\max_n(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{a}_n) = N/M$, and $\boldsymbol{G} = \boldsymbol{I}$. This is true when $\boldsymbol{A}$ is the random partial DFT matrix and the signature waveforms are orthogonal. If we scale σ^2 by M/N, then the right-hand sides of (29) and (32) are identical to the corresponding quantities in Theorem 4 of [22]. Hence, in this case, Theorem 1 has the same conditions as those of Theorem 4 in [22]. However, Theorem 4 in [22] only guarantees detecting the correct sparsity pattern of $\boldsymbol{b}$ (equivalently, the correct active users), whereas Theorem 1 guarantees correct detection of not only the active users but their symbols as well. Theorem 1 is also applied to more general colored noise.

Remarks:

1) The term on the right-hand side of (29) and (32) is bounded by $1 \leq \max_n(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{a}_n) \leq 1 + (N-1)\mu^2$.
2) There is a noise phase-transition effect. Define the minimum signal-to-noise ratio (SNR) as

$$\mathsf{SNR}_{\min} = \frac{|r_{\min}|^2}{\sigma^2\lambda_{\max}(\boldsymbol{G}^{-1})} \tag{35}$$

where $\lambda_{\max}(\boldsymbol{G}^{-1})$ captures the noise amplification effect in the subspace projection due to nonorthogonal signature waveforms. Conditions (29) and (32) suggest that for the RDD and RDDF detectors to have P_e as small as (30), we need at least

$$\mathsf{SNR}_{\min} > 8\log N. \tag{36}$$

This means that if the minimum SNR is not sufficiently high, then these algorithms cannot attain small probability of symbol error. We illustrate this effect via numerical examples in Section V-5 (a similar effect can be observed in standard MUD detectors).

3) In Theorem 1, the condition of having a small probability of symbol error for the RDDF detector is weaker than for the RDD detector. Intuitively, the iterative approach of DF removes the effect of strong users iteratively, which helps the detection of weaker users.

D. Bounding Probability of Symbol Error of RDD and RDDF

Theorem 1 provides a condition on how small μ has to be to achieve a small probability of symbol error. We can eliminate the constant α and rewrite Theorem 1 in an equivalent form that gives explicit error bounds for the RDD and RDDF detectors. Define

$$\beta_1 \triangleq \frac{[1-(2K-1)\mu|r_{\max}|/|r_{\min}|]^2}{\max_n(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{a}_n)}$$
$$\beta_2 \triangleq \frac{[1-(2K-1)\mu]^2}{\max_n(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{a}_n)}. \tag{37}$$

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 (29) in Theorem 1 is nonnegative. For the same reason, for the RDDF detector, we have assumed that $1-(2K-1)\mu > 0$. By Remark (1) and (37), $\beta_1 \leq 1$ and $\beta_2 \leq 1$. We can prove the following corollary from Theorem 1 (see [28] for details).

Corollary 1: Under the setting of Theorem 1, with the definitions (35) and (37), the probability of symbol error for the RDD detector is upper bounded by

$$P_{e,\mathrm{RDD}} \leq \frac{2N}{\sqrt{\pi}}\left[\frac{\mathsf{SNR}_{\min}}{2}\cdot\beta_1\right]^{-1/2} e^{-\frac{1}{4}\frac{\mathsf{SNR}_{\min}}{2}\cdot\beta_1} \tag{38}$$

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

$$P_{e,\mathrm{RDDF}} \leq \frac{2N}{\sqrt{\pi}}\left[\frac{\mathsf{SNR}_{\min}}{2}\cdot\beta_2\right]^{-1/2}\cdot e^{-\frac{1}{4}\frac{\mathsf{SNR}_{\min}}{2}\beta_2} \tag{39}$$

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

For the decorrelating detector in a conventional MUD, a commonly used performance measure is the probability of error of each user [4], which is given by

$$\mathbb{P}\{\hat{b}_n \neq b_n\} = Q\left(|r_n|/(\sigma\sqrt{[\boldsymbol{G}^{-1}]_{nn}})\right) \tag{40}$$

where $Q(x) = \int_x^\infty (1/\sqrt{2\pi})e^{-z^2/2}dz$ is the Gaussian tail probability. Using (40) and the union bound, we obtain

$$\begin{aligned} P_e = \mathbb{P}\{\hat{\boldsymbol{b}} \neq \boldsymbol{b}\} &\leq \sum_{n=1}^{N} \mathbb{P}\{\hat{b}_n \neq b_n\} \leq NQ\left(\sqrt{\mathrm{SNR}_{\min}}\right) \\ &\leq \frac{N}{2\sqrt{\pi}}\left[\frac{\mathrm{SNR}_{\min}}{2}\right]^{-1/2} e^{-\frac{\mathrm{SNR}_{\min}}{2}} \end{aligned} \tag{41}$$

where we also used the fact that $|r_n|/[\sigma\sqrt{[\boldsymbol{G}^{-1}]_{nn}}] \geq \sqrt{\mathrm{SNR}_{\min}}$ and $Q(x)$ is decreasing in x, as well as the bound [1] Since conventional MUD is not concerned with active user detection and the errors are due to symbol detection, it only makes sense to compare (41) to (38) and (39) when $K = N$ and $M = N$. Under this setting, $\beta_1 = 1$ and $\beta_2 = 1$, and the bounds on P_e (38) of RDD and (39) of RDDF are larger than the bound (41) of the decorrelating detector. This is because when deriving bounds for symbol detection error in the proof of Theorem 1, we consider the probability of (34), which requires two-side tail probability of a Gaussian random variable. In contrast, in a conventional MUD, without active-user detection, only the one-sided tail probability of the Gaussian random variable $\mathbb{P}\{\boldsymbol{w} > \tau\}$ is required because we use binary modulation. Nevertheless, obtaining a tighter bound for a symbol detection error is not necessary in RD-MUD because when $K < N$, the active user detection error dominates the symbol detection error.

By letting the noise variance σ^2 go to zero in (38) and (39) for the RDD and RDDF detectors, we can derive the following corollary from Theorem 1 (a proof of this corollary for the RDD detector 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).

E. Lower Bound on the Number of Correlators

Theorem 1 is stated for any matrix $\boldsymbol{A}$. By substitution of the expression for coherence of a given $\boldsymbol{A}$ 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 symbol error. For example, the coherence of the random partial DFT matrix can be bounded in probability (easily provable by the complex Hoeffding's inequality [33]):

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

$$\mu < [4(2\log N + c)/M]^{1/2} \tag{42}$$

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

Lemma 1 together with Theorem 1 implies that for the partial DFT matrix to attain a small probability of symbol error, the number of correlators needed by the RDD and RDDF detectors is on the order of $\log N$. This is much smaller than that required by the conventional MUD using an MF-bank, which is on the order of N.

Corollary 2 together with the Welch bound imply that, for the RDD and RDDF detectors 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 property of a matrix have a "quadratic bottleneck" [32]: the number of measurements is on the order of K^2. Nevertheless, the coherence property is easy to check for a given matrix, and it is a convenient measure of the user interference level in the detection subspace, as we demonstrated in the proof of Theorem 1.

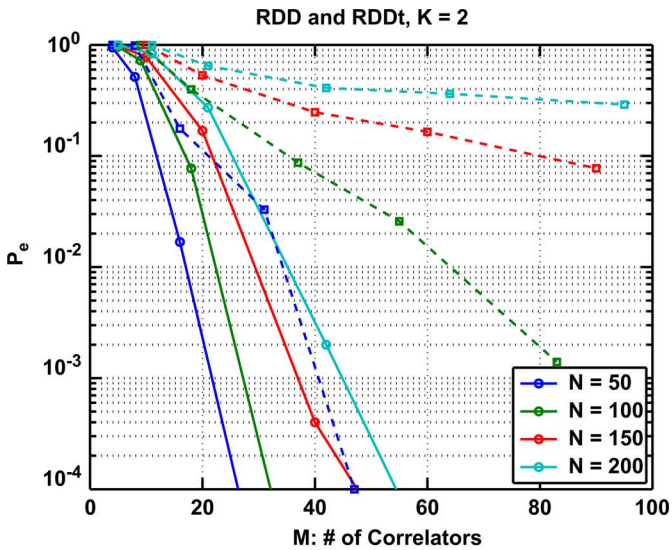
V. Numerical Examples

As an illustration of the performance of RD-MUD, we present some numerical examples. We first generate 10^5 partial random DFT matrices and choose the matrix that has the smallest coherence as $\boldsymbol{A}$. Then, using the fixed $\boldsymbol{A}$, we obtain results from 5×10^5 Monte Carlo trials. For each trial, we generate a Gaussian random noise vector and random bits: $b_n \in \{-1, 1\}$, $n \in \mathcal{I}$ with probability 1/2. In this setting, the conventional decorrelating detector has P_e equal to that of the RDD when $M = N$.

