Multiuser Detection in Asynchronous On–Off Random Access Channels Using Lasso

Lorne Applebaum, Waheed U. Bajwa, Marco F. Duarte, and Robert Calderbank

***Abstract*—This paper considers on–off random access channels where users transmit either a one or a zero to a base station. Such channels represent an abstraction of control channels used for scheduling requests in third-generation cellular systems and uplinks in wireless sensor networks deployed for target detection. This paper introduces a novel convex-optimization-based scheme for multiuser detection (MUD) in asynchronous on–off random access channels that does not require knowledge of the delays or the instantaneous received signal-to-noise ratios of the individual users at the base station. For any fixed number of temporal signal space dimensions N and maximum delay τ in the system, the proposed scheme can accommodate $M \lesssim \exp(O(N^{1/3}))$ total users and $k \lesssim N/\log M$ active users in the system—a significant improvement over the $k \leq M \lesssim N$ scaling suggested by the use of classical matched-filtering-based approaches to MUD employing orthogonal signaling. Furthermore, the computational complexity of the proposed scheme differs from that of a similar oracle-based scheme with perfect knowledge of the user delays by at most a factor of $\log(N+\tau)$. Finally, the results presented in here are non-asymptotic, in contrast to related previous work for synchronous channels that only guarantees that the probability of MUD error at the base station goes to zero asymptotically in M.**

I. Introduction

In wireless systems, the term *random access* commonly refers to the scenario in which a number of users vie to simultaneously communicate with a base station (access point) in an uncoordinated fashion. In this paper, we are interested in studying *on–off random access channels*, which are characterized by the fact that the users transmit either a "one" or a "zero" to the base station (BS). Such channels represent an abstraction that arises frequently in many applications. In third-generation cellular systems, for example, control channels that are used for scheduling requests can be modeled as on–off random access channels; in this case, users requesting permissions to send data to the BS can be thought of as transmitting 1's and inactive users can be thought of as transmitting 0's. Similarly, uplinks in wireless sensor networks deployed for target detection can also be modeled as on–off random access channels; in this case, sensors that detect a target can be made to transmit 1's and sensors that have nothing to report can be thought of as transmitting 0's.

The primary objective of the BS in on–off random access channels is to reliably detect the identity of the active users (i.e., users that transmit 1's) in polynomial time. The two biggest impediments to this goal are that (i) random access channels tend to be asynchronous in nature and (ii) it is quite difficult, if not impossible, for the BS to know the instantaneous received signal-to-noise ratio (SNR) of each individual user. Given a fixed number of temporal signal space dimensions N, the system-design goal therefore is to simultaneously maximize the total number of users M and the expected number of active users k that the system can handle *without* requiring knowledge of the delays or the instantaneous received SNRs of the individual users at the BS.

In this paper, we propose a novel convex-optimization-based scheme for multiuser detection (MUD) in asynchronous on–off random access channels that does not require knowledge of the delays or the instantaneous received SNRs of the individual users at the BS. In particular, for any fixed number of temporal signal space dimensions N and maximum delay τ in the system, we rigorously establish that the proposed scheme successfully carries out the MUD with high probability as long as the total number of users $M \lesssim \exp(O(N^{1/3}))$ and the expected number of active users $k \lesssim N/\log M$. In order to put the significance of this result into context, note that classical matched-filtering-based approaches to MUD using orthogonal signaling dictate that $k \leq M \lesssim N$, which severely limits the total number of users that can be handled by the system for a given N. In addition, we also present an efficient implementation of the proposed MUD scheme based on the *fast Fourier transform* (FFT) that ensures that the computational complexity of the proposed scheme at worst differs by a factor of $\log(N+\tau)$ from an *oracle-based* scheme that has perfect knowledge of the user delays.

In regards to previous work, we note that Fletcher et al. [1] have also recently studied the problem of MUD in on–off random access channels. However, the results in [1]—while similar in spirit to the ones in here—are limited by the facts that [1]: (i) assumes perfect synchronization among the M users, which is hard to guarantee in practical settings for large M; (ii) assumes that instantaneous received SNRs of the individual users are available to the BS in certain cases, which is difficult—if not impossible—to justify for the case of *fading* random access channels; and (iii) only guarantees that the probability of error P_{err} at the BS goes to zero asymptotically in M, which does not shed light on the scaling of P_{err}. Finally, while preparing this paper, we became aware of [2] that also considers on–off random access in the context of configuration in ad-hoc wireless networks, and makes assumptions about the channel model that are similar to [1].

This work was completed at Princeton University. Three of the authors, WUB, MFD, and RC, are now with Duke University. The work is supported in part by NSF under grant DMS-0914892, by ONR under grant N00014-08-1-1110, and by AFOSR under grants FA9550-09-1-0643 and FA9550-09-1-0551. MFD is also supported by NSF Supplemental Funding DMS-0439872 to UCLA-IPAM, P.I. R. Caflisch. (E-mails: lappleba@princeton.edu, {w.bajwa, marco.duarte, robert.calderbank}@duke.edu)

II. Problem Formulation

In this section, we formulate the problem of MUD in asynchronous on–off random access channels, along with the accompanying assumptions. To begin, we assume that there are a total of M users in the system that communicate with the BS using packets of duration T and (two-sided) bandwidth W; in other words, the total number of temporal signal space dimensions (degrees of freedom) in the wireless system are $N = TW$. Further, we assume that users communicate using spread spectrum waveforms of the form

$$x_i(t) = \sqrt{\mathcal{E}_i} \sum_{n=0}^{N-1} x_n^i \, g(t - nT_c), \; t \in [0, T) \tag{1}$$

where $g(t)$ is a unit-energy prototype pulse ($\int |g(t)|^2 dt = 1$), $T_c \approx \frac{1}{W}$ is the *chip duration*, $\mathcal{E}_i$ denotes the transmit power of the i-th user, and

$$\mathbf{x}_i = \begin{bmatrix} x_0^i & x_1^i & \dots & x_{N-1}^i \end{bmatrix}^{\mathrm{T}}, \; i = 1, \dots, M \tag{2}$$

is the N-length (real- or complex-valued) *codeword* of unit energy ($\|\mathbf{x}_i\|_2 = 1$) assigned to the i-th user.

The key feature of on–off random access channels that distinguishes them from the more commonly studied multiple-access channels in network information theory (and related multiuser-detection problems) is the assumption that only a small number of *random* users communicate 1's with the BS at any time instant. Specifically, we assume (without loss of generality) that on average a total of k of the M users transmit 1's at time $t = 0$, resulting in the following expression for the received signal at the BS

$$y(t) = \sum_{i=1}^{M} h_i \delta_i x_i(t - \tau_i) + w(t). \tag{3}$$

Here, $h_i \in \mathbb{C}$ and $\tau_i \in \mathbb{R}_+$ are the *channel fading coefficient* and the *delay* associated with the i-th user, respectively, $w(t)$ is complex additive white Gaussian noise (AWGN) introduced by the receiver circuitry, and $\{\delta_i\}$ are independent 0–1 Bernoulli random variables that model the random activation of the M users in the system in the sense that $\Pr(\delta_i = 1) = k/M$. Note that one of the major differences between [1], [2] and the setup in here is that it is assumed in [1], [2] that $\max_{i,j}(\tau_i - \tau_j) < T_c$ whereas we do not make any such assumption here since it is nearly impossible to satisfy this condition for large-enough values of M. Finally, we assume that the transmissions of the users undergo independent Rayleigh fading in the sense that the h_i's are independently distributed as $\mathcal{CN}(0, \rho_i^2)$.

