Fundamental limits of many-user MAC with finite payloads and fading

Suhas S Kowshik and Yury Polyanskiy, *Senior Member, IEEE*

***Abstract*—Consider a (multiple-access) wireless communication system where users are connected to a unique base station over a shared-spectrum radio links. Each user has a fixed number k of bits to send to the base station, and his signal gets attenuated by a random channel gain (quasi-static fading). In this paper we consider the many-user asymptotics of Chen-Chen-Guo'2017, where the number of users grows linearly with the blocklength. Differently, though, we adopt a per-user probability of error (PUPE) criterion (as opposed to classical joint-error probability criterion). Under PUPE the finite energy-per-bit communication is possible, and we are able to derive bounds on the tradeoff between energy and spectral efficiencies. We reconfirm the curious behaviour (previously observed for non-fading MAC) of the possibility of almost perfect multi-user interference (MUI) cancellation for user densities below a critical threshold. Further, we demonstrate the suboptimality of standard solutions such as orthogonalization (i.e., TDMA/FDMA) and treating interference as noise (i.e. pseudo-random CDMA without multi-user detection). Notably, the problem treated here can be seen as a variant of support recovery in compressed sensing for the unusual definition of sparsity with one non-zero entry per each contiguous section of 2^k coordinates. This identifies our problem with that of the sparse regression codes (SPARCs) and hence our results can be equivalently understood in the context of SPARCs with sections of length 2^{100}. Finally, we discuss the relation of the almost perfect MUI cancellation property and the replica-method predictions.**

***Index Terms*—Finite blocklength, many-user MAC, per-user probability of error, approximate message passing, replica-method**

I. Introduction

We clearly witness two recent trends in the wireless communication technology: the increasing deployment density and miniaturization of radio-equipped sensors. The first trend results in progressively worsening interference environment, while the second trend puts ever more stringent demands on communication energy efficiency. This suggests a bleak picture for the future networks, where a chaos of packet collisions and interference contamination prevents reliable connectivity.

This paper is part of a series aimed at elucidating the fundamental tradeoffs in this new "dense-networks" regime of communication, and on rigorously demonstrating suboptimality of state-of-the-art radio-access solutions (ALOHA, orthogonalization, or FDMA/TDMA [1], and treating interference as noise, or TIN). This suboptimality will eventually lead to dramatic consequences. For example, environmental impact of billions of toxic batteries getting depleted at 1/10 or 1/100 of the planned service time is easy to imagine. In order to future-proof our systems, we should avoid locking in on outdated and unfixable multiple-access architectures causing tens of dB losses in energy efficiency. The information-theoretic analysis in this paper demonstrates that the latter is indeed unavoidable (with state-of-the-art schemes). However, our message is in fact optimistic, as we also demonstrate existence of protocols which are partially immune to the increase of the sensor density.

Specifically, in this paper we consider a problem of K nodes communicating over a frame-synchronized multiple-access channel. When K is fixed and the frame size n (which we will also call "blocklength" or the "number of degrees of freedom") is taken to infinity we get the classical regime [2], in which the fundamental limits are given by well-known mutual information expressions. A new regime, deemed *many-access*, was put forward by Chen, Chen and Guo [3]. In this regime the number of nodes K grows with blocklength n. It is clear that the most natural scaling is linear: $K = \mu n, n \to \infty$, corresponding to the fact that in time n there are linearly many users that will have updates/traffic to send [4]. That is, if each device wakes up once in every T seconds and transmits over a frame of length t, then in time (proportional to) t there are $K \approx t/T$ users where t is large enough for this approximation to hold but small that no device wakes up twice. Further, asymptotic results obtained from this linear scaling have been shown to approximately predict behavior of the fundamental limit at finite blocklength, e.g. at $n = 30000$ and $K <= 300$ [4, 5]. The analysis of [3] focused on the regime of infinitely large payloads (see also [6] for a related massive MIMO MAC analysis in this setting). In contrast [4] proposed to focus on a model where each of the $K = \mu n$ nodes has only finitely many bits to send. In this regime, it turned out, one gets the relevant engineering trade-offs. Namely, the communication with finite energy-per-bit is possible as $n \to \infty$ and the optimal energy-per-bit depends on the user density μ. For this to happen, however, a second crucial departure from the classical MAC model was needed: the per-user probability of error, *PUPE*, criterion [4].

This material is based upon work supported by the National Science Foundation under Grant No CCF-17-17842 and a grant from Skoltech–MIT Joint Next Generation Program (NGP). This work was also supported in part by the MIT-IBM Watson AI Lab.

S. S. Kowshik and Y. Polyanskiy are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

This paper was presented in part at the 2019 IEEE International Symposium on Information Theory (ISIT) [1].

[1] In this paper we do not distinguish between TDMA, FDMA or any other orthogonalization strategy. From here on, all orthogonalization strategies will be referred to as "TDMA"

These two modifications (the scaling $K = \mu n$ and the PUPE) were investigated in the case of the AWGN channel in [4, 5, 7]. We next describe the main discovery of that work. The channel model is:

$$Y^n = \sum_{i=1}^{K} X_i + Z^n , \qquad Z^n \sim \mathcal{CN}(0, I_n) , \tag{1}$$

and $X_i = f_i(W_i) \in \mathbb{C}^n$ is the codeword of i-th user corresponding to $W_i \in [2^k]$ chosen uniformly at random. The system is said to have PUPE ϵ if there exist decoders $\hat{W}_i = \hat{W}_i(Y^n)$ such that

$$P_{e,u} = \frac{1}{K} \sum_{i=1}^{K} \mathbb{P}\left[W_i \neq \hat{W}_i\right] \leq \epsilon . \tag{2}$$

The energy-per-bit is defined as

$$\frac{E_b}{N_0} = \frac{1}{k} \sup_{i \in [K], w \in [2^k]} \|f_i(w)\|^2 .$$

The goal in [4, 7] was to characterize the asymptotic limit

$$\mathcal{E}^*(\mu, k, \epsilon) \triangleq \limsup_{n \to \infty} \inf \frac{E_b}{N_0} \tag{3}$$

where infimum is taken over all possible encoders $\{f_i\}$ and decoders $\{\hat{W}_i\}$ achieving the PUPE ϵ for $K = \mu n$ users. (Note that this problem may be recast in the language of compressed sensing and sparse regression codes (SPARCs) – see Section II-A below.)

To predict how $\mathcal{E}^*(\mu, \epsilon)$ behaves, first consider a naive Shannon-theoretic calculation [8]: if K users want to send k bits in n degrees of freedom, then their sum-power P_{tot} should satisfy

$$n \log(1 + P_{tot}) = kK .$$

In turn, the sum-power $P_{tot} = \frac{kK}{n} \frac{E_b}{N_0}$. Overall, we get

$$\mathcal{E}^* \approx \frac{2^{\mu k} - 1}{k\mu} .$$

This turns out to be a correct prediction, but only in the large-μ regime. The true behavior of the fundamental limit is roughly given by

$$\mathcal{E}^*(\mu, k, \epsilon) \approx \max\left(\frac{2^{\mu k} - 1}{k\mu}, \mathcal{E}_{\text{s.u.}}\right) , \tag{4}$$

where $\mathcal{E}_{\text{s.u.}} = \mathcal{E}_{\text{s.u.}}(k, \epsilon)$ does not depend on μ and corresponds to the single-user minimal energy-per-bit for sending k bits with error ϵ, for which a very tight characterization is given in [9]. In particular, with good precision for $k \geq 10$ we have

$$\mathcal{E}_{\text{s.u.}}(k, \epsilon) = \frac{1}{2k} \left(\mathcal{Q}^{-1}\left(2^{-k}\right) - \mathcal{Q}^{-1}\left(1 - \epsilon\right)\right)^2 \tag{5}$$

where $\mathcal{Q}$ is the complementary CDF of the standard normal distribution: $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$.