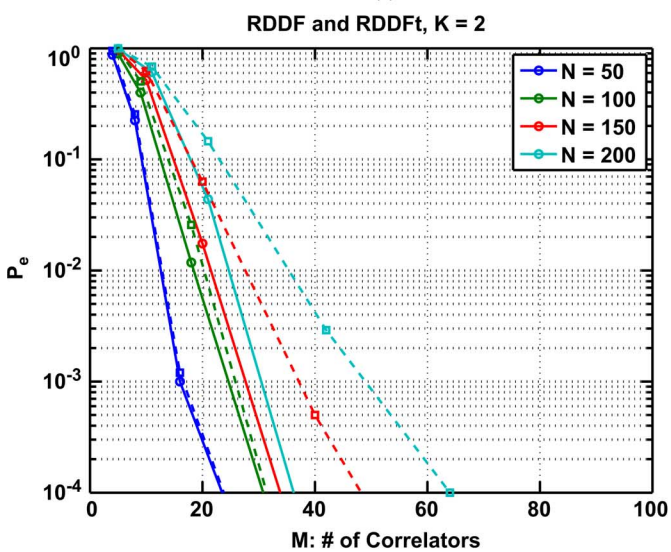
1) P_e Versus M, as N Increases: Fig. 4 shows the P_e of the RDD, RDDF, RDDt, and RDDFt detectors as a function of M, for fixed $K = 2$, and different values of N. The amplitudes $r_n = 1$ for all n, the noise variance is $\sigma^2 = 0.005$, and $\boldsymbol{G} = \boldsymbol{I}$, which corresponds to $\mathsf{SNR}_{\min} = 23$ dB. For each combination of N and M, we numerically search to find the best values for parameters ξ and ϵ. The values of ξ range from 0.78 to 0.92, increase for larger N and decrease for larger M for the RDDt detector. The values of ϵ range from 0.50 to 0.80, increase for larger N and decrease for larger M for the RDDFt detector. The RDD and RDDF detectors can achieve small P_e for M much smaller than N; the RDDt and RDDFt have some sacrifice in performance due to their lack of knowledge of K. This degradation becomes more pronounced for larger values of N.

2) P_e Versus M, as K Increases: Fig. 5(a) demonstrates the P_e of the RDD, RDDF, RDDt, and RDDFt detectors as a function of M, for a fixed $N = 100$, and different values of K. For each combination of M and K, we numerically search to find the best values for the parameters ξ and ϵ. Here, ξ ranges from 0.68 to 0.80 and ϵ ranges from 0.32 to 0.70. The amplitudes $r_n = 1$ for all n, the noise variance is $\sigma^2 = 0.005$, and $\boldsymbol{G} = \boldsymbol{I}$, which corresponds to $\mathsf{SNR}_{\min} = 23$ dB. Clearly, the number of correlators needed to obtain small P_e increases as K increases.

3) Comparison of Random Matrices $\boldsymbol{A}$: We compare the P_e of the RDD and RDDF detectors when the Gaussian random matrices, the random partial DFT matrices, or the Kerdock codes are used for $\boldsymbol{A}$. In Fig. 6(a), the P_e of the Gaussian random matrix converges to a value much higher than that of the partial DFT matrix, when M increases to N. In this example, $N = 100$, $K = 6$, the amplitudes $r_n = 1$ for all n,

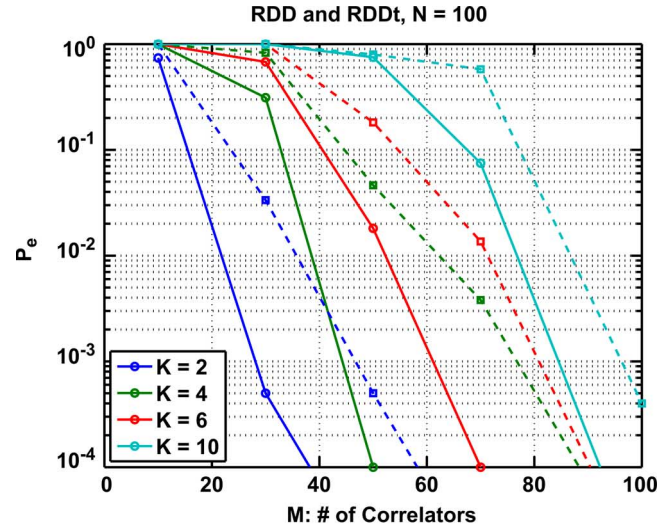


(a)

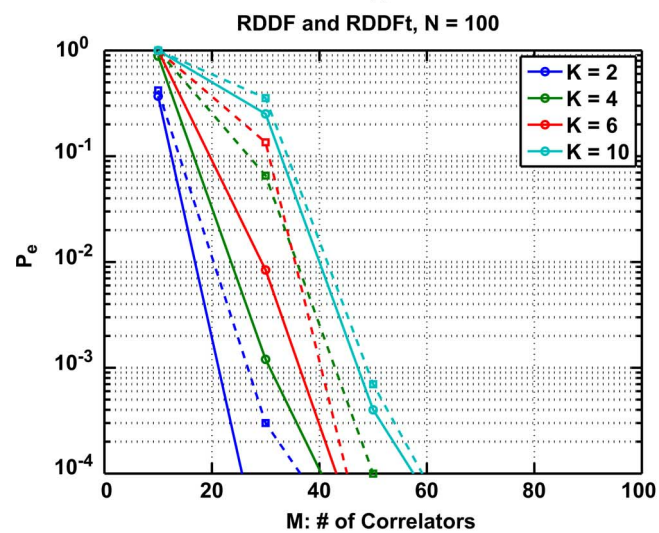


(b)

Fig. 4. P_e versus M for $K = 2$ and different N. The amplitudes $r_n = 1$ for all n, the noise variance is $\sigma^2 = 0.005$, and $\boldsymbol{G} = \boldsymbol{I}$, which corresponds to $\mathsf{SNR}_{\min} = 23$ dB. (a) RDD and RDDt, where the solid lines correspond to RDD and the dashed lines correspond to RDDt. (b) RDDF and RDDFt, where the solid lines correspond to RDDF and the dashed lines correspond to RDDFt.



(a)



(b)

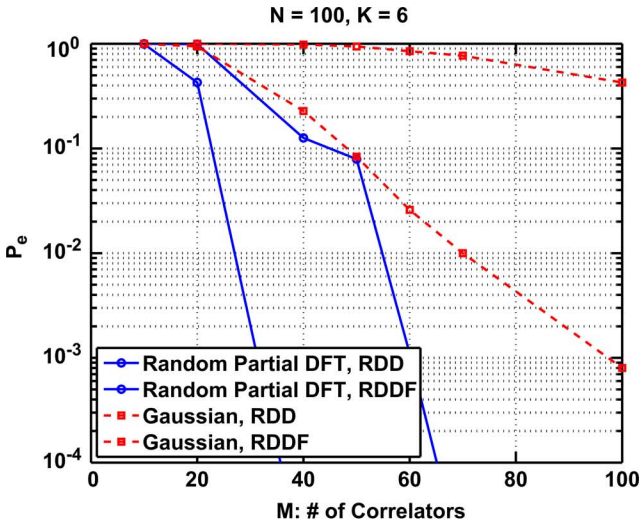
Fig. 5. P_e versus M for $N = 100$ and different K. The amplitudes $r_n = 1$ for all n, the noise variance is $\sigma^2 = 0.005$, and $\boldsymbol{G} = \boldsymbol{I}$. (a) RDD and RDDt, where the solid lines correspond to RDD and the dashed lines correspond to RDDt; and (b) RDDF and RDDFt, where the solid lines correspond to RDDF and the dashed lines correspond to RDDFt.

the noise variance is $\sigma^2 = 0.005$, and $\boldsymbol{G} = \boldsymbol{I}$. In Fig. 6(b), Kerdock codes outperform both the partial DFT and the Gaussian random matrices because of their good coherence properties. This behavior can be explained as follows. For larger N and relatively small M, it becomes harder to select a partial DFT matrix with small coherence by random search, whereas the Kerdock codes can be efficiently constructed and they obtain the Welch lower bound on coherence by design. For fixed N and M, Kerdock codes can support a large number of active users, as demonstrated in Fig. 7. In this example, the coherence of the Kerdock code is $\mu = 0.0312$, which is much smaller than the coherence $\mu = 0.0480$ obtained by choosing from 10^5 random partial DFT matrices. Kerdock codes are tight frames [31], [34] meaning that $\boldsymbol{G} = N/M\boldsymbol{I}$ so that no prewhitening is needed.