Next, we define the individual *discrete delays* $\tau_i' \in \mathbb{Z}_+$ as $\tau_i' \stackrel{def}{=} \left\lfloor \frac{\tau_i}{T_c} \right\rfloor$ and define the *maximum discrete delay* $\tau \in \mathbb{Z}_+$ in the system as $\tau \stackrel{def}{=} \max_i \tau_i'$. It is easy to see that the received signal $y(t)$ at the BS can be sampled at the chip rate to obtain the equivalent discrete representation

$$\mathbf{y} = \sum_{i=1}^{M} h_i \delta_i \sqrt{\mathcal{E}_i} \, \tilde{\mathbf{x}}_i + \mathbf{w}. \tag{4}$$

Here, the (complex) AWGN vector $\mathbf{w}$ is distributed as $\mathcal{CN}(\mathbf{0}_{N+\tau}, \mathbf{I}_{N+\tau})$ (in other words, the instantaneous received SNR of the active users is $\mathcal{E}_i |h_i|^2$) and the vectors $\tilde{\mathbf{x}}_i \in \mathbb{C}^{N+\tau}$ are defined as

$$\tilde{\mathbf{x}}_i = \begin{bmatrix} \mathbf{0}_{\tau_i'}^{\mathrm{T}} & \mathbf{x}_i^{\mathrm{T}} & \mathbf{0}_{\tau-\tau_i'}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \; i = 1, \dots, M. \tag{5}$$

The goal of any MUD scheme in asynchronous on–off random access channels is to obtain an estimate $\widehat{\mathcal{I}}$ of the set of active users $\mathcal{I} \stackrel{def}{=} \{i : \delta_i = 1\}$ from the $(N + \tau)$-dimensional vector $\mathbf{y}$ without knowledge of the set of delays $\{\tau_i'\}$ or the set of channel coefficients $\{h_i\}$ at the BS. In particular, for the sake of this exposition, we are interested in characterizing three key aspects of our proposed scheme for MUD in asynchronous on–off random access channels:

1) the computational complexity of the solution,
2) the probability of error, $P_{err} \stackrel{def}{=} \Pr(\widehat{\mathcal{I}} \neq \mathcal{I})$, and
3) the relationship between the number of temporal signal space dimensions N, the maximum (discrete) delay τ in the system, the total number of users M that can be accommodated by the BS, and the average number of active users k in the system.

In this regard, the only assumptions we allow ourselves to make here are that (i) the maximum delay τ (or an upper bound on τ) is known at the BS and (ii) each user has knowledge of the SNR at which its transmitted signal arrives at the BS (in other words, the i-th user knows $|h_i|$). Note that both these assumptions are quite reasonable from a practical perspective; in particular, if one assumes that the BS transmits a beacon signal before the users start transmitting then the last assumption simply follows because of reciprocity between the downlink and the uplink.

III. Multiuser Detection Using Lasso

In this section, we describe our proposed approach to MUD in asynchronous on–off random access channels that is based on the mixed-norm convex optimization program known as the lasso [3]. The lasso was first proposed in the statistics literature for linear regression in underdetermined settings. In [1], the lasso has been suggested as a potential method for MUD in *synchronous* on–off random access channels. However, extending the ideas of [1] to the asynchronous case using the standard lasso formulation seems very difficult. In contrast, while the MUD scheme proposed in this paper is based on the lasso, we present a rather nonconventional usage of the lasso that is specific to the problem at hand and one of our major contributions indeed is establishing that this formulation is guaranteed to yield successful MUD with high probability. It is also worth mentioning here that the analysis carried out in the paper in this regard might also be of independent interest to researchers working on configuration (neighbor discovery) in ad-hoc wireless networks and sensor networks.

A. Main Result

In order to make use of the lasso for MUD in asynchronous on-off random access channels, we first rewrite (4) in the following matrix–vector product form

$$\mathbf{y} = \underbrace{\begin{bmatrix}\tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_2 & \dots & \tilde{\mathbf{x}}_M\end{bmatrix}}_{\widetilde{\mathbf{X}}}\tilde{\boldsymbol{\beta}} + \mathbf{w} \tag{6}$$

where the i-th entry of the vector $\tilde{\boldsymbol{\beta}} \in \mathbb{C}^M$ is described as $\tilde{\beta}_i \stackrel{def}{=} h_i\delta_i\sqrt{\mathcal{E}_i}$. Note that despite the fact that the above expression appears superficially similar to the standard lasso formulation, we cannot use the lasso to obtain an estimate of the set of active users $\mathcal{I}$ from (6) since the $(N+\tau)\times M$ matrix $\widetilde{\mathbf{X}}$ in (6) is unknown due to the asynchronous nature of the problem. In order to overcome this obstacle, we first define $(N+\tau)\times(\tau+1)$ *Toeplitz matrices* $\mathbf{X}_i$ as follows

$$\mathbf{X}_i = \begin{bmatrix}\mathbf{x}_i & & 0\\ & \ddots & \\ 0 & & \mathbf{x}_i\end{bmatrix},\ i = 1,\dots,M \tag{7}$$

and observe that we can equivalently write (6) in the form

$$\mathbf{y} = \underbrace{\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_M\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1^{\mathrm{T}} & \boldsymbol{\beta}_2^{\mathrm{T}} & \dots\boldsymbol{\beta}_M^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}}_{\mathbf{X}\boldsymbol{\beta}} + \mathbf{w} \tag{8}$$

where $\mathbf{X}$ is now an $(N+\tau)\times M(\tau+1)$ *known* matrix and the vector $\boldsymbol{\beta}\in\mathbb{C}^{M(\tau+1)}$ is a concatenation of M vectors, each of length $(\tau+1)$, whose entries are given by

$$\beta_{i,j} = \tilde{\beta}_i 1_{\{\tau_i'=j-1\}},\quad i=1,\dots,M,\ j=1,\dots,\tau+1. \tag{9}$$

We can now make use of this notation to describe the proposed lasso-based scheme for MUD in asynchronous on–off random access channels.[1]

Algorithm 1 Multiuser Detection in Asynchronous On–Off Random Access Channels Using Lasso

Inputs

1) The chip-rate sampled vector $\mathbf{y}$
2) Set of N-dimensional codewords $\{\mathbf{x}_i\}_{i=1}^M$
3) Maximum discrete delay τ in the system
4) A regularization parameter λ for the lasso

Compute the matrix $\mathbf{X}$ described in (8) using $\{\mathbf{x}_i\}$ and τ

$$\widehat{\boldsymbol{\beta}} \leftarrow \underset{\mathbf{b}\in\mathbb{C}^{M(\tau+1)}}{\arg\min}\ \tfrac{1}{2}\big\|\mathbf{y}-\mathbf{X}\mathbf{b}\big\|_2^2 + \lambda\|\mathbf{b}\|_1 \qquad \text{(LASSO)}$$

$$\widehat{\mathcal{I}} \leftarrow \{i : \|\widehat{\boldsymbol{\beta}}_i\|_0 > 0\}$$

Return $\widehat{\mathcal{I}}$ as an estimate of the set of active users $\mathcal{I}$

We are now ready to state the main result of this paper, which bounds the probability of error of Algorithm 1 and specifies the corresponding relationship between the system parameters τ, k, N, and M.

[1]Algorithm 1 acts as a hybrid between the standard lasso and the group lasso [4]. Specifically, it is clear from the problem formulation that the group lasso is ill-suited for the specified MUD problem since each of the sub-vectors $\{\boldsymbol{\beta}_i\}$ in (8) has at most one nonzero entry. On the other hand, we are only interested in detecting the active users and need not estimate their delays; hence, the group nature of the detection criterion in the definition of $\widehat{\mathcal{I}}$.

Theorem 1. *Suppose that the M codewords $\{\mathbf{x}_i \in \mathbb{C}^N\}_{i=1}^M$ are drawn independently from a* binary$(\pm 1/\sqrt{N}, \mathbf{I}_N)$ *distribution and pick the parameter $\lambda = 2\sqrt{2\log(M\sqrt{\tau+1})}$. Further, let the transmit powers of the active users satisfy*

$$\mathcal{E}_i > \frac{128\log(M\sqrt{\tau+1})}{|h_i|^2},\ i\in\mathcal{I}. \tag{10}$$

Then Algorithm 1 successfully carries out multiuser detection with $P_{err} \le 13\big(M(\tau+1)\big)^{-1} + 4M^{-1} + 2\exp\Big(-\frac{\sqrt{NM}}{8}\Big)$ if

$$M \le \frac{\exp\big(c_1(\tau+1)^{-2/3}N^{1/3}\big)}{\tau+1} \qquad \textit{and} \tag{11}$$

$$k \le \frac{c_2 N}{(\tau+1)\log\big(M(\tau+1)\big)}. \tag{12}$$

Here, the constants $c_1, c_2 > 0$ are independent of the problem parameters.

