

Compressive Link Acquisition in Multiuser Communications

Xiao Li, Andrea Rueetschi, Anna Scaglione, *Fellow, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

Abstract—An important sensing operation is to detect the presence of specific signals with unknown transmission parameters. This task, referred to as “link acquisition,” is typically a sequential search over the transmitted signal space. Recently, the use of sparsity in similar estimation or detection problems has received considerable attention. These works typically focus on the benefits of compressed sensing, but not generally on the cost brought by sparse recovery. Our goal is to examine the tradeoff in complexity and performance when using sparse recovery with compressed or uncompressed samples. To do so, we propose a compressive sparsity aware (CSA) acquisition scheme, where a compressive multi-channel sampling (CMS) front-end is followed by a sparsity regularized likelihood ratio test (SR-LRT) module. The CSA scheme borrows insights from the models studied in sub-Nyquist sampling and finite rate of innovation (FRI) signals. We further optimize the CMS front-end by maximizing the average Kullback–Leibler distance of all the hypotheses in the SR-LRT. We compare the CSA scheme vis-à-vis other popular alternatives in terms of performance and complexity. Simulations suggest that one can use the CSA scheme to scale down the implementation cost with greater flexibility than other alternatives. However, we find that they both have overall complexities that scale linearly with the search space. Furthermore, it is shown that compressive measurements used in the SR-LRT lead to a performance loss when noise prevails, while providing better performance in spite of the compression when noise is mild.

Index Terms—Compressed sensing, detection and estimation, Kullback–Leibler distance, multiuser communications.

I. INTRODUCTION

ONE of the critical receiver tasks in a multiuser scenario, referred to as *link acquisition*, is that of detecting the presence of signals, and identifying the *link parameters* (e.g., delays, carrier offsets) of an *unknown subset* \mathcal{I} of active components out of I possible sources. Similar to [1], [2], we consider the case where the active users \mathcal{I} transmit known and distinct training preambles $\phi_i(t)$ to the destination. Usually these

preambles are designed to be fairly long so that their energy harvested at the receiver can rise above the noise. Being completely agnostic about the existing sources, the receiver tests the received signal $x(t)$ in the initial phase until it detects the presence of such signals in order to establish the active links. This requires accumulating observations and repeating the test sequentially. The signal features extracted from the link information are essential for the receiver to determine if it can decode data [3], [4] and refine the link parameters using mid-ambles and decoded data after the initial training phase.

A. Related Works on Link Acquisition of Multiuser Signals

We can categorize link acquisition algorithms into three main groups. The first category acquires a *sufficient statistic* by directly sampling $x(t)$ at (or above) the Nyquist rate. Given the set of active users \mathcal{I} , the likelihood function associated with the sufficient statistic is then exploited to detect the presence of signals and determine the link parameters. We refer to such techniques as Direct Sampling (DS) methods (e.g. [5]–[8]).

A second approach [4], [9], [10] facilitates the search of both the *active set* \mathcal{I} and link parameters by Matched Filtering (MF). This approach compares the filtered outputs of $x(t)$ from a bank of filters constructed by shifting and modulating the preamble $\phi_i(t - \tau)e^{-i\omega t}$, each of which matches a sufficiently wide collection of points in the full parameter set $\mathcal{T} \times \mathcal{F}$ where \mathcal{T} and \mathcal{F} are the delay and Doppler spread respectively. MF is a prevalent choice in hardware implementations because of its simplicity. The MF approach can be implemented in the digital domain, where samples are projected onto the sampled version of $\phi_i(t - \tau)e^{-i\omega t}$, or in the analog domain, where the receiver performs filtering operations onto the templates $\phi_i(t - \tau)e^{-i\omega t}$ directly in hardware. Specific details on these architectures are provided in Section III.

Classical algorithms take little advantage of the low dimensionality of the received signal space in storing and processing the observations to improve the performance or reduce complexity. Recently, there have been advances in exploiting *sparsity*, or the low dimensionality of the signal space, to improve receiver performance. One class of papers suggests using sparse signal recovery for the purpose of either detection or estimation. For instance, assuming that the signal is present, the results in [1], [2], [4], [11] deal with identification of the active users and/or estimation of signal parameters by creating a dictionary from the known templates $\phi_i(t)$ and viewing the signal $x(t)$ as a sparse linear combination of these templates. Without knowledge of signal presence, the proposed detection schemes in [12]–[16] use generic compressed measurements to detect the presence of certain signals within a specific observation window, starting from an abstract discrete model. We refer

Manuscript received December 09, 2011; revised September 15, 2012 and January 21, 2013; accepted February 19, 2013. Date of publication April 12, 2013; date of current version May 23, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Milica Stajanovic. This work was supported by the National Science Foundation under Grant CCF-1011811 and NeTS-ProWin-0626751. The work of Y. Eldar was supported in part by the Ollendorf Foundation and by the Israel Science Foundation under Grant 170/10.

X. Li, A. Rueetschi, and A. Scaglione are with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616 USA (e-mail: eceli@ucdavis.edu; andrear@ucdavis; ascaglione@ucdavis.edu).

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

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

Digital Object Identifier 10.1109/TSP.2013.2258014

to this general class of methods as the Sparsity-Aware (SA) approach. In these papers, delays and carrier offsets are not explicitly considered and the discrete observations are treated independently as a single snapshot from certain linear models, upon which SA algorithms are applied.

B. Multiuser Signals With Finite Rate of Innovation (FRI)

What is often neglected in existing DS and SA approaches is the acquisition of informative low rate discrete samples from the analog domain. As we mentioned, preamble sequences are usually fairly long and the receiver needs to sample the signal $x(t)$ at a fast rate and store these samples prior to processing. This can become a bottleneck in designing preambles so that they have the appropriate energy gain to rise above noise.

Reducing the sampling rate and the associated overhead incurred at the analog-to-digital (A/D) front-end is the concern of another line of studies [17]–[21] on signals with Finite Rate of Innovations (FRI) [22]. In general, an FRI signal has a *sparse* parametric representation. Given the preamble $\phi_i(t)$ for each user $i \in \mathcal{I}$ traveling through R channels, the class of signals $x(t)$ lies in a subspace with no more than $|\mathcal{I}|R$ dimensions, where each dimension has three unknowns (e.g., delay, carrier offset, channel coefficient), irrespective of its bandwidth and duration. The premier objective of FRI sampling is A/D conversion at sub-Nyquist rates for signal reconstructions. This objective is different from the interest of link acquisition in statistical inference. In this paper, we wish to harness similar benefits of the FRI architecture, while mitigating the detection performance losses due to reduced observations in the presence of noise.

We formulate the link acquisition problem as a *Sparsity Regularized Likelihood Ratio Test* (SR-LRT) using samples from a *Compressive Multi-channel Sampling* (CMS) architecture. To enhance acquisition performance, we optimize the front-end by choosing sampling kernels that maximize a probability divergence measure of all the hypotheses in the test. We refer to the proposed scheme as the *Compressive Sparsity Aware* (CSA) scheme. Specifically, we discuss in this paper:

- 1) a unified low-rate A/D conversion front-end using the proposed CMS architecture;
- 2) a SR-LRT that uses compressive samples from the CMS architecture for sequential joint detection and estimation;
- 3) the optimization of the CMS architecture for maximum average Kullback-Leibler (KL) distance of the SR-LRT;
- 4) the comparison of the proposed CSA scheme with the DSA and MF scheme in terms of performance and costs.

Our work bridges the results pertaining to sparsity-aware estimation/detection [1], [2], [4], [12]–[16] the literature on analog compressed sensing and sub-Nyquist sampling [4], [8], [10], [17], [18], [23] and FRI sampling [19], [20], [22] such that sampling and acquisition operations are considered jointly.

To measure the benefits of the proposed CSA scheme, we analyze the practical trade-off between the implementation costs in physically acquiring samples and those invested computationally in sparse recovery. This is important to clarify the potential benefits of sub-Nyquist architectures in communication receivers. These methods often benefit from the denoising capabilities of SA algorithms (as demonstrated in [1], [2], [4],

[12]–[16]) but must lose sensitivity due to the fact that they do not use sufficient statistics for the receiver inference.

The question we consider is, therefore, *what is there to gain: implementation costs or performance?* Our numerical experiments indicate that the main advantage of the proposed scheme is that it enables the designer to find an adequate operating point for link acquisition such that processing requirements and complexity of the receiver can be reduced to an acceptable level without significantly sacrificing acquisition performance compared with the MF architecture. We also confirm numerically that the optimized CMS architecture acquires highly informative samples for the SR-LRT in terms of estimation and detection performance.

C. Notation and Paper Organization

We denote vectors and matrices by boldface lower-case and boldface upper-case symbols and the set of real (complex) numbers by \mathbb{R} (\mathbb{C}). We denote sets by calligraphic symbols, where the intersection and the union of two sets \mathcal{A} and \mathcal{B} are written as $\mathcal{A} \cap \mathcal{B}$ and $\mathcal{A} \cup \mathcal{B}$ respectively. The operator $|\mathcal{A}|$ on a discrete (continuous) set takes the cardinality (measure) of the set. The magnitude of a complex number x is denoted by $|x| = \sqrt{xx^*}$, where x^* is the conjugate. The transpose, conjugate transpose, and inverse of a matrix \mathbf{X} are denoted by \mathbf{X}^T , \mathbf{X}^H and \mathbf{X}^{-1} , respectively. The symbol \otimes represents Kronecker product. The inner product between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{N \times 1}$ and between two continuous functions $f(t), g(t)$ in $L_2(\mathbb{C})$ are defined accordingly as $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^N y_n^* x_n$ and $\langle f(t), g(t) \rangle = \int_{-\infty}^{\infty} g^*(t) f(t) dt$. The \mathbf{W} -weighted ℓ_2 -norm of a vector \mathbf{x} with a positive definite matrix \mathbf{W} is denoted by $\|\mathbf{x}\|_{\mathbf{W}} = \sqrt{\mathbf{x}^H \mathbf{W} \mathbf{x}}$, and the conventional ℓ_2 -norm is written as $\|\mathbf{x}\|$. The L_2 -norm of a continuous-time signal $f(t) \in L_2(\mathbb{C})$ is $\|f(t)\| = \sqrt{\langle f(t), f(t) \rangle}$.

The paper is organized as follows. Section II introduces the received signal model. We discuss related works on link acquisition in Section III. The CMS architecture we propose is introduced in Section IV. Using the compressive samples obtained from the CMS architecture, we develop the SR-LRT module for CSA link acquisition in Section V. We then optimize the CMS front-end in Section VI. Simulations demonstrating the performance are presented in Section VII. The CSA scheme is compared against the MF scheme in terms of storage and computational costs in Section VIII.

II. SIGNAL MODEL FOR LINK ACQUISITION

In every communication standard, a key control sequence in the training phase is the initial preamble. Each user $i \in \mathcal{I}$ from the unknown active set transmits a specific preamble $\phi_i(t)$ to the receiver. This transmission is followed by the mid-ambles and payload. A common preamble in multiuser communications is a linearly pulse modulated sequence with a chip rate $1/T$ close to the signal bandwidth

$$\phi_i(t) = \sum_{m=0}^{M-1} a_i[m] g(t - mT). \quad (1)$$

Here $g(t)$ is the pulse shaping filter (chip) and $\{a_i[m]\}_{m=0}^{M-1}$ is typically a long preamble sequence $M \gg 1$ for each user.

A. Received Signal Model

The observation at the receiver can be written as

$$x(t) = \sum_{i \in \mathcal{I}} \sum_{r=1}^R h_{i,r} \phi_i(t - t_{i,r}) e^{i\omega_{i,r}t} + v(t), \quad (2)$$

where $t_{i,r}$ is the unknown propagation delay of the i th user in the r th multipath, $|\omega_{i,r}| \leq \omega_{\max}$ is the Doppler frequency upper bounded by the maximum Doppler spread ω_{\max} , and $h_{i,r}$ is the channel coefficient. Without loss of generality, we assume that the maximum multipath order R is known and the noise component $v(t)$ is a white Gaussian process with $\mathbb{E}\{v(t)v^*(s)\} = \sigma^2\delta(t-s)$. Our problem is to detect the presence of the active user set \mathcal{I} and the corresponding link parameters $\{h_{i,r}, t_{i,r}, \omega_{i,r}\}$ for $i \in \mathcal{I}$ and $r = 1, \dots, R$.

Since the propagation delays $t_{i,r}$ are unknown and possibly large, the typical A/D front-end for link acquisition is sequential. The acquisition scheme produces test statistics every D units of time, where D is the shift in the time reference for detections. At every shift $t = nD$, the receiver decides whether the signal $x(t)$ is present at or after $t = nD$. For convenience, we denote $t_0 = \min\{t_{i,r}\}_{i \in \mathcal{I}, r=1, \dots, R}$ as the delay of the first arrival path among all users. Let

$$\ell = \left\lfloor \frac{t_0}{D} \right\rfloor \quad (3)$$

be the shift that matches best with signal arrival and

$$\tau_{i,r} = t_{i,r} - \ell D \geq 0 \quad (4)$$

be the *composite delay*, where $0 \leq \tau_{i,r} \leq \tau_{\max}$ and τ_{\max} is the composite delay spread. Note that $\tau_{i,r} = (t_0 - \ell D) + (t_{i,r} - t_0)$, where the first term is the fractional delay within $[0, D)$ while the second term is a multipath delay relative to the first arrival path, which is bounded by the channel delay spread $\tilde{\tau}_{\max}$. This implies that $\tau_{i,r} \leq D + \tilde{\tau}_{\max}$. Given the multipath delay spread $\tilde{\tau}_{\max}$ and the shift size D , we can obtain the composite delay spread $\tau_{\max} = D + \tilde{\tau}_{\max}$ as the search space to fully capture the signal at least in the ℓ th shift.

This allows us to express (2) equivalently as

$$x(t) = \sum_{i \in \mathcal{I}} \sum_{r=1}^R h_{i,r} \phi_i(t - \ell D - \tau_{i,r}) e^{i\omega_{i,r}t} + v(t). \quad (5)$$

After these considerations, it is clear that the search spaces of delays and Doppler frequencies for each shift n are respectively $\mathcal{T} \triangleq [0, \tau_{\max}]$ and $\mathcal{F} \triangleq [-\omega_{\max}, \omega_{\max}]$.

B. Goal of Link Acquisition

Link acquisition is typically formulated as composite hypothesis tests with unknown link parameters, where the likelihood ratio between the signal hypothesis and the noise hypothesis is the test statistic for the detection task. Note that there could be multiple values of $n \neq \ell$ that lead to valid positive detections, where for a given ℓ , the relative composite delay with respect to the n th shift would be

$$\tau_{i,r}^{(n)} = \tau_{i,r} + (\ell - n)D. \quad (6)$$

Therefore, when the signal is captured in an earlier shift $n < \ell$, the relative composite delay would be greater than $\tau_{i,r}$, and if it is captured in a later shift $n > \ell$ the relative composite delay would be smaller than $\tau_{i,r}$. In order to single out the best reference shift, the receiver will have to compare a sequence of N_0 test statistics after the first positive detection at $n = N_\eta$, and choose the particular shift ℓ_* that maximizes the likelihood ratio. We call ℓ_* the *maximum likelihood ratio* (MLR) shift. The look-ahead horizon N_0 can be chosen considering the type of sampling kernels, the preambles $\phi_i(t)$'s, and the delay spread τ_{\max} , making reasonable approximations about the duration of the signals.

Definition 1: Link acquisition refers to

- 1) locating the MLR shift ℓ_* ;
- 2) identifying the set of active users \mathcal{I} in the ℓ_* th shift;
- 3) resolving the delay-Doppler pairs $\{\hat{\tau}_{i,r}, \hat{\omega}_{i,r}\}$ for $i \in \mathcal{I}$ and $r = 1, \dots, R$.