In all, results of [4, 5, 7] suggest that the minimal energy-per-bit has a certain "inertia": as the user density μ starts to climb from zero up, initially the energy-per-bit should stay the same as in the single-user $\mu = 0$ limit. In other words, optimal multiple-access architectures should be able to *almost perfectly cancel all multi-user interference (MUI)*, achieving an essentially single-user performance for each user, *provided the user density is below a critical threshold*. Note that this is much better than orthogonalization, which achieves the same effect at the expense of shortening the available (to each user) blocklength by a factor of $\frac{1}{K}$. Quite surprisingly, standard approaches to multiple-access such as TDMA and TIN[2], while having an optimal performance at $\mu \to 0$ demonstrated a significant suboptimality for $\mu > 0$ regime. In particular, no "inertia" was observed and the energy-per-bit for those suboptimal architectures is always a monotonically increasing function of the user density μ. This opens the (so far open) quest for finding a future-proof MAC architecture that would achieve $\mathcal{E}_{\text{s.u.}}$ energy-per-bit for a strictly-positive $\mu > 0$. A thorough discussion of this curious behavior and its connections to replica-method predicted phase transitions is contained in Section V-A.

(We note that in this short summary we omitted another important part of [4]: the issue of random-access – i.e. communicating when the identities/codebooks of active users are unknown a priori. We mention, however, that for the random-access version of the problem, there are a number of low-complexity (and quite good performing) algorithms that are available [10–16]. See [17–20] for more recent developments.)

The *contribution* of this paper is in demonstrating the same almost perfect MUI cancellation effect in a much more practically relevant communication model, in which the ideal unit power-gains of (1) are replaced by random (but static) fading gain coefficients. We consider two cases of the channel state information: known at the receiver (CSIR) and no channel state information (noCSI).

Key technical ideas: For handling the noCSI case we employ the subspace projection decoder similar to the one proposed in [21], which can be seen as a version of the maximum-likelihood decoding (without prior on fading coefficients) – an idea often used in support recovery literature [22–24]. Another key idea is to decode only a subset of users corresponding to the strongest channel gains – a principle originating from Shamai-Bettesh [25]. While the randomness of channel gains increases the energy-per-bit requirements, in a related paper we find [10] an unexpected advantage: the inherent randomization helps the decoder disambiguate different users and improves performance of the belief propagation decoder. Our second achievability bound improves projection decoder in the the low user density (low spectral efficiency) regime by applying the Approximate Message Passing (AMP) algorithm [26]. The rigorous analysis of its performance is made possible by results in [24, 27]. On the converse side, we leverage the recent finite blocklength results for the noCSI channel from [21, 28].

The paper is organized as follows. In Section II we formally define the problem and the fundamental limits. In Section II-A relation with compressed sensing is discussed. In Section III as a warm-up we discuss the classical regime (K–fixed, $n \to \infty$) under the PUPE criterion. We show that our projection de-

[2]Note that pseudo-random CDMA systems without multi-user detection and large load factor provide an efficient implementation of TIN. So throughout our discussions, conclusions about TIN also pertain to CDMA systems of this kind.

coder achieves the best known achievability bound in this setting [25]. (We also note that for the quasi-static fading channel model the idea of PUPE is very natural, and implicitly appears in earlier works, e.g. [25, 29], where it is conflated with the outage probability.) After this short warm-up we go to our main Section IV, which contains rigorous achievability and converse bounds for the $K = \mu n, n \to \infty$ scaling regime. Some numerical evaluations are presented Section V, where we also compare our bounds with the TDMA and TIN. Finally, in Section V-A we discuss the effect of almost perfect MUI cancellation and its relation to other phase transitions in compressed sensing.

A. Notations

Let $\mathbb{N}$ denote the set of natural numbers. For $n \in \mathbb{N}$, let $\mathbb{C}^n$ denote the n–dimensional complex Euclidean space. Let $S \subset \mathbb{C}^n$. We denote the projection operator or matrix on to the subspace *spanned* by S as P_S and its orthogonal complement as $P_S^{\perp}$. For $0 \leq p \leq 1$, let $h_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ and $h(p) = -p\ln(p) - (1-p)\ln(1-p)$, with $0\ln 0$ defined to be 0. We denote by $\mathcal{N}(0,1)$ and $\mathcal{CN}(0,1)$ the standard normal and the standard circularly symmetric complex normal distributions, respectively. $\mathbb{P}$ and $\mathbb{E}$ denote probability measure and expectation operator respectively. For $n \in \mathbb{N}$, let $[n] = \{1, 2, ..., n\}$. log denotes logarithm to base 2. Lastly, $\|\cdot\|$ represents the standard euclidean norm.

II. System Model

Fix an integer $K \geq 1$ – the number of users. Let $\{P_{Y^n|X^n} = P_{Y^n|X_1^n, X_2^n, ..., X_K^n} : \prod_{i=1}^K \mathcal{X}_i^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$ be a multiple access channel (MAC). In this work we consider only the quasi-static fading AWGN MAC: the channel law $P_{Y^n|X^n}$ is described by

$$Y^n = \sum_{i=1}^K H_i X_i^n + Z^n \tag{6}$$

where $Z^n \sim \mathcal{CN}(0, I_n)$, and $H_i \overset{iid}{\sim} \mathcal{CN}(0,1)$ are the fading coefficients which are independent of $\{X_i^n\}$ and Z^n. Naturally, we assume that there is a maximum power constraint:

$$\|X_i^n\|^2 \leq nP. \tag{7}$$

We consider two cases: 1) no channel state information (no-CSI): neither the transmitters nor the receiver knows the realizations of channel fading coefficients, but they both know the law; 2) channel state information only at the receiver (CSIR): only the receiver knows the realization of channel fading coefficients. The special case of (6) where $H_i = 1, \forall i$ is called the Gaussian MAC (GMAC).

In the rest of the paper we drop the superscript n unless it is unclear.

Definiton 1. *An $((M_1, M_2, ..., M_K), n, \epsilon)_U$ code for the MAC $P_{Y^n|X^n}$ is a set of (possibly randomized) maps $\{f_i : [M_i] \to \mathcal{X}_i^n\}_{i=1}^K$ (the encoding functions) and $g : \mathcal{Y}^n \to \prod_{i=1}^K [M_i]$ (the decoder) such that if for $j \in [K]$, $X_j = f_j(W_j)$ constitute the input to the channel and W_j is chosen uniformly (and independently of other W_i, $i \neq j$) from $[M_j]$ then the average (per-user) probability of error satisfies*

$$P_{e,u} = \frac{1}{K}\sum_{j=1}^K \mathbb{P}\left[W_j \neq (g(Y))_j\right] \leq \epsilon \tag{8}$$

where Y is the channel output.

We define an $((M_1, M_2, ..., M_K), n, \epsilon)_J$ code similarly, where $P_{e,u}$ is replaced by the usual joint error

$$P_{e,J} = \mathbb{P}\left[\bigcup_{j\in[K]} \left\{W_j \neq (g(Y))_j\right\}\right] \leq \epsilon \tag{9}$$

Further, if there are cost constraints, we naturally modify the above definitions such that the codewords satisfy the constraints.

Remark 1. *Note that in* (8)*, we only consider the* average *per-user probability. But in some situations, it might be relevant to consider* maximal *per-user error (of a codebook tuple) which is the maximum of the probability of error of each user. Formally, let $\mathcal{C}_{[K]} = \{\mathcal{C}_1, ..., \mathcal{C}_K\}$ denote the set of codebooks. Then*

$$\begin{aligned} P_{e,u}^{\max} &= P_{e,u}^{\max}(\mathcal{C}_{[K]}) \\ &= \max\left\{\mathbb{P}\left[W_1 \neq \hat{W}_1\right], ..., \mathbb{P}\left[W_K \neq \hat{W}_K\right]\right\} \end{aligned} \tag{10}$$

where the probabilities are with respect to the channel and possibly random encoding and decoding functions. In this paper we only consider the fundamental limits with respect to $P_{e,u}$ and PUPE always refers to this unless otherwise noted. But we note here that for both asymptotics and FBL the difference is not important. See appendix C for a discussion on this – there we show that by random coding $\mathbb{E}\left[P_{e,u}^{\max}\right]$ is asymptotically equal to $\mathbb{E}\left[P_{e,u}\right]$ (expectations are over random codebooks).