The proof of this theorem is provided in Section IV. The implications of the scaling behavior outlined in (11) and (12) are quite positive in the important special case of fixed-bandwidth spread spectrum waveforms and a base station serving a bounded geographic region. Specifically, Theorem 1 signifies that—for any fixed number of temporal signal space dimensions N and maximum delay τ in the system—the proposed MUD scheme can accommodate $M \lesssim \exp(O(N^{1/3}))$ total users and $k \lesssim N/\log M$ active users in the system. This is a significant improvement over the $k \le M \lesssim N$ scaling suggested by the use of classical matched-filtering-based approaches to MUD employing orthogonal signaling.

We conclude our discussion of Theorem 1 by noting that $k \lesssim N/\log M$ scaling has also been suggested in [1] for the case of MUD in synchronous on–off random access channels using the lasso. In contrast, Theorem 1 establishes that the MUD scheme proposed here for asynchronous on–off random access channels has the ability to achieve roughly the same scaling of the system parameters k, N, and M as that reported in [1] for the ideal case of synchronous channels.

B. Computational Complexity

Theorem 1 helps us characterize the performance of Algorithm 1 for MUD in asynchronous on–off random access channels but fails to shed any light on the issue of computational complexity of the proposed scheme. However, note that the lasso is a well-studied program in the statistics literature and—thanks to its convex nature—there exist a number of extremely fast (polynomial-time) implementations of the unconstrained version of the lasso specified in (LASSO); see, e.g., [5].

In this regard, note that the computational complexity of the implementations of (LASSO) such as SpaRSA [5] is determined—to a large extent—by the complexity of the matrix–vector multiplications $\mathbf{X}\mathbf{b}$ and $\mathbf{X}^{\mathrm{H}}\mathbf{y}$. It therefore seems that Algorithm 1 increases the computational complexity of the matrix–vector multiplications from $O(NM)$, corresponding to the case of perfectly-known user delays [cf. (6)], to $O(NM(\tau+1))$. This observation, however, ignores the fact that the matrix $\mathbf{X}$ in (8) has a Toeplitz-block structure. Specifically, note that if we write $\mathbf{b} \in \mathbb{C}^{M(\tau+1)}$ as $\mathbf{b} = \begin{bmatrix}\mathbf{b}_1^{\mathrm{T}} & \dots & \mathbf{b}_M^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ then it follows from elementary signal processing that

$$\mathbf{X}\mathbf{b} = \sum_{i=1}^{M} \mathcal{F}_{N+\tau}^{-1}\Big(\mathcal{F}_{N+\tau}(\mathbf{x}_i) \odot \mathcal{F}_{N+\tau}(\mathbf{b}_i)\Big) \tag{13}$$

where $\mathcal{F}_n(\cdot)$ and $\mathcal{F}_n^{-1}(\cdot)$ denote the FFT implementation of the n-point discrete Fourier transform (DFT) and the n-point inverse DFT of a sequence, respectively, while $\odot$ denotes pointwise multiplication. Similarly, if we use $(\cdot)[n_1 : n_2]$ to denote the n_1-th to n_2-th elements of a vector and $(\cdot)^{-}$ to denote the time-reversed version of a vector, then it follows from routine calculations that $\forall\ i = 1, \dots, M$, we have

$$\mathbf{X}^{\mathrm{H}}\mathbf{y}[i(\tau+1)-\tau : i(\tau+1)] = \mathcal{F}_{2N+\tau-1}^{-1}\Big(\mathcal{F}_{2N+\tau-1}(\mathbf{x}_i^{-})\odot \odot \mathcal{F}_{2N+\tau-1}(\mathbf{y})\Big)\big[N : N+\tau\big]. \tag{14}$$

It therefore follows from the complexity of the FFT that the matrix–vector multiplications $\mathbf{X}\mathbf{b}$ and $\mathbf{X}^{\mathrm{H}}\mathbf{y}$ in Algorithm 1 can in fact be carried out using only $O(NM\log(N+\tau))$ operations as opposed to $O(NM(\tau+1))$ operations. This suggests that the computational complexity of Algorithm 1 at worst differs by a factor of $\log(N+\tau)$ from an oracle-based scheme that has perfect knowledge of the user delays.

IV. Proof of the Main Result

In this section, we provide a proof of Theorem 1. To begin, we develop some notation to facilitate the forthcoming analysis. Throughout this section, we use $\mathbf{X}_{\mathcal{B}}$ to denote the *block subdictionary* of $\mathbf{X}$ obtained by collecting the Toeplitz blocks of $\mathbf{X}$ corresponding to the indices of the active users; in other words, we have $\mathbf{X}_{\mathcal{B}} \stackrel{def}{=} \big[\mathbf{X}_i : i \in \mathcal{I}\big]$. In addition, we use $\mathbf{X}_{\mathcal{S}}$ to denote the $(N+\tau)\times|\mathcal{I}|$ submatrix obtained by collecting the columns of $\mathbf{X}$ corresponding to the nonzero entries of $\boldsymbol{\beta}$, while we use $\boldsymbol{\beta}_{\mathcal{S}}$ to denote the $|\mathcal{I}|$-dimensional vector comprising of the nonzero entries of $\boldsymbol{\beta}$. Finally, we use $\mathrm{sgn}(\cdot)$ for elementwise *signum* function, where $\mathrm{sgn}(z) \stackrel{def}{=} z/|z|$ for any $z \in \mathbb{C}$.

The basic idea behind the proof of Theorem 1 follows from the proof of [6, Theorem 1.3]. Specifically, using $\mathcal{S} = supp(\boldsymbol{\beta})$ to denote the set of the locations of the nonzero entries of $\boldsymbol{\beta}$, we have from [6, Lemma 3.4] that the lasso solution $\widehat{\boldsymbol{\beta}} \stackrel{def}{=} \boldsymbol{\beta} + \mathbf{h}$ satisfies $\mathbf{h}_{\mathcal{S}^c} = 0$ and

$$\mathbf{h}_{\mathcal{S}} = (\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}[\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{w} - \lambda\,\mathrm{sgn}(\boldsymbol{\beta}_{\mathcal{S}})] \tag{15}$$

if $\min_{i\in\mathcal{S}}|\beta_i| > 4\lambda$ and the following five conditions are met:

- $\mathcal{C}_1$ – Invertibility condition: $\|(\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\|_2 \le 2$.
- $\mathcal{C}_2$ – Noise stability: $\|(\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{w}\|_\infty \le \lambda$.
- $\mathcal{C}_3$ – Complementary noise stability:
$$\|\mathbf{X}_{\mathcal{S}^c}^{\mathrm{H}}(\mathbf{I} - \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{H}})\mathbf{w}\|_\infty \le \tfrac{\lambda}{\sqrt{2}}.$$
- $\mathcal{C}_4$ – Size condition: $\|(\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\mathrm{sgn}(\boldsymbol{\beta}_{\mathcal{S}})\|_\infty \le 3$
- $\mathcal{C}_5$ – Complementary size condition:
$$\|\mathbf{X}_{\mathcal{S}^c}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\mathrm{sgn}(\boldsymbol{\beta}_{\mathcal{S}})\|_\infty \le \tfrac{1}{4}.$$

Further, it trivially follows in this case that $supp(\widehat{\boldsymbol{\beta}}) \equiv \mathcal{S}$, which guarantees that $\widehat{\mathcal{I}} = \mathcal{I}$. Our goal then is to consider the probability of each one of these conditions *not being met* under the assumptions of Theorem 1 and the proof of the theorem would then simply follow from the union bound. The requisite analysis in this regard frequently requires a bound on the maximum inner products between the columns of $\mathbf{X}$ and a bound on the spectral norm of $\mathbf{X}$, and the following two lemmas help us specify these two bounds.