Usually, the preamble signals $\phi_i(t)$ have large energy, so that they can rise above the receiver noise. Given that the average power is constant, the $\phi_i(t)$ typically last much longer (i.e., M is large) than subsequent mid-ambles or spreading codes that modulate data. For a typical wireless application such as GPS or IS-95/IMT-2000, transmitters continuously send out preamble sequences with length on the order of $M = 20 \times 1023$ [24] or $M = 32768$ [25], respectively. This means that in order to detect the presence of such preambles and acquire the synchronization parameters, architectures using DS, MF or SA approaches would have to store a large amount of data to process in a sequential manner. This phase is crucial to properly initialize any channel tracking that ensues. In Section III, we provide details on the A/D architectures and the corresponding post-processing for conventional link acquisition. We then present the proposed CMS architecture and the CSA acquisition scheme in Section IV.

III. EXISTING ARCHITECTURES FOR LINK ACQUISITION

For future use, we let the Nyquist rate of the signal $x(t)$ be $f_{\text{NYQ}} = 2W + \omega_{\max}/\pi$ with W being the maximum single-sided bandwidth of $\phi_i(t)$, $i = 1, \dots, I$.

A. Direct Sampling (DS)

In DS schemes, the received analog signal $x(t)$ is sampled at rate $f_s = 1/T_s \geq f_{\text{NYQ}}$, resulting in samples $c_{\text{DS}}[w] = \langle x(t), \delta(t - wT_s) \rangle = x(wT_s)$. Then DS schemes use the most recent W Nyquist samples for every shift of $D = NT_s$

$$\mathbf{c}_{\text{DS}}[n] = [c_{\text{DS}}[nN], \dots, c_{\text{DS}}[nN + (W - 1)]]^T \quad (7)$$

to perform detection. This is a sliding window operation of W samples where in each shift, the most obsolete N samples are replaced with the latest N samples, where we choose $W \geq N$ such that no samples are missed between shifts.

Based on (2), the samples $\mathbf{c}_{\text{DS}}[n]$ can be expressed as

$$\mathbf{c}_{\text{DS}}[n] = \mathbf{\Phi}_{\mathcal{J}}(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}) \mathbf{h}_{\mathcal{J}} + \mathbf{v}[n], \quad (8)$$

where the parameters

$$\boldsymbol{\tau}_{\mathcal{J}} \triangleq [\dots, \tau_{j,1}^{(n)}, \dots, \tau_{j,R}^{(n)}, \dots]^T \quad (9)$$

$$\boldsymbol{\omega}_{\mathcal{J}} \triangleq [\cdots, \omega_{j,1}, \cdots, \omega_{j,R}, \cdots]^T \quad (10)$$

$$\mathbf{h}_{\mathcal{J}} \triangleq [\cdots, h_{j,1}, \cdots, h_{j,R}, \cdots]^T \quad (11)$$

represent the $|\mathcal{J}|R$ residual delays, Doppler and channel coefficients corresponding to the set of users $\mathcal{J} \subseteq \{1, \cdots, I\}$ in the n th shift. The set of active users \mathcal{J} varies with n depending on which user component is captured in that shift. We omit the argument n to make the notations lighter. The vector $\mathbf{v}[n]$ contains the noise samples $[\mathbf{v}[n]]_w = v(nD + wT_s)$, and $\Phi_{\mathcal{J}}(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}})$ is a $W \times |\mathcal{J}|R$ sub-matrix of the complete $W \times IR$ matrix $\Phi(\boldsymbol{\tau}, \boldsymbol{\omega})$, from which we extract columns $j \in \mathcal{J}$. The full matrix $\Phi(\boldsymbol{\tau}, \boldsymbol{\omega})$ is defined by

$$[\Phi(\boldsymbol{\tau}, \boldsymbol{\omega})]_{w, (i-1)R+r} \triangleq \phi_i(wT_s - L_gT - \tau_{i,r}) e^{i\omega_{i,r}wT_s},$$

where L_gT is the duration of the pulse $g(t)$ on one side¹.

Using the n th shift of observations, link acquisition amounts to the following composite hypothesis test

$$\begin{aligned} \mathcal{H}_{\mathcal{J}} : \mathbf{c}_{\text{DS}}[n] &= \Phi_{\mathcal{J}}(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}) \mathbf{h}_{\mathcal{J}} + \mathbf{v}[n], \\ \mathcal{H}_{\emptyset} : \mathbf{c}_{\text{DS}}[n] &= \mathbf{v}[n], \end{aligned}$$

with unknown parameters $\mathcal{J}, \boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}$ and $\mathbf{h}_{\mathcal{J}}$. The Generalized Likelihood Ratio Test (GLRT) is typically used, which requires solving a non-linear least squares (NLLS) problem

$$\{\hat{\mathcal{I}}, \hat{\boldsymbol{\tau}}_{\hat{\mathcal{I}}}, \hat{\boldsymbol{\omega}}_{\hat{\mathcal{I}}}, \hat{\mathbf{h}}_{\hat{\mathcal{I}}}\} = \arg \min_{\mathcal{J}, \boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}, \mathbf{h}_{\mathcal{J}}} \|\mathbf{c}_{\text{DS}}[n] - \Phi_{\mathcal{J}}(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}) \mathbf{h}_{\mathcal{J}}\|^2,$$

over all possible $\mathcal{J}, (\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}) \in \mathcal{T}^{|\mathcal{J}|} \times \mathcal{F}^{|\mathcal{J}|}, \mathbf{h}_{\mathcal{J}} \in \mathbb{C}^{|\mathcal{J}|}$ to compute the generalized likelihood ratio

$$\eta_{\text{DS}}(n) = \frac{\mathbb{P}(\mathcal{H}_{\hat{\mathcal{I}}})}{\mathbb{P}(\mathcal{H}_{\emptyset})} \quad (12)$$

with estimates $\{\hat{\mathcal{I}}, \hat{\boldsymbol{\tau}}_{\hat{\mathcal{I}}}, \hat{\boldsymbol{\omega}}_{\hat{\mathcal{I}}}, \hat{\mathbf{h}}_{\hat{\mathcal{I}}}\}$ obtained at every shift $t = nD$. The expression of the generalized likelihood ratio is given in [26] for cases with known or unknown noise variance σ^2 . Without loss of generality, we consider the most general case where σ^2 is unknown. In this case, the ratio is obtained as

$$\eta_{\text{DS}}(n) = \frac{\|\mathbf{c}_{\text{DS}}[n]\|^{2W}}{\left\| \mathbf{c}_{\text{DS}}[n] - \Phi_{\hat{\mathcal{I}}}(\hat{\boldsymbol{\tau}}_{\hat{\mathcal{I}}}, \hat{\boldsymbol{\omega}}_{\hat{\mathcal{I}}}) \hat{\mathbf{h}}_{\hat{\mathcal{I}}} \right\|^{2W}}. \quad (13)$$

Using this ratio as the test statistics, the receiver checks if the statistic satisfies $\eta_{\text{DS}}(n) \geq \eta_0$ for some properly chosen threshold η_0 . Denote the first shift that passes the GLRT as

$$N_{\eta} \triangleq \min_n \left\{ \arg \eta_{\text{DS}}(n) \geq \eta_0 \right\}. \quad (14)$$

The MLR shift ℓ_{\star} is then given by

$$\ell_{\star} = \arg \max_n \eta_{\text{DS}}(n), \quad n = N_{\eta}, \cdots, N_{\eta} + N_0. \quad (15)$$

¹A pulse has finite durations only if $g(t)$ is not bandlimited. If the pulse is bandlimited, the pulse $g(t)$ consists of side-lobes of length T . The parameter L_g in this case is chosen to provide modest approximation errors.

The test above is intractable in general because there are 2^I hypotheses at each shift to explore, and there is an NLLS problem to solve for each hypothesis. Therefore in practice, DS schemes either deal with the known user case $\mathcal{J} = \mathcal{I}$ or assume the full set $\mathcal{J} = \{1, \cdots, I\}$ during detection, followed by NLLS for that specific user set. When the active users \mathcal{I} are unknown, alternatives are *Matched Filtering* (MF) and *Sparsity-Aware* (SA) approaches, as described next.

B. Matched Filtering (MF)

The MF receiver is widely used in practice. It correlates the signal $x(t)$ with a filterbank constructed from the *MF templates* $\phi_i(t) e^{ik\Delta\omega t}$. Since the size of the filterbank has to be finite, it is usually assumed that $\tau_{i,r} \approx q_{i,r} \Delta\tau$ and $\omega_{i,r} \approx k_{i,r} \Delta\omega$ for some integers $q_{i,r}$ and $k_{i,r}$ with a certain resolution $\Delta\tau = \tau_{\max}/Q$ and $\Delta\omega = \omega_{\max}/K$. The search spaces for the MF receiver then become $\mathcal{Q} = \{0, 1, \cdots, Q-1\}$ and $\mathcal{K} = \{-K, \cdots, K\}$, which is the discrete counterpart of the continuous search space $\mathcal{T} \times \mathcal{F}$.

The MF obtains the decision statistics by passing $x(t)$ through a bank of $P_{\text{MF}} = I|\mathcal{K}|$ MF templates, and sampling the outputs every $\Delta\tau$. To be consistent with the sequential structure in (5) and the DS method, the MF shifts its templates every $D = NT_s$, and samples the outputs every $\Delta\tau = T_s \leq 1/f_{\text{NYQ}}$. The MF output corresponding to the i th user at the k th discrete frequency $\omega = k\Delta\omega$ is obtained as

$$c_{i,k}[w] = \left\langle x(t), \phi_i(t - wT_s) e^{ik\Delta\omega(t - wT_s)} \right\rangle. \quad (16)$$

Oftentimes, the filtering process is implemented in the digital domain using the samples $\mathbf{c}_{\text{DS}}[n]$ in (7). For consistency, we proceed with the description in the analog domain. At the n th shift, the samples used for detection can be stacked into an $I|\mathcal{K}| \times |\mathcal{Q}|$ exhaustive MF output array

$$\mathbf{C}_{\text{MF}}[n] = \begin{bmatrix} \vdots & \cdots & \vdots \\ c_{1,k}[nN] & \cdots & c_{1,k}[nN + Q - 1] \\ \vdots & \cdots & \vdots \\ c_{I,k}[nN] & \cdots & c_{I,k}[nN + Q - 1] \\ \vdots & \cdots & \vdots \end{bmatrix}. \quad (17)$$

Then, the MF receiver uses $\mathbf{C}_{\text{MF}}[n]$ as test statistics and detects the active user set at the n th shift as

$$\hat{\mathcal{I}} = \{i : |c_{i,k}[nN + q]| \geq \rho_i, \quad \forall i, k, q\},$$

where ρ_i is the chosen detection threshold for each user. The shift that triggers the first positive detection is denoted as

$$N_{\eta} \triangleq \min_n \left\{ \arg |c_{i,k}[nN + q]| \geq \rho_i, \quad \forall i, k, q \right\}. \quad (18)$$

The MLR shift is obtained by locating the maximum output

$$\ell_{\star} \triangleq \arg \max_n \left\{ \max_{i,k,q} |c_{i,k}[nN + q]| \right\}, \quad n = N_{\eta}, \cdots, N_{\eta} + N_0.$$

Given the multipath order R , the delay-Doppler pairs at the n th shift are pinpointed by the R strongest outputs for all detected users $i \in \hat{\mathcal{I}}$ over the search space $k \in \mathcal{K}$ and $q \in \mathcal{Q}$. For convenience, we denote the strongest path by $(k_{i,1}, q_{i,1})$ for the i th user $i \in \hat{\mathcal{I}}$ and rank the outputs by magnitudes

$$\begin{aligned} |c_{i,k_{i,1}}[\ell_*N + q_{i,1}]| &> |c_{i,k_{i,2}}[\ell_*N + q_{i,2}]| > \dots \\ &> |c_{i,k_{i,R}}[\ell_*N + q_{i,R}]| > \dots \end{aligned} \quad (19)$$

for each user $i \in \hat{\mathcal{I}}$. The delay-Doppler pairs are identified as

$$\mathcal{M}_i \triangleq \{(k_{i,r}, q_{i,r}) : r = 1, \dots, R\}, \quad (20)$$

which give the following link parameters

$$\hat{\tau}_{i,r} = q_{i,r} \Delta\tau, \quad \hat{\omega}_{i,r} = k_{i,r} \Delta\omega, \quad (k_{i,r}, q_{i,r}) \in \mathcal{M}_i.$$

Although the MF approach shows an advantage in its post-processing and implementation, it has a few drawbacks:

- i) the size of the MF filterbank scales with the number of users I and the parameter set $|\mathcal{K}|$;
- ii) digital implementation requires high rate processing, increasing storage and pipelining² cost;
- iii) the MF samples $c_{i,k}[\cdot]$ contain interference from different users and multipath components;

Usually the effect of interference (iii) is mitigated by using wideband pulses $g(t)$. During data detection, multipath interferences are dealt with using RAKE type receivers and the multiuser interference is tackled either by using linear multiuser receivers or, in some cases, successive interference cancellation (SIC) or even maximum likelihood multiuser detection [3]. Typically, the complexities of these schemes for data detection grow rapidly with the size of the MF filterbank (i) and the sampling rate (ii). Since data detection is conducted after link acquisition, the uncertainties about the set of active users, their delays and Doppler frequencies have already been resolved and therefore, these tasks become more manageable.

The MF receiver is a popular choice for multiuser acquisition [3], for example, in GPS receivers [27] or CDMA receivers [25]. Its comparison with the CSA scheme using the CMS architecture we propose in this paper is insightful because, although the MF front-end requires a large filterbank, the post-processing of its outputs is very simple. One could certainly perform more complex post-processing to enhance its performance. For instance, the Orthogonal Matching Pursuit (OMP) algorithm used in our CSA scheme can be applied on the MF outputs for better results. However, in that case, as illustrated in Section VIII, the resulting scheme will have much higher storage cost and computational complexity, which render the merits of the MF approach meaningless. More importantly, the OMP algorithm can be applied to the Nyquist samples as in SA methods, making the MF stage superfluous³.

²Pipelining refers to timely processing of the samples that stream into the system per unit of time.

³Strictly speaking, the OMP technique performs a MF stage in its first iteration. The subsequent iterations can be viewed as applying successive interference cancellation (SIC) in multiuser communications.

C. Sparsity Aware (SA) Approach

Instead of simply observing the MF outputs, many recent works have proposed the idea of *compressed sensing* or *sparse recovery* to solve estimation and detection problems. For the purpose of *user identification and parameter estimation*, one approach is to approximate (8) by a sparse model with a dictionary constructed from the ensemble of possible templates $\phi_i(t)$ [1], [4] and/or discretized delays τ [2], [28] (similar to the MF templates), where the joint recovery of active users and unknown parameters is relaxed as a *sparse estimation* problem. These sparse methods, which we call the *Direct Sparsity Aware* (DSA) scheme, usually require Nyquist samples and assume that the signal is already present (i.e., the MLR shift ℓ_x is known). For clarity, DSA scheme should not be confused with DS schemes, because DS schemes do not resort to sparsity approach based on discretization of analog parameters and require a non-linear search over the parameter space $\mathcal{T} \times \mathcal{F}$.

On the other hand, aiming at *signal presence detection* rather than identifying the active users and recovering the parameters, [12]–[16] reduce the number of samples required for the test by using a linear *compressor* on the block of given discrete observations. The compressor used in this class of detection schemes can in fact be implemented using our proposed CMS architecture in the digital domain, therefore we also categorize this method as the CSA scheme discussed in this paper to avoid confusion. Last but not least, the DSA scheme can also be regarded as a special case of the proposed CSA scheme with a compressor that is an identity matrix.

A distinctive difference between the CSA scheme we propose and those in [12]–[16] is that our CSA scheme unifies the sequential signal detection, user identification and parameter estimation by using the compressive samples obtained from a flexible multi-rate A/D architecture. On the other hand, the CSA schemes in [12]–[16] directly start from an abstract discrete model that is already sampled. Furthermore, the sampling kernels in the proposed CMS architecture are optimized with respect to the estimation and detection performance in terms of the average KL distance of the hypotheses in the SR-LRT.