A. Connection to compressed sensing and sparse regression codes

The system model and coding problem considered in this work (see eqn. (6)) can be cast as a support recovery problem in compressed sensing. Suppose we have K users each with a codebook of size M and blocklength n. Let A_i be the $n \times M$ matrix consisting of the codewords of user i as columns. Then the codeword transmitted by the user can be represented as $X_i = A_i W_i$ where $W_i \in \{0,1\}^M$ with a single nonzero entry. Since each codeword is multiplied by a scalar random gain H_i, we let $U_i = H_i W_i$ which is again a 1 sparse vector of length M. Finally the received vector Y can be represented as

$$Y = \sum_{i\in[K]} H_i X_i + Z = AU + Z \tag{11}$$

where $A = [A_1, \cdots, A_K]$ is $n \times KM$ matrix obtained by concatenating the codebooks, $U = [U_1^T, \cdots, U_K^T]^T$ is the $MK \times 1$ length vector denoting the codewords (and fading gains) of each user. In our problem, the vector U has a *block-sparse* structure, namely U has K sections, each of length M, and there is only a single non-zero entry in each section. (Majority of compressed sensing literature focuses on the *non-block-sparse* case, where U has just K non-zero entries, which can be spread arbitrarily inside KM positions.) Decoding of the codewords, then, is equivalent to the support recovery problem under the block-sparse structure, a problem considered in compressed sensing. In our setup, we keep M fixed and let $K, n \to \infty$ with constant $\mu = K/n$. Hence $1/M$ is the sparsity rate and $M\mu$ is the measurement rate.

This connection is not new and has been observed many times in the past [30, 31]. In [31] the authors consider a the exact support recovery problem in the case when the vector U is just sparse (with or without random gains). This corresponds to the random access version of our model where the users share a same codebook [32]. They analyze the fundamental limits in terms of the rate (i.e., ratio of logarithm of signal size to number of measurements) necessary and sufficient to ensure exact recovery in both cases when sparsity is fixed and growing with the signal size. For the fixed sparsity case and U having only $0, 1$ entries, this fundamental limit is exactly the symmetric capacity of an AWGN multiple access channel with same codebook (with non colliding messages). With fading gains, they recover the outage capacity of quasi-static MAC [33, 34] (but with same codebook).

In [30], the authors discuss necessary and sufficient conditions for the exact and approximate support recovery (in Hamming distortion), and L_2 signal recovery with various conditions on signal X and matrix A (deterministic versus random, discrete versus continuous support etc.). These results differ from ours in the sense that they are not for block sparse setting and more importantly, they do not consider approximate support recovery with Hamming distortion when the entries of the support of the signal are sampled from a continuous distribution, which is the case we analyze. Hence our results are not directly comparable.

Work [24] comes closest to our work in terms of the flavor of results of achievability. As pointed out in [24] itself, many other works like [30] focus on the necessary and sufficient scalings (between sparsity, measurements and signal dimension) for various forms of support recovery. But the emphasis in [24] and this work is on the precise constants associated with these scalings. In particular, the authors in [24] consider the approximate support recovery (in Hamming distortion) problem when the entries in the support of the signal come from a variety of distributions. They analyze various algorithms, including matched filter and AMP, to find the minimum measurement rate required to attain desired support distortion error in terms of signal to noise ratio and other parameters. Furthermore, they compare these results to that of the optimal decoder predicted by the replica method [35].

The result on using replica method in [24] is not directly applicable since our signal has block sparse (as opposed to i.i.d.) coordinates. But the AMP analysis presented there can be extended to our setting. Because of the generality of the analysis in [27], it turns out to be possible to derive rigorous claims (and computable expressions) on the performance of the (scalar) AMP even in the block sparse setting. This is the content of Section IV-C below. Unlike the achievability side, for the converse we cannot rely on bounds in [24] proven for the i.i.d. coordinates of X. Even ignoring the difference between the structural assumptions on X, we point out also that our converse bounds leverage finite-length results from [28], which makes them tighter than the genie-based bounds in [36].

The block-sparse assumption, however, comes very naturally in the area of SPARCs [37–39]. The section error rate (SER) of a SPARC is precisely our PUPE. The vector AMP algorithm has been analyzed for SPARC with i.i.d Gaussian design matrix in [39] and for the spatially-coupled matrix in [40] but for the AWGN channel (i.e., when non-zero entries of U in (11) are all 1). In[41], heuristic derivation of state evolution of the vector-AMP decoder for spatially-coupled SPARCS was presented for various signal classes (this includes our fading scenario). However, the the resulting fixed point equations may not be possible to solve for our block size as it amounts to computing 2^{100} dimensional integrals (and this also prevents evaluation of replica-method predictions from [41]).

III. Classical regime: K fixed, $n \to \infty$

In this section, we focus on the channel under classical asymptotics where K is fixed (and large) and $n \to \infty$. Further, we consider two distinct cases of joint error and per-user error. We show that subspace projection decoder (14) achieves a) ϵ–capacity region ($C_{\epsilon,J}$) for the joint error and b) the best known bound for ϵ–capacity region $C_{\epsilon,PU}$ under per-user error. This motivates using projection decoder in the many-user regime.

A. Joint error

A rate tuple $(R_1, ..., R_K)$ is said to be ϵ–achievable [34] for the MAC if there is a sequence of codes whose rates are asymptotically at least R_i such that joint error is asymptotically smaller than ϵ. Then the ϵ–capacity region $C_{\epsilon,J}$ is the closure of the set of ϵ–achievable rates. For our channel (6), the $C_{\epsilon,J}$ does not depend on whether or not the channel state information (CSI) is available at the receiver since the fading coefficients can be reliably estimated with negligible rate penalty as $n \to \infty$ [33][25]. Hence from this fact and using [34, Theorem 5] it is easy to see that, for $0 \le \epsilon < 1$, the ϵ–capacity region is given by

$$C_{\epsilon,J} = \{R = (R_1, ..., R_K) : \forall i, R_i \ge 0 \text{ and } P_0(R) \le \epsilon\} \tag{12}$$

where the outage probability $P_0(R)$ is given by

$$P_0(R) = \mathbb{P}\left[\bigcup_{S\subset[K], S\neq\emptyset} \left\{\log\left(1 + P\sum_{i\in S}|H_i|^2\right) \le \sum_{i\in S} R_i\right\}\right] \tag{13}$$

Next, we define a subspace projection based decoder, inspired from [21]. The idea is the following. Suppose there were no additive noise. Then the received vector will lie in the subspace spanned by the sent codewords no matter what the fading coefficients are. To formally define the decoder, let C denote a set of vectors in $\mathbb{C}^n$. Denote P_C as the orthogonal projection operator onto the subspace spanned by

C. Let $P_C^{\perp} = I - P_C$ denote the projection operator onto the orthogonal complement of $span(C)$ in $\mathbb{C}^n$.

Let $\mathcal{C}_1, ..., \mathcal{C}_K$ denote the codebooks of the K users respectively. Upon receiving Y from the channel the decoder outputs $g(Y)$ which is given by

$$g(Y) = \left(f_1^{-1}(\hat{c}_1), ..., f_K^{-1}(\hat{c}_K)\right)$$
$$(\hat{c}_1, ...\hat{c}_K) = \arg \max_{(c_i \in \mathcal{C}_i)_{i=1}^K} \left\| P_{\{c_i : i \in [K]\}} Y \right\|^2 \quad (14)$$

where f_i are the encoding functions.