Lemma 1. *Given any fixed $\varsigma > 0$, the Toeplitz-block matrix $\mathbf{X}$ described in* (8) *satisfies*

$$\mu(\mathbf{X}) \stackrel{def}{=} \max_{(i,j)\neq(i',j')} \big|\langle \mathbf{x}_{i,j}, \mathbf{x}_{i',j'}\rangle\big| \le \varsigma \tag{16}$$

with probability exceeding $1 - 2M^2(\tau+1)^2 e^{-\frac{N\varsigma^2}{4}}$. Here, $\mathbf{x}_{i,j}$ denotes the j-th column of the Toeplitz matrix $\mathbf{X}_i$.

Proof: The proof of this lemma is a consequence of the bound on the worst-case coherence μ of random Toeplitz matrices [7, Theorem 3.5] and the Hoeffding inequality [8]. Specifically, note that we can write

$$\mu(\mathbf{X}) = \max\Big\{\max_{j\neq j'}\big|\langle \mathbf{x}_{i,j}, \mathbf{x}_{i,j'}\rangle\big|, \max_{i\neq i'}\big|\langle \mathbf{x}_{i,j}, \mathbf{x}_{i',j'}\rangle\big|\Big\}.$$

Further, note that the proof of Theorem 3.5 in [7] implies that $\big|\langle \mathbf{x}_{i,j}, \mathbf{x}_{i,j'}\rangle\big| \le \varsigma$ with probability exceeding $1 - 4e^{-\frac{N\varsigma^2}{4}}$ for any $j \neq j'$. Finally, since the product of two independent binary random variables is again a binary random variable, it can also be shown using the Hoeffding inequality that $\big|\langle \mathbf{x}_{i,j}, \mathbf{x}_{i',j'}\rangle\big| \le \varsigma$ with probability exceeding $1 - 2e^{-\frac{N\varsigma^2}{2}}$ for any $i \neq i'$. It therefore follows from the union bound that $\mu(\mathbf{X}) \le \varsigma$ with probability exceeding $1 - 2M^2(\tau+1)^2 e^{-\frac{N\varsigma^2}{4}}$. This completes the proof of the lemma. ∎

Lemma 2. *The spectral norm of the Toeplitz-block matrix $\mathbf{X}$ described in* (8) *satisfies*

$$\|\mathbf{X}\|_2 \stackrel{def}{=} \sqrt{\lambda_{max}\left(\mathbf{X}^{\mathrm{H}}\mathbf{X}\right)} \le 26\sqrt{\tau+1}\left(1 + \sqrt{\frac{M}{N}}\right) \tag{17}$$

with probability exceeding $1 - e^{-\frac{\sqrt{NM}}{8}}$.

Proof: We first recall that the spectral norm is invariant under column-interchange operations. Now define $\boldsymbol{\Phi} \stackrel{def}{=} \begin{bmatrix}\mathbf{x}_1 & \dots & \mathbf{x}_M\end{bmatrix}$ and $\boldsymbol{\Psi} \stackrel{def}{=} \begin{bmatrix}\boldsymbol{\Phi}_0 & \boldsymbol{\Phi}_1 & \dots & \boldsymbol{\Phi}_\tau\end{bmatrix}$, where each block $\boldsymbol{\Phi}_i$ is an $(N+\tau)\times M$ matrix that is constructed by prepending and appending $\boldsymbol{\Phi}$ with i rows and $(\tau - i)$ rows of all zeros, respectively. It is then easy to see that $\|\mathbf{X}\|_2 = \|\boldsymbol{\Psi}\|_2$ and $\|\boldsymbol{\Phi}_0\|_2 = \dots = \|\boldsymbol{\Phi}_\tau\|_2 = \|\boldsymbol{\Phi}\|_2$. Further, note that we can write for any $M(\tau+1)$-dimensional vector $\mathbf{z} = \begin{bmatrix}\mathbf{z}_0^{\mathrm{T}} & \mathbf{z}_1^{\mathrm{T}} & \dots \mathbf{z}_\tau^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$

$$\frac{\|\boldsymbol{\Psi}\mathbf{z}\|_2}{\|\mathbf{z}\|_2} \stackrel{(a)}{\le} \frac{\sum_{i=0}^{\tau}\|\boldsymbol{\Phi}_i\mathbf{z}_i\|_2}{\|\mathbf{z}\|_2} \le \frac{\|\boldsymbol{\Phi}\|_2\sum_{i=0}^{\tau}\|\mathbf{z}_i\|_2}{\|\mathbf{z}\|_2} \stackrel{(b)}{\le} \frac{\sqrt{\tau+1}\|\boldsymbol{\Phi}\|_2\|\mathbf{z}\|_2}{\|\mathbf{z}\|_2} = \sqrt{\tau+1}\|\boldsymbol{\Phi}\|_2 \tag{18}$$

where (a) follows from the definition of $\mathbf{\Psi}$ and the triangle inequality, while (b) follows from the Cauchy–Schwarz inequality. It therefore follows from the previous discussion and (18) that $\|\mathbf{X}\|_2 \leq \sqrt{\tau+1}\|\mathbf{\Phi}\|_2$.

In order to complete the proof, notice that $\mathbf{\Phi}$ is an $N \times M$ random matrix whose entries are independently distributed as $\mathsf{binary}(\pm 1/\sqrt{N})$. It can therefore be established, similar to [9, Proposition 2.4], that $\|\mathbf{\Phi}\|_2 \leq 26\left(1+\sqrt{\frac{M}{N}}\right)$ with probability exceeding $1-e^{-\frac{\sqrt{NM}}{8}}$. ■

Note that Lemma 1 implies that the event

$$\mathcal{G}_1 = \left\{\mu(\mathbf{X}) \leq \sqrt{\frac{12\log\big(M(\tau+1)\big)}{N}}\right\} \tag{19}$$

holds with probability exceeding $1-2\big(M(\tau+1)\big)^{-1}$. Similarly, Lemma 2 implies that the event

$$\mathcal{G}_2 = \left\{\|\mathbf{X}\|_2 \leq 52\sqrt{\frac{M(\tau+1)}{N}}\right\} \tag{20}$$

holds with probability exceeding $1-e^{-\frac{\sqrt{NM}}{8}}$. The rest of the analysis in this section is carried out by implicitly conditioning on these two events.

A. Invertibility Condition

In order to establish the invertibility condition, we will make use of the following proposition from [10].

Proposition 1 ([10])**.** *Fix* $q = 2\log\big(M(\tau+1)\big)$ *and define the block coherence*

$$\mu_B(\mathbf{X}) \stackrel{def}{=} \max_{1\leq i,j\leq M} \|\mathbf{X}_i^H\mathbf{X}_j - 1_{\{i=j\}}\mathbf{I}\|_2.$$

Then, for $\mathbb{E}_q Z \stackrel{def}{=} [\mathbb{E}|Z|^q]^{1/q}$ *and* $\delta \stackrel{def}{=} k/M$*, we have the following bound*

$$\mathbb{E}_q\|\mathbf{X}_\mathcal{B}^{\mathrm{H}}\mathbf{X}_\mathcal{B} - \mathbf{I}\|_2 \leq 20\mu_B(\mathbf{X})\log\big(M(\tau+1)\big) + \delta\|\mathbf{X}\|_2^2 + 9\sqrt{\delta\log\big(M(\tau+1)\big)\big(1+\tau\mu(\mathbf{X})\big)}\|\mathbf{X}\|_2. \tag{21}$$

Now note that, since we are conditioning on $\mathcal{G}_1$ and $\mathcal{G}_2$, it follows from (11), (12), (19), and (20) that

$$\mu(\mathbf{X}) \leq \frac{1}{c'(\tau+1)\log\big(M(\tau+1)\big)}, \quad \text{and} \tag{22}$$

$$\|\mathbf{X}\|_2^2 \leq \frac{1}{c'\delta\log\big(M(\tau+1)\big)} \tag{23}$$

for $c' \stackrel{def}{=} 6000$ as long as the constants c_1 and c_2 in (11) and (12) are appropriately chosen. It therefore follows from the definition of the block coherence, (22), and the linear algebra fact $\|\cdot\|_2 \leq \sqrt{\|\cdot\|_1\|\cdot\|_\infty}$ [11] that

$$\mu_B(\mathbf{X}) \leq \frac{1}{c'\log\big(M(\tau+1)\big)}. \tag{24}$$

Consequently, substituting (22), (23), and (24) into (21) yields $\mathbb{E}_q\|\mathbf{X}_\mathcal{B}^{\mathrm{H}}\mathbf{X}_\mathcal{B} - \mathbf{I}\|_2 < \frac{1}{4}$.