IV. COMPRESSIVE SEQUENTIAL LINK ACQUISITION

A. Compressive Multichannel Sampling (CMS)

We propose to use the A/D front-end in Fig. 1, typical in FRI sampling [17], [19], [20], [22], [23]. The signal is sampled every $t = nD$ by a P -channel filterbank

$$c_p[n] \triangleq \langle x(t), \psi_p(t - nD) \rangle, \quad p = 1, \dots, P. \quad (21)$$

We call this architecture the Compressive Multichannel Sampling (CMS) module, which forms the A/D conversion front-end of the proposed CSA acquisition scheme.

Note that (21) can also be implemented in the digital domain by performing linear projections of the discrete signal $\mathbf{c}_{DS}[n]$ in (8). This means that the CMS architecture becomes part of the post-processing of the Nyquist samples of $x(t)$, which lowers the storage and computation requirements as illustrated

in Section VIII. Similar derivations can be done in discrete time, but the advantage of using the analog description is that we do not necessarily have to target bandlimited signals. In the FRI literature [17]–[20], [22], [23], in the absence of noise, the sampling rate required for the unique reconstruction of the signal in (5) is the *number of degrees of freedom* of the signal $x(t)$ per shift D , equal to the number of unknowns $\{\tau_{i,r}, \omega_{i,r}, h_{i,r}\}_{i \in \mathcal{I}, r=1, \dots, R}$. This amounts to $P_{\text{FRI}} = 3|\mathcal{I}|R$, which can be much less than what is needed in the MF approach $P_{\text{FRI}} \ll P_{\text{MF}} = I|\mathcal{K}|$ when the number of active users is not large $|\mathcal{I}| \ll I$, or the number of multipaths R is much less than the dimension of the search space for Doppler $|\mathcal{K}|$. However, since the estimation and detection are performed in the presence of noise, the value of P needs to be increased in general to enhance the sensitivity of the receiver. This gives the option of trading off accuracy with storage cost and computational complexity, by adjusting the number of samples P to process between $P_{\text{FRI}} \leq P \leq P_{\text{MF}}$.

Note that for different schemes, we need different post-processing to produce final decisions for different receivers. For instance, the MF scheme has very simple post-processing at the cost of handling exhaustive MF samples, while the CSA scheme can tune the number of measurements to handle less data by spending a higher premium for sparse recovery. Therefore, we discuss this trade-off in detail in Section VIII.

B. CMS Observation Model

Similar to [1], [2], we follow the analog description of (5) but discretize the parameters as in the MF approach. For notational convenience, we introduce the triple-index coefficient

$$\alpha_{i,k,q} = \sum_{j \in \mathcal{I}} \sum_{r=1}^R h_{j,r} \delta[i-j] \delta[k-k_{j,r}] \delta[q-q_{j,r}], \quad (22)$$

for $k \in \mathcal{K}, q \in \mathcal{Q}$ as an indicator of whether the i th user is transmitting and whether there exists a link at a certain delay $\tau = q\Delta\tau$ with a certain carrier offset $\omega = k\Delta\omega$ in the window. Note that $\alpha_{i,k,q} = 0$ except when $k = k_{i,r}$ and $q = q_{i,r}$ for $i \in \mathcal{I}$. Denoting the MF template as $\phi_{i,k,q}(t) \triangleq \phi_i(t - q\Delta\tau) e^{ik\Delta\omega t}$, the signal in (5) can be approximated as

$$x(t) = \sum_{i=1}^I \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}} \alpha_{i,k,q} e^{ik\Delta\omega \ell D} \phi_{i,k,q}(t - \ell D) + v(t). \quad (23)$$

Clearly, $x(t)$ has at most $|\mathcal{I}|R$ active components due to the sparsity of $\alpha_{i,k,q}$. To facilitate notations in our derivations, we introduce the triplet index (i, k, q) and define the length- $I|\mathcal{K}||\mathcal{Q}|$ link vector $\boldsymbol{\alpha}[\ell]$ at the ℓ th shift as

$$[\boldsymbol{\alpha}[\ell]]_{(i,k,q)} \triangleq [\boldsymbol{\alpha}[\ell]]_{(i-1)|\mathcal{K}||\mathcal{Q}| + (k+K-1)|\mathcal{Q}| + q} \quad (24)$$

$$= \alpha_{i,k,q} \quad (25)$$

for any $i = 1, \dots, I, k \in \mathcal{K}$ and $q \in \mathcal{Q}$. We define the associated delay-Doppler set for the i th user at the ℓ th shift

$$\mathcal{A}_i^{(\ell)} \triangleq \{(k, q) : |\alpha_{i,k,q}| > 0, k \in \mathcal{K}, q \in \mathcal{Q}\}, \quad (26)$$

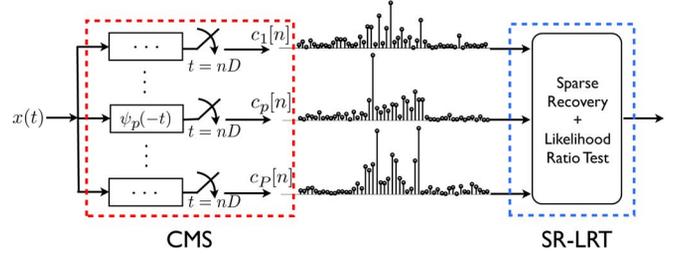


Fig. 1. The compressive samples obtained in the CMS architecture and the SR-LRT in the proposed CSA acquisition scheme.

from which we extract the delays $\tau_{i,r} = q_{i,r}\Delta\tau$ and carrier offsets $\omega_{i,r} = k_{i,r}\Delta\omega$ for the active users $i \in \mathcal{I}$ if $\mathcal{A}_i^{(\ell)} \neq \emptyset$.

We consider sampling kernels $\psi_p(t)$ that are linear combinations of MF templates [23]. We specify the observation model for the CMS architecture in Fig. 1 using the link vector $\boldsymbol{\alpha}[\ell]$ in the theorem below.

Theorem 1: Suppose that we choose sampling kernels $\{\psi_p(t)\}_{p=1}^P$ as linear combinations of the MF templates

$$\psi_p(t) = \sum_{i=1}^I \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}} b_{p,(i,k,q)} \phi_{i,k,q}(t), \quad p = 1, \dots, P. \quad (27)$$

The length- P sample vector $\mathbf{c}[n] \triangleq [c_1[n], \dots, c_P[n]]^T$ taken at shift $t = nD$, can then be expressed as

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}_{\phi\phi}[n - \ell]\boldsymbol{\Gamma}[\ell]\boldsymbol{\alpha}[\ell] + \boldsymbol{\nu}[n], \quad (28)$$

where $\boldsymbol{\alpha}[\ell]$ is the link vector at the ℓ th shift and

- 1) \mathbf{B} is a $P \times I|\mathcal{K}||\mathcal{Q}|$ matrix with $[\mathbf{B}]_{p,(i,k,q)} \triangleq b_{p,(i,k,q)}$;
- 2) $\mathbf{M}_{\phi\phi}[n - \ell]$ is an $I|\mathcal{K}||\mathcal{Q}| \times I|\mathcal{K}||\mathcal{Q}|$ matrix with

$$[\mathbf{M}_{\phi\phi}[n - \ell]]_{(i',k',q'),(i,k,q)} \triangleq R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n - \ell)D],$$

where $R_{\phi_{i',k',q'}\phi_{i,k,q}}[\Delta t] = R_{\phi_{i',\phi_i}}^{(k-k')}[(q' - q)\Delta\tau + \Delta t] \times e^{ik\Delta\omega(\Delta t - q'\Delta\tau)}$ and $R_{\phi_{i',\phi_i}}^{(k-k')}(\Delta t)$ is the ambiguity function

$$R_{\phi_{i',\phi_i}}^{(k-k')}(\Delta t) = \int \phi_{i'}^*(t) \phi_i(t - \Delta t) e^{i(k-k')\Delta\omega t} dt. \quad (29)$$

- 3) $\boldsymbol{\Gamma}[\ell] \triangleq \mathbf{I}_I \otimes \mathbf{E}[\ell] \otimes \mathbf{I}_{|\mathcal{Q}|}$ with

$$\mathbf{E}[\ell] \triangleq \text{diag}[e^{iK\Delta\omega\ell D}, \dots, e^{-iK\Delta\omega\ell D}]; \quad (30)$$

- 4) $\boldsymbol{\nu}[n] \triangleq [\nu_1[n], \dots, \nu_P[n]]^T$ is the filtered Gaussian noise vector with zero mean and covariance

$$\mathbf{R}_{\nu\nu} = \sigma^2 \mathbf{R}_{\psi\psi} \quad \text{with} \quad \mathbf{R}_{\psi\psi} = \mathbf{B}\mathbf{M}_{\phi\phi}[0]\mathbf{B}^H. \quad (31)$$

Proof: See Appendix A. ■

The freedom in choosing \mathbf{B} allows us to optimize acquisition performance. Before discussing the details of optimization in Section VI, we further simplify the model in Theorem 1.

C. CMS Sequential Acquisition Model

Theorem 1 describes the general model of the samples $\mathbf{c}[n]$ obtained in the n th shift with respect to the link vector $\boldsymbol{\alpha}[\ell]$.

However, the exact shift ℓ is unknown to the receiver. As mentioned earlier, determining the exact shift is not necessary to recover the link parameters, as long as the shift is properly aligned with the signal and produces a positive detection maximizing the likelihood ratio. In the following, we transform the observation model $\mathbf{c}[n]$ in Theorem 1 to an equivalent model. The equivalent model is stated with respect to a modified link vector $\boldsymbol{\alpha}[n]$ at the n th shift, which contains entries that are shifted with a relative placement of $(n - \ell)$ in relation to $\boldsymbol{\alpha}[\ell]$. The reason for this is that we can use a time-invariant system matrix instead of a time-variant one $\mathbf{M}_{\phi\phi}[n - \ell]$ for the purpose of sequential detection.

Theorem 2: Let $D = N\Delta\tau$ for some integer $N \in \mathbb{Z}$. The outputs $\mathbf{c}[n]$ of the compressive samplers can be re-written as

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}\boldsymbol{\Gamma}[n]\boldsymbol{\alpha}[n] + \boldsymbol{\nu}[n], \quad (32)$$

where $\mathbf{M} \triangleq \mathbf{M}_{\phi\phi}[0]$ and $\boldsymbol{\alpha}[n]$ is the n th shift link vector

$$[\boldsymbol{\alpha}[n]]_{(i,k,q)} \triangleq \alpha_{i,k,q+(n-\ell)N}. \quad (33)$$

Proof: See Appendix B. ■

Corollary 1: Let the *delay-Doppler sets* at the n th shift be

$$\mathcal{A}_i^{(n)} \triangleq \left\{ (k, q) : \left| [\boldsymbol{\alpha}[n]]_{(i,k,q)} \right| \neq 0, k \in \mathcal{K}, q \in \mathcal{Q} \right\} \quad (34)$$

for $i = 1, \dots, I$. Then for any $(k, q) \in \mathcal{A}_i^{(\ell)}$ at the ℓ th shift (26), we have $(k, q + (\ell - n)N) \in \mathcal{A}_i^{(n)}$ at the n th shift.

Using the modified sets $\mathcal{A}_i^{(n)}$, the number of delay-Doppler pairs included at the n th shift equals $\sum_{i=1}^I |\mathcal{A}_i^{(n)}|$. It is obvious that $|\mathcal{A}_i^{(n)}| \leq |\mathcal{A}_i^{(\ell)}|$ for any i . At the n th shift, if a positive detection is declared and $|\mathcal{A}_i^{(n)}| = |\mathcal{A}_i^{(\ell)}|$ for all i , then the modified link vector $\boldsymbol{\alpha}[n]$ carries equivalent link information as the link vector $\boldsymbol{\alpha}[\ell]$ at the ℓ th shift. Therefore, we use the model in Theorem 2 and re-state the goal of *link acquisition*

- 1) locating the MLR shift ℓ_* ;
- 2) identifying the set of active users $\widehat{\mathcal{I}}$ indicated by the *delay-Doppler set* $\mathcal{A}_i^{(\ell_*)} \neq \emptyset$;
- 3) resolving the delay-Doppler pairs in the ℓ_* th window $\mathcal{A}_i^{(\ell_*)} \subseteq \mathcal{K} \times \mathcal{Q}$ for $i \in \widehat{\mathcal{I}}$.

For better representation and comparison of the individual support set $\mathcal{A}_i^{(n)}$ in relation to the original support set $\mathcal{A}_i^{(\ell)}$, we introduce the full *user-delay-Doppler sets* for the link vectors $\boldsymbol{\alpha}[n]$ and $\boldsymbol{\alpha}[\ell]$ respectively

$$\mathcal{A}_n \triangleq \left\{ (i, k, q) : (k, q) \in \mathcal{A}_i^{(n)}, i \in \mathcal{I} \right\}, \quad (35)$$

$$\mathcal{A}_\ell \triangleq \left\{ (i, k, q) : (k, q) \in \mathcal{A}_i^{(\ell)}, i \in \mathcal{I} \right\}. \quad (36)$$

In the following sections, we express the link vector explicitly with respect to the full *user-delay-Doppler set* \mathcal{A}_n and combine the phase rotation matrix $\boldsymbol{\Gamma}[n]$ at the n th shift as

$$\boldsymbol{\beta}_{\mathcal{A}_n} \triangleq \boldsymbol{\Gamma}[n]\boldsymbol{\alpha}[n]. \quad (37)$$

We call $\boldsymbol{\beta}_{\mathcal{A}_n} = [\dots, \beta_{i,k,q}, \dots]^T$ the *modified link vector* and note that it is also a $|\mathcal{I}R$ -sparse vector.

V. SPARSITY REGULARIZED LIKELIHOOD RATIO TEST

We now develop an SR-LRT detection algorithm that tackles the link acquisition problem exploiting the compressive observation model given in Theorem 2. The goal of link acquisition is to pinpoint the true set $\mathcal{S}_n = \mathcal{A}_n$ by performing a multiple composite hypothesis test

$$\mathcal{H}_{\mathcal{S}_n} : \quad \mathbf{c}[n] = \mathbf{B}\mathbf{M}\boldsymbol{\beta}_{\mathcal{S}_n} + \boldsymbol{\nu}[n] \quad (38)$$

over all possible \mathcal{S}_n at every shift $t = nD$. Note that the detection of signal presence is incorporated in this test by setting $\mathcal{S}_n = \emptyset$ as the null hypothesis. Given a set \mathcal{S}_n , the amplitudes $\boldsymbol{\beta}_{\mathcal{S}_n}$ and noise variance σ^2 are unknown and treated as nuisance parameters. Link acquisition is to detect the full *user-delay-Doppler set* \mathcal{S}_n for all possible $\mathcal{H}_{\mathcal{S}_n}$. Following the GLRT rationale, the test finds the set \mathcal{S}_n that gives the maximum log-likelihood associated with each hypothesis

$$\log \mathbb{P}(\mathcal{H}_{\mathcal{S}_n} | \boldsymbol{\beta}_{\mathcal{S}_n}, \sigma^2) \propto -P \log \sigma^2 - \frac{1}{\sigma^2} \|\mathbf{c}[n] - \mathbf{B}\mathbf{M}\boldsymbol{\beta}_{\mathcal{S}_n}\|_{\mathbf{R}_{\psi\psi}^{-1}}^2 \quad (39)$$

in the presence of unknown parameters $\boldsymbol{\beta}_{\mathcal{S}_n}$ and σ^2 .

Note that when $\mathbf{B} = \mathbf{I}$, the samples $\mathbf{c}[n]$ are equivalent to the outputs of the MF approach. This implies that the samples $\mathbf{c}[n]$ obtained in the CMS architecture, using only P sampling kernels in (27), are equivalent to a linearly compressed version of the exhaustive MF output. The difference is that the samples here are obtained directly from the A/D architecture, instead of linearly compressing the exhaustive MF filterbank outputs as in [14], [15], where the latter would be much more complex. On the other hand, the difference in post-processing between the MF and CMS architectures is that MF allows to simply pick the hypothesis corresponding to the largest magnitude in the output $\mathbf{c}[n]$ as the detection result, while a more sophisticated detection scheme is necessary when using the compressive samples $\mathbf{c}[n]$.