In this section, we show that using spherical codebook with projection decoding, $C_{\epsilon,J}$ of the K–MAC is achievable. We prove the following theorem

Theorem III.1 (Projection decoding achieves $C_{\epsilon,J}$)**.** *Let $R \in C_{\epsilon,J}$ of* (6). *Then R is ϵ–achievable through a sequence of codes with the decoder being the projection decoder* (14).

Proof. We generate codewords iid uniformly on the power sphere and show that (14) yields a small $P_{e,J}$. See appendix A for details. □

Remark 2. *Note that [31] also analyzed capacity region of the quasi-static MAC, but under the same codebook requirement, for the joint error probability (as opposed to PUPE), and with a different decoder.*

B. Per-user error

In this subsection, we consider the case of per-user error under the classical setting. Further, we assume availability of CSI at receiver (CSIR) which again can be estimated with little penalty.

The ϵ–capacity region for the channel under per-user error, $C_{\epsilon,PU}$ is defined similarly as $C_{\epsilon,J}$ but with per-user error instead of joint error. $C_{\epsilon,PU}$ is unknown, but the best lower bound is given by the Shamai-Bettesh capacity bound [25]: given a rate tuple $R = (R_1, ..., R_K)$, an upper bound on the per-user probability of error under the channel (6), as $n \to \infty$, is given by

$$\begin{aligned} P_{e,u} &\leq P_e^S(R) \\ &\equiv 1 - \frac{1}{K}\mathbb{E} \sup \Bigg\{ |D| : D \subset [K], \forall S \subset D, S \neq \emptyset, \\ &\quad \sum_{i \in S} R_i < \log\left(1 + \frac{P\sum_{i\in S}|H_i|^2}{1 + P\sum_{i \in D^c}|H_i|^2}\right)\Bigg\} \quad (15) \end{aligned}$$

where the maximizing set, among all those that achieve the maximum, is chosen to contain the users with largest fading coefficients. The corresponding achievability region is

$$C_{\epsilon,PU}^{S.B} = \left\{R : P_e^S(R) \leq \epsilon\right\} \quad (16)$$

and hence it is an inner bound on $C_{\epsilon,PU}$.

We note that, in [25], only the symmetric rate case i.e, $R_i = R_j \,\forall i,j$ is considered. So (15) is the extension of that result to the general non-symmetric case.

Here, we show that the projection decoding (suitably modified to use CSIR) achieves the same asymptotics as (15) for per-user probability of error i.e., achieves the Shamai-Bettesh capacity bound. Next we describe the modification to the projection decoder to use CSIR.

Let $\{\mathcal{C}_i\}_{i=1}^K$ denote the codebooks of the K users with $|\mathcal{C}_i| = M_i$. We have a maximum power constraint given by (7). Using the idea of joint decoder from [25], our decoder works in 2 stages. The first stage finds the following set

$$\begin{aligned} D \in \arg\max \Bigg\{ &|D| : D \subset [K], \forall S \subset D, S \neq \emptyset, \\ &\sum_{i\in S} R_i < \log\left(1 + \frac{P\sum_{i\in S}|H_i|^2}{1 + P\sum_{i\in D^c}|H_i|^2}\right)\Bigg\} \quad (17) \end{aligned}$$

where D is chosen to contain users with largest fading coefficients. The second stage is similar to (14) but decodes only those users in D. Formally, let ? denote an error symbol. The decoder output $g_D(Y) \in \prod_{i=1}^K \mathcal{C}_i$ is given by

$$(g_D(Y))_i = \begin{cases} f_i^{-1}(\hat{c}_i) & i \in D \\ ? & i \notin D \end{cases}$$
$$(\hat{c}_i)_{i\in D} = \arg \max_{(c_i \in \mathcal{C}_i)_{i \in D}} \left\| P_{\{c_i : i \in D\}} Y \right\|^2 \quad (18)$$

where f_i are the encoding functions. Our error metric is the average per-user probability of error (9).

The following theorem is the main result of this section.

Theorem III.2. *For any $R \in C_{\epsilon,PU}^{S.B}$ there exists a sequence of codes with projection decoder* (17)(18) *with asymptotic rate R such that the per-user probability of error is asymptotically smaller than ϵ*

Proof. We generate iid (complex) Gaussian codebooks $\mathcal{CN}(0, P' I_n)$ with $P' < P$ and show that for $R \in C_{\epsilon,PU}^{S.B}$, (18) gives small $P_{e,u}$. See appendix A for details. □

In the case of symmetric rate, an outer bound on $C_{\epsilon,PU}$ can be given as follows.

Proposition 1. *If the symmetric rate R is such that $P_{e,u} \leq \epsilon$, then*

$$\begin{aligned} R \leq \min\Bigg\{ &\frac{1}{K(\theta - \epsilon)}\mathbb{E}\left[\log_2\left(1 + P \min_{\substack{S\subset[K] \\ |S| = \theta K}} \sum_{i\in S}|H_i|^2\right)\right], \\ &\log_2\left(1 - P\ln(1-\epsilon)\right)\Bigg\}, \forall \theta \in (\epsilon, 1] \quad (19) \end{aligned}$$

Proof. The first of the two terms in the min in (19) follows from Fano's inequality (see (136), with $\mu = K/n$, $M = 2^{nR}$ and taking $n \to \infty$). The second is a single-user based converse using a genie argument. See appendix A-C for details. □

Remark 3. *We note here that the second term inside the minimum in* (19) *is the same as the one we would obtain if we used strong converse for the MAC. To be precise, let $\{|H_{(1)}| > |H_{(2)}| > ... > |H_{(K)}|\}$ denote the order statistics of the fading coefficients. If $R > \log(1 + P|H_{(t)}|^2)$ then, using a Genie that reveals the codewords (and fading gains) of $t-1$ users corresponding to $t-1$ largest fading coefficients, it can be seen that $P_{e,u} \geq \frac{K-t+1}{K}$. Setting $t = \theta K$ and* *considering the limit as $K \to \infty$ (with $P = P_{tot}/K$) we obtain $S \leq -P_{tot} \log_2(1-\epsilon)$ which is same as that obtained from the second term in* (19) *under these limits.*

C. Numerical evaluation

First notice that $C_{\epsilon,J}$ (under joint error) tends to $\{0\}$ as $K \to \infty$ because, it can be seen, for the symmetric rate, by considering that order statistics of the fading coefficients that $P_0(R) \to 1$ for $R_i = O(1/K)$. $C_{\epsilon,PU}$, however, is more interesting. We evaluate trade-off between system spectral efficiency and the minimum energy-per-bit required for a target per-user error for the symmetric rate, in the limit $K \to \infty$ and power scaling as $O(1/K)$.

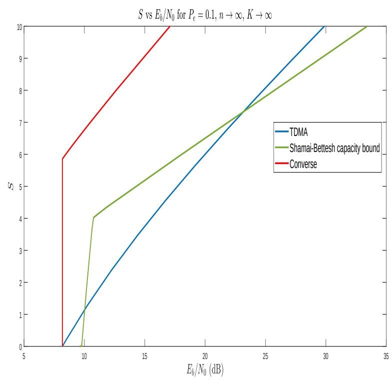

Fig. 1: S vs E_b/N_0 for per-user error $\epsilon \leq 0.1$, $n \to \infty$ and then $K \to \infty$

In the above figure we have also presented the performance of TDMA. That is, if we use orthogonalization then for any number of users K (not necessarily large), we have

$$\epsilon = \mathbb{P}\left[R > 1/K \log(1 + KP|H|^2)\right] \tag{20}$$

where ϵ is the PUPE. Thus the sum-rate vs E_b/N_0 formula for orthogonalization is

$$E_b/N_0 = \frac{2^S - 1}{S} \frac{1}{-\ln(1-\epsilon)} \tag{21}$$

where S is the sum-rate or the spectral efficiency.

We see that orthogonalization is suboptimal under the PUPE criterion. The reason is that it fails to exploit the multi-user diversity by allocating resources even to users in deep fades. Indeed, under orthogonalized setting the resources allocated to a user that happens to experience a deep fade become completely wasted, while non-orthogonal schemes essentially adapt to the fading realization: the users in deep fades create very little interference for the problem of decoding strong users. This is the effect stemming from the PUPE criterion for error rate.