Finally, notice that $\mathbf{X}_\mathcal{S}$ is a submatrix of $\mathbf{X}_\mathcal{B}$ and therefore we trivially have that $\|\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S} - \mathbf{I}\|_2 \leq \|\mathbf{X}_\mathcal{B}^{\mathrm{H}}\mathbf{X}_\mathcal{B} - \mathbf{I}\|_2$. It can then be easily seen from the Markov inequality that

$$\Pr(\|\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S} - \mathbf{I}\|_2 > 1/2) \leq 2^q(\mathbb{E}_q\|\mathbf{X}_\mathcal{B}^{\mathrm{H}}\mathbf{X}_\mathcal{B} - \mathbf{I}\|_2)^q \stackrel{(a)}{\leq} \big(M(\tau+1)\big)^{-2\log 2} \tag{25}$$

where (a) follows from the fact that $\mathbb{E}_q\|\mathbf{X}_\mathcal{B}^{\mathrm{H}}\mathbf{X}_\mathcal{B} - \mathbf{I}\|_2 < \frac{1}{4}$. We have now established that $\|\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S}\|_2 \in [1/2, 3/2]$ with high probability; that is, $\|(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}\|_2 > 2$ with probability

$$\Pr(\mathcal{C}_1^c\big|\mathcal{G}_1, \mathcal{G}_2) \leq \big(M(\tau+1)\big)^{-2\log 2}. \tag{26}$$

B. Noise Stability

In order to establish the noise-stability condition, we first condition on $\mathcal{C}_1$ (the invertibility condition). Next, we denote the j-th column of $\mathbf{X}_\mathcal{S}(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}$ by $\mathbf{z}_j$ and note that

$$\|(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{w}\|_\infty = \max_{1\leq j\leq|\mathcal{S}|}|\langle\mathbf{z}_j, \mathbf{w}\rangle|. \tag{27}$$

Further, since the noise vector $\mathbf{w}$ is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I})$, we also have that $\langle\mathbf{z}_j, \mathbf{w}\rangle \sim \mathcal{CN}(0, \|\mathbf{z}_j\|_2^2)$. Finally, note that conditioned on $\mathcal{C}_1$, we have the upper bound

$$\|\mathbf{z}_j\|_2 \leq \|\mathbf{X}_\mathcal{S}(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}\|_2 \leq \|\mathbf{X}_\mathcal{S}\|_2\|(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}\|_2 \leq \sqrt{2}.$$

The rest of the argument now follows easily from bounds on the maximum of a collection of arbitrary (complex) Gaussian random variables. Specifically, it can be seen from the previous discussion and [12, Lemma 6] that

$$\Pr\left(\|(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{w}\|_\infty \geq \sqrt{2}t\big|\mathcal{C}_1\right) \leq \frac{4Me^{-t^2/2}}{\sqrt{2\pi}t}.$$

We substitute $t = \lambda/\sqrt{2}$ in the above expression to obtain

$$\frac{4Me^{-\lambda^2/4}}{\sqrt{\pi}\lambda} = \frac{2}{M(\tau+1)\sqrt{2\pi\log(M\sqrt{\tau+1}\,)}}.$$

Summarizing, we have that the noise stability condition fails to hold with probability at most

$$\Pr(\mathcal{C}_2^c\big|\mathcal{C}_1) \leq \frac{2}{M(\tau+1)\sqrt{2\pi\log(M\sqrt{\tau+1}\,)}}. \tag{28}$$

C. Complementary Noise Stability

In order to establish the complementary noise-stability condition, we use ideas similar to the ones used in the previous section. To begin with, we again condition on the event $\mathcal{C}_1$ and use $\mathbf{P}_{\mathbf{X}_\mathcal{S}} \stackrel{def}{=} \mathbf{X}_\mathcal{S}(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}\mathbf{X}_\mathcal{S}^{\mathrm{H}}$ to denote the orthogonal projector onto the column span of $\mathbf{X}_\mathcal{S}$. Next, we use $\mathbf{z}_j$ to denote the j-th column of $(\mathbf{I}-\mathbf{P}_{\mathbf{X}_\mathcal{S}})\mathbf{X}_{\mathcal{S}^c}$ and note that

$$\|\mathbf{X}_{\mathcal{S}^c}^{\mathrm{H}}(\mathbf{I}-\mathbf{P}_{\mathbf{X}_\mathcal{S}})\mathbf{w}\|_\infty = \max_{1\leq j\leq|\mathcal{S}^c|}|\langle\mathbf{z}_j, \mathbf{w}\rangle|. \tag{29}$$

Finally, given that $\mathbf{P}_{\mathbf{X}_\mathcal{S}}$ is a projection matrix and the columns of $\mathbf{X}$ have unit norm, we have that

$$\|\mathbf{z}_j\|_2 = \|(\mathbf{I}-\mathbf{P}_{\mathbf{X}_\mathcal{S}})\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2 \leq 1 \tag{30}$$

where $\mathbf{e}_j$ denotes the j-th canonical basis vector.

It is now easy to see that, since $\langle \mathbf{z}_j, \mathbf{w} \rangle$ is also distributed as $\mathcal{CN}(0, \|\mathbf{z}_j\|_2^2)$, we can make use of [12, Lemma 6] to obtain

$$\Pr\left(\|\mathbf{X}_{\mathcal{S}^c}^{\mathrm{H}}(\mathbf{I}-\mathbf{P}_{\mathbf{X}_\mathcal{S}})\mathbf{w}\|_\infty \geq t \big| \mathcal{C}_1\right) \leq \frac{4M(\tau+1)e^{-t^2/2}}{\sqrt{2\pi}t}.$$

We substitute $t = \lambda/\sqrt{2}$ in the above expression to obtain

$$\frac{4M(\tau+1)e^{-\lambda^2/4}}{\sqrt{\pi}\lambda} \leq \frac{2}{M\sqrt{2\pi\log(M\sqrt{\tau+1})}}. \quad (31)$$

Summarizing, we have that the complementary noise stability condition fails to hold with probability at most

$$\Pr(\mathcal{C}_3^c | \mathcal{C}_1) \leq \frac{2}{M\sqrt{2\pi\log(M\sqrt{\tau+1})}}. \quad (32)$$