A. Sequential Estimation for Link Acquisition

The GLRT requires estimating $\boldsymbol{\beta}_{\mathcal{S}_n}$ and σ^2 for every possible \mathcal{S}_n at every shift $t = nD$. For every hypothesis $\mathcal{S}_n \neq \emptyset$, the estimate of $\boldsymbol{\beta}_{\mathcal{S}_n}$ under colored Gaussian noise $\boldsymbol{\nu}[n]$ with covariance $\mathbf{R}_{\nu\nu} = \sigma^2 \mathbf{R}_{\psi\psi}$ is

$$\widehat{\boldsymbol{\beta}}_{\mathcal{S}_n} \triangleq \arg \min_{\boldsymbol{\beta}_{\mathcal{S}_n}} \|\mathbf{c}[n] - \mathbf{B}\mathbf{M}\boldsymbol{\beta}_{\mathcal{S}_n}\|_{\mathbf{R}_{\psi\psi}^{-1}}^2. \quad (40)$$

The “hat” notation $\widehat{(\cdot)}$ on the vector $\boldsymbol{\beta}_{\mathcal{S}_n}$ refers to the estimates of the amplitudes on the support \mathcal{S}_n . The total number of such estimates scales with the number of hypothesis which in this case is $2^{|\mathcal{K}||\mathcal{Q}|}$, resulting in an NP-hard combinatorial estimation problem. Instead, we solve this combinatorial problem in a “soft” fashion at every shift $t = nD$ similar to [2]

$$\widehat{\boldsymbol{\beta}} \triangleq \arg \min_{\boldsymbol{\beta}} \|\mathbf{c}[n] - \mathbf{B}\mathbf{M}\boldsymbol{\beta}\|_{\mathbf{R}_{\psi\psi}^{-1}}^2 + \lambda \cdot f(\boldsymbol{\beta}), \quad (41)$$

where λ is some regularization parameter and $f(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_0$ or $f(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ are the sparsity regularization constraint. If the $\|\cdot\|_0$ constraint is imposed, then the problem is approximately solved via greedy methods such as orthogonal matching pursuit (OMP) [29]. When $\|\cdot\|_1$ norm is used, this problem can be

solved via convex programs [30]. Generally speaking, as discussed in [2] and [30], the required number of samples P for sparse recovery in the noiseless case scales logarithmically with the length- $I|\mathcal{K}||\mathcal{Q}|$ link vector $P \propto |\mathcal{I}|R \log I|\mathcal{K}||\mathcal{Q}|$, which increases if discretizations are made finer.

From the solution of (41), we extract the full *user-delay-Doppler set* $\hat{\mathcal{A}}_n$ as explained in Section V-B. With the estimated set of active users $\hat{\mathcal{I}}$ and individual *delay-Doppler set* $\hat{\mathcal{A}}_i^{(n)}$, we have the truncated estimate of the link vector $\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}_n}$ and the estimated noise variance

$$\hat{\sigma}_{\hat{\mathcal{A}}_n}^2 = \left\| \mathbf{c}[n] - \mathbf{B}\mathbf{M}\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}_n} \right\|_{\mathbf{R}_{\psi\psi}^{-1}}^2 / P. \quad (42)$$

Since the formulation in (41) is no longer maximum likelihood due to the sparsity regularization, we call it the Sparsity-Regularized Likelihood Ratio Test (SR-LRT).

B. User-Delay-Doppler Set Extraction

Given the soft estimate $\hat{\boldsymbol{\beta}}$ in (41) at every shift $t = nD$, the estimated user-delay-Doppler set $\hat{\mathcal{A}}_n$ is extracted depending on the application scenarios below.

1) *Unknown, Random Number of Active Users \mathcal{I}* : In random access communications the receiver has no knowledge of who is active, nor any expectation on the number of active components. Using the estimate $\hat{\boldsymbol{\beta}}$, we identify the active users

$$\hat{\mathcal{I}} \triangleq \left\{ i : \max_{k,q} |\hat{\beta}_{i,k,q}|^2 \geq \rho_i, k \in \mathcal{K}, q \in \mathcal{Q} \right\}, \quad (43)$$

where ρ_i is a chosen threshold for that specific user to be considered present, usually set as a fraction of the magnitude of the amplitudes in $\hat{\boldsymbol{\beta}}$. Then for each detected active user $i \in \hat{\mathcal{I}}$, we take R strongest paths in $\hat{\beta}_{i,k,q}$ with respect to $k \in \mathcal{K}$ and $q \in \mathcal{Q}$ to be the active set $\hat{\mathcal{A}}_i^{(n)}$ for each user $i \in \hat{\mathcal{I}}$.

2) *Partial Knowledge on Active Users \mathcal{I}* : This scenario corresponds to environments where all users are active, however only a certain subset is likely to be detectable by the receiver. GPS receivers are an example. Specifically, there are a total of $I = 24$ quasi-stationary GPS satellites moving around the earth and the active satellites in the field-of-view of a specific GPS receiver are unknown. However, the GPS receiver is informed that at any point in space there should be $|\mathcal{I}| = 4$ strongest signals from satellites, and it attempts to find such signals, along with their delay-Doppler parameters for triangularization. In this case a positive detection corresponds to having at least four components detected and we can interpret this case as fixing $|\mathcal{I}|$ for the receiver detection. In general, we identify the users $i \in \hat{\mathcal{I}}$ as those with the $|\mathcal{I}|$ strongest amplitudes $|\hat{\beta}_{i,k,q}|$ with respect to $i = 1, \dots, I$ in $\hat{\boldsymbol{\beta}}$. Then we take R strongest paths in $|\hat{\beta}_{i,k,q}|$ with respect to k and q to be the active set $\hat{\mathcal{A}}_i^{(n)}$ for each user $i \in \hat{\mathcal{I}}$.

3) *Known Active Users \mathcal{I}* : This is the simplest scenario including multi-antenna and cooperative systems, where the receiver is aware of the active sources, i.e., \mathcal{I} is known. This case is trivial because we do not need to identify the active users. The active set $\hat{\mathcal{A}}_i^{(n)}$ for each user $i \in \mathcal{I}$ chosen as the R strongest components in $|\hat{\beta}_{i,k,q}|$ with respect to \mathcal{K} and \mathcal{Q} .

C. Sequential Detection for Link Acquisition

Substituting $\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}_n}$ and $\hat{\sigma}_{\hat{\mathcal{A}}_n}^2$ back to (39), the generalized likelihood ratio can be computed as

$$\eta_{\text{CSA}}(n) \triangleq \frac{\mathbb{P} \left(\mathcal{H}_{\hat{\mathcal{A}}_n} | \hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}_n}, \hat{\sigma}_{\hat{\mathcal{A}}_n}^2 \right)}{\mathbb{P} \left(\mathcal{H}_{\emptyset} | \hat{\sigma}_{\emptyset}^2 \right)} = \frac{\| \mathbf{c}[n] \|_{\mathbf{R}_{\psi\psi}^{-1}}^{2P}}{\| \mathbf{c}[n] - \mathbf{B}\mathbf{M}\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}_n} \|_{\mathbf{R}_{\psi\psi}^{-1}}^{2P}}.$$

Denote the first window that passes the above test as

$$N_\eta \triangleq \min \left\{ n : \arg \eta_{\text{CSA}}(n) \geq \eta_0 \right\}. \quad (44)$$

As mentioned in (15), the MLR window is located as the window that maximizes the likelihood ratio

$$\ell_* = \arg \max_n \eta_{\text{CSA}}(n), \quad n = N_\eta, \dots, N_\eta + N_0. \quad (45)$$

Accordingly, from the link vector $\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}_{\ell_*}}$ in the ℓ_* th window, we can extract the delay-Doppler pairs

$$\hat{\tau}_{i,r} = q_{i,r} \Delta \tau, \quad \hat{\omega}_{i,r} = k_{i,r} \Delta \omega, \quad (k, q) \in \hat{\mathcal{A}}_i^{(\ell_*)}, \quad i \in \hat{\mathcal{I}}. \quad (46)$$

VI. OPTIMIZATION OF COMPRESSIVE FRONT-END

The link acquisition performance depends on the ability of the SR-LRT to differentiate between different hypotheses \mathcal{H}_{S_n} . In this section, we seek a criterion to optimize the sampling kernels $\{\psi_p(t)\}_{p=1}^P$ in the CMS architecture by designing the matrix \mathbf{B} . The metric we exploit is the weighted average of the Kullback-Leibler (KL) distances between any \mathcal{H}_{S_n} in (38). Since every possible pattern for S_n is independent of n , here we omit the subscript for convenience.

A. Motivation of Maximizing the Average KL Distance

In choosing the KL distance we are motivated by the Chernoff-Stein's lemma [31], whose statement indicates that the probability of confusing \mathcal{H}_S and $\mathcal{H}_{S'}$ decreases exponentially with the pair-wise KL distance between them. By defining $\mathcal{G}(\mathbf{B}) \triangleq \mathbf{M}^H \mathbf{B}^H (\mathbf{B}\mathbf{M}\mathbf{B}^H)^{-1} \mathbf{B}\mathbf{M}$, the *pair-wise KL distance* between any \mathcal{H}_S and $\mathcal{H}_{S'}$ depends on \mathbf{B} as

$$\mathbb{D}(\mathcal{H}_S \| \mathcal{H}_{S'}; \mathbf{B}) = \frac{(\boldsymbol{\beta}_S - \boldsymbol{\beta}_{S'})^H \mathcal{G}(\mathbf{B}) (\boldsymbol{\beta}_S - \boldsymbol{\beta}_{S'})}{\sigma^2}. \quad (47)$$

The minimum pair-wise KL distance determines the *worst-case* performance in terms of missed detection error in the Neyman-Pearson framework. Furthermore, when the received noise is Gaussian, the *pair-wise KL distance* has the same expression as the Chernoff information under the Bayesian framework. Thus the minimum KL distance is also an effective measure for the Bayesian detection error. As a result, a *sufficient* condition to guarantee detection performance is to maximize the minimum of all the pair-wise KL distances

$$\mathbf{B} = \arg \max_{\mathbf{B}} \inf \mathbb{D}(\mathcal{H}_S \| \mathcal{H}_{S'}; \mathbf{B}), \quad \forall S \neq S'. \quad (48)$$

According to (47) and [30], the above criterion is equivalent to maximizing the lower bound on the Restricted Isometry Property (RIP) of the matrix $\mathcal{G}(\mathbf{B})$, which is intractable to analyze and difficult to solve for. Therefore, we maximize the *average*

KL distance as an alternative, because it is a *necessary* condition to maximize the minimum pair-wise KL distance.

To define the average KL distance, we associate each pair of sets $\mathcal{S} \neq \mathcal{S}'$ with the weight $\gamma_{\mathcal{S},\mathcal{S}'}$. Furthermore, we associate the nuisance amplitudes $\boldsymbol{\beta}_{\mathcal{S}}$ with a continuous weighting function $\mathbb{P}(\boldsymbol{\beta}_{\mathcal{S}})$ for any \mathcal{S} . Under these assumptions, the weighted average of all *pair-wise* KL distances is defined as

$$\bar{\mathbb{D}} = \sum_{\mathcal{S},\mathcal{S}'} \gamma_{\mathcal{S},\mathcal{S}'} \int \int \mathbb{P}(\boldsymbol{\beta}_{\mathcal{S}}) \mathbb{P}(\boldsymbol{\beta}_{\mathcal{S}'}) \mathbb{D}(\mathcal{H}_{\mathcal{S}} \parallel \mathcal{H}_{\mathcal{S}'}; \mathbf{B}) d\boldsymbol{\beta}_{\mathcal{S}} d\boldsymbol{\beta}_{\mathcal{S}'}$$

Proposition 1: Given a set of normalized weights $\gamma_{\mathcal{S},\mathcal{S}'}$ for every distinct pair $\mathcal{S} \neq \mathcal{S}'$, and a continuous weighting function $\mathbb{P}(\boldsymbol{\beta}_{\mathcal{S}}) = \prod_{(i,k,q) \in \mathcal{S}} \mathbb{P}(\beta_{i,k,q})$ over the amplitudes with $\int \boldsymbol{\beta}_{\mathcal{S}} \mathbb{P}(\boldsymbol{\beta}_{\mathcal{S}}) d\boldsymbol{\beta}_{\mathcal{S}} = \mathbf{0}$ and $\int |\beta_{i,k,q}|^2 \mathbb{P}(\beta_{i,k,q}) d\beta_{i,k,q} = \sigma_{\beta}^2$, the average KL distance $\bar{\mathbb{D}}$ is equal to

$$\bar{\mathbb{D}} = \frac{\sigma_{\beta}^2}{\sigma^2} \text{Tr} \left[\mathbf{M}^H \mathbf{B}^H (\mathbf{B} \mathbf{M} \mathbf{B}^H)^{-1} \mathbf{B} \mathbf{M} \right]. \quad (49)$$

Proof: See Appendix C. \blacksquare

The way we choose the weights $\gamma_{\mathcal{S},\mathcal{S}'}$ and weighting functions $\mathbb{P}(\boldsymbol{\beta}_{\mathcal{S}})$ is equivalent to assuming a uniform distribution on the users, delays and Dopplers together with i.i.d. Gaussian priors on the amplitudes in $\boldsymbol{\beta}_{\mathcal{S}}$ in a Bayesian framework. According to [32], the pair-wise KL distance has the same expression as the Chernoff information, which determines the Bayesian detection error exponent. However, if any of the pair-wise KL distance is zero, then the two hypotheses $\mathcal{H}_{\mathcal{S}}$ and $\mathcal{H}_{\mathcal{S}'}$ are indistinguishable for that particular pair of \mathcal{S} and \mathcal{S}' . A non-zero pair-wise KL distance between arbitrary pairs of \mathcal{S} and \mathcal{S}' with $|\mathcal{S}|, |\mathcal{S}'| \leq s$ requires $\text{spark}[\mathcal{G}(\mathbf{B})] \geq 2s$, where $\text{spark}[\cdot]$ is the kruskal rank of a matrix.

We note that the average KL distance metric defined here does not automatically ensure that $\mathbb{D}(\mathcal{H}_{\mathcal{S}} \parallel \mathcal{H}_{\mathcal{S}'}; \mathbf{B}) > 0$ for all $\mathcal{S} \neq \mathcal{S}'$. In fact, it is possible that specific choices of the number of samplers P and the dictionary $\{\phi_{i,k,q}(t)\}_{i=1, \dots, I}^{k \in \mathcal{K}, q \in \mathcal{Q}}$ (i.e., the Gram matrix \mathbf{M}) lead to indistinguishable sparsity patterns [30] such that $\text{spark}[\mathcal{G}(\mathbf{B})] \leq |\mathcal{S}| + |\mathcal{S}'|$. In other words, the design of \mathbf{B} cannot cure intrinsic problems caused by the choice of P or the Gram matrix \mathbf{M} , which are given in the optimization. Intrinsic problems caused by the Gram matrix \mathbf{M} is typically handled by optimizing the transmit sequences $\phi_i(t)$ irrespective of the receiver such that \mathbf{M} becomes diagonally dominated. This implies a well localized ambiguity function for each of the $\phi_i(t)$ and low cross-correlation between $\phi_i(t)$'s with different delays and Doppler. Gold sequences used in GPS and M-sequences used in spread spectrum communications, for example, are known to have good properties in this regard. The design of these sequences is well investigated [33] and we do not aim to cover here.

B. Optimization of Compressive Samplers

Given P and \mathbf{M} , we propose an optimal \mathbf{B} that maximizes the average KL distance $\bar{\mathbb{D}}$ if there is a unique solution to the optimization; when there are multiple solutions that yield identical average KL distance $\bar{\mathbb{D}}$, we choose the matrix \mathbf{B} that gives the least occurrence of events $\mathbb{D}(\mathcal{H}_{\mathcal{S}} \parallel \mathcal{H}_{\mathcal{S}'}; \mathbf{B}) = 0$. We use the results in the following lemma for our optimization.