IV. Many user MAC: $K = \mu n$, $n \to \infty$

This is our main section. We consider the linear scaling regime where the number of users K scales with n, and $n \to \infty$. We are interested in the tradeoff of minimum E_b/N_0 required for the PUPE to be smaller than ϵ, with the user density μ ($\mu < 1$). So, we fix the message size k. Let $S = k\mu$ be the spectral efficiency.

We focus on the case of different codebooks, but under symmetric rate. So if M denotes the size of the codebooks, then $S = \frac{K \log M}{n} = \mu \log M$. Hence, given S and μ, M is fixed. Let $P_{tot} = KP$ denote the total power. Therefore denoting by $\mathcal{E}$ the energy-per-bit, $\mathcal{E} = E_b/N_0 = \frac{nP}{\log_2 M} = \frac{P_{tot}}{S}$. For finite E_b/N_0, we need finite P_{tot}, hence we consider the power P decaying as $O(1/n)$.

Let $\mathcal{C}_j = \{c_1^j, ..., c_M^j\}$ be the codebook of user j, of size M. The power constraint is given by $\left\|c_i^j\right\|^2 \leq nP = \mathcal{E} \log_2 M, \forall j \in [K], i \in [M]$. The collection of codebooks $\{\mathcal{C}_j\}$ is called an $(n, M, \epsilon, \mathcal{E}, K)$–code if it satisfies the power constraint described before, and the per-user probability of error is smaller than ϵ. Then, we can define the following fundamental limit for the channel

$$\mathcal{E}^*(M, \mu, \epsilon) = \lim_{n\to\infty} \inf \left\{\mathcal{E} : \exists (n, M, \epsilon, \mathcal{E}, K = \mu n) - code\right\}.$$

We make an important remark here that all the following results also hold for maximal per-user error (PUPE-max) (10) as discussed in appendix C.

A. No-CSI

In this subsection, we focus on the no-CSI case. The difficulty here is that, a priori, we do not know which subset of the users to decode. We have the following theorem.

Theorem IV.1. *Consider the channel* (6) *(no-CSI) with $K = \mu n$ where $\mu < 1$. Fix the spectral efficiency S and target probability of error (per-user) ϵ. Let $M = 2^{S/\mu}$ denote the size of the codebooks and $P_{tot} = KP$ be the total power. Fix $\nu \in (1-\epsilon, 1]$. Let $\epsilon' = \epsilon - (1-\nu)$. Then if $\mathcal{E} > \mathcal{E}^*_{no-CSI} = \sup_{\frac{\epsilon'}{\nu} < \theta \leq 1} \sup_{\xi \in [0, \nu(1-\theta)]} \frac{P_{tot,\nu}(\theta,\xi)}{S}$, there exists a sequence of $(n, M, \epsilon_n, \mathcal{E}, K = \mu n)$ codes such that $\limsup_{n\to\infty} \epsilon_n \leq \epsilon$, where, for $\frac{\epsilon'}{\nu} < \theta \leq 1$ and $\xi \in [0, \nu(1-\theta)]$,*

$$P_{tot,\nu}(\theta, \xi) = \frac{\hat{f}(\theta,\xi)}{1 - \hat{f}(\theta,\xi)\alpha\left(\xi + \nu\theta, \xi + 1 - \nu(1-\theta)\right)} \tag{22}$$

$$\hat{f}(\theta, \xi) = \frac{f(\theta)}{\alpha(\xi, \xi + \nu\theta)} \tag{23}$$

$$f(\theta) = \frac{\frac{1 + \delta_1^*(1 - V_\theta)}{V_\theta} - 1}{1 - \delta_2^*} \tag{24}$$

$$V_\theta = e^{-\tilde{V}_\theta} \tag{25}$$

$$\tilde{V}_\theta = \delta^* + \frac{\theta\mu\nu \ln M}{1 - \mu\nu} + \frac{1 - \mu\nu(1-\theta)}{1 - \mu\nu} h\left(\frac{\theta\mu\nu}{1 - \mu\nu(1-\theta)}\right) + \frac{\mu(1 - \nu(1-\theta))}{1 - \mu\nu} h\left(\frac{\theta\nu}{1 - \nu(1-\theta)}\right) \tag{26}$$

$$\delta^* = \frac{\mu h(1 - \nu(1-\theta))}{1 - \mu\nu} \tag{27}$$

$$c_\theta = \frac{2V_\theta}{1 - V_\theta} \tag{28}$$

$$q_\theta = \frac{\mu h(1-\nu(1-\theta))}{1-\mu\nu(1-\theta)} \tag{29}$$

$$\delta_1^* = q_\theta(1+c_\theta) + \sqrt{q_\theta^2(c_\theta^2+2c_\theta)+2q_\theta(1+c_\theta)} \tag{30}$$

$$\delta_2^* = \inf\Big\{x : 0 < x < 1, -\ln(1-x) - x > \frac{\mu h(1-\nu(1-\theta))}{1-\mu\nu(1-\theta)}\Big\} \tag{31}$$

$$\alpha(a,b) = a\ln(a) - b\ln(b) + b - a. \tag{32}$$

Hence $\mathcal{E}^* \leq \mathcal{E}^*_{no-CSI}$.

Proof Idea. Before we present the full proof, the main ideas are presented here. Also, over the course, we explain the quantities that are present in the statement of the theorem. We start with choosing independent random Gaussian codebooks for all users. That is, for each message of each user there is an independent complex Gaussian $\mathcal{CN}(0, P'I_n)$ codeword where $P' < P$. The choice $P' < P$ is to ensure we can control the maximum power constraint violation events.

For simplicity we will consider $\nu = 1$. Here ν represents the fraction of users that the decoder can choose to decode. Due to random coding, we can assume that a particular tuple of codewords $(c_1, c_2, \cdots, c_K)$ were transmitted i.e., the received vector at the decoder is $Y = \sum_{i=1}^K H_i c_i + Z$. Then the decoder performs subspace projection decoding. The idea is that in the absence of noise, the received vector lies in the subspace spanned by the K codewords. Since we assume $\mu = K/n < 1$, and the K codewords are linearly independent, we can uniquely decode them by projecting the received vector onto various K dimensional subspaces formed by taking a codeword from each of the codebooks. Formally,

$$\{\hat{c}_i : i \in [K]\} = \arg\max_{(c_i \in \mathcal{C}_i : i \in [K])} \left\|P_{\{c_i : i \in S\}} Y\right\|^2 \tag{33}$$

Notice that the PUPE is given by

$$P_e = \frac{1}{K}\sum_{i\in[K]} \mathbb{P}\left[c_i \neq \hat{c}_i\right].$$

We will bound this error with the probability of events F_t– event that exactly t users were misdecoded. That is

$$P_e \leq \epsilon + \mathbb{P}\Bigg[\bigcup_{t > \epsilon K} F_t\Bigg] \tag{34}$$

Hence it is enough to find conditions under which the second term (call it p_1) in the above display goes to 0 in our scaling. To analyze F_t, we consider subsets $S \subset [K]$ with $|S| = t$ and a choice of *incorrect* codewords $(c_i' \in \mathcal{C}_i : i \in S)$ where $c_i' \neq c_i$, and bound F_t as union (over S and $(c_i' : i \in [S])$) of events $\left\{\left\|P_{c'_{[S]}, c_{[[K]\setminus S]}} Y\right\|^2 > \left\|P_{c_{[[K]]}} Y\right\|^2\right\}$. With abuse of notation, denote this set as $F(S,t)$.

Let $c_{[S]} = \{c_i : i \in S\}$, similarly we have $H_{[S]}$. We make a crucial observation that, conditioned on $c_{[K]}$, $H_{[K]}$ and Z, the random variable $\left\|P_{c'_{[S]}, c_{[[K]\setminus S]}} Y\right\|^2$ can be written as $\left\|P_{c_{[K\setminus S]}} Y\right\|^2 + \left\|P^\perp_{c_{[K\setminus S]}} Y\right\|^2 \mathrm{Beta}(t, n-K)$ where $\mathrm{Beta}(a,b)$ is a beta distributed random variable with parameters a and b.