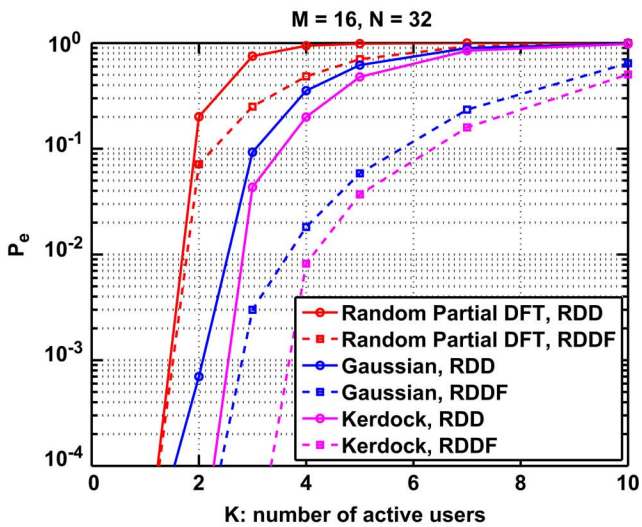
4) P_e Versus M, as SNR *Changes:* Consider a case where $\mathsf{SNR}_{\min}$ changes by fixing $\boldsymbol{G} = \boldsymbol{I}$, $r_n = 1$ for all n, and varying σ^2 (see Fig. 8). For comparison, we also consider the conventional decorrelating detector, which corresponds to the RDD detector with $M = N$. Assume $N = 100$ and $K = 2$. Note that there is a noise phase-transition effect in Fig. 5, which is discussed in the Remarks of Section IV-C.

5) Near-Far Problem, $\boldsymbol{G} = \boldsymbol{I}$: To illustrate the performance of the RDD and RDDF detectors in the presence of the near-far problem, we choose r_n uniformly random from $[1, 1.5]$ for active users. Assume $N = 100$, $K = 2$, and $\sigma^2 = 0.005$. In Fig. 9, RDDF significantly outperforms RDD.

6) P_e Versus M, Performance of the Noise Whitening Transform: Next, we consider practical signature waveforms in CDMA systems. There are many choices for signature sequences and the Gold code is one that is commonly used [35]. For signature sequences $\{s_{n\ell}\}$, the signature waveforms



(a)



(b)

Fig. 6. (a) P_e versus M of the RDD and RDDF detectors using random partial DFT versus Gaussian random matrices for $N = 100$ and $K = 6$ (Kerdock codes require dimensions of $M = 2^m$ for $m = 4, 6, \ldots$ and hence are not presented here). (b) P_e versus K of the RDD and RDDF detectors using Gaussian random matrices, random partial DFT matrices, and Kerdock codes of size 16×256 (arbitrarily select 32 columns for 32 users), for $N = 32$ and $M = 16$. In both examples, the amplitudes $r_n = 1$ for all n, the noise variance is $\sigma^2 = 0.005$, and $\boldsymbol{G} = \boldsymbol{I}$.

are generated by $s_n(t) = \sum_{\ell=0}^{L-1} s_{n\ell} p(t - \ell T_c)$, where L is the sequence length, $T_c \ll T$ is the chip duration, and the sequences $\{s_{n\ell}\}$ are modulated by unit-energy chip waveform $p(t)$ with $\int |p(t)|^2 dt = 1$ and $\int p(t - \ell T_c) p(t - kT_c) dt = 0$, $\ell \neq k$. For Gold codes, we choose $m = 10$ (with length $L = 2^{10} - 1 = 1023$ and 1025 possible codewords) [36]. We use 100 Gold codes to support $N = 100$ users. The Gram matrix of the Gold code is given by

$$\boldsymbol{G} = \frac{L+1}{L} \boldsymbol{I}_{N \times N} - \frac{1}{L} \boldsymbol{1}\boldsymbol{1}^\top \tag{43}$$

which has two distinct eigenvalues. In this example, $\lambda_1 = (N+1)/N = 1.0010$, $\lambda_2 = (L - N + 1)/L = 0.8768$, $\lambda_{\max}(\boldsymbol{G}^{-1}) = 1.1405$, and hence, the signature waveforms are nearly orthogonal. We also consider a simulated

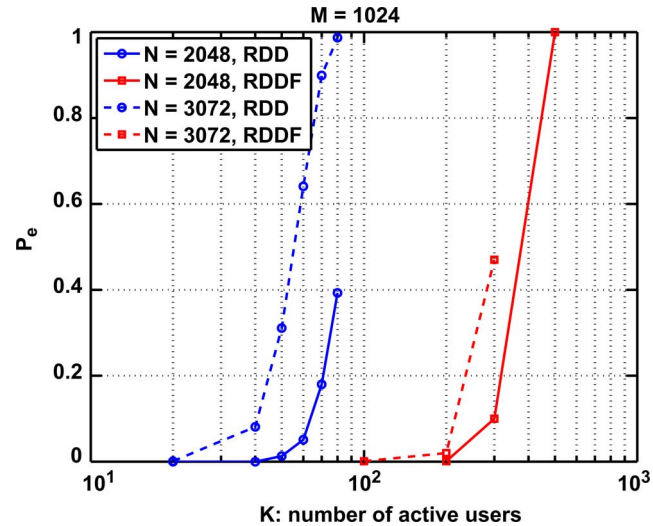


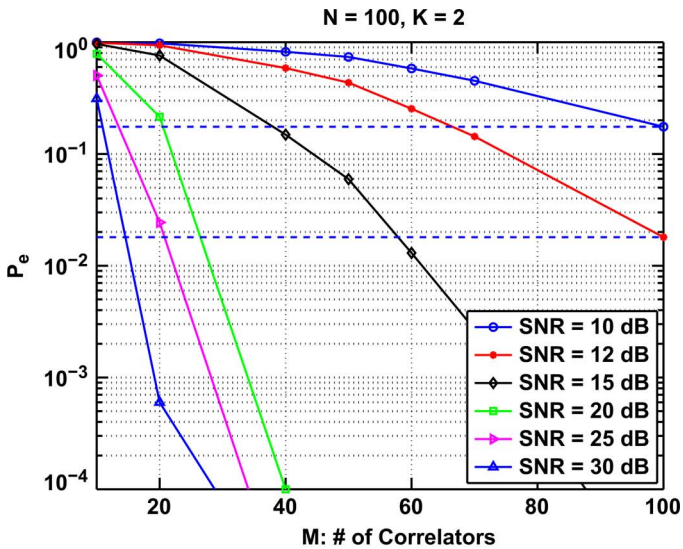
Fig. 7. Performance of RDD and RDDF detectors, when using Kerdock codes for $\boldsymbol{A}$ with $M = 1024$, P_e versus K for various N when amplitudes r_n uniformly random in $[1, 1.5]$, $\boldsymbol{G} = \boldsymbol{I}$, and $\sigma^2 = 0.005$.

$\boldsymbol{G} = \boldsymbol{U}\mathrm{diag}\{1/400, 2/400, \ldots, 100/400\}\boldsymbol{U}^\top$ for a randomly generated unitary matrix $\boldsymbol{U} \in \mathbb{R}^{100 \times 100}$, and hence $\lambda_{\max}(\boldsymbol{G}^{-1}) = 400$ which is much larger than that of the Gold codes. In Fig. 10(a) and (b), when the signature waveforms are nearly orthogonal, the noise whitening transform does not reduce P_e much. Fig. 10(c) and (d) shows that the performance of the RDD and RDDF detectors can be significantly improved by the noise whitening transform for large M. We also verified that using the noise whitening transform cannot achieve the probability of error that is obtained with orthogonal signature waveforms $\boldsymbol{G} = \boldsymbol{I}$. This is because the noise whitening transform distorts the signal component.

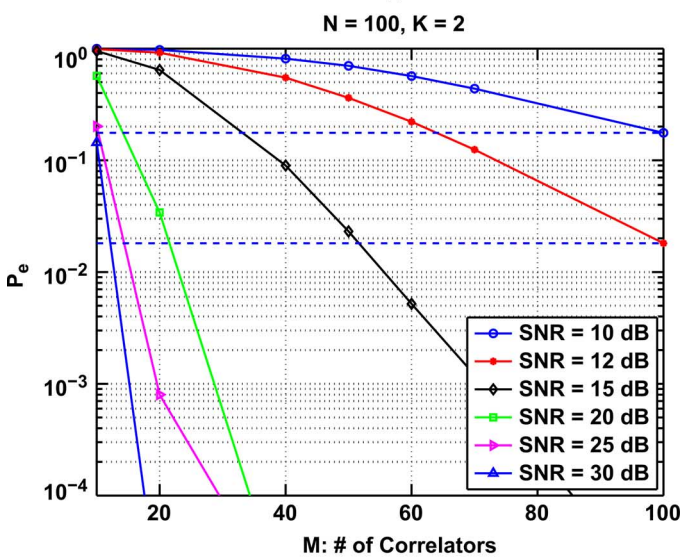
7) P_e Versus M, RD-MUD Linear Detectors: To compare performance of the RD-MUD linear detectors, we consider two sets of schemes. The first are one-step methods: using (13) for active user detection followed by symbol detection using (14) (corresponds to RDD), (20) (corresponds to RD-MMSE), or (21) (corresponds to RD-LS), respectively. The second set of schemes detects active users and symbols iteratively: the RDDF detector, the modified RDDF detector, modified by replacing the symbol detection by the RD-LS detector (21) on the detected support in each iteration $\mathcal{I}^{(k)}$, and the modified RDDF detector, modified by replacing the symbol detection by the MMSE detector (20) on the detected support in each iteration $\mathcal{I}^{(k)}$. Assume $N = 100$, $K = 2$, $r_n = 1$ for all n, and $\sigma^2 = 0.005$. Again we consider Gold codes as defined in Section V-6. As showed in Table I, iterative methods including RDDF outperform the one-step methods including RDD. However, the difference between various symbol detection methods is very small, since active user detection error dominates the symbol detection error. By examining the conditional probability of symbol error $\mathbb{P}\{\hat{\boldsymbol{b}} \neq \boldsymbol{b} | \hat{\mathcal{I}} = \mathcal{I}\}$, in Fig. 11, we see that both RD-LS and RD-MMSE detectors have an advantage over sign detection.