Lemma 1: (Ratio Trace Maximization [34]) Given a pair of $L \times L$ positive semi-definite matrices (\mathbf{S}, \mathbf{G}) and an $L \times P$ full column rank matrix \mathbf{W} , the ratio trace problem is

$$\mathbf{W}^{\text{opt}} = \arg \max_{\mathbf{W}} \text{Tr} [(\mathbf{W}^H \mathbf{S} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{G} \mathbf{W}]. \quad (50)$$

The optimal $\mathbf{W}^{\text{opt}} = [\mathbf{w}_1^{\text{opt}}, \dots, \mathbf{w}_P^{\text{opt}}]$ is given by the generalized eigenvectors $\{\mathbf{w}_p^{\text{opt}}\}_{p=1}^P$ corresponding to P largest generalized eigenvalues of the pair (\mathbf{S}, \mathbf{G}) with $P \leq \text{rank}[\mathbf{S}]$.

The optimal \mathbf{B} is identified in the following theorem⁴:

Theorem 3: Let the eigendecomposition be $\mathbf{M} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^H$ where $\boldsymbol{\Sigma} = \text{diag}[\sigma_1, \dots, \sigma_{|\mathcal{K}| |\mathcal{Q}|}]$ is the eigenvalue matrix in descending order and \mathbf{U} is the eigenvector matrix of \mathbf{M} . Denote the $P \leq \text{rank}[\mathbf{M}]$ principal eigenvectors as

$$\mathcal{U} = \{ \mathbf{U}_P = [\mathbf{u}_1, \dots, \mathbf{u}_P] : \quad (51)$$

$$\mathbf{M} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^H, \mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{|\mathcal{K}| |\mathcal{Q}|}] \}. \quad (52)$$

Let $\bar{\mathbf{E}}_P$ be an arbitrary non-singular $P \times P$ matrix. When \mathbf{U}_P is unique, the matrix $\mathbf{B}_{\star} = \bar{\mathbf{E}}_P \mathbf{U}_P^H$ is chosen to maximize the average KL distance $\bar{\mathbb{D}}$. When \mathbf{U}_P is not unique, we choose $\hat{\mathbf{U}}_P = \max_{\mathbf{U}_P \in \mathcal{U}} \text{spark}(\mathbf{U}_P^H)$ to maximize the average KL distance and minimize the occurrence of events $\mathbb{D}(\mathcal{H}_{\mathcal{S}} \parallel \mathcal{H}_{\mathcal{S}'}; \mathbf{B}) = 0$ for $\mathcal{S} \neq \mathcal{S}'$.

Proof: See Appendix D. \blacksquare

Note that, as long as the preamble sequences do not change, the optimal matrix \mathbf{B}_{\star} and the corresponding sampling kernels $\psi_p(t)$ are pre-computed only once and their design does not contribute to the running cost of the receiver operations. If the projections on the sampling kernels are implemented in the digital domain instead of being analog filters, then the samples of $\psi_p(t)$ are placed in the static memory that contains the receiver signal processing algorithms.

If the principal eigenvectors are unique, the choice of \mathbf{B}_{\star} spreads out the pairwise KL distances, but $\mathbb{D}(\mathcal{H}_{\mathcal{S}} \parallel \mathcal{H}_{\mathcal{S}'}; \mathbf{B}) = 0$ is possible for some choice of \mathcal{S} and \mathcal{S}' . If $\text{spark}(\mathbf{M}) \geq 2|\mathcal{I}|R$ and $P = \text{rank}[\mathbf{M}]$, then it is ensured that $\bar{\mathbb{D}}$ is maximized and $\mathbb{D}(\mathcal{H}_{\mathcal{S}} \parallel \mathcal{H}_{\mathcal{S}'}; \mathbf{B}) > 0$ for $\mathcal{S} \neq \mathcal{S}'$ with $|\mathcal{S}|, |\mathcal{S}'| \leq |\mathcal{I}|R$. On the other hand, an extreme example where the eigenvectors are not unique is when $\{\phi_{i,k,q}(t)\}_{i=1, \dots, I}^{k \in \mathcal{K}, q \in \mathcal{Q}}$ form an orthogonal basis such that $\mathbf{M} = \mathbf{I}$. In this case, Theorem 3 is analogous to the fundamental criterion in compressed sensing that aims to find a matrix with $\text{spark}(\mathbf{U}_P^H) \geq 2|\mathcal{I}|R$ that guarantees the recovery of any $|\mathcal{I}|R$ -sparse vectors.

Remark: The number $|\mathcal{U}|$ of possible eigen-decompositions of the given matrix \mathbf{M} may be quite large. For the extreme case when $\mathbf{M} = \mathbf{I}$, an arbitrary unitary matrix will be a possible choice. Fortunately, it is well known in compressed sensing that partial unitary matrices (such as the partial DFT matrix) have good compressive sensing properties (mutual coherence), thus this would not entail much loss if the matrix does not have exactly the maximum spark. On the other hand, as long as the number $|\mathcal{U}|$ is small, a finite search is also possible. More importantly, this task only needs to be done once and off-line.

⁴Note that the computation of the weights is done offline, and does not add complexity to the online processing.

VII. NUMERICAL RESULTS

In this section, we compare the CSA acquisition scheme using the CMS architecture against alternative methods. The first alternative is the DSA approach discussed in Section III-C, which processes the uncompressed Nyquist-rate samples $c_{DS}[n]$ using sparse recovery methods. The comparison with the DSA scheme can also be viewed as a comparison with [2], [28], which propose the SA method for CDMA users using direct sampling.⁵ Another alternative is the MF receiver discussed in Section III-B, which also processes uncompressed samples but does not exploit the underlying signal sparsity.

To benchmark CSA against the DSA and MF approaches, we simulate the link acquisition of a single receiver plugged in a network populated by $I = 10$ users, out of which $|\mathcal{I}| = 4$ are randomly chosen to be actively transmitting. The user signature codes, $a_i[m]$'s in (1), belong to a set of M-sequences [35], which are quasi-orthogonal BPSK sequences of length $M = 255$ with unit power ($|a_i[m]|^2 = 1$). Due to the user dislocation, mobility and possible scattering, each path of this asynchronous multi-user channel is characterized by the triplet $\{h_{i,r}, t_{i,r}, \omega_{i,r}\}_{i \in \mathcal{I}, r \in \{1, \dots, R\}}$ with $R = 2$, where $\{h_{i,r}\}$ are Rayleigh distributed, $h_{i,r} \sim \mathcal{CN}(0, 1/|\mathcal{I}|R)$, and uncorrelated normalized fading coefficients $\mathbb{E}\{h_{i,r}h_{i',r'}^*\} = \delta[i - i']\delta[r - r'] \cdot 1/|\mathcal{I}|R$. The random delays $\{t_{i,r}\}$ are the sum of: (i) a time of arrival $t_0 = \min_{i,r} t_{i,r}$ that is uniformly distributed over an interval that spans the duration of the preamble $t_0 \in \mathcal{U}(0, MT)$, and of (ii) multipath delays that are uniformly distributed within an interval $(t_0, t_0 + \tilde{\tau}_{\max})$ where $\tilde{\tau}_{\max}$ is the multipath channel delay spread. Consequently, all the arrival times are within a window of duration $t_0 + MT + \tilde{\tau}_{\max}$. The random frequency offsets $\{\omega_{i,r}\}$ are uniformly distributed, $\omega_{i,r} \in \mathcal{U}(-\omega_{\max}, \omega_{\max})$, over a range delimited at each side by the maximum Doppler spread ω_{\max} . As we simulate underspread channel conditions, we choose ω_{\max} such that $\omega_{\max}\tilde{\tau}_{\max} \ll 2\pi$. Thus, for a multipath delay spread of $\tilde{\tau}_{\max} = 4T$, the choice of $\omega_{\max} = 2.5 \cdot 10^{-3} \times 2\pi/T$ is comparable to a 25 kHz offset for 1 MHz signals.

We compare CSA with DSA and MF at the same resolution: $\Delta\omega = \omega_{\max}/5 = 0.5 \times 2\pi/T$ and $\Delta\tau = T/2$ and thus $\mathcal{K} = \{-5, \dots, 5\}$. Given a multipath delay spread of $\tilde{\tau}_{\max} = 4T$ and a shift of size $D = 10T$, the delay space \mathcal{Q} accounts for a composite delay spread of $\tau_{\max} \geq D + \tilde{\tau}_{\max} = 14T$ and therefore $\mathcal{Q} = \{0, \dots, 27\}$. This parameter discretization leads to a multi-user time-frequency grid of $I|\mathcal{K}||\mathcal{Q}| = 3080$ elements for the CSA and DSA schemes.

We test the CSA using three different numbers of sampling channels $P = \{60, 80, 100\}$. The DSA uses Nyquist-rate samples per shift, which corresponds to using the whole spreading code duration of 255 samples. The MF scheme uses a $I|\mathcal{K}| = 110$ -channel filterbank and performs $I|\mathcal{K}||\mathcal{Q}| = 3080$ projections per shift with $|\mathcal{Q}|$ samples on each branch. In the simulations, the CSA compressive samples are generated from the same Nyquist-rate samples used for the other receivers, by projecting them onto the digitized version of the sampling kernels

⁵There are some subtle differences in the model. For example, the online sequential detection and unknown frequency offsets were not considered in [2], [28]. Hence, strictly speaking DSA is a generalized version of [2], [28], where the same basic idea is expanded to handle a larger set of hypotheses.

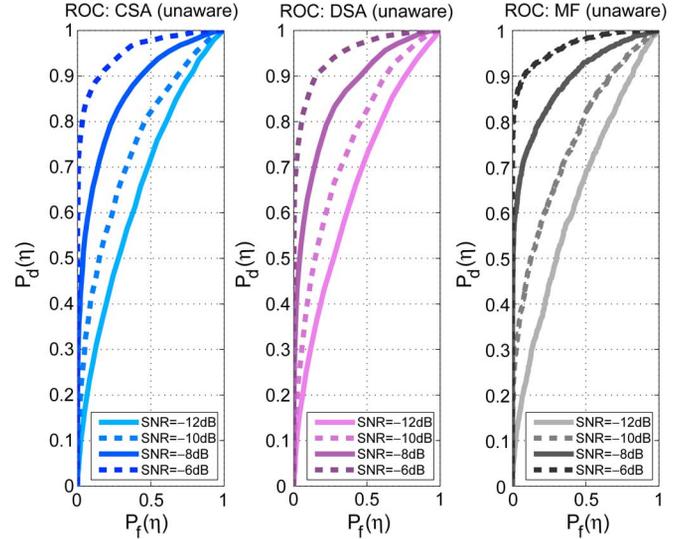


Fig. 2. ROC curves for the order unaware receiver using CSA with $P = 80$ channels (left), DSA using M uncompressed Nyquist observations (middle), and MF (right); tested at SNR = $\{-12, -10, -8, -6\}$ dB.

$\psi_p(mT_s)$, $p = 1, \dots, P$, $m = 0, \dots, M - 1$. The CSA simulation recovers the link parameters by solving (41) with the OMP algorithm [29], which is a popular choice to approximate the solution of a sparse problem [36]. To motivate our selection of OMP, we refer to Section VIII for an empirical evaluation of the OMP against two well-known ℓ_1 minimizers, SpaRSA [37] and ℓ_1 -Homotopy [38], [39].

A. Signal Detection Performance

We first test the case of completely unknown active user sets, as discussed in Section V-B-1. In Fig. 2, all receivers are unaware of the random set \mathcal{I} of active users. Specifically, receivers consider as active components those that are found to have a signal strength that is at least 30% of the strongest components they estimate, i.e. in (43) $\rho_i = \max_{k,q} |\beta_{i,k,q}|^2/3$. If no possible component meets this requirement, the channel is declared idle. To first compare the sensitivity of the different receivers to active components, we define a signal hypothesis \mathcal{H}_1 corresponding to all the non-idle channel hypotheses, i.e. $\hat{\mathcal{A}}_{\ell^*} \neq \emptyset$. Then, the detection sensitivity is measured in terms of the receiver operating characteristic (ROC) curve, tracing the probability of detection $\mathbb{P}_d(\eta_0) = \mathbb{P}(\eta_{\text{CSA}}(\ell^*) \geq \eta_0 | \mathcal{H}_1)$, against the probability of false alarm $\mathbb{P}_f(\eta_0) = \mathbb{P}(\eta_{\text{CSA}}(\ell^*) \geq \eta_0 | \mathcal{H}_\emptyset)$ when the channel is actually idle. Note that a positive detection may correspond to an incorrect identification of the specific users that are active. Thus, Section VII-B shows the rate of correct detection of active components for the same simulation scenario.

As can be observed, although the CSA receiver exploits less than $P/M = 80/255 \approx 1/3$ of the Nyquist-rate samples, the results from Fig. 2 show a modest degradation of the ROC compared to the MF receiver (less than 0.1 measured at $\mathbb{P}_f(\eta_0) = 0.1$ and SNR = -8 dB). As expected, since the DSA can leverage the additional observations to enhance its sensitivity, a growing gap $\mathbb{P}_d(\eta_{\text{DSA}}) - \mathbb{P}_d(\eta_{\text{CSA}})$ is observable as the SNR increases (measured at $\mathbb{P}_f(\eta_0) = 0.1$) between the SNR = -6 dB and the SNR = -12 dB curves.

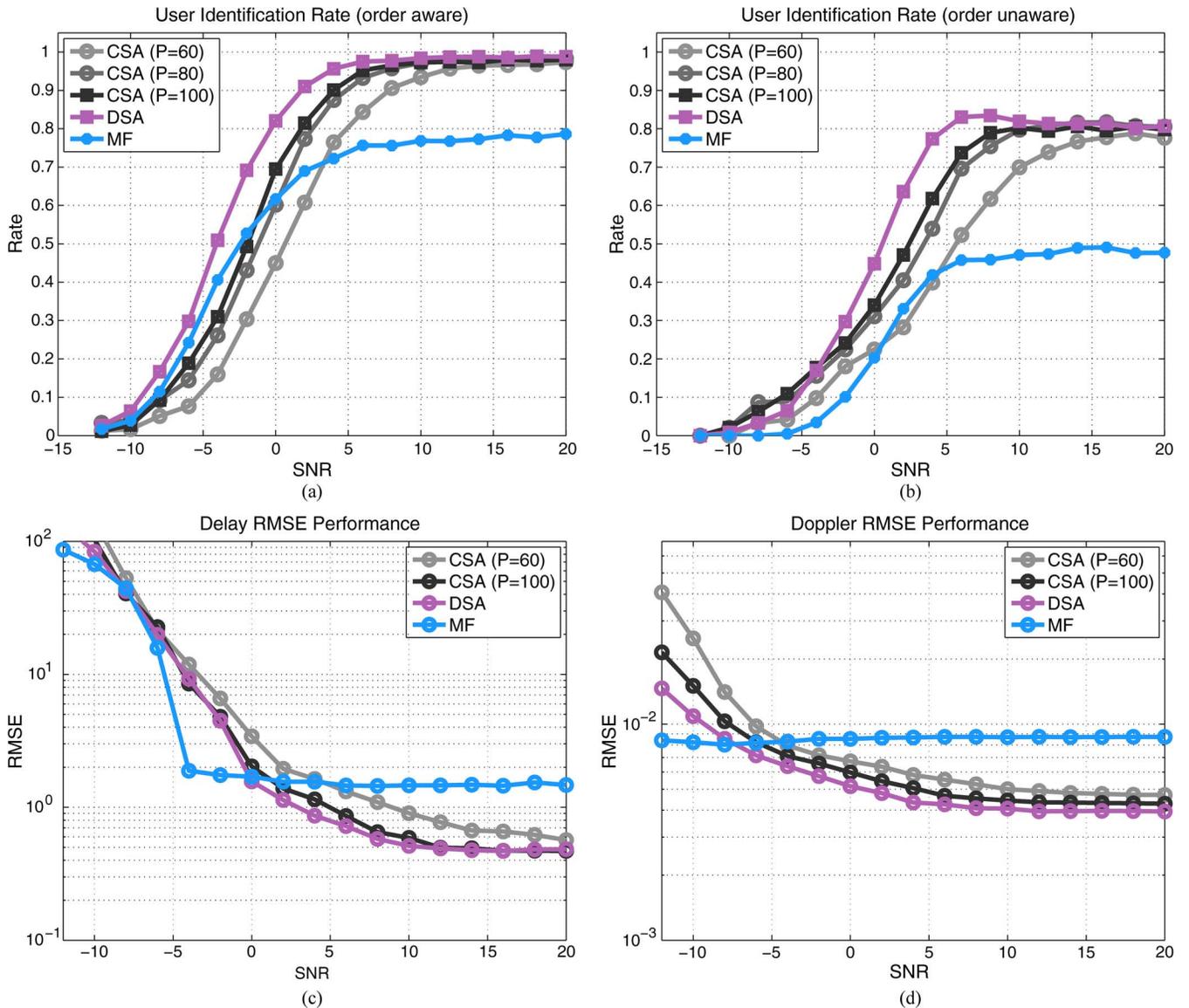


Fig. 3. (a) User identification rate $\mathbb{P}(\hat{\mathcal{I}} = \mathcal{I})$ in the order aware scenario for the MF (blue) receiver, the DSA receiver (red), the CSA receiver (grey shades) with $P = \{60, 80, 100\}$. (b) User identification rate $\mathbb{P}(\hat{\mathcal{I}} = \mathcal{I})$ in the order unaware scenario for the MF (blue) receiver, the DSA receiver (red), or the CSA receiver (grey shades) with $P = \{60, 80, 100\}$. (c) RMSE(τ) against SNR for the CSA scheme (grey shades) with $P = \{60, 100\}$, by the DSA (red) or by the MF (blue). (d) RMSE(ω) against SNR for the CSA scheme (grey shades) with $P = \{60, 100\}$, by the DSA (red) or by the MF (blue).