Let $G_S = \frac{\left\|P^\perp_{c_{[[K]]}} Y\right\|^2}{\left\|P^\perp_{c_{[K\setminus S]}} Y\right\|^2}$. Then we show that

$$\begin{aligned}
&\mathbb{P}\left[F(S,t) | c_{[K]}, H_{[K]}, Z\right] \\
&= \mathbb{P}\left[\mathrm{Beta}(n-K, t) < G_S | c_{[K]}, H_{[K]}, Z\right] \\
&\leq \binom{n-K+t-1}{t-1} (G_S)^{n-K}
\end{aligned} \tag{35}$$

Next, we use the idea of random coding union (RCU) bound [42] to get

$$\mathbb{P}\Bigg[\bigcup_t F_t\Bigg] \leq \mathbb{E}\Bigg[\min\Bigg\{1, \sum_{t,S} \mathbb{P}\left[F(S,t)|c_{[K]}, H_{[K]}, Z\right]\Bigg\}\Bigg] \tag{36}$$

Let $\theta = t/K$, which is the fraction of misdecoded users. Now, by thresholding the value of G_S (this threshold is parameterized by a $\delta > 0$) we get from (36) a sum of an exponentially decaying term with combinatorial factors and the probability that G_S violates this threshold for some S and t (call this probability p_2). Choosing the right threshold (δ^* and corresponding threshold value V_θ in the theorem) the first term vanishes (in the limit) and we are left with p_2.

This is analyzed by conditioning on $c_{[K]}$ and $H_{[K]}$ along with using concentration of non-central chi-squared distributed variables (see claim 3). We follow similar procedure to above (using RCU and thresholding) multiple times to obtain thresholds parameterized by δ_1^* and δ_2^* to vanish combinatorial factors (like q_θ in the theorem which is the exponent of a binomial coefficient) and finally we are left with the *bottleneck* term:

$$\begin{aligned}
&\limsup_n P_e \leq \epsilon + \\
&\limsup_n \mathbb{P}\Bigg[\bigcup_{t,S}\Bigg\{P' \sum_{i\in[S]} |H_i|^2 < g\left(\delta_1^*, \delta_2^*, \delta_3^*, M, \mu, \theta\right)\Bigg\}\Bigg]
\end{aligned} \tag{37}$$

where g is some specific function. In essence, this bottleneck term is precisely the event that $> \epsilon$ fraction of users are outside the Gaussian capacity region!

Next step is to replace $\cup_S$ with $\min_S$ and use the convergence of order statistics of fading coefficients i.e., $|H_{(1)}| > \cdots > |H_{(K)}|$:

$$\begin{aligned}
\limsup_n P_e \leq \epsilon + \limsup_n \mathbb{P}\Bigg[\bigcup_t \Bigg\{P' \sum_{i=K-t+1}^{K} |H_{(i)}|^2 < \\
g\left(\delta_1^*, \delta_2^*, \delta_3^*, M, \mu, t/K\right)\Bigg\}\Bigg]
\end{aligned} \tag{38}$$

Then we show that, for $t = \theta K$, $\frac{1}{K}\sum_{i=K-t+1}^K |H_{(i)}|^2 \to \int_{1-\theta}^1 F^{-1}_{|H|^2}(1-\gamma)\, d\gamma \equiv \alpha(1-\theta, 1)$ in probabilty as $n \to \infty$. Hence the bottleneck term becomes deterministic in the limit. The choice P_{tot} such that this terms vanishes is precisely the one given in the statement of the theorem. □

Proof. The proof uses random coding. Let each user generate a Gaussian codebook of size M and power $P' < P$ independently such that $KP' = P'_{tot} < P_{tot}$. Let W_j denote the random (in $[M]$) message of user j. So, if $\mathcal{C}_j = \{c_i^j : i \in [M]\}$ is the codebook of user j, he/she transmits $X_j = c^j_{W_j} 1\left\{ \left\| c^j_{W_j} \right\|^2 \le nP \right\}$. For simplicity let $(c_1, c_2, ..., c_K)$ be the sent codewords. Hence the received vector is $Y = \sum_{i\in[K]} H_i c_i + Z$ where Z is the noise vector. Fix $\nu \in (1-\epsilon, 1]$. Let $K_1 = \nu K$ be the number of users that are decoded. Since there is no knowledge of CSIR, it is not possible to, a priori, decide what set to decode. Instead, the decoder searches of all K_1 sized subsets of $[M]$. Formally, let ? denote an error symbol. The decoder output $g_D(Y) \in \prod_{i=1}^k \mathcal{C}_i$ is given by

$$\left[\hat{S}, (\hat{c}_i)_{i\in\hat{S}}\right] = \arg \max_{\substack{S\subset[K] \\ |S|=K_1}} \max_{(c_i\in\mathcal{C}_i)_{i\in S}} \left\| P_{\{c_i : i \in S\}} Y \right\|^2$$

$$(g_D(Y))_i = \begin{cases} f_i^{-1}(\hat{c}_i) & i \in \hat{S} \\ ? & i \notin \hat{S} \end{cases} \tag{39}$$

where f_i are the encoding functions. The probability of error (averaged over random codebooks) is given by

$$P_e = \frac{1}{K} \sum_{j=1}^{K} \mathbb{P}\left[W_j \neq \hat{W}_j\right] \tag{40}$$

where $\hat{W}_j = (g(Y))_j$ is the decoded message of user j.

We perform a change of measure to $X_j = c^j_{W_j}$. Since P_e is the expectation of a non-negative random variable bounded by 1, this measure change adds a total variation distance which can bounded by $p_0 = K\mathbb{P}\left[\frac{\chi_2(2n)}{2n} > \frac{P}{P'}\right] \to 0$ as $n \to \infty$, where $\chi_2(d)$ is the distribution of sum of squares of d iid standard normal random variables (the chi-square distribution). The reason is as follows. If we have two random vectors U_1 and U_2 on a the same probability space such that $U_1 = U_2 1[U_2 \in E]$, where E is a Borel set, then for any Borel set A, we have

$$\begin{aligned} &|\mathbb{P}[U_1 \in A] - \mathbb{P}[U_2 \in A]| \\ &= |1[0 \in A]\mathbb{P}[U_2 \in E^c] - \mathbb{P}[U_2 \in A \cap E^c]| \\ &\le \mathbb{P}[U_2 \in E^c]. \end{aligned} \tag{41}$$

Henceforth we only consider the new measure.

Let $\epsilon > 1-\nu$ and $\epsilon' = \epsilon - (1-\nu)$. Now we have

$$\begin{aligned} P_e &\le \epsilon + \mathbb{P}\left[\frac{1}{K}\sum_{j=1}^{K} 1[W_j \neq \hat{W}_j] > \epsilon\right] \\ &= \epsilon + \mathbb{P}\left[\sum_{j=1}^{K} 1[W_j \neq \hat{W}_j] > K\epsilon' + K - K_1\right] \\ &= \epsilon + p_1. \end{aligned} \tag{42}$$

where

$$p_1 = \mathbb{P}\left[\bigcup_{t=\epsilon' K}^{\nu K} \left\{ \sum_{j=1}^{K} 1[W_j \neq \hat{W}_j] = K - K_1 + t \right\}\right].$$