VI. Conclusion and Discussions

We have developed an RD-MUD structure, assuming symbol-rate synchronization, which decreases the number of correlators at the front end of an MUD receiver by exploiting



(a)



(b)

Fig. 8. Performance of RDD and RDDF detectors when $r_n = 1$ for all n, $\boldsymbol{G} = \boldsymbol{I}$, and various σ^2, where we denote $\mathsf{SNR} = 10\log_{10}(r_n^2/\sigma^2)$ dB. The dashed lines show P_e for the conventional decorrelating detectors at the corresponding SNR.

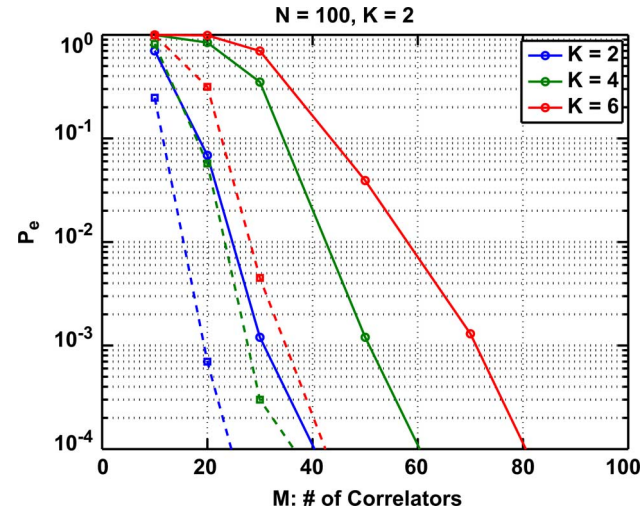


Fig. 9. Comparison of RDD and RDDF in the presence of near-far problem, with amplitudes r_n uniformly random in $[1, 1.5]$, $N = 100$, $K = 2$, $\sigma^2 = 0.005$, and $\boldsymbol{G} = \boldsymbol{I}$. The solid lines correspond to RDD and the dashed lines correspond to RDDF.

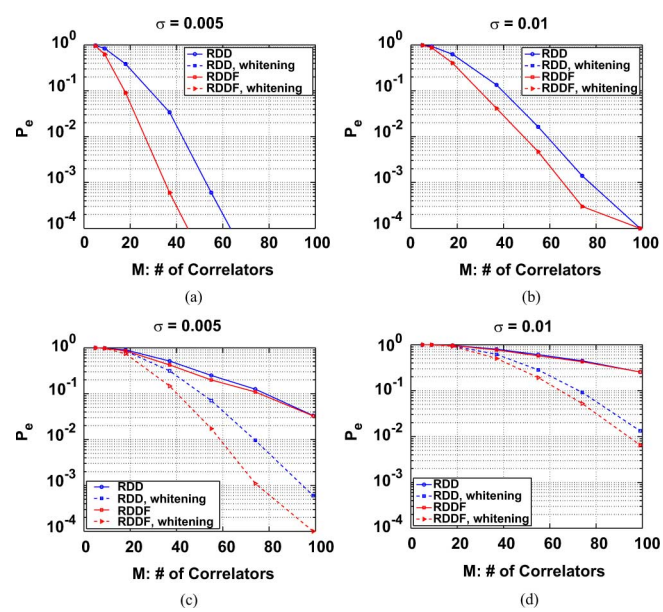


Fig. 10. Comparison of RDD and RDDF detectors with and without noise whitening when $N = 100$, $K = 2$, amplitudes r_n uniformly random in $[1, 1.5]$, and the following settings for $\boldsymbol{G}$ and σ^2: (a) Gold codes with $\lambda_{\max}(\boldsymbol{G}^{-1}) = 1.1405$, $\sigma^2 = 0.005$, (b) same Gold codes as in (a) but $\sigma^2 = 0.01$, (c) simulated $\boldsymbol{G}$ with $\lambda_{\max}(\boldsymbol{G}^{-1}) = 400$, $\sigma^2 = 0.005$, and (d) same simulated $\boldsymbol{G}$ as in (c) but $\sigma^2 = 0.01$.

the fact that the number of active users is typically much smaller than the total number of users in the system. The front end of the RD-MUD is motivated by analog CS and it 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 $\boldsymbol{A}$, 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 studied in detail two such detectors. The RDD detector, which is a linear detector that combines subspace projection along with thresholding for active user detection and RDDF detector, which is a nonlinear detector that combines DF-MP for active user detection. We have shown that to achieve a desired probability of symbol error, the number of correlators used by RD-MUD can be much smaller than that used by 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 $\boldsymbol{A}$ and the RDD and RDDF detectors are used for detection, the RD-MUD front end requires a number of correlators proportional to log of the number of users, whereas the conventional MF-bank front end requires a number of correlators equal to the number of users in the system. We obtained theoretical performance guarantees for the RDD and RDDF detectors in terms of the coherence of $\boldsymbol{A}$, which are validated via numerical examples.

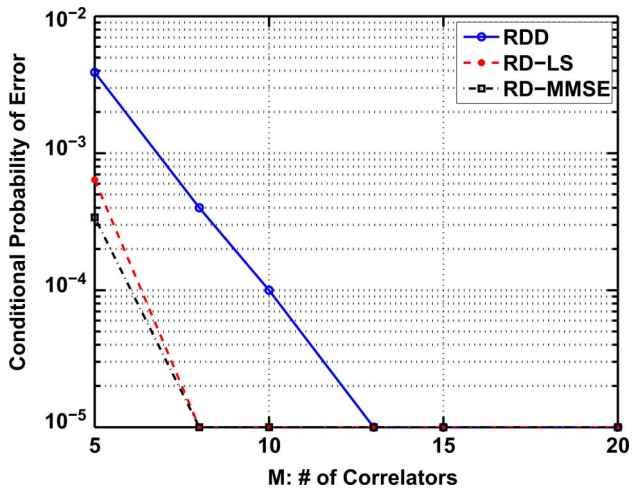


Fig. 11. Comparison of $\mathbb{P}\{\hat{\boldsymbol{b}} \neq \boldsymbol{b} | \hat{\mathcal{I}} = \mathcal{I}\}$ for RDD, RD-LS, and RD-MMSE in the same setting as that in Fig. 10.

TABLE I
$\mathbb{P}\{\hat{\boldsymbol{b}} \neq \boldsymbol{b} | \hat{\mathcal{I}} = \mathcal{I}\}$ VERSUS M, $N = 100$, $K = 2$

	M			
Methods	5	9	18	37
RDD	0.9780	0.8400	0.3857	0.0342
RD-LS	0.9780	0.8400	0.3857	0.0342
RD-MMSE	0.9779	0.8400	0.3857	0.0342
RDDF	0.9527	0.6248	0.0905	0.0006
modified RDDF with LS	0.9526	0.6247	0.0905	0.0006
modified RDDF with MMSE	0.9526	0.6247	0.0905	0.0006

In contrast to other work exploiting compressed sensing techniques for multiuser detection, our work has several distinctive features: 1) we consider analog received multiuser signals; 2) we consider front-end complexity, which is the number of filters/correlators at the front end to perform the analog-to-discrete conversion; 3) the noise is added in the analog domain prior to processing of the front end so that the output noise vector can be colored due to front-end filtering; and 4) we modify several conventional compressed sensing estimation algorithms to make them applicable for symbol detection and study their probability-of-symbol-error performance.

Our results are based on binary modulation and can be extended to higher order modulation with symbols taking more possible values. In this case, however, the conditions to guarantee correct symbol detection may be stronger than the conditions to guarantee correct active user detection. 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 $\boldsymbol{A}$. Our results assume a synchronous channel model. Extending the ideas of this work to asynchronous channels, perhaps using the methods developed in [37] for time-delay recovery from low-rate samples, is a topic of future research.