B. User Identification and Parameter Estimation Performance

In the simulations shown in Figs. 3(a) and 3(b), we measure the detection performance of $\hat{\mathcal{I}}$ by the rate of successful identification $\mathbb{P}(\hat{\mathcal{I}} = \mathcal{I})$. In these simulations, the threshold η_0 is set to a level such that $\mathbb{P}_f(\eta_0) \leq 0.1$, and thus we trace the curve at the point $\mathbb{P}_f(\eta_0) = 0.1$ on the figures. The two sets of figures correspond to, respectively, the case where a receiver has partial knowledge of the active user components (as in the GPS receiver discussed in Section V-B-2) and the exact same case examined previously in Fig. 2, where the receivers are unaware of the user \mathcal{I} . In the second case, for successful detection, not only the elements of the sets have to be consistent $\hat{\mathcal{I}} \subseteq \mathcal{I}$ but also their cardinality needs to be identical $|\hat{\mathcal{I}}| = |\mathcal{I}|$. Instead for the first case, if the receiver has partial knowledge of the active components, then what matters is that the components are correctly identified, but their number is known ahead of time.

With a relatively short training sequence, we can see in Fig. 3(a) that the CSA in the first case identifies the active user set with large probability (0.96 at SNR = 20 dB). The MF has worse performance due to the multi-user interference and to the presence of unresolvable paths. In fact, the MF receiver is unable to isolate the multi-path arrivals that fall within the same symbol period and, due to the presence of different Dopplers, its side-lobes may contribute negatively to the correlation, masking other active components. In contrast, the OMP algorithm in the CSA scheme cancels the contributions from paths detected in previous iterations, before updating the projections to search for other components (the OMP processing steps are summarized in Section VIII). It is evident, however, that a low SNR, the CSA scheme suffers from a loss due to the compression (-1 dB at the rate 0.6 with $P = 100$). This is understood by observing the performance of the DSA receiver as well.

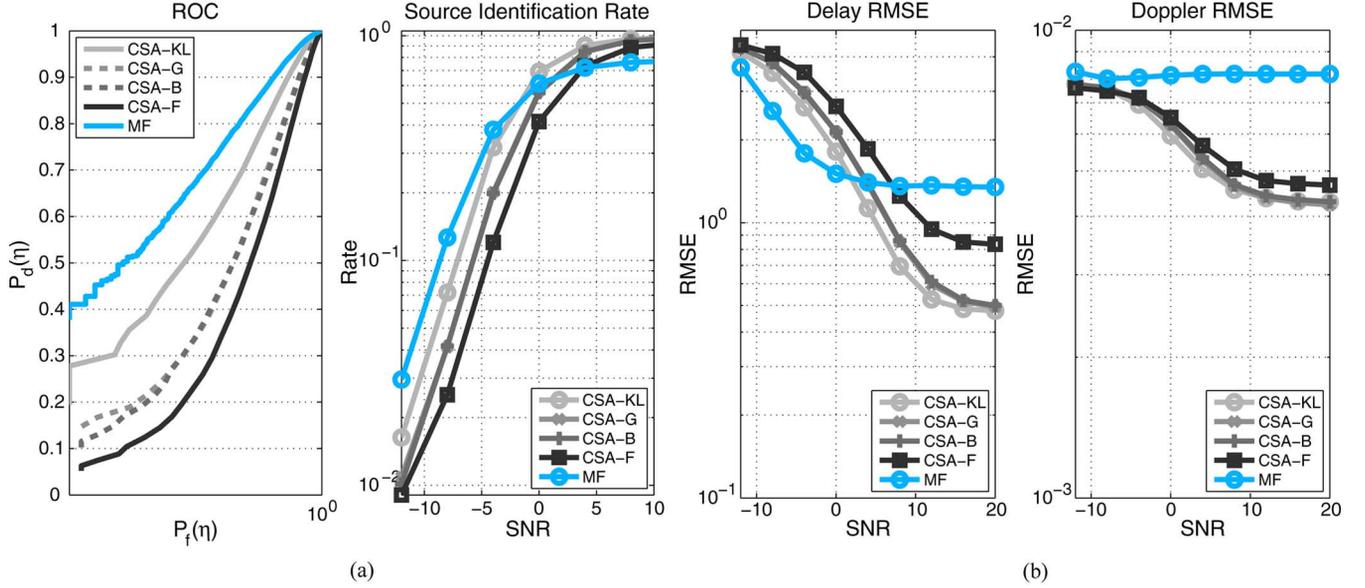


Fig. 4. (a) ROC curve at SNR = -8 dB and user identification rate $\mathbb{P}(\hat{\mathcal{I}} = \mathcal{I})$ of the CSA scheme, against the MF receiver and different choices of \mathbf{B} 's. (b) Delay and Doppler estimation RMSE of the CSA scheme, against different random designs of \mathbf{B} and the MF.

By processing uncompressed samples with sparse recovery, the DSA curve combines the best of both worlds and, thus, its performance bounds the user identification rate for both the MF and the CSA receivers in both examples. As shown in Fig. 3(b), the performance degrades when the receivers do not have side information on $|\mathcal{I}|$ (a difference of -0.13 for the CSA receiver against -0.3 of the MF at SNR = 20 dB). This is due to the cardinality mismatch, $\{|\hat{\mathcal{I}}| \neq |\mathcal{I}|\}$, that occurs while estimating the order.

The accuracy of the recovered set $\hat{\mathcal{A}}_{\ell^*}$ is evaluated by the root mean square error (RMSE) of $\tau_{i,r}$ and $\omega_{i,r}$ that are associated with the correctly identified users $\mathcal{I} = \hat{\mathcal{I}}$. Thus,

$$\text{RMSE}(\boldsymbol{\tau}) \triangleq \sqrt{\frac{1}{|\hat{\mathcal{I}} \cap \mathcal{I}|} \frac{1}{R} \sum_{r=1}^R \sum_{i \in \{\hat{\mathcal{I}} \cap \mathcal{I}\}} (\tau_{i,r} - \hat{q}_{i,r} \Delta \tau)^2}$$

$$\text{RMSE}(\boldsymbol{\omega}) \triangleq \sqrt{\frac{1}{|\hat{\mathcal{I}} \cap \mathcal{I}|} \frac{1}{R} \sum_{r=1}^R \sum_{i \in \{\hat{\mathcal{I}} \cap \mathcal{I}\}} (\omega_{i,r} - \hat{k}_{i,r} \Delta \omega)^2}$$

are the average RMSE of the delay parameters the Doppler frequencies respectively.

To verify the accuracy of the parameter estimates of $\tau_{i,r}$ and $\omega_{i,r}$, we trace the RMSE's of the order-aware case. Once again we observe, from Fig. 3(c) and Fig. 3(d), that the performance of $\text{RMSE}(\boldsymbol{\tau})$ and $\text{RMSE}(\boldsymbol{\omega})$ is enhanced by the detector that better leverages the presence of the multi-path. In fact, at SNR = 20 dB, the accuracy of the CMS, with $P = 100$, and the DSA approach the grid resolution, i.e. $\text{RMSE}(\boldsymbol{\tau}) \approx \Delta \tau$ and $\text{RMSE}(\boldsymbol{\omega}) \approx \Delta \omega$. Oppositely, the contribution of the unresolvable paths to the correlation, in either frequency or time, adversely affects the parameter selection. Not canceling the previously selected components contributes to a large error after the selection of the dominant paths as the same arrivals are likely to be selected more than once by the presence of correlated components. These errors impact the highest

resolution since at SNR = 20 dB: $\text{RMSE}(\boldsymbol{\tau}) > 2\Delta \tau$ and $\text{RMSE}(\boldsymbol{\omega}) > 2\Delta \omega$. Instead, at low SNR, the performance is bounded by the maximum error given by the search space which is a function of ω_{\max} and τ_{\max} , respectively, due to the early detection resulting from heavy noise.

C. Optimality of Compressive Samplers

In this subsection, we briefly compare the performances of the CSA scheme using a $P = 100$ -channel CMS architecture with the optimal samplers versus other random projection schemes in compressed sensing. The ROC curve and the user identification rate $\mathbb{P}(\hat{\mathcal{I}} = \mathcal{I})$ in Fig. 4(a) show that the optimal sampling kernel, denoted by CSA-KL, exhibits better performance than random designs of \mathbf{B} using matrices whose entries are Gaussian (CSA-G), Bernoulli (CSA-B), or randomly selected rows of a DFT matrix (CSA-F). It can also be observed from Fig. 4(b) that the RMSE of the delay and Doppler estimates are improved.

VIII. COST ANALYSIS

In this section, we analyze the implementation costs of the MF and the proposed CSA scheme in terms of storage requirement and computational complexity. The analysis is performed in two regimes respectively: the *analog implementation*, that corresponds to what the paper describes mathematically in detail, and a *digital implementation*, which would be necessary if the compressive samplers $\psi_p(t)$ cannot be implemented as analog filters. The CMS architecture in this case emulates our simulations, where Nyquist samples of $x(t)$ are projected onto digital filters matched to the samples of $\psi_p(t)$.

The metric to evaluate *storage requirements* is chosen as the A/D hardware cost, measured as the size of the filterbank which is also the buffer size of the A/D samples. The *computational complexity* is evaluated by the *total number of additions and multiplications*, and by the average run time (a 64-bit i7 920 CPU at 2.67 GHz). In the following, we first settle on the sparse

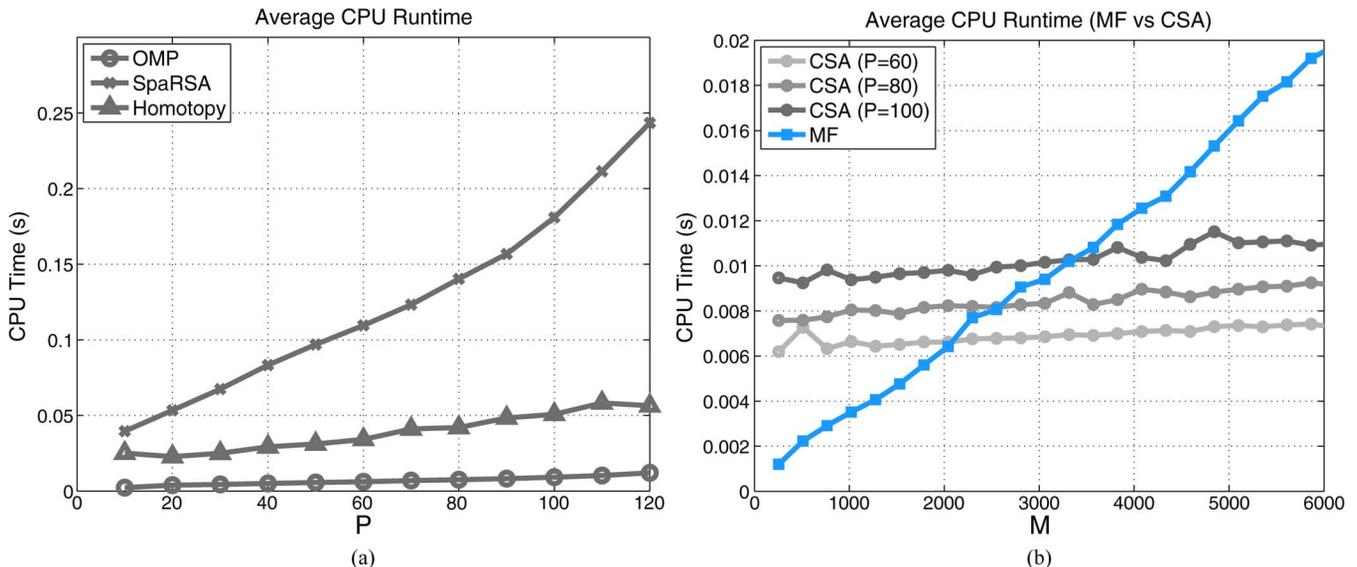


Fig. 5. (a) Average CPU run time for the CSA receiver using different sparse recovery solvers (OMP, SpaRSA, ℓ_1 -Homotopy), as a function of P . (b) Average CPU runtime of the CSA receiver using $P = \{60, 80, 100\}$ compressed observations against the MF receiver. The average runtime is measured against the preamble length M .

recovery solver for the SR-LRT in the CSA scheme, and then continue our comparison using the chosen solver.

A. Sparse Recovery Solver: The OMP Algorithm

The CSA receiver spends its greatest effort in solving the optimization (41). Fast ℓ_1 minimizers like SpaRSA [37] or ℓ_1 -Homotopy [38], [39] are often the methods of choice. The former greatly reduces the complexity by approximating the Hessian of the gradient descent by a diagonal matrix, whereas the latter inverts a system of equations whose number of unknowns, at each iteration, remains restricted to the non-zero elements of the sparse vector estimate. Greedy algorithms like the OMP [29], are efficient approximations to the solution of sparse problems as well [36]. The OMP algorithm iteratively detects the strongest element in the sparse vector and removes its contribution in the next iteration; thus, the number of iterations required by OMP is bounded by the maximum components that can possibly exist, which in our case is $|\mathcal{I}|R$.

The average CPU run time spent in solving (41) with different solvers is illustrated in Fig. 5(a) traced against P , where the OMP algorithm shows significantly less computation time. Thus, in Fig. 5(b) we further compare the average CPU run time of the CSA receiver using OMP against the MF receiver, the implementation details of which will be discussed in the following subsection. OMP has smaller complexity primarily because it stops as soon as all the strong entries have been detected. In contrast, ℓ_1 -Homotopy and SpaRSA do not limit the search to a single set, but rather explore the feasible set by selecting and de-selecting elements of the support vector (ℓ_1 -Homotopy), or by shrinking it through a gradient descent (SpaRSA), until a desired convergence criterion has been met.