D. Size Condition

In order to establish the size condition, we first write

$$\begin{aligned}
&\|(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}\mathrm{sgn}(\boldsymbol{\beta}_\mathcal{S})\|_\infty \\
&\quad \overset{(a)}{\leq} \|\big((\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}-\mathbf{I}\big)\mathrm{sgn}(\boldsymbol{\beta}_\mathcal{S})\|_\infty + \|\mathrm{sgn}(\boldsymbol{\beta}_\mathcal{S})\|_\infty \\
&\quad = \|\big((\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}-\mathbf{I}\big)\mathrm{sgn}(\boldsymbol{\beta}_\mathcal{S})\|_\infty + 1 \quad (33)
\end{aligned}$$

where (a) follows from the triangle inequality. Next, we once again use $\mathbf{z}_j$ to denote the j-th column of $\big((\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}-\mathbf{I}\big)$, which simply implies that $\|\big((\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S})^{-1}-\mathbf{I}\big)\mathrm{sgn}(\boldsymbol{\beta}_\mathcal{S})\|_\infty = \max_{1\leq j\leq|\mathcal{S}|}|\langle \mathbf{z}_j, \mathrm{sgn}(\boldsymbol{\beta}_\mathcal{S})\rangle|$. Now define $\mathbf{A} = \big(\mathbf{X}_\mathcal{S}^{\mathrm{H}}\mathbf{X}_\mathcal{S}-\mathbf{I}\big)$ and condition on the event $\mathcal{C}_1$. Then it follows from the von Neumann series (cf. [6, p. 2171]) that $\|\mathbf{z}_j\|_2 \leq 2\|\mathbf{A}\mathbf{e}_j\|_2$. Further, since $\mathbf{X}_\mathcal{S}$ is a submatrix of $\mathbf{X}_\mathcal{B}$, we have $\|\mathbf{A}\mathbf{e}_j\|_2 \leq \|(\mathbf{X}_\mathcal{B}^{\mathrm{H}}\mathbf{X}_\mathcal{B}-\mathbf{I})\mathbf{e}_{j'}\|_2$, where j' is such that the j'-th column of $\mathbf{X}_\mathcal{B}$ matches the j-th column of $\mathbf{X}_\mathcal{S}$.

Finally, define the diagonal matrix $\mathbf{Q} \overset{def}{=} \mathrm{diag}(\delta_1, \ldots, \delta_M)$ with the "random activation variables" $\{\delta_i\}$ on the diagonal and define a new matrix $\mathbf{R} = \mathbf{Q}\otimes\mathbf{I}_{\tau+1}$, where $\otimes$ denotes the Kronecker product. Next, use the notation $\mathbf{H} \overset{def}{=} (\mathbf{X}^{\mathrm{H}}\mathbf{X}-\mathbf{I})$ and notice that $\|(\mathbf{X}_\mathcal{B}^{\mathrm{H}}\mathbf{X}_\mathcal{B}-\mathbf{I})\mathbf{e}_{j'}\|_2 = \|\mathbf{R}\mathbf{H}\mathbf{e}_{j''}\|_2$, where j'' is such that the j''-th column of $\mathbf{X}$ matches the j-th column of $\mathbf{X}_\mathcal{S}$. In addition, note that $\mathbf{H}$ has a block structure that can be expressed as follows

$$\mathbf{H} = \begin{bmatrix}\mathbf{H}_1 & \mathbf{H}_2 & \ldots & \mathbf{H}_M\end{bmatrix} = \begin{bmatrix}\mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \ldots & \mathbf{H}_{1,M}\\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \ldots & \mathbf{H}_{2,M}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{H}_{M,1} & \mathbf{H}_{M,2} & \ldots & \mathbf{H}_{M,M}\end{bmatrix} \quad (34)$$

where $\mathbf{H}_{i,j} = \mathbf{X}_i^{\mathrm{H}}\mathbf{X}_j - 1_{\{i=j\}}\mathbf{I}$, $1 \leq i,j \leq M$, and $\mathbf{H}_i = [\mathbf{H}_{1,i}^H \ \ldots \ \mathbf{H}_{M,i}^H]^H$. We now define two blockwise norms on $\mathbf{H}$ as follows:

- $\|\mathbf{H}\|_{B,1} \overset{def}{=} \max_{1\leq i\leq M}\|\mathbf{H}_i\|_2$, and
- $\|\mathbf{H}\|_{B,2} \overset{def}{=} \max_{1\leq i,j\leq M}\|\mathbf{H}_{i,j}\|_2$.

Then it follows from the preceding discussion and the structure of the block matrix $\mathbf{H}$ that

$$\|\mathbf{z}_j\|_2 \leq 2\|\mathbf{A}\mathbf{e}_j\|_2 \leq 2\|\mathbf{R}\mathbf{H}\mathbf{e}_{j''}\|_2 \leq 2\|\mathbf{R}\mathbf{H}\|_{B,1}. \quad (35)$$

Our next goal then is to provide a bound on $\|\mathbf{R}\mathbf{H}\|_{B,1}$ and for this we resort to [10, Lemma 5].

Proposition 2 ([10]). *For $q \geq 2\log M$ and $\delta = k/M$, we have that*

$$\mathbb{E}_q\|\mathbf{R}\mathbf{H}\|_{B,1} \leq 2^{1.5}\sqrt{q}\|\mathbf{H}\|_{B,2} + \sqrt{\delta}\|\mathbf{H}\|_{B,1}. \quad (36)$$

Now notice from the definition of $\mathbf{H}$ and $\|\cdot\|_{B,2}$ that $\|\mathbf{H}\|_{B,2} \equiv \mu_B(\mathbf{X}) \leq (\tau+1)\mu(\mathbf{X})$. In addition, we have from the definition of $\mathbf{H}$ and $\|\cdot\|_{B,1}$ that

$$\begin{aligned}
\|\mathbf{H}\|_{B,1} &\overset{(b)}{\leq} \max_{1\leq i\leq M}\|\mathbf{X}^{\mathrm{H}}\mathbf{X}_i\|_2 + \|\mathbf{I}_{\tau+1}\|_2\\
&\overset{(c)}{\leq} \sqrt{1+\tau\mu(\mathbf{X})}\|\mathbf{X}\|_2 + 1 \quad (37)
\end{aligned}$$

where (b) follows from the definition of the spectral norm and the triangle inequality, while (c) mainly follows from the fact that $\|\mathbf{X}_i\|_2 \leq \sqrt{1+\tau\mu(\mathbf{X})}$ because of the Geršgorin disc theorem [11]. We can now fix $q = 2\log M$ and make use of the above bounds to conclude from Proposition 2 that

$$\begin{aligned}
\mathbb{E}_q\|\mathbf{R}\mathbf{H}\|_{B,1} \leq 4(\tau+1)\mu(\mathbf{X})\sqrt{\log M} + \\
+ \sqrt{\delta(1+\tau\mu(\mathbf{X}))}\|\mathbf{X}\|_2 + \sqrt{\delta}. \quad (38)
\end{aligned}$$

We can now substitute (22) and (23) into the above expression to obtain $\mathbb{E}_q\|\mathbf{R}\mathbf{H}\|_{B,1} \leq \gamma_0$ with

$$\begin{aligned}
\gamma_0 \overset{def}{=} &\frac{4}{c'\sqrt{\log(M(\tau+1))}} + \\
&+ \frac{2}{\sqrt{c'\log(M(\tau+1))}}\sqrt{1+\frac{1}{c'\log(M(\tau+1))}}. \quad (39)
\end{aligned}$$

In order to establish the size condition, we now define the event $\mathcal{E} = \{\max_{1\leq j\leq|\mathcal{S}|}\|\mathbf{z}_j\|_2 < \gamma\}$ and make use of the Markov inequality along with (35) and the preceding discussion to obtain

$$\begin{aligned}
\Pr(\mathcal{E}^c) &\leq \gamma^{-q}\big[\mathbb{E}_q \max_{1\leq j\leq|\mathcal{S}|}\|\mathbf{z}_j\|_2\big]^q\\
&\leq \left(\frac{2}{\gamma}\mathbb{E}_q\|\mathbf{R}\mathbf{H}\|_{B,1}\right)^q \leq \left(\frac{2\gamma_0}{\gamma}\right)^q. \quad (40)
\end{aligned}$$