APPENDIX A
DERIVATION OF RD-MUD MMSE

Given the active user index set $\hat{\mathcal{I}}$ obtained from (13), we define $\boldsymbol{W} = \boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}}^2\boldsymbol{A}_{\hat{\mathcal{I}}}^H + \sigma^2\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H$, and $\bar{\boldsymbol{M}} = \boldsymbol{R}_{\hat{\mathcal{I}}}\boldsymbol{A}_{\hat{\mathcal{I}}}^H\boldsymbol{W}^{-1}$. We want to show that $\bar{\boldsymbol{M}} = \arg\min_{\boldsymbol{M}} \mathbb{E}\{\|\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y}\|^2\}$. Using the same method for deriving the conventional MMSE detector of the MF-bank [1], we assume that $\boldsymbol{b}_{\hat{\mathcal{I}}}$ has a distribution that is uncorrelated with the noise $\boldsymbol{w}$ and that $\mathbb{E}\{\boldsymbol{b}_{\hat{\mathcal{I}}}\boldsymbol{b}_{\hat{\mathcal{I}}}^H\} = \boldsymbol{I}$. Based on $\hat{\mathcal{I}}$, we refer to the model (12), and write the MSE as $\mathbb{E}\{\|\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y}\|^2\} = \mathrm{tr}(\mathbb{E}\{(\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y})(\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y})^H\})$. Now we expand

$$\begin{aligned}
&\mathbb{E}\{(\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y})(\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y})^H\} \\
&= \mathbb{E}\{\boldsymbol{b}_{\hat{\mathcal{I}}}\boldsymbol{b}_{\hat{\mathcal{I}}}^H\} - \mathbb{E}\{\boldsymbol{b}_{\hat{\mathcal{I}}}\boldsymbol{y}^H\}\boldsymbol{M}^H - \boldsymbol{M}\mathbb{E}\{\boldsymbol{y}\boldsymbol{b}_{\hat{\mathcal{I}}}^H\} \\
&\quad + \boldsymbol{M}\mathbb{E}\{\boldsymbol{y}\boldsymbol{y}^H\}\boldsymbol{M}^H \\
&= \boldsymbol{I} + \boldsymbol{M}(\boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}}^2\boldsymbol{A}_{\hat{\mathcal{I}}}^H + \sigma^2\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H)\boldsymbol{M}^H \\
&\quad - \boldsymbol{R}_{\hat{\mathcal{I}}}\boldsymbol{A}_{\hat{\mathcal{I}}}^H\boldsymbol{M}^H - \boldsymbol{M}\boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}}.
\end{aligned} \tag{44}$$

It can be verified that $\boldsymbol{M}\boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}} = \boldsymbol{M}\boldsymbol{W}\bar{\boldsymbol{M}}^H$. Hence, from (44), we have

$$\begin{aligned}
&\mathbb{E}\{(\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y})(\boldsymbol{b}_{\hat{\mathcal{I}}} - \boldsymbol{M}\boldsymbol{y})^H\} \\
&= \boldsymbol{I} + \boldsymbol{M}\boldsymbol{W}\boldsymbol{M}^H - \bar{\boldsymbol{M}}\boldsymbol{W}\boldsymbol{M}^H - \boldsymbol{M}\boldsymbol{W}\bar{\boldsymbol{M}}^H \\
&= \boldsymbol{I} - \bar{\boldsymbol{M}}\boldsymbol{W}\bar{\boldsymbol{M}}^H + (\boldsymbol{M} - \bar{\boldsymbol{M}})\boldsymbol{W}(\boldsymbol{M} - \bar{\boldsymbol{M}})^H \\
&= \boldsymbol{I} - \boldsymbol{R}_{\hat{\mathcal{I}}}\boldsymbol{A}_{\hat{\mathcal{I}}}^H\boldsymbol{W}^{-1}\boldsymbol{A}_{\hat{\mathcal{I}}}\boldsymbol{R}_{\hat{\mathcal{I}}} \\
&\quad + (\boldsymbol{M} - \bar{\boldsymbol{M}})\boldsymbol{W}(\boldsymbol{M} - \bar{\boldsymbol{M}})^H.
\end{aligned} \tag{45}$$

Since $\boldsymbol{W}$ is a positive semidefinite matrix, the trace of the second term in (45) is always nonnegative. Therefore, the matrix $\boldsymbol{M}$ that minimizes the MSE is $\bar{\boldsymbol{M}}$.

APPENDIX B
PROOF OF THEOREM 1

The proof of Theorem 1 for both the RDD and RDDF detectors is closely related. We therefore begin by proving several lemmas that are useful for both results.

First, we prove that the random event $\mathcal{G}$ defined in (34) occurs with high probability, where τ is defined in (28). Then, we show that when $\mathcal{G}$ occurs, both algorithms can detect the active users and their symbols. The proofs follow the arguments in [22] with modifications to account for the fact that $\boldsymbol{w}$ 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.

Lemma 2: 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 (34) occurs with probability at least one minus (30).

Proof: The random variables $\{\boldsymbol{a}_n^H\boldsymbol{w}\}_{n=1}^N$ are jointly Gaussian, with means equal to zero, variances σ_n^2 equal to $\sigma^2\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H\boldsymbol{a}_n$. Define

$$\hat{\tau} \stackrel{\triangle}{=} \sigma[2(1+\alpha)\log N]^{1/2} \cdot \left[\max_n(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H\boldsymbol{a}_n)\right]^{1/2} \tag{46}$$

and an event $\hat{\mathcal{G}} \stackrel{\triangle}{=} \left\{\max_{1\leq n\leq N}|\boldsymbol{a}_n^H\boldsymbol{w}| < \hat{\tau}\right\}$. Using Sidak's lemma [38], we have

$$\begin{aligned}
\mathbb{P}\left\{\hat{\mathcal{G}}\right\} &= \mathbb{P}\left\{|\boldsymbol{a}_1^H\boldsymbol{w}| < \hat{\tau}, \ldots, |\boldsymbol{a}_N^H\boldsymbol{w}| < \hat{\tau}\right\} \\
&\geq \prod_{n=1}^N \mathbb{P}\{|\boldsymbol{a}_n^H\boldsymbol{w}| < \hat{\tau}\}.
\end{aligned} \tag{47}$$

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

$$\mathbb{P}\{|\boldsymbol{a}_n^H\boldsymbol{w}| < \hat{\tau}) = 1 - 2Q\left(\frac{\hat{\tau}}{\sigma_n}\right\}. \tag{48}$$

Using the bound on $Q(x)$: $Q(x) \leq (x\sqrt{2\pi})^{-1}e^{-x^2/2}$, (48) can be bounded as $\mathbb{P}\{|\boldsymbol{a}_n^H\boldsymbol{w}| < \hat{\tau}\} \geq 1 - \eta_n$, where $\eta_n \triangleq \sqrt{2/\pi}(\sigma_n/\hat{\tau})e^{-\hat{\tau}^2/(2\sigma_n^2)}$. Define $\sigma_{\max} \triangleq \max_n \sigma_n = \sigma\left[\max_n(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H\boldsymbol{a}_n)\right]^{1/2}$, $\eta_{\max} \triangleq \sqrt{2/\pi}(\sigma_{\max}/\hat{\tau})e^{-\hat{\tau}^2/(2\sigma_{\max}^2)}$. Since $\sigma_{\max}/\hat{\tau} = [2(1+\alpha)\log N]^{-1/2}$ by the definition of $\hat{\tau}$, we have $\eta_{\max} = \sqrt{2/\pi}[2(1+\alpha)\log N]^{-1/2}e^{-(1+\alpha)\log N}$. 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 (47) as

$$\begin{aligned}\mathbb{P}\{\hat{\mathcal{G}}\} &\geq \prod_{n=1}^{N}(1-\eta_n) \geq (1-\eta_{\max})^N \geq 1 - N\eta_{\max}\\ &= 1 - N^{-\alpha}[\pi(1+\alpha)\log N]^{-1/2}\end{aligned} \tag{49}$$

which holds for any $\eta_{\max} \leq 1$ and $N \geq 1$. Next, we show that $\hat{\tau} \leq \tau$. Note that

$$\begin{aligned}&\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H\boldsymbol{a}_n\\ &\leq \|\boldsymbol{A}^H\boldsymbol{a}_n\|^2\lambda_{\max}(\boldsymbol{G}^{-1})\\ &\leq [\max_n\left(\boldsymbol{a}_n^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{a}_n\right)]\lambda_{\max}(\boldsymbol{G}^{-1}).\end{aligned} \tag{50}$$

From inequality (50) and definitions (28) for τ and (46) for $\hat{\tau}$, we obtain $\hat{\tau} \leq \tau$. Hence

$$\begin{aligned}&\mathbb{P}\{\mathcal{G}\} = \mathbb{P}\{\max_n |\boldsymbol{a}_n^H\boldsymbol{w}| < \tau\}\\ &\geq \mathbb{P}\{\max_n |\boldsymbol{a}_n^H\boldsymbol{w}| < \hat{\tau}\} = \mathbb{P}\{\hat{\mathcal{G}}\}.\end{aligned} \tag{51}$$

Combining (49) and (51), we conclude that $P(\mathcal{G})$ is greater than one minus the expression (30), as required. ■

The next lemma shows that, under appropriate conditions, ranking the inner products between $\boldsymbol{a}_n$ and $\boldsymbol{y}$ is an effective method of detecting the set of active users. The proof of this lemma is adapted from Lemma 3 in [22] to account for the fact that the signal vector $\boldsymbol{y}$ here can be complex as $\boldsymbol{A}$ can be complex. Since the real part contains all the useful information, to prove this lemma, we follow the proof for Lemma 3 in [22] while using the following inequality whenever needed: $|\Re[\boldsymbol{a}_n^H\boldsymbol{a}_m]| \leq |\boldsymbol{a}_n^H\boldsymbol{a}_m| \leq \mu$ for $n \neq m$, and $|\Re[\boldsymbol{a}_n^H\boldsymbol{w}]| \leq |\boldsymbol{a}_n^H\boldsymbol{w}|$. The proofs are omitted due to space limitations. Details of the proof can be found in [28].