Let $F_t = \left\{\sum_{j=1}^{K} 1[W_j \neq \hat{W}_j] = K - K_1 + t\right\}$. Let $c_{[S]} \equiv \{c_i : i \in S\}$ and $H_{[S]} \equiv \{H_i : i \in S\}$, where $S \subset [K]$. Conditioning on $c_{[K]}, H_{[K]}$ and Z, we have

$$\begin{aligned} &\mathbb{P}\left[F_t | c_{[K]}, H_{[K]}, Z\right] \\ &\le \mathbb{P}\Bigg[\exists S \subset [K] : |S| = K - K_1 + t, \exists S_1 \subset S : |S_1| = t, \\ &\quad \exists \{c'_i \in \mathcal{C}_i : i \in S_1, c'_i \neq c_i\} : \left\| P_{c'_{[S_1]}, c_{[[K]\setminus S]}} Y \right\|^2 > \\ &\quad \max_{\substack{S_2\subset S \\ |S_2|=t}} \left\| P_{c_{[S_2]}, c_{[[K]\setminus S]}} Y \right\|^2 \Bigg| c_{[K]}, H_{[K]}, Z \Bigg] \\ &\le \mathbb{P}\Bigg[\bigcup_{\substack{S\subset[K] \\ |S|=K-K_1+t}} \bigcup_{\substack{S_1\subset S \\ |S_1|=t}} \bigcup_{\substack{\{c'_i\in\mathcal{C}_i: \\ i\in S_1, c'_i\neq c_i\}}} \\ &\quad F(S, S_2^*, S_1, t) \Bigg| c_{[K]}, H_{[K]}, Z \Bigg] \end{aligned} \tag{43}$$

where

$$F(S, S_2^*, S_1, t) = \left\{ \left\| P_{c'_{[S_1]}, c_{[[K]\setminus S]}} Y \right\|^2 > \left\| P_{c_{[S_2^*]}, c_{[[K]\setminus S]}} Y \right\|^2 \right\}$$

and $S_2^* \subset S$ is a possibly random (depending only on $H_{[K]}$) subset of size t, to be chosen later. Next we will bound $\mathbb{P}\left[F(S, S_2^*, S_1, t) | c_{[K]}, H_{[K]}, Z\right]$.

For the sake of brevity, let $A_0 = c_{[S_2^*]} \cup c_{[[K]\setminus S]}$, $A_1 = c_{[[K]\setminus S]}$ and $B_1 = c'_{[S_1]}$. We have the following claim.

Claim 1. *For any $S_1 \subset S$ with $|S_1| = t$, conditioned on $c_{[K]}$, $H_{[K]}$ and Z, the law of $\left\| P_{c'_{[S_1]}, c_{[[K]\setminus S]}} Y \right\|^2$ is same as the law of $\|P_{A_1} Y\|^2 + \|(I - P_{A_1})Y\|^2 \mathrm{Beta}(t, n - K_1)$ where $\mathrm{Beta}(a, b)$ is a beta distributed random variable with parameters a and b.*

Proof. Let us write $V = span\{A_1, B_1\} = A \oplus B$ where $A \perp B$ are subspaces of dimension $K_1 - t$ and t respectively, with $A = span(A_1)$ and B is the orthogonal complement of A_1 in V. Hence $\|P_V Y\|^2 = \|P_A Y\|^2 + \|P_B Y\|^2$ (by definition, $P_A = P_{A_1}$). Now we analyze $\|P_B Y\|^2$. We can further write $P_B Y = P_B P_A^\perp Y$. Observe that the subspace B is the span of $P_A^\perp B_1$, and, conditionally, $P_A^\perp B_1 \sim \mathcal{CN}^{\otimes|S|}(0, P' P_A^\perp)$ which is the product measure of $|S|$ complex normal vectors in a subspace of dimension $n - K_1 + t$. Hence, the conditional law of $\left\|P_B P_A^\perp Y\right\|^2$ is the law of squared length of projection of a fixed $n - K_1 + t$ dimensional vector of length $\|(I - P_A)Y\|^2$ onto a (uniformly) random t dimensional subspace.

Further, the law of the squared length of the orthogonal projection of a fixed unit vector in $\mathbb{C}^d$ onto a random t–dimensional subspace is same as the law of the squared length of the orthogonal projection of a random unit vector in $\mathbb{C}^d$ onto a fixed t–dimensional subspace, which is $\mathrm{Beta}(t, d - t)$ (see for e.g. [43, Eq. 79]): that is, if u is a unit random vector in $\mathbb{C}^d$ and L is a fixed t dimensional subspace, then $\mathbb{P}\left[\|P_L u\|^2 \le x\right] = \mathbb{P}\left[\frac{\sum_{i=1}^t |Z_i|^2}{\sum_{i=1}^d |Z_i|^2} \le x\right] = F_\beta(x; t, n - K_1)$

where $Z_i \overset{iid}{\sim} \mathcal{CN}(0,1)$ and $F_\beta(x;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x w^{a-1}(1-w)^{b-1} dw$ denotes the CDF of the beta distribution with parameters a and b. Hence the conditional law of $\left\|P_B P_A^\perp Y\right\|^2$ is $\|(I-P_A)Y\|^2 \operatorname{Beta}(t, n-K_1)$.

□

Therefore we have,

$$\begin{aligned}
&\mathbb{P}\left[F(S, S_2^*, S_1, t) | c_{[K]}, H_{[K]}, Z\right] \\
&= \mathbb{P}\left[Beta(n-K_1, t) < G_S | c_{[K]}, H_{[K]}, Z\right] \\
&= F_\beta\left(G_S; n-K_1, t\right)
\end{aligned} \tag{44}$$

where

$$G_S = \frac{\|Y\|^2 - \|P_{A_0} Y\|^2}{\|Y\|^2 - \|P_{A_1} Y\|^2}. \tag{45}$$

Since $t \geq 1$, we have

$$F_\beta\left(G_S; n-K_1, t\right) \leq \binom{n-K_1+t-1}{t-1} G_S^{n-K_1}.$$

Let us denote $\bigcup_{t=\epsilon' K}^{\nu K}$ as $\bigcup_t$, $\bigcup_{\substack{S \subset [K] \\ |S| = K-K_1+t}}$ as $\bigcup_{S,K_1}$, and $\bigcup_t \bigcup_{\substack{S \subset [K] \\ |S| = K-K_1+t}}$ as $\bigcup_{t,S,K_1}$; similarly for $\sum$ and $\bigcap$ for the ease of notation. Using the above claim, we get,

$$\begin{aligned}
&\mathbb{P}\left[F_t | c_{[K]}, H_{[K]}, Z\right] \leq \\
&\sum_{S,K_1} \binom{K-K_1+t}{t} M^t \binom{n-K_1+t-1}{t-1} G_S^{n-K_1}.
\end{aligned} \tag{46}$$

Therefore p_1 can be bounded as

$$\begin{aligned}
p_1 &= \mathbb{P}\left[\bigcup_t F_t\right] \\
&\leq \mathbb{E}\left[\min\left\{1, \sum_{t,S,K_1} \binom{K-K_1+t}{t} M^t \cdot \right.\right. \\
&\qquad \left.\left. \binom{n-K_1+t-1}{t-1} G_S^{n-K_1}\right\}\right] \\
&= \mathbb{E}\left[\min\left\{1, \sum_{t,S,K_1} e^{(n-K_1)s_t} M^t G_S^{n-K_1}\right\}\right]
\end{aligned} \tag{47}$$

where $s_t = \frac{\ln\left(\binom{K-K_1+t}{t}\binom{n-K_1+t-1}{t-1}\right)}{n-K_1}$.