Finally, we use $Z \overset{def}{=} \max_{1\leq j\leq|\mathcal{S}|}|\langle\mathbf{z}_j, \mathrm{sgn}(\boldsymbol{\beta}_\mathcal{S})\rangle|$ and conclude that

$$\begin{aligned}
\Pr(Z \geq t) &\leq \Pr(Z \geq t|\mathcal{E}) + \Pr(\mathcal{E}^c)\\
&\overset{(d)}{\leq} 2Me^{-t^2/2\gamma^2} + (2\gamma_0/\gamma)^q \quad (41)
\end{aligned}$$

where (d) is a consequence of the complex Bernstein inequality [13, Proposition 16] and the union bound. The condition is now established from (33) by setting $t = 2$ in the above expression. Further, set

$$\gamma \leq \sqrt{\frac{2}{(1+2\log 2)\log M}} \quad (42)$$

which leads to $2Me^{-2/\gamma^2} \leq 2M^{-2\log 2}$ and

$$\frac{\gamma_0}{\gamma} \leq \frac{2(\sqrt{1+c'}+2)}{0.9155c'} < 1/4. \tag{43}$$

Therefore, we obtain that $\Pr(\mathcal{E}^c) \leq (1/2)^q \leq M^{-2\log 2}$ and thus we have that the size condition does not hold with probability at most

$$\Pr(\mathcal{C}_4^c | \mathcal{C}_1) \leq 3M^{-2\log 2}. \tag{44}$$

E. Complementary Size Condition

In order to establish the complementary size condition, we proceed similar to the case of the "size condition" and define the vector $\mathbf{z}_j$ as $\mathbf{z}_j = (\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j$. It can then be easily seen that $\|\mathbf{X}_{\mathcal{S}^c}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\mathrm{sgn}(\boldsymbol{\beta}_{\mathcal{S}})\|_\infty = \max_{1\leq j\leq|\mathcal{S}^c|}|\langle \mathbf{z}_j, \mathrm{sgn}(\boldsymbol{\beta}_{\mathcal{S}})\rangle|$. Now condition on the event $\mathcal{C}_1$ and notice that $\|\mathbf{z}_j\|_2 \leq 2\|\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2$, $j=1,\dots,|\mathcal{S}^c|$.

We now define $\mathbf{X}_{\mathcal{B}^c} \overset{def}{=} [\mathbf{X}_i : i \in \mathcal{I}^c]$ and consider the set of indices $\mathcal{T}_1 \overset{def}{=} \{j' : \mathbf{X}_{\mathcal{S}^c}\mathbf{e}_{j'} \in \mathbf{X}_{\mathcal{B}^c}\}$. It is then easy to argue by making use of the notation developed in Section IV-D that if $j \in \mathcal{T}_1$ then

$$\begin{aligned}\|\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2 &\leq \max_{i\in\mathcal{I}^c}\|\mathbf{X}_{\mathcal{B}}^{\mathrm{H}}\mathbf{X}_i\|_2 \\ &= \|\mathbf{X}_{\mathcal{B}}^{\mathrm{H}}\mathbf{X}_{\mathcal{B}^c}\|_{B,1} \overset{(a)}{\leq} \|\mathbf{RH}\|_{B,1}\end{aligned} \tag{45}$$

where (a) follows from the fact that $\mathbf{X}_{\mathcal{B}}^{\mathrm{H}}\mathbf{X}_{\mathcal{B}^c}$ is a submatrix of $\mathbf{RH}$. We therefore have from the discussion following Proposition 2 and the Markov inequality that $\forall\, j \in \mathcal{T}_1$ and for $q = 2\log M$ and $\gamma > 0$

$$\Pr(\|\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2 > \gamma) \leq \frac{[\mathbb{E}_q\|\mathbf{RH}\|_{B,1}]^q}{\gamma^q} \leq \left(\frac{\gamma_0}{\gamma}\right)^q. \tag{46}$$

Finally, the argument involving $j \in \mathcal{T}_1^c$ is a little more involved but follows along similar lines. Specifically, fix any $j \in \mathcal{T}_1^c$ and define $i' \in \mathcal{I}$ to be such that $\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j$ is a column of $\mathbf{X}_{i'}$. Next, define $\tilde{\mathbf{x}}_{\mathcal{S}\cap i'}$ to be the column of $\mathbf{X}_{\mathcal{S}}$ that lies within the Toeplitz block $\mathbf{X}_{i'}$ and $\widetilde{\mathbf{X}}_{\mathcal{S}\setminus i'}$ to be the submatrix constructed by removing the column $\tilde{\mathbf{x}}_{\mathcal{S}\cap i'}$ from $\mathbf{X}_{\mathcal{S}}$. Then, if we use the notation $\mathbf{X}_{\mathcal{B}\setminus i'} \overset{def}{=} [\mathbf{X}_i : i \in \mathcal{B}\setminus\{i'\}]$, it can be verified that for any $j \in \mathcal{T}_1^c$ we have

$$\begin{aligned}\|\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2^2 &= \|\widetilde{\mathbf{X}}_{\mathcal{S}\setminus i'}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2^2 + |\tilde{\mathbf{x}}_{\mathcal{S}\cap i'}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j|^2 \\ &\leq \max_{i'\in\mathcal{I}}\|\mathbf{X}_{\mathcal{B}\setminus i'}^{\mathrm{H}}\mathbf{X}_{i'}\|_2^2 + \mu^2(\mathbf{X}) \\ &\overset{(b)}{\leq} \|\mathbf{RH}\|_{B,1}^2 + \mu^2(\mathbf{X})\end{aligned} \tag{47}$$

where (b) again makes use of the fact that the spectral norm of a matrix is an upper bound for the spectral norm of any of its submatrices. We therefore once again obtain from the discussion following Proposition 2 and the Markov inequality that $\forall\, j \in \mathcal{T}_1^c$ and for $q = 2\log M$ and $\gamma > 0$

$$\begin{aligned}\Pr(\|\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2 > \gamma) &\leq \Pr\left(\|\mathbf{RH}\|_{B,1} > \sqrt{\gamma^2 - \mu^2(\mathbf{X})}\right) \\ &\leq \left(\frac{\gamma_0}{\sqrt{\gamma^2-\mu^2}}\right)^q.\end{aligned} \tag{48}$$

We can now define the event $\mathcal{E} = \{\|\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}^c}\mathbf{e}_j\|_2 \leq \gamma\}$ and use the notation $Z \overset{def}{=} \max_{1\leq j\leq|\mathcal{S}^c|}|\langle \mathbf{z}_j, \mathrm{sgn}(\boldsymbol{\beta}_{\mathcal{S}})\rangle|$ to conclude from (46) and (48) that

$$\begin{aligned}\Pr(Z \geq t) &\leq \Pr(Z \geq t|\mathcal{E}) + \Pr(\mathcal{E}^c) \\ &\overset{(c)}{\leq} 2M(\tau+1)e^{-t^2/8\gamma^2} + (\gamma_0/\gamma)^q + \\ &\quad + (\gamma_0/\sqrt{\gamma^2-\mu^2(\mathbf{X})})^q\end{aligned} \tag{49}$$

where (c) follows from [13, Proposition 16] and the union bound. The condition is now established by setting $t = \frac{1}{4}$ in the above expression. Further, set

$$\gamma \leq \frac{1}{\sqrt{128(1+2\log 2)\log(M(\tau+1))}} \tag{50}$$

which yields $2M(\tau+1)e^{-1/128\gamma^2} \leq 2(M(\tau+1))^{-2\log 2}$ and

$$\frac{\gamma_0}{\sqrt{\gamma^2-\mu^2}} \leq \frac{\frac{2\sqrt{1+c'}}{c'} + \frac{4}{c'}}{\sqrt{0.0572^2 - 1/c'^2}} < 1/2.$$

Therefore, we obtain that $\Pr(\mathcal{E}^c) \leq 2(\gamma_0/\sqrt{\gamma^2-\mu^2})^q \leq 2(1/2)^q \leq 2(M(\tau+1))^{-2\log 2}$ and thus we have that the size condition does not hold with probability at most

$$\Pr(\mathcal{C}_5^c|\mathcal{C}_1) \leq 4\big(M(\tau+1)\big)^{-2\log 2}. \tag{51}$$

F. Proof of Theorem 1

The proof of Theorem 1 follows from the preceding discussion by taking a union bound over all the respective conditions and removing the conditionings: $\Pr((\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4 \cap \mathcal{C}_5)^c) \leq 2\Pr(\mathcal{G}_1^c) + 2\Pr(\mathcal{G}_2^c) + 2\Pr(\mathcal{C}_1^c|\mathcal{G}_1,\mathcal{G}_2) + \Pr(\mathcal{C}_2^c|\mathcal{C}_1) + \Pr(\mathcal{C}_3^c|\mathcal{C}_1) + \Pr(\mathcal{C}_4^c|\mathcal{C}_1) + \Pr(\mathcal{C}_5^c|\mathcal{C}_1)$. Consequently, we obtain that the probability of error is upper bounded by $13\big(M(\tau+1)\big)^{-2\log 2} + 4M^{-1}\big(2\pi\log(M\sqrt{\tau+1})\big)^{-1/2} + 2e^{-\frac{\sqrt{NM}}{8}}$.