Lemma 3: 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 (27), and suppose that

$$|r_{\min}| - (2K-1)\mu|r_{\max}| \geq 2\tau. \tag{52}$$

Then, if the event $\mathcal{G}$ of (34) occurs, we have $\min_{n\in\mathcal{I}} |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]| > \max_{n\notin\mathcal{I}} |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]|$. If, rather than (52), a weaker condition holds

$$|r_{\max}| - (2K-1)\mu|r_{\max}| \geq 2\tau. \tag{53}$$

Then, if the event $\mathcal{G}$ of (34) occurs, we have $\max_{n\in\mathcal{I}} |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]| > \max_{n\notin\mathcal{I}} |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]|$.

The following lemma demonstrates that the sign detector can effectively detect transmitted symbols for the RDD and RDDF detectors. This lemma bounds the second term in P_e that has not been considered in [22].

Lemma 4: Let $\boldsymbol{b}$ be a vector with $b_n \in \{1, -1\}$, for $n \in \mathcal{I}$ and $b_n = 0$ otherwise, and let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{R}\boldsymbol{b}+\boldsymbol{w}$ for a Gaussian noise vector $\boldsymbol{w}$ with zero mean and covariance $\sigma^2\boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H$. Suppose that

$$|r_{\min}| - (K-1)\mu|r_{\max}| \geq \tau. \tag{54}$$

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

$$\mathbf{sgn}(r_n\Re[\boldsymbol{a}_n^H\boldsymbol{y}]) = b_n, \qquad n \in \mathcal{I}. \tag{55}$$

If, instead of (54), a weaker condition

$$|r_{\max}| + |r_{\min}| - 2(K-1)\mu|r_{\max}| \geq 2\tau \tag{56}$$

holds, then under the event $\mathcal{G}$, we have $\mathbf{sgn}(r_{n_1}\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]) = b_{n_1}$, for

$$n_1 = \arg\max_n |\Re[\boldsymbol{a}_n^H\boldsymbol{y}]|. \tag{57}$$

Proof: To detect correctly, for $b_n = 1$, $\Re[r_n\boldsymbol{a}_n^H\boldsymbol{y}]$ has to be positive, and for $b_n = -1$, $\Re[r_n\boldsymbol{a}_n^H\boldsymbol{y}]$ has to be negative.

First assume $b_n = 1$. We expand $\Re[r_n\boldsymbol{a}_n^H\boldsymbol{y}]$, find the lower bound and the condition such that the lower bound is positive. Substituting in the expression for $\boldsymbol{y}$, using the inequality that $x + y + z \geq x - |y| - |z|$, under the event $\mathcal{G}$, we obtain

$$\begin{aligned}&\Re[r_n\boldsymbol{a}_n^H\boldsymbol{y}]\\ &= |r_n|^2 + \sum_{m\neq n} b_m r_n r_m \Re\left[\boldsymbol{a}_n^H\boldsymbol{a}_m\right] + r_n\Re\left[\boldsymbol{a}_n^H\boldsymbol{w}\right]\\ &\geq |r_n||r_{\min}| - \sum_{m\neq n}|r_n||r_m||\Re\left[\boldsymbol{a}_n^H\boldsymbol{a}_m\right]|\\ &\quad -|r_n||\Re\left[\boldsymbol{a}_n^H\boldsymbol{w}\right]|\\ &> |r_n|\left[|r_{\min}| - (K-1)\mu|r_{\max}| - \tau\right].\end{aligned} \tag{58}$$

From (58), $\Re[r_n\boldsymbol{a}_n^H\boldsymbol{y}] > 0$ for $n \in \mathcal{I}$ if (54) holds and $b_n = 1$.

Similarly, we can show for $b_n = -1$, under event $\mathcal{G}$, if (54) holds, $\Re[r_n\boldsymbol{a}_n^H\boldsymbol{y}] < 0$. Hence, if (54) holds, we obtain (55).

Recall that n_0 is the index of the largest gain: $|r_{n_0}| = |r_{\max}|$. Due to (57), we have

$$|\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]| \geq |\Re[\boldsymbol{a}_{n_0}^H\boldsymbol{y}]|. \tag{59}$$

We will show that under the event $\mathcal{G}$, once (56) holds, then $\mathbf{sgn}(r_{n_1}\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]) \neq b_{n_1}$ leads to a contradiction to (59). First assume $b_{n_1} = 1$. If $\hat{b}_{n_1} = \mathbf{sgn}(r_{n_1}\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]) \neq b_{n_1}$, then

$$\begin{aligned}&\hat{b}_{n_1}\\ &= \mathbf{sgn}\left(r_{n_1}^2 + \sum_{m\neq n_1} b_m r_{n_1} r_m \Re\left[\boldsymbol{a}_{n_1}^H\boldsymbol{a}_m\right] + r_{n_1}\Re\left[\boldsymbol{a}_{n_1}^H\boldsymbol{w}\right]\right)\\ &= -1.\end{aligned} \tag{60}$$

So, the expression inside the $\mathbf{sgn}$ operator of (60) must be negative. Since $r_{n_1}^2 > 0$, we must have

$$\sum_{m\neq n_1} b_m r_{n_1} r_m \Re\left[\boldsymbol{a}_{n_1}^H\boldsymbol{a}_m\right] + r_{n_1}\Re\left[\boldsymbol{a}_{n_1}^H\boldsymbol{w}\right] < 0. \tag{61}$$

Multiplying the left-hand side of (59) by $|r_{n_1}|$, and using the equality $|x|\cdot|y| = |xy|$, we obtain

$$\begin{aligned} &|r_{n_1}||\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]| \\ &= |r_{n_1}|\left|r_{n_1} + \sum_{m\neq n_1} b_m r_m \Re[\boldsymbol{a}_{n_1}^H\boldsymbol{a}_m] + \Re[\boldsymbol{a}_{n_1}^H\boldsymbol{w}]\right| \\ &= \left|r_{n_1}^2 + \sum_{m\neq n_1} b_m r_{n_1} r_m \Re[\boldsymbol{a}_{n_1}^H\boldsymbol{a}_m] + r_{n_1}\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{w}]\right|. \end{aligned} \tag{62}$$

Due to (60), the last line of (62) inside the $|\cdot|$ operator is negative. Using the fact that $r_{n_1}^2 > 0$ and (61), and the identity $|x+y| = -(x+y) = |y| - x$ when $x+y<0$ and $y<0$, under the event $\mathcal{G}$, we obtain that

$$\begin{aligned} &|r_{n_1}||\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]| \\ &= \left|\sum_{m\neq n_1} b_m r_{n_1} r_m \Re\left[\boldsymbol{a}_{n_1}^H\boldsymbol{a}_m\right] + r_{n_1}\Re\left[\boldsymbol{a}_{n_1}^H\boldsymbol{w}\right]\right| - r_{n_1}^2 \\ &< |r_{n_1}|(K-1)\mu|r_{\max}| + |r_{n_1}|\tau - |r_{n_1}||r_{\min}| \\ &= |r_{n_1}|[(K-1)\mu|r_{\max}| + \tau - |r_{\min}|]. \end{aligned} \tag{63}$$

On the other hand, multiply the right-hand side of (59) by $|r_{n_1}|$. Similarly, using the equality $|x|\cdot|y| = |xy|$ and triangle inequality, under the event $\mathcal{G}$, we obtain

$$\begin{aligned} &|r_{n_1}||\Re[\boldsymbol{a}_{n_0}^H\boldsymbol{y}]| \\ &= \left|r_{n_1} r_{n_0} b_{n_0} + \sum_{m\neq n_0} b_m r_{n_1} r_m \Re\left[\boldsymbol{a}_{n_0}^H\boldsymbol{a}_m\right] + r_{n_1}\Re\left[\boldsymbol{a}_{n_0}^H\boldsymbol{w}\right]\right| \\ &> |r_{n_1}|[|r_{\max}| - (K-1)\mu|r_{\max}| - \tau]. \end{aligned} \tag{64}$$

Combining (63) and (64), we have that once (56) holds, if $b_{n_1} = 1$, then $\mathbf{sgn}(r_{n_1}\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]) = -1$ leads to $|\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]| < |\Re[\boldsymbol{a}_{n_0}^H\boldsymbol{y}]|$, which contradicts (59), and hence, $\mathbf{sgn}(r_{n_1}\Re[\boldsymbol{a}_{n_1}^H\boldsymbol{y}]) = 1$. 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 2–4. Lemma 2 ensures that the event $\mathcal{G}$ occurs with probability at least as high as one minus (30). Whenever $\mathcal{G}$ occurs, Lemma 3 guarantees, by using (13), that the RDD detector can correctly detect active users under the condition (29), i.e., $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\}$. Finally, whenever $\mathcal{G}$ occurs, Lemma 4 guarantees that, based on the correct support of active users, their transmitted symbols can be detected correctly under the condition (54), i.e., $\mathcal{G} \subset \{\hat{b}_n = b_n,\ n\in\mathcal{I}\}$. Clearly, condition (54) is weaker than (29), since (29) can be written as $|r_{\min}| - (K-1)\mu|r_{\max}| \geq \tau + (\tau + K\mu|r_{\max}|) > \tau$, and hence, if (29) holds, then (54) also holds. In summary, under condition (29), $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\boldsymbol{b}} = \boldsymbol{b}\}$, and $1 - P_e = P(\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\boldsymbol{b}} = \boldsymbol{b}\}) \geq P(\mathcal{G})$, which is greater than one minus (30), which concludes the proof for the RDD detector.