TABLE I
COMPLEXITY OF ALGORITHM 1 FOR THE CSA RECEIVERS

CSA Receiver	Storage (analog)	Complexity (analog)	Storage (digital)	Complexity (digital)
(CSA.1)	$\mathcal{O}(P)$	0	$\mathcal{O}(MP)$	$\mathcal{O}(MP)$
(CSA.2)	0	$\mathcal{O}(I \mathcal{K} \mathcal{Q}) + \mathcal{O}(P^3)$	0	$\mathcal{O}(I \mathcal{K} \mathcal{Q}) + \mathcal{O}(P^3)$
(CSA.3)	$\mathcal{O}(P)$	$\mathcal{O}(P)$	$\mathcal{O}(P)$	$\mathcal{O}(P)$
Total	$\mathcal{O}(P)$	$\mathcal{O}(I \mathcal{K} \mathcal{Q}) + \mathcal{O}(P^3)$	$\mathcal{O}(MP)$	$\mathcal{O}(I \mathcal{K} \mathcal{Q}) + \mathcal{O}(P^3)$

TABLE II
COMPLEXITY OF ALGORITHM 2 FOR THE MF RECEIVERS

MF Receiver	Storage (analog)	Complexity (analog)	Storage (digital)	Complexity (digital)
(MF.1)	$\mathcal{O}(I \mathcal{K} \mathcal{Q})$	0	$\mathcal{O}(MI \mathcal{K} \mathcal{Q})$	$\mathcal{O}(MI \mathcal{K} \mathcal{Q})$
(MF.2)	0	$\mathcal{O}(I \mathcal{K} \mathcal{Q})$	0	$\mathcal{O}(I \mathcal{K} \mathcal{Q})$
Total	$\mathcal{O}(I \mathcal{K} \mathcal{Q})$	$\mathcal{O}(I \mathcal{K} \mathcal{Q})$	$\mathcal{O}(MI \mathcal{K} \mathcal{Q})$	$\mathcal{O}(MI \mathcal{K} \mathcal{Q})$

B. Complexities of the CSA Scheme v.s. MF Scheme

Using the OMP algorithm for sparse recovery, we summarize the steps of the CSA and MF schemes in Algorithms 1 and 2 respectively. The CSA scheme has 4 steps **CSA.1**~**CSA.4** and the MF scheme has 3 steps **MF.1**~**MF.3**. Based on the algorithm descriptions, we provided the order of storage cost and computational complexities in Tables I and II. Storage accounts for a *data path storage* component, dynamically updated with the streaming data that correspond to new observations to be processed, and for a *static component*, that stores filters or sampling kernels parameters needed to perform signal processing on the data.

It is seen in Tables I and II that both the CSA receiver and MF receiver have computational complexities that scale linearly with the dimension of the search space $I|\mathcal{K}||\mathcal{Q}|$. However, the *data path storage* of the CSA receivers is greatly reduced. Another storage gain is found in the case of digital implementation, because there are fewer projections to be made on the streaming samples. Thus, unless the sampling kernels are synthesized on the fly, a smaller amount of static memory is required to store the samples of $\psi_p(t)$.

Algorithm 1: CSA Scheme

- (CSA.1) obtain compressive samples $\mathbf{c}[n]$ at the n th shift;
(CSA.2) initialize $\boldsymbol{\beta}^0 = \mathbf{0}$, $\mathcal{A}_n^0 = \emptyset$, $\bar{\mathbf{S}}_0 = \mathbf{BM}$, $\mathbf{S}_0 = \emptyset$, $j = 1$ and run the OMP algorithm;
(OMP.1) remove interference $\boldsymbol{\delta}^j = \mathbf{c}[n] - \mathbf{S}_{j-1}[\boldsymbol{\beta}^{j-1}]_{\mathcal{A}_n^{j-1}}$;
(OMP.2) projection $\boldsymbol{\xi}^j = \bar{\mathbf{S}}_{j-1}^T \boldsymbol{\delta}^j$, $\boldsymbol{\xi}^j = [\cdots, \xi_{i,k,q}^j, \cdots]^T$;
(OMP.3) detection $\mathcal{A}_n^j = \mathcal{A}_n^{j-1} \cup \{(i, k, q)\}$ with

$$(i, k, q) = \arg \max_{i,k,q} \left| \xi_{i,k,q}^j \right|^2;$$

- (OMP.4) update $\mathbf{S}_j = [\mathbf{BM}]_{(\mathcal{A}_n^j)}$ and $\bar{\mathbf{S}}_j = [\mathbf{BM}]_{\overline{(\mathcal{A}_n^j)}}$;
(OMP.5) update the link vector

$$[\boldsymbol{\beta}^j]_{\mathcal{A}_n^j} = (\mathbf{S}_j^T \mathbf{S}_j)^{-1} \mathbf{S}_j^T \mathbf{c}[n] \quad (53)$$

$$[\boldsymbol{\beta}^j]_{\overline{\mathcal{A}_n^j}} = \mathbf{0}; \quad (54)$$

- (OMP.6) stop if either $j = |\mathcal{I}|R$ or $\|\mathbf{c}[n] - \mathbf{BM}\boldsymbol{\beta}^j\| < \epsilon$, and set $j = j + 1$.

- (CSA.3) Evaluate the likelihood ratio $\eta_{\text{CSA}}(n)$ and check if it exceeds η_0 .

- (CSA.4) If yes, then extract components accordingly (order-aware, order-unaware).

Algorithm 2: MF Scheme

- (MF.1) obtain the sample array $\mathbf{C}_{\text{MF}}[n]$ in (17) from the MF filterbank;

- (MF.2) identify the maximum output and check if it exceeds ρ_i for all $i = 1, \dots, I$;

- (MF.3) If yes, then extract the delay-Doppler set for each active user as in Section III.

When implemented in the digital domain, the CSA receiver also leads to a great reduction in computational complexity, with an approximate ratio with the MF receiver complexity of

$$\frac{MP + I|\mathcal{K}||\mathcal{Q}| + P^3}{MI|\mathcal{K}||\mathcal{Q}|} \approx \frac{P}{I|\mathcal{K}||\mathcal{Q}|} + \frac{P^3}{MI|\mathcal{K}||\mathcal{Q}|}. \quad (55)$$

When the preamble sequence is long and the search space is large $MI|\mathcal{K}||\mathcal{Q}| \gg P^3$, this ratio becomes less than 1 and the CSA architecture leads to computational savings while, as seen in simulations, maintaining comparable performance. This explains why the CSA receiver in the simulation (see Fig. 5(b)) considerably outperforms the MF receiver in terms of average run time for large $M \geq 3000$ with $P = 60, 80, 100$. On the other hand, when M is small (e.g. $M = 255$ in Section VII), the MF receiver has less computation time for $M < 3000$ against

$P = 60, 80, 100$ as in Fig. 5(b), but such a short preamble does not provide sufficient processing gain for reliable link acquisition, as can be clearly seen from the numerical results (e.g., see Fig. 3(a)). Thus, when M is small, the gain of the CSA receiver also lies in the superior acquisition performance demonstrated by the numerical results, except for the low SNR region where the CSA is not sufficiently sensitive.

IX. CONCLUSIONS

In this paper, we proposed the CSA link acquisition scheme using a unified CMS architecture and a SR-LRT module for multiuser signals. This scheme uses a sequential SR-LRT that jointly detects signal presence and recovers the active users with their link parameters. We optimized the CMS architecture to maximize the average Kullback-Leibler distance among the hypotheses tested in the SR-LRT and show that, with the optimal compressive samplers we propose, the receiver detection outperforms those with conventional compressed sensing alternatives. Furthermore, through the numerical comparison of the proposed architecture with the DSA scheme and the MF approach, we have shown that the CSA receiver can scale down its processing storage and complexity with greater flexibility, while maintaining satisfactory performance.

APPENDIX A
PROOF OF THEOREM 1

Substituting (23) into (21), we have

$$c_p[n] = \sum_{i=1}^I \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}} \alpha_{i,k,q} e^{ik\Delta\omega\ell D} \langle \phi_{i,k,q}(t-\ell D), \psi_p(t-nD) \rangle + \langle v(t), \psi_p(t-nD) \rangle. \quad (56)$$

Define the $P \times I|\mathcal{K}||\mathcal{Q}|$ matrix

$$[\mathbf{M}_{\psi\phi}[n-\ell]]_{p,(i,k,q)} = R_{\psi_p\phi_{i,k,q}}[(n-\ell)D] \quad (57)$$

$$\triangleq \langle \phi_{i,k,q}(t-\ell D), \psi_p(t-nD) \rangle \quad (58)$$

and denote $v_p[n] \triangleq \langle v(t)\psi_p(t-nD) \rangle$ as the sample of filtered noise, whose covariance can be obtained as

$$\mathbb{E}\{v_p[n]v_p^*[n]\} = \sigma^2 \langle \psi_p(t), \psi_p(t) \rangle$$

using $\mathbb{E}\{v(t)v^*(s)\} = \sigma^2\delta(t-s)$. Therefore, the noise vector $\boldsymbol{\nu}[n] \triangleq [\nu_1[n], \dots, \nu_P[n]]^T$ has a covariance matrix obtained as $\mathbf{R}_{\nu\nu} = \mathbb{E}\{\boldsymbol{\nu}[n]\boldsymbol{\nu}^H[n]\} = \sigma^2\mathbf{R}_{\psi\psi}$ where

$$[\mathbf{R}_{\psi\psi}]_{p,p'} \triangleq \langle \psi_p(t), \psi_{p'}(t) \rangle$$

is the Gram matrix of the kernels $\psi_p(t)$'s.

Denote $\mathbf{c}[n] \triangleq [c_1[n], \dots, c_P[n]]^T$ as the length- P vector containing the samples acquired from the CMS filterbank at time $t = nD$. Given the link vector $\boldsymbol{\alpha}[\ell] = [\cdots, \alpha_{i,k,q}, \cdots]^T$ at the ℓ th shift as (24), we then have the observation model

$$\mathbf{c}[n] = \mathbf{M}_{\psi\phi}[n-\ell]\boldsymbol{\Gamma}[\ell]\boldsymbol{\alpha}[\ell] + \boldsymbol{\nu}[n]. \quad (59)$$

Using a sampling kernel constructed as

$$\psi_p(t) = \sum_{i=1}^I \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}} b_{p,(i,k,q)} \phi_{i,k,q}(t), \quad (60)$$

the cross-correlation $R_{\psi_p \phi_{i,k,q}}[(n-\ell)D]$ in (57) is given as

$$\begin{aligned} R_{\psi_p \phi_{i,k,q}}[(n-\ell)D] \\ = \sum_{i'=1}^I \sum_{q' \in \mathcal{Q}} \sum_{k' \in \mathcal{K}} b_{p,(i',k',q')} R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D], \end{aligned} \quad (61)$$

where

$$R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D] = \langle \phi_{i,k,q}(t-\ell D), \phi_{i',k',q'}(t-nD) \rangle.$$

With a change of variable $t' = t - nD - q'\Delta\tau$, the correlation can be expressed as

$$\begin{aligned} R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D] = e^{ik\Delta\omega(n-\ell)D} e^{-jk\Delta\omega q'\Delta\tau} \\ \times R_{\phi_{i',k',q'} \phi_{i,k,q}}^{(k-k')}[(q'-q)\Delta\tau + (n-\ell)D], \end{aligned} \quad (62)$$

where $R_{\phi_{i',k',q'} \phi_{i,k,q}}^{(k-k')}(\Delta t)$ is the ambiguity function

$$R_{\phi_{i',k',q'} \phi_{i,k,q}}^{(k-k')}(\Delta t) = \int \phi_{i'}^*(t) \phi_i(t - \Delta t) e^{-i(k-k')\Delta\omega t} dt. \quad (63)$$

From (61), $\mathbf{M}_{\psi\phi}[n-\ell]$ in (57) can be re-written as

$$\mathbf{M}_{\psi\phi}[n-\ell] = \mathbf{B}\mathbf{M}_{\phi\phi}[n-\ell], \quad (64)$$

where

$$[\mathbf{M}_{\phi\phi}[n-\ell]]_{(i',k',q'),(i,k,q)} = R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D].$$

Then the observation model can be re-written as

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}_{\phi\phi}[n-\ell]\mathbf{\Gamma}[\ell]\mathbf{\alpha}[n] + \mathbf{v}[n]. \quad (65)$$

Finally, the Gram matrix of $\psi_p(t)$'s is obtained accordingly as

$$\mathbf{R}_{\psi\psi} = \mathbf{B}\mathbf{M}_{\phi\phi}[0]\mathbf{B}^H, \quad (66)$$

which gives the noise covariance as $\mathbf{R}_{vv} = \sigma^2 \mathbf{B}\mathbf{M}_{\phi\phi}[0]\mathbf{B}^H$.

APPENDIX B PROOF OF THEOREM 2

From (56), each sample $c_p[n]$ from the CMS architecture $p = 1, \dots, P$ can be expressed as

$$\begin{aligned} c_p[n] = \sum_{i'=1}^I \sum_{k' \in \mathcal{K}} \sum_{q' \in \mathcal{Q}} b_{p,(i',k',q')} \sum_{i=1}^I \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}} \alpha_{i,k,q} e^{ik\Delta\omega\ell D} \\ \times R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D] + v_p[n]. \end{aligned} \quad (67)$$

The summation $\sum_{q \in \mathcal{Q}} \alpha_{i,k,q} e^{ik\Delta\omega\ell D} R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D]$ can be adjusted with respect to the relative time index $[n-\ell]$ by re-writing the correlation in (62) as

$$\begin{aligned} R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D] \\ = e^{ik\Delta\omega(n-\ell)D} e^{-jk\Delta\omega q'\Delta\tau} R_{\phi_{i',k',q'} \phi_{i,k,q}}^{(k-k')} \\ \times [(q'-q)\Delta\tau + (n-\ell)D] \\ = e^{ik\Delta\omega(n-\ell)D} e^{-jk\Delta\omega q'\Delta\tau} R_{\phi_{i',k',q'} \phi_{i,k,q}}^{(k-k')} \\ \times [(q' - [q + (\ell-n)N])\Delta\tau] \\ = e^{ik\Delta\omega(n-\ell)D} R_{\phi_{i',k',q'} \phi_{i,k,q+(\ell-n)N}}[0]. \end{aligned} \quad (68)$$

Without loss of generality, let $D/\Delta\tau = N \in \mathbb{Z}$. With a change of variable $q'' = q + (\ell-n)N$ and substituting the equivalent correlation in (68), we have

$$\begin{aligned} \sum_{q \in \mathcal{Q}} \alpha_{i,k,q} e^{ik\Delta\omega\ell D} R_{\phi_{i',k',q'} \phi_{i,k,q}}[(n-\ell)D] \\ \stackrel{(28)}{=} \sum_{q \in \mathcal{Q}} \alpha_{i,k,q} e^{ik\Delta\omega n D} R_{\phi_{i',k',q'} \phi_{i,k,q+(\ell-n)N}}[0] \\ = \sum_{q'' \in \mathcal{Q}} \alpha_{i,k,q''+(n-\ell)N} e^{ik\Delta\omega n D} R_{\phi_{i',k',q'} \phi_{i,k,q''}}[0]. \end{aligned}$$

With the re-formulation, (67) is re-written as below

$$\begin{aligned} c_p[n] = \sum_{i'=1}^I \sum_{k' \in \mathcal{K}} \sum_{q' \in \mathcal{Q}} b_{p,(i',k',q')} \sum_{i=1}^I \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}} \alpha_{i,k,q+(n-\ell)N} \\ \times e^{ik\Delta\omega n D} R_{\phi_{i',k',q'} \phi_{i,k,q}}[0] + v_p[n]. \end{aligned}$$

By letting $\mathbf{M} \triangleq \mathbf{M}_{\phi\phi}[0]$ and defining the shifted link vector $\mathbf{\alpha}[n]$ at the n th shift as

$$[\mathbf{\alpha}[n]]_{(i,k,q)} \triangleq \alpha_{i,k,q+(n-\ell)N}, \quad (69)$$

the observation model can be equivalently re-written as

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}\mathbf{\Gamma}[n]\mathbf{\alpha}[n] + \mathbf{v}[n]. \quad (70)$$