V. Numerical Results and Discussion

We conclude this paper by making use of Monte Carlo simulations to validate and discuss the results reported in here. The simulation setup corresponds to a total of $M = 3000$ users communicating to the base station using codewords of length $N = 1024$ that are drawn independently from a binary$(\pm 1/\sqrt{N}, \mathbf{I}_M)$ distribution. In the following, user activity is generated using independent 0–1 Bernoulli random variables $\{\delta_i\}$ such that $\Pr(\delta_i = 1) = k/M$ for a given k. Further, for a given maximum delay τ, the individual user delays $\{\tau_i\}$ are generated once at random for each experiment and then fixed for the remainder of the experiment in keeping with the fact that our results hold uniformly over all possible $\{\tau_i\}$. The implementation of Algorithm 1 uses the SpaRSA package [5] in order to solve (LASSO) and includes the modifications described in Section III-B for speeding up the matrix–vector multiplications $\mathbf{Xb}$ and $\mathbf{X}^{\mathrm{H}}\mathbf{y}$. Finally, we use the performance metric of *average number of detection errors* in our simulations, which corresponds to the cardinality of the set $(\mathcal{I}\setminus\widehat{\mathcal{I}}) \cup (\widehat{\mathcal{I}}\setminus\mathcal{I})$ averaged over the independent trials.

The first numerical experiment that we carry out corresponds to studying the performance of the proposed MUD

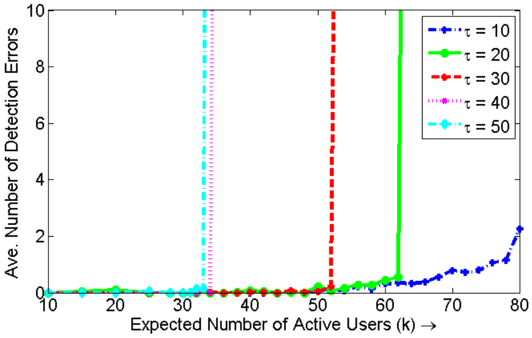

Fig. 1. Performance of Algorithm 1 as a function of the expected number of active users k for five different values of the maximum delay τ.

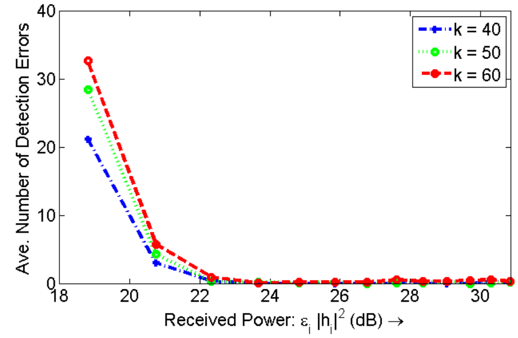

Fig. 2. Performance of Algorithm 1 as a function of the instantaneous received SNRs of the active users for $\tau = 19$ and three different values of k.

scheme as a function of the expected number of active users k in the system. In this experiment, it is assumed that the users know the magnitudes of their respective channel fading coefficients $|h_i|$ and control their powers so that the transmit power requirement described in (10) is satisfied. The results of this experiment are reported in Fig. 1 for five different values of the maximum delay τ. There are two important remarks that can be made concerning Fig. 1. First, note that Theorem 1 suggests that—for appropriate values of k and τ—the product τk should be approximately constant in order for the proposed scheme to successfully carry out MUD (cf. (12)). In this regard, it can be seen from Fig. 1 that indeed the points at which the curves begin to diverge from the horizontal axis tend to be when $\tau k \approx 1500$. This suggests that the scaling relationship described by (12) in Theorem 1 is accurate.

Second, Fig. 1 also helps put the novelty of this work into perspective. Specifically, note that one could have simply considered the problem of MUD in asynchronous on–off random access channels in the context of the recent literature on compressed sensing. Indeed, the expression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{w}$ in (8) describes the problem of recovering a sparse signal $\boldsymbol{\beta}$ from linear measurements in the presence of noise. However, if one were to naively apply the theory of compressed sensing to this problem then one would expect the performance of Algorithm 1 to improve with increasing τ. This is because the number of rows of the matrix $\mathbf{X}$ in (8) increases as τ is increased. However, as indicated by the results of this paper, the Toeplitz-block structure of $\mathbf{X}$ does not allow for such an improvement. On the contrary, increasing the range of possible delays poses a more difficult MUD problem, and thus the expected number of active users for which MUD succeeds decreases, as can be seen from Fig. 1.

The second numerical experiment that we carry out corresponds to studying the performance of the proposed MUD scheme as a function of the instantaneous received SNRs $\{\mathcal{E}_i|h_i|^2\}$ of the active users. The results of this experiment are reported in Fig. 2 for three different values of the expected number of active users k and with the maximum delay τ set to 19. The main conclusion that can be drawn from Fig. 2 in this regard is that the power condition of (10) in Theorem 1 is possibly overly restrictive. Specifically, note that the power condition (10) for the specified parameters reduces to $\mathcal{E}_i|h_i|^2 \gtrapprox 31$ dB for every active user. On the other hand, Fig. 2 shows that Algorithm 1 carries out successful MUD even when the instantaneous received SNRs of the active users are significantly below 31 dB. Finally, Fig. 2 also helps us verify that the theory is correct in predicting that the transmit power required for successful MUD does not depend upon the expected number of active users k.

In conclusion, simulation results confirm that our proposed scheme successfully carries out MUD in asynchronous on–off random access channels and that the theoretical guarantees provided in Theorem 1 are nearly optimal in terms of the scaling relationship between k, τ, M, and N. In the future, we plan to extend this work by designing deterministic codewords appropriate for this application and by analyzing the (detection) outage rates of individual users arising because of fading and fixed transmit power constraints.

References

[1] A. Fletcher, S. Rangan, and V. Goyal, "On-off random access channels: A compressed sensing framework," submitted [arXiv:0903.1022v2].

[2] D. Guo and J. Luo, "Compressed neighbor discovery for wireless networks," submitted [private communication].

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, pp. 267–288, 1996.

[4] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. Ser. B*, pp. 49–67, 2006.

[5] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Processing*, pp. 2479–2493, Jul. 2009.

[6] E. Candès and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Ann. Statist.*, pp. 2145–2177, Oct. 2009.

[7] W. U. Bajwa, "New information processing theory and methods for exploiting sparsity in wireless systems," Ph.D. dissertation, 2009.

[8] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, pp. 13–30, Mar. 1963.

[9] M. Rudelson and R. Vershynin, "Non-asymptotic theory of random matrices: Extreme singular values," in *Proc. Int. Congr. of Mathematicians*, Hyderabad, India, Aug. 2010.

[10] W. U. Bajwa, R. Calderbank, and M. F. Duarte, "On the conditioning of random block subdictionaries," Tech. Rep. TR-2010-06, Dept. of Computer Science, Duke University, Durham, NC, Sep. 2010.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 1st ed. Cambridge, U.K.: Cambridge University Press, 1985.

[12] W. U. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," *J. Commun. Netw.*, pp. 289–307, Aug. 2010.

[13] J. Tropp, "The random paving property for uniformly bounded matrices," *Studia Math.*, pp. 67–82, 2008.