The proof for RDDt is similar to that for an RDD detector and inspired by the proof of Theorem 1 in [30]. Using similar arguments to Lemma 3, we can demonstrate that, when the number of active users $K \leq K_0$, when $\mathcal{G}$ occurs

$$\begin{aligned} \min_{n\in\mathcal{I}} |\mathbb{R}[\boldsymbol{a}_n^H\boldsymbol{y}]| &> |r_{\min}| - (K-1)\mu|r_{\max}| - \tau \\ &\geq |r_{\min}| - (K_0-1)\mu|r_{\max}| - \tau \end{aligned} \tag{65}$$

and

$$\max_{n\notin\mathcal{I}} |\mathbb{R}[\boldsymbol{a}_n^H\boldsymbol{y}]| < K\mu|r_{\max}| + \tau \leq K_0\mu|r_{\max}| + \tau. \tag{66}$$

If (52) holds for $K = K_0$, we can choose a threshold ξ such that $K_0\mu|r_{\max}| + \tau < \xi < |r_{\min}| - (K_0-1)\mu|r_{\max}| - \tau$. Then, $\min_{n\in\mathcal{I}} |\mathbb{R}[\boldsymbol{a}_n^H\boldsymbol{y}]| > \xi$ and $\max_{n\notin\mathcal{I}} |\mathbb{R}[\boldsymbol{a}_n^H\boldsymbol{y}]| < \xi$, and hence, for such ξ, the RDDt detector can correctly detect the active users with high probability. Since when (52) holds, (54) is true, from Lemma 4 we know the symbol can be correctly detected with high probability as well.

We now prove the performance guarantee for the RDDF detector adopting the technique used in proving Theorem 4 in [22]. First, we show that whenever $\mathcal{G}$ occurs, the RDDF detector correctly detects an active user in the first iteration, which follows from Lemmas 2 and 3. Note that (32) implies (53), and therefore, by Lemma 3, we have that by choosing the largest $|\Re[\boldsymbol{a}_n^H\boldsymbol{y}]|$, the RDDF detector can detect a correct user in 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 (32) also implies (56), since (32) 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 4, 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., $\mathcal{G} \subset \{\mathcal{I}^{(1)} \subset \mathcal{I}, b_{n_1}^{(1)} = b_{n_1}\}$.

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 $k-1$ 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 $|\Re[\boldsymbol{a}_n^H\boldsymbol{v}^{(k-1)}]|$. Using the same notations as those in Section III-B2 and by definition of $\boldsymbol{v}^{(k-1)}$, we have

$$\boldsymbol{v}^{(k-1)} = \boldsymbol{AR}(\boldsymbol{b} - \boldsymbol{b}^{(k-1)}) + \boldsymbol{w} = \boldsymbol{ARx}^{(k-1)} + \boldsymbol{w} \tag{67}$$

where $\boldsymbol{x}^{(k-1)} \triangleq \boldsymbol{b} - \boldsymbol{b}^{(k-1)}$. This vector has support $\mathcal{I}/\mathcal{I}^{(k-1)}$ and has at most $K-k+1$ nonzero elements, since $\boldsymbol{b}^{(k-1)}$ 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 $\boldsymbol{v}^{(k-1)}$ is a noisy measurement of the vector $\boldsymbol{ARx}^{(k-1)}$. The data model in (67) for the kth iteration is identical to the data model in the first iteration with $\boldsymbol{b}$ replaced by $\boldsymbol{x}^{(k-1)}$ (with a smaller sparsity $K-k+1$ rather than K), $\mathcal{I}$ replaced by $\mathcal{I}/\mathcal{I}^{(k-1)}$, and $\boldsymbol{y}$ replaced by $\boldsymbol{v}^{(k-1)}$. Let $|r_{\max}^{(k)}| \triangleq \max_{n\in\mathcal{I}/\mathcal{I}^{(k-1)}} |r_n|$. By assumption, $k-1$ active users with largest gains have been correctly detected in the first $k-1$ rounds, and hence, $|r_{\max}^{(k)}| = |r^{(k)}|$. Since

$$|r^{(k)}| \geq |r_{\min}| \tag{68}$$

we have that under condition (32), this model (67) also satisfies the requirement (53). Consequently, by Lemma 3, we have that under the event $\mathcal{G}$, $\max_{n\in\mathcal{I}/\mathcal{I}^{(k-1)}} |\Re[\boldsymbol{a}_n^H\boldsymbol{v}^{(k-1)}]| > \max_{n\in(\mathcal{I}/\mathcal{I}^{(k-1)})^c} |\Re[\boldsymbol{a}_n^H\boldsymbol{v}^{(k-1)}]|$. Therefore, in the kth

iteration, the RDDF detector can detect an active user correctly, i.e., $\mathcal{G} \subset \{\mathcal{I}^{(k)} \subset \mathcal{I}\}$, and no index of active users that has been detected before will be chosen again. On the other hand, since (32) can be written as $|r_{\min}| \geq 2\tau/[1-(2K-1)\mu]$, from (68), this implies $|r^{(k)}| \geq 2\tau/[1-(2K-1)\mu]$, and hence, $|r^{(k)}| - (2K-1)\mu|r^{(k)}| \geq 2\tau$, and consequently, $|r^{(k)}| - (2K-2)\mu|r^{(k)}| + |r_{\min}| \geq 2\tau$. Consequently, condition (56) is true for (67). Then, by Lemma 4, we have that under the event $\mathcal{G}$, $\mathbf{sgn}(r_{n_k}\Re[\boldsymbol{a}_{n_k}^H \boldsymbol{v}^{(k-1)}]) = b_{n_k}$, 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.,

$$\begin{aligned}\mathcal{G} &\subset \cup_{k=1}^{K}\{\mathcal{I}^{(k)} \subset \mathcal{I}, b_{n_k}^{(k)} = b_{n_k}\} \\ &= \{\mathcal{I}^{(K)} = \mathcal{I}, b_n^{(K)} = b_n, n \in \mathcal{I}^{(K)}\}.\end{aligned} \tag{69}$$

Note that condition (53) is weaker than (32), since (32) 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 (32), from (69), $\mathcal{G} \subset \{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\boldsymbol{b}} = \boldsymbol{b}\}$, and $1 - P_e = \mathbb{P}\{\{\hat{\mathcal{I}} = \mathcal{I}\} \cap \{\hat{\boldsymbol{b}} = \boldsymbol{b}\}\} \geq \mathbb{P}\{\mathcal{G}\}$ which is greater than one minus (30), which concludes the proof for the RDDF detector.

The proof for RDDFt follows the aforementioned proof for RDDF with one more step. Note that when we have correctly detected all active users in K rounds, from (67), the residual $\boldsymbol{v}^K = \boldsymbol{w}$ contains only noise. Hence, when $\mathcal{G}$ occurs, $\|\boldsymbol{A}^H\boldsymbol{v}^K\|_\infty = \|\boldsymbol{A}^H\boldsymbol{w}\|_\infty = \max_{1\leq n\leq N}|\boldsymbol{a}_n^H\boldsymbol{w}| < \tau$, from Lemma 2. On the other hand, in the kth round, $k = 1, \ldots, K$, from (67), we have that when $\mathcal{G}$ occurs

$$\begin{aligned}&\|\boldsymbol{A}^H\boldsymbol{v}^{(k-1)}\|_\infty \\ &= \max_{1\leq n\leq N}\left|\sum_{m\in\mathcal{I}/\mathcal{I}^{(k-1)}} r_m b_m \boldsymbol{a}_n^H\boldsymbol{a}_m + \boldsymbol{a}_n^H\boldsymbol{w}\right| \qquad (70)\\ &> |r_{\max}^{(k)}| - (K-k)\mu|r_{\max}^{(k)}| - \tau > 0. \qquad (71)\end{aligned}$$

The expression in (71) is positive, when (32) holds (recall that (32) is also required to detect correct active users): because when $|r_{\min}| - (2K-1)\mu|r_{\min}| > 2\tau$, since $|r^{(k)}| \geq |r_{\min}|$, $|r^{(k)}| - (2K-1)\mu|r^{(k)}| > 2\tau$, and hence, $|r^{(k)}| - (2K-2k-1)\mu|r^{(k)}| > 2\tau \geq 0$. Therefore when (32) holds, we can choose $\epsilon < \min_{k=1}^{K}\{|r^{(k)}|[1-(K-k)\mu] - \tau\} < r_{\min} - \tau$. Therefore, under the condition (32), when $\mathcal{G}$ occurs, we can choose $\tau < \epsilon < \min_{k=1}^{K}\{|r^{(k)}|[1-(K-k)\mu]| - \tau\}$, so that $\|\boldsymbol{A}^H\boldsymbol{v}^{(k-1)}\|_\infty > \epsilon$, $k = 1, \ldots, K$, and $\|\boldsymbol{A}^H\boldsymbol{v}^{(K)}\|_\infty < \epsilon$. Finally, using similar arguments as for RDDF that (32) guarantees (56), RDDFt can also correctly detect the symbols with high probability.