APPENDIX C PROOF OF PROPOSITION 1

The pair-wise KL distance in (47) can be re-written with the trace operator $\text{Tr}(\cdot)$ below

$$\mathbb{D}(\mathcal{H}_S \parallel \mathcal{H}_{S'}; \mathbf{B}) = \frac{1}{\sigma^2} \text{Tr} \left[\mathbf{M}^H \mathbf{B}^H (\mathbf{B}\mathbf{M}\mathbf{B}^H)^{-1} \mathbf{B}\mathbf{M}\mathbf{R}_{\boldsymbol{\beta}_S, \boldsymbol{\beta}_{S'}} \right],$$

where $\mathbf{R}_{\boldsymbol{\beta}_S, \boldsymbol{\beta}_{S'}} \triangleq (\boldsymbol{\beta}_S - \boldsymbol{\beta}_{S'}) (\boldsymbol{\beta}_S - \boldsymbol{\beta}_{S'})^H$. Then the average pair-wise KL distance $\overline{\mathbb{D}}$ becomes

$$\overline{\mathbb{D}} = \frac{1}{\sigma^2} \text{Tr} \left[\mathbf{M}^H \mathbf{B}^H (\mathbf{B}\mathbf{M}\mathbf{B}^H)^{-1} \mathbf{B}\mathbf{M}\mathbf{R} \right],$$

where $\mathbf{R} \triangleq \sum_S \sum_{S'} \gamma_{S,S'} \mathbf{R}_{S,S'}$ and $\mathbf{R}_{S,S'}$ is the averaged covariance matrix of $\boldsymbol{\beta}_S$ over the amplitudes

$$\mathbf{R}_{S,S'} = \int \int \mathbb{P}(\boldsymbol{\beta}_S) \mathbb{P}(\boldsymbol{\beta}_{S'}) \mathbf{R}_{\boldsymbol{\beta}_S, \boldsymbol{\beta}_{S'}} d\boldsymbol{\beta}_S d\boldsymbol{\beta}_{S'}.$$

Given $\mathbb{P}(\boldsymbol{\beta}_S) = \prod_{(i,k,q) \in \mathcal{S}} \mathbb{P}(\beta_{i,k,q})$ with $\int \boldsymbol{\beta}_S \mathbb{P}(\boldsymbol{\beta}_S) d\boldsymbol{\beta}_S = \mathbf{0}$ and $\int |\beta_{i,k,q}|^2 \mathbb{P}(\beta_{i,k,q}) d\beta_{i,k,q} = \sigma_\beta^2$, the averaged matrix $\mathbf{R}_{S,S'}$ is diagonal. Furthermore, if the set of weights $\gamma_{S,S'}$ are constant for all S, S' and the individual weighting function $\mathbb{P}(\beta_{i,k,q})$ is identical for all i, k, q , it also satisfies $\mathbf{R} = \sigma_\beta^2 \mathbf{I}$ because the summation over S, S' is symmetric, and hence produces equal sum. Thus the result follows.

APPENDIX D

PROOF OF THEOREM 3

By analogy with Lemma 1, we have $\mathbf{S} = \mathbf{M}$ and $\mathbf{G} = \mathbf{M}\mathbf{M}^H$ in (49). Let

$$\mathbf{B} \triangleq [\mathbf{b}_1 \ \cdots \ \mathbf{b}_P]^H, \quad (71)$$

where \mathbf{b}_p is a length- $J|\mathcal{K}||\mathcal{Q}|$ column vector with $\mathbf{b}_p = \mathbf{w}_p$. In this setting, according to Lemma 1, the optimal \mathbf{b}_p is chosen as the generalized eigenvector of the matrix pair (\mathbf{S}, \mathbf{G}) such that $\mathbf{M}\mathbf{b}_p = \lambda_p \mathbf{M}\mathbf{M}^H \mathbf{b}_p$. Using the eigen-decomposition of $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H$ and the property $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, we have

$$\boldsymbol{\Sigma}\mathbf{U}^H \mathbf{b}_p = \lambda_p \boldsymbol{\Sigma}\mathbf{U}^H \mathbf{U}^H \mathbf{b}_p, \quad p = 1, \dots, P. \quad (72)$$

If we choose $\mathbf{b}_p = \mathbf{u}_p$, where \mathbf{u}_p is the p th column in the matrix \mathbf{U} , then the above relationship holds for all $p = 1, \dots, P$ as long as $P \leq \text{rank}(\boldsymbol{\Sigma})$ because $\mathbf{u}_i^H \mathbf{u}_j = \delta[i - j]$. This gives

$$\text{L.H.S.} : \sigma_p \mathbf{U}^H \mathbf{u}_p = \sigma_p \mathbf{e}_p,$$

$$\text{R.H.S.} : \lambda_p \boldsymbol{\Sigma}\mathbf{U}^H \mathbf{U}^H \mathbf{u}_p = \lambda_p \sigma_p^2 \mathbf{e}_p,$$

with generalized eigenvalue $\lambda_p = 1/\sigma_p$, where $\sigma_p > 0$ is the p th eigenvalue in $\boldsymbol{\Sigma}$ and \mathbf{e}_p is the canonical basis with 1 in the p th entry and 0 otherwise. Denote the principal eigenvalue and eigenvector matrix as $\boldsymbol{\Sigma}_P$ and \mathbf{U}_P . Then the optimal \mathbf{B} is chosen as

$$\mathbf{B} = \boldsymbol{\Xi}_P \mathbf{U}_P^H, \quad (73)$$

where $\boldsymbol{\Xi}_P$ is an arbitrary non-singular $P \times P$ matrix. According to (49), this choice gives

$$\bar{\mathbb{D}} = \frac{\sigma_\beta^2}{\sigma^2} \text{Tr} \left(\underbrace{\boldsymbol{\Sigma}_P^H \boldsymbol{\Xi}_P^H \boldsymbol{\Xi}_P^{-H}}_{=\mathbf{I}} \boldsymbol{\Sigma}_P^{-1} \underbrace{\boldsymbol{\Xi}_P^{-1} \boldsymbol{\Xi}_P}_{=\mathbf{I}} \boldsymbol{\Sigma}_P \right) = \frac{\sigma_\beta^2}{\sigma^2} \sum_{p=1}^P \sigma_p,$$

which is independent of $\boldsymbol{\Xi}_P$. If the principal eigenvectors \mathbf{U}_P are unique, the above \mathbf{B} maximizes the average KL distance $\bar{\mathbb{D}}$. This choice of \mathbf{B} in general spreads out the individual KL distance, while the occurrence of the events $\mathbb{D}(\mathcal{H}_S || \mathcal{H}_{S'}; \mathbf{B}) = 0$ is analyzed below. So is the case when \mathbf{U}_P is not unique.

Now we examine the occurrence of $\mathbb{D}(\mathcal{H}_S || \mathcal{H}_{S'}; \mathbf{B}) = 0$. Let $\boldsymbol{\beta}_{S \cup S'} = (\boldsymbol{\beta}_S - \boldsymbol{\beta}_{S'})$ be a sparse vector with $|\mathcal{S}|, |\mathcal{S}'| \leq s$, and $s \leq |\mathcal{I}|R$. Substituting $\mathbf{B} = \boldsymbol{\Xi}_P \mathbf{U}_P^H$ back to (47) and simplifying the expression, the individual KL distance is

$$\mathbb{D}(\mathcal{H}_S || \mathcal{H}_{S'}; \mathbf{B}) = \frac{1}{\sigma^2} \boldsymbol{\beta}_{S \cup S'}^H \mathbf{U}_P \boldsymbol{\Sigma}_P \mathbf{U}_P^H \boldsymbol{\beta}_{S \cup S'}, \quad (74)$$

$$\forall \mathcal{S} \neq \mathcal{S}', |\mathcal{S}|, |\mathcal{S}'| \leq s. \quad (75)$$

Since $\boldsymbol{\beta}_{S \cup S'}$ is a $2s$ -sparse vector, thus $\mathbb{D}(\mathcal{H}_S || \mathcal{H}_{S'}; \mathbf{B})$ is bounded away from zero as long as any $2s$ -sparse vectors do not fall into the null space of the matrix \mathbf{U}_P^H , which implicitly implies $P \geq 2s$. In order to minimize the occurrence of the event $\mathbb{D}(\mathcal{H}_S || \mathcal{H}_{S'}; \mathbf{B}) = 0$ given a certain level of signal sparsity s , it is equivalent to maximizing the kruskal rank of the matrix \mathbf{U}_P^H such that the matrix \mathbf{B} can recover any s -sparse vector $\boldsymbol{\beta}_S$ with s as large as possible. This is consistent with the popular results in compressed sensing, therefore when the solution obtained from the optimization is not unique, one can use this as a criterion to choose the best candidate from the solutions of \mathbf{B} that maximize the average KL distance.

REFERENCES

- [1] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," Mar. 2009 [Online]. Available: [Online] <http://arxiv.org/abs/0903.1022>
- [2] L. Applebaum, W. U. Bajwa, M. F. Duarte, and R. Calderbank, "Asynchronous code-division random access using convex optimization," *Elsevier Phys. Commun.*, vol. 5, no. 2, pp. 129–147, Jun. 2012.
- [3] S. Verdú, *Multuser Detection*. Cambridge, MA, USA: Cambridge Univ. Press, 1998.
- [4] Y. Xie, Y. C. Eldar, and A. Goldsmith, "Reduced-dimension multiuser detection," *IEEE Trans. Inf. Theory*, 2013, accepted for publication.
- [5] B. Sirkeci-Mergen and A. Scaglione, "Signal acquisition for cooperative transmissions in multi-hop ad-hoc networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2004, vol. 2, pp. ii–345.
- [6] S. Buzzi, A. De Maio, and M. Lops, "Code-aided blind adaptive new user detection in DS/CDMA systems with fading time-dispersive channels," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2637–2649, 2003.
- [7] L. Qiu, Y. Huang, and J. Zhu, "Fast acquisition scheme and implementation of PRACH in WCDMA system," in *Proc. IEEE Veh. Technol. Conf.*, 2001, vol. 3, pp. 1701–1705.
- [8] R. Wang and H. Li, "Decorrelating multiuser code-timing estimation for long-code CDMA with bandlimited chip waveforms," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2369–2381, 2005.
- [9] E. Fishler, A. Haimovich, R. Blum, L. Cimini, Jr, D. Chizhik, and R. Valenzuela, "Spatial diversity in RADARS-models and detection performance," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 823–838, 2006.
- [10] Z. Tian and G. Giannakis, "A GLRT approach to data-aided timing acquisition in UWB Radios-Part I: Algorithms," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2956–2967, 2005.
- [11] H. Zhu and G. Giannakis, "exploiting sparse user activity in multiuser detection," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 454–465, 2011.
- [12] M. Duarte, M. Davenport, M. Wakin, and R. Baraniuk, "Sparse signal detection from incoherent projections," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2006, vol. 3, pp. III-305–III-308.
- [13] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, 2010.
- [14] J. Haupt and R. Nowak, "Compressive sampling for signal detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2007, vol. 3, pp. III-1509.
- [15] Z. Wang, G. Arce, and B. Sadler, "Subspace compressive detection for sparse signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2008, pp. 3873–3876.
- [16] J. Paredes, Z. Wang, G. Arce, and B. Sadler, "Compressive matched subspace detection," in *Proc. Eur. Signal Process. Conf.*, 2009, pp. 120–124.
- [17] 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, 2010.
- [18] W. Bajwa, K. Gedalyahu, and Y. Eldar, "Identification of parametric underspread linear systems and super-resolution radar," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2548–2561, 2011.
- [19] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.

- [20] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, 2005.
- [21] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1491–1504, 2011.
- [22] J. Urigüen, Y. C. Eldar, P. L. Dragotti, and Z. Ben-Haim, "Sampling at the rate of innovation: Theory and applications," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [23] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, 2009.
- [24] S. Qaisar and A. Dempster, "Cross-correlation performance assessment of global positioning system (gps) l_1 and l_2 civil codes for signal acquisition," *IET Radar Sonar Navig.*, vol. 5, no. 3, pp. 195–203, 2011.
- [25] J. Kim, S. Sarin, M. Yasunaga, and H. Oh, "Robust noncoherent PN-code acquisition for CDMA communication systems," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 278–286, 2001.
- [26] S. Kay, *Fundamentals of Statistical Signal Processing: Volumes I & II*. Englewood Cliffs, NY, USA: Prentice-Hall PTR, 1998.
- [27] X. Li, A. Rueetschi, Y. C. Eldar, and A. Scaglione, "GPS signal acquisition via compressive multichannel sampling," *Phys. Commun.*, vol. 5, no. 2, pp. 173–184, 2012.
- [28] A. Daniele, G. Emanuele, G. Georgios, and B. L. Marco *et al.*, "Sparsity-aware estimation of CDMA system parameters," *EURASIP J. Adv. Signal Process.*, vol. 2010, 2010, doi: 10.1155/2010/417981.
- [29] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th IEEE Asilomar Conf. Signals Syst. Comput.*, 1993, pp. 40–44.
- [30] É. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [31] T. M. Cover, J. A. Thomas, and J. Wiley *et al.*, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.
- [32] D. Bajovic, B. Sinopoli, and J. Xavier, "Robust linear dimensionality reduction for hypothesis testing with application to sensor selection," in *Proc. IEEE 47th Annu. Allerton Conf. Commun. Control Comput.*, 2009, pp. 363–370.
- [33] J. Klauder, "The design of radar signals having both high range resolution and high velocity resolution," *Bell Syst. Tech. J.*, vol. 39, pp. 809–820, 1960.
- [34] R. Duda, P. Hart, and D. Stork, *Pattern Classification and Scene Analysis. Part I, Pattern Classification*. New York, NY, USA: Wiley, 2010.
- [35] G. Buracas and G. Boynton, "Efficient design of event-related fMRI EXPERiments using M-sequences," *Neuroimage*, vol. 16, no. 3, pp. 801–813, 2002.
- [36] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [37] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [38] M. Salman Asif and J. Romberg, "Dynamic updating for minimization," *IEEE J. Selt. Topics Signal Process.*, vol. 4, no. 2, pp. 421–434, 2010.
- [39] D. Malioutov, M. Cetin, and A. Willsky, "Homotopy continuation for sparse signal representation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2005, vol. 5, p. v-733.



Andrea Rueetschi received the Diploma degree in physical electronics from the University of Neuchâtel, Switzerland, in 1997. In 2008, he joined the graduate program at Cornell University and transferred to the University of California, Davis, CA, USA, in January 2009.

He worked, until 2007, as an ASIC Design Engineer for Philips Semiconductors, RF Micro Devices, and Marvell Semiconductors developing system-on-chip solutions for wireless communication.



Anna Scaglione (F'10) received the Laurea (M.Sc.) and Ph.D. degrees from the University of Rome "La Sapienza," Rome, Italy, in 1995 and 1999, respectively.

She is currently a Professor in electrical and computer engineering at the University of California, Davis, CA, USA, which she joined in 2008. She was previously at Cornell University, Ithaca, NY, USA, from 2001, where she became Associate Professor in 2006. Prior to joining Cornell, she was an Assistant Professor, from 2000 to 2001, at the University

of New Mexico, Albuquerque, NM, USA. Her current research focuses on communication and wireless networks, sensors' systems for monitoring, control and energy management and network science. Her expertise is in the broad area of signal processing for communication systems and networks.

Dr. Scaglione served as Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2002 to 2005. She has served on the Editorial Board of the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2008 and was an Area Editor from 2008 to 2011. She received the IEEE Signal Processing Transactions Best Paper Award in 2000, the NSF Career Award in 2002, and is a corecipient of the Ellersick Best Paper Award (MILCOM 2005) and the 2013 IEEE Donald G. Fink Prize Paper Award.



Yonina C. Eldar (S'98–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, MA, USA, 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, Israel. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT and a Visiting Professor at Stanford University, Stanford, CA, USA. 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 Technions 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. 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 Bio Imaging 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*.



Xiao Li received the B.Eng. degree, in 2007, from Sun Yat-Sen (Zhongshan) University, China, and the M.Phil. degree, in 2009, from the University of Hong Kong. He has been working towards the Ph.D. degree at the University of California, Davis, CA, USA, since 2009.

His research interests are in the theoretical and algorithmic studies in signal processing and optimizations, statistical learning and inferences for high dimensional data, distributed optimizations and adaptive algorithms, as well as their applications in communications, networked systems, and smart grid.