This completes the proof of Theorem 1.

ACKNOWLEDGMENT

The authors would like to thank Robert Calderbank and Lorne Applebaum for providing helpful suggestions with the numerical example regarding Kerdock codes.

REFERENCES

[1] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[2] C. Schlegal and A. Grant, *Coordinated Multiuser Communications*. New York, NY, USA: Springer-Verlag, 2006.
[3] *Advances in Multiuser Detection*, ser. Wiley Series in Telecommunications and Signal Processing, M. L. Honig, Ed. New York, NY: Wiley, 2009.
[4] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 123–136, Jan. 1989.
[5] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 85–96, Jan. 1986.
[6] M. K. Varanasi, "Decision feedback multiuser detection: A systematic approach," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 219–240, Jan. 1999.
[7] L. Applebaum, W. Bajwa, M. Duarte, and R. Calderbank, "Asynchronous code-division random access using convex optimization," *Phys. Commun.*, vol. 5, no. 2, pp. 129–149, Jun. 2012.
[8] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," IEEE Trans. Inf. Theory Mar. 2010 [Online]. Available: arXiv:0903.1022v2
[9] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, Aug. 2009.
[10] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.
[11] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5424, Dec. 2006.
[12] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[13] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, Jan. 2009.
[14] Y. Jin, Y.-H. Kim, and B. D. Rao, "Support recovery of sparse signals," IEEE Trans. Inf. Theory Mar. 2010 [Online]. Available: arXiv:1003.0888v1
[15] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4080, Sep. 2006.
[16] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 454–465, Feb. 2011.
[17] E. Arias-Castro and Y. C. Eldar, "Noise folding in compressed sensing," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 478–481, Jun. 2011.
[18] R. Gribonval, B. Mailhe, H. Rauhut, K. Schnass, and P. Vandergheynst, "Average case analysis of multichannel thresholding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Honolulu, HI, USA, 2007, pp. II-853–II-856.
[19] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, pp. 265–274, 2009.
[20] 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 Syst.*, Nov. 1993, pp. 40–44.
[21] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[22] 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, no. 10, pp. 5030–5043, Oct. 2010.
[23] 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. Commun. Control, Comput.*, Monticello, IL, USA, Sep. 2010, pp. 130–137.
[24] J. van de Beek, P. O. Borjesson, M. Boucheret, D. Landstrom, J. M. Arenas, P. Odling, C. Ostberg, M. Wahlqvist, and S. K. Wilson, "A time and frequency synchronization scheme for multiuser OFDM," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1900–1914, Nov. 1999.

[25] M. Morelli, "Timing and frequency synchronization for the uplink of an OFDMA system," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 296–306, Feb. 2004.
[26] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.
[27] E. Biglieri and M. Lops, "Multiuser detection in dynamic environment—Part I: User identification and data detection," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3158–3170, Sep. 2007.
[28] Y. Xie, "Statistical signal detection with multisensor and sparsity," Ph.D. dissertation, Dept. Elect. Comput. Eng., Stanford University, Stanford, CA, USA, 2011.
[29] Y. Xie, Y. C. Eldar, and A. Goldsmith, "Reduced-dimension multiuser detection," in *48th Annu. Allerton Conf. Comm., Control, Comput.*, Sep. 2010, pp. 584–589.
[30] W. U. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," *J. Commun. Netw.*, vol. 12, no. 4, pp. 289–307, 2010.
[31] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel, "$\mathbb{Z}^4$-Kerdock codes, orthogonal spreads, and extremal Euclidean line-sets," *Proc. Lond. Math. Soc.*, vol. 75, no. 2, pp. 436–480, 1997.
[32] M. Fornasier and H. Rauhut, "Compressive sensing," in *Handbook of Mathematical Methods in Imaging*, O. Scherzer, Ed. New York, NY, USA: Springer-Verlag, 2011, ch. 2.
[33] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 13–30, Mar. 1963.
[34] O. Christensen, *An Introduction to Frames and Riesz Bases*. Cambridge, MA, USA: Birkhauser, 2002.
[35] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.
[36] X. Wang and H. Ge, "Spreading codes enabling warp converging Wiener filters for multiuser detection in CDMA systems," in *Proc. IEEE Int. Conf. Commun.*, 2008, pp. 866–870.
[37] K. Gedalyahu and Y. C. Eldar, "Time-delay estimation from low-rate samples: A union of subspaces approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3017–3031, Jun. 2010.
[38] Z. Sidak, "Rectangular confidence regions for the means of multivariate normal distributions," *J. Amer. Statist. Assoc.*, vol. 62, no. 318, pp. 626–633, Jun. 1967.

Yao Xie (S'04–M'12) is a Research Scientist in the Electrical and Computer Engineering Department at Duke University, Durham, NC. She received the Ph.D. degree in Electrical Engineering (minor in Mathematics) from Stanford University, Stanford, CA, in January 2012. She received the best paper award at the Asilomar conference on Signals, Systems, and Computers, in 2005. In 2007, she was awarded the Pan Wen Yuan scholarship. From 2007 to 2010, she held the General Yao-Wu Wang Stanford Graduate Fellowship. Her research interests include sequential statistical inference, change-point detection, big data analysis, with applications in wireless communications, (sensor and social) networks, medical and astronomical imaging.

Yonina C. Eldar (S'9–M'02–SM'07–F'12) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering both from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002.

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

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in electrical engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009 the Technion's Award for Excellence in Teaching, in 2010 the Michael Bruno Memorial Award from the Rothschild Foundation, and in 2011 the Weizmann Prize for Exact Sciences. In 2012 she was elected to the Young Israel Academy of Science and to the Israel Committee for Higher Education, and elected an IEEE Fellow. She received several best paper awards together with her research students and colleagues. She is a Signal Processing Society Distinguished Lecturer, a member of the IEEE BioImaging Signal Processing technical committee, an Associate Editor for the SIAM Journal on Imaging Sciences, and Editor in Chief of Foundations and Trends in Signal Processing. In the past, she was a member of the IEEE Signal Processing Theory and Methods technical committee, and served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the EURASIP Journal of Signal Processing, and the SIAM Journal on Matrix Analysis and Applications.

Andrea Goldsmith (S'90–M'93–SM'99–F'05) is the Stephen Harris professor in the School of Engineering and a professor of Electrical Engineering at Stanford University. She was previously on the faculty of Electrical Engineering at Caltech. She co-founded Accelera, Inc., which develops software-defined wireless network technology, and Quantenna Communications Inc., which develops high-performance WiFi chipsets. She has previously held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. Dr. Goldsmith is a Fellow of the IEEE and of Stanford, and she has received several awards for her work, including the IEEE Communications Society and Information Theory Society joint paper award, the National Academy of Engineering Gilbreth Lecture Award, the IEEE Wireless Communications Technical Committee Recognition Award, the Alfred P. Sloan Fellowship, and the Silicon Valley/San Jose Business Journal's Women of Influence Award. She is author of the book "Wireless Communications" and co-author of the books "MIMO Wireless Communications" and "Principles of Cognitive Radio," all published by Cambridge University Press. She received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from U.C. Berkeley.

Dr. Goldsmith is currently on the Steering Committee for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and has previously served as editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, the *Journal on Foundations and Trends in Communications and Information Theory and in Networks*, the IEEE *Transactions on Communications*, and the *IEEE Wireless Communications Magazine*. Dr. Goldsmith participates actively in committees and conference organization for the IEEE Information Theory and Communications Societies and has served on the Board of Governors for both societies. She has been a Distinguished Lecturer for both societies, served as the President of the IEEE Information Theory Society in 2009, founded and chaired the student committee of the IEEE Information Theory society, and currently chairs the Emerging Technology Committee and is a member of the Strategic Planning Committee in the IEEE Communications Society. At Stanford she received the inaugural University Postdoc Mentoring Award, served as Chair of its Faculty Senate, and currently serves on its Faculty Senate and on its Budget Group.