Now we can bound the binomial coefficient [44, Ex. 5.8] as

$$\begin{aligned}
&\binom{n-K_1+t-1}{t-1} \\
&\leq \sqrt{\frac{n-K_1+t-1}{2\pi(t-1)(n-K_1)}} e^{(n-K_1+t-1)h\left(\frac{t-1}{n-K_1+t-1}\right)} \\
&= O\left(\frac{1}{\sqrt{n}}\right) e^{n(1-\mu\nu(1-\theta))h\left(\frac{\theta\mu\nu}{1-\mu\nu(1-\theta)}\right)}.
\end{aligned} \tag{48}$$

Similarly,

$$\begin{aligned}
&\binom{K-K_1+t}{t} \\
&\leq O\left(\frac{1}{\sqrt{n}}\right) e^{n\mu(1-\nu(1-\theta))h\left(\frac{\theta\nu}{1-\nu(1-\theta)}\right)}
\end{aligned} \tag{49}$$

Let $r_t = s_t + \frac{t \ln M}{n-K_1}$. For $\delta > 0$, define $\tilde{V}_{n,t} = r_t + \delta$ and $V_{n,t} = e^{-\tilde{V}_{n,t}}$. Let E_1 be the event

$$\begin{aligned}
E_1 &= \bigcap_{t,S,K_1} \left\{-\ln G_S - r_t > \delta\right\} \\
&= \bigcap_{t,S,K_1} \left\{G_S < V_{n,t}\right\}.
\end{aligned} \tag{50}$$

Let $p_2 = \mathbb{P}\left[\bigcup_{t,S,K_1} \{G_S > V_{n,t}\}\right]$. Then

$$\begin{aligned}
&p_1 \\
&\leq \mathbb{E}\left[\min\left\{1, \sum_{t,S,K_1} e^{(n-K_1)r_t} G_S^{n-K_1}\right\}(1[E_1] + 1[E_1^c])\right] \\
&\leq \mathbb{E}\left[\sum_{t,S,K_1} e^{-(n-K_1)\delta}\right] + p_2 \\
&= \sum_t \binom{K}{K-K_1+t} e^{-(n-K_1)\delta} + p_2.
\end{aligned} \tag{51}$$

Observe that, for $t = \theta K_1 = \theta\nu K$,

$$\begin{aligned}
s_t = &\frac{1-\mu\nu(1-\theta)}{1-\mu\nu} h\left(\frac{\theta\mu\nu}{1-\mu\nu(1-\theta)}\right) + \\
&\frac{\mu(1-\nu(1-\theta))}{1-\mu\nu} h\left(\frac{\theta\nu}{1-\nu(1-\theta)}\right) - O\left(\frac{\ln(n)}{n}\right)
\end{aligned}$$

and $r_t = s_t + \frac{\theta\mu\nu}{1-\mu\nu} \ln M$. Therefore $n \to \infty$ with θ fixed, we have

$$\lim_{n \to \infty} \tilde{V}_{n,\theta\nu\mu} = \tilde{V}_\theta \tag{52}$$

where $\tilde{V}_\theta$ is given in (26).

Now, note that, for $1 < t < K_1$,

$$\begin{aligned}
&\binom{K}{K-K_1+t} \\
&\leq \sqrt{\frac{K}{2\pi(K-K_1+t)(K_1-t)}} e^{Kh\left(\frac{K-K_1+t}{K}\right)}.
\end{aligned} \tag{53}$$

Hence choosing $\delta > \frac{Kh\left(\frac{K-K_1+t}{K}\right)}{n-K_1}$ will ensure that the first term in (51) goes to 0 as $n \to \infty$. So for $t = \theta K_1 = \theta\nu K$, we need to have

$$\delta > \delta^*. \tag{54}$$

where δ^* is given in (27).

Let us bound p_2. Let $\hat{Z} = Z + \sum_{i \in S \setminus S_2^*} H_i c_i$. We have

Claim 2.

$$\begin{aligned}
p_2 &= \mathbb{P}\left[\bigcup_{t,S,K_1} \{G_S > V_{n,t}\}\right] \\
&\leq \mathbb{P}\left[\bigcup_{t,S,K_1} \left\{\left\|(1-V_{n,t})P_{A_1}^\perp \hat{Z} - V_{n,t} P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i\right\|^2\right.\right. \\
&\qquad \left.\left. \geq V_{n,t}\left\|P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i\right\|^2\right\}\right].
\end{aligned} \tag{55}$$

Proof. See appendix B. □

Let $\chi_2'(\lambda, d)$ denote the non-central chi-squared distributed random variable with non-centrality λ and degrees of freedom d. That is, if $W_i \sim \mathcal{N}(\mu_i, 1), i \in [d]$ and $\lambda = \sum_{i\in[d]} \mu_i^2$, then $\chi_2'(\lambda, d)$ has the same distribution as that of $\sum_{i\in[d]} W_i^2$. We have the following claim.

Claim 3. *Conditional on $H_{[K]}$ and A_0,*

$$\left\| P_{A_1}^{\perp}\left(\hat{Z} - \frac{V_{n,t}}{1-V_{n,t}} \sum_{i\in S_2^*} H_i c_i\right)\right\|^2 \sim \left(1 + P' \sum_{i\in S\setminus S_2^*} |H_i|^2\right) \frac{1}{2}\chi_2'\left(2F, 2n'\right) \tag{56}$$

where

$$F = \frac{\left\| \frac{V_{n,t}}{1-V_{n,t}} P_{A_1}^{\perp} \sum_{i\in S_2^*} H_i c_i \right\|^2}{\left(1 + P' \sum_{i\in S\setminus S_2^*} |H_i|^2\right)} \tag{57}$$

$$n' = n - K_1 + t. \tag{58}$$

Hence its conditional expectation is

$$\mu = n' + F. \tag{59}$$

Proof. See appendix B. □

Now let

$$T = \frac{1}{2}\chi_2'(2F, 2n') - \mu \tag{60}$$

$$U = \frac{V_{n,t}}{(1-V_{n,t})} \frac{\left\| P_{A_1}^{\perp} \sum_{i\in S_2^*} H_i c_i \right\|^2}{\left(1 + P' \sum_{i\in S\setminus S_2^*} |H_i|^2\right)} - n' \tag{61}$$

$$U^1 = \frac{1}{1-V_{n,t}} \left(V_{n,t} W_S - 1\right) \tag{62}$$

where

$$W_S = \left(1 + \frac{\left\| P_{A_1}^{\perp} \sum_{i\in S_2^*} H_i c_i \right\|^2}{n' \left(1 + P' \sum_{i\in S\setminus S_2^*} |H_i|^2\right)}\right).$$

Notice that $U = n'U^1$ and $F = \frac{V_{n,t}}{1-V_{n,t}} n'(1+U^1)$.

Then we have

$$\begin{aligned}
&\text{RHS of (55)} \\
&= \mathbb{P}\Bigg[\bigcup_{t,S,K_1} \Bigg\{ \Bigg\| P_{A_1}^{\perp}\hat{Z} - \frac{V_{n,t}}{(1-V_{n,t})} P_{A_1}^{\perp} \sum_{i\in S_2^*} H_i c_i \Bigg\|^2 \\
&\quad - \mu \geq U \Bigg\}\Bigg] \\
&= \mathbb{P}\Bigg[\bigcup_{t,S,K_1} \{T \geq U\}\Bigg].
\end{aligned} \tag{63}$$

Now, let $\delta_1 > 0$, and $E_2 = \cap_{t,S,K_1} \left\{U^1 > \delta_1\right\}$. Taking expectations over E_1 and its complement, we have

$$\begin{aligned}
&\mathbb{P}\Bigg[\bigcup_{t,S,K_1} \{T \geq U\}\Bigg] \\
&\leq \sum_{t,S,K_1} \mathbb{P}\left[T > U, U^1 > \delta_1\right] + \mathbb{P}\left[E_2^c\right] \\
&= \sum_{t,S,K_1} \mathbb{E}\left[\mathbb{P}\left[T > U \middle| H_{[K]}, A_0\right] 1[U^1 > \delta_1]\right] \\
&\quad + \mathbb{P}\left[E_2^c\right]
\end{aligned} \tag{64}$$

which follows from the fact that $\{U^1 > \delta_1\} \in \sigma(H_{[K]}, A_0)$. To bound this term, we use the following concentration result from [45, Lemma 8.1].

Lemma IV.2 ([45])**.** *Let $\chi = \chi_2'(\lambda, d)$ be a non-central chi-squared distributed variable with d degrees of freedom and non-centrality parameter λ. Then $\forall x > 0$*

$$\begin{aligned}
&\mathbb{P}\left[\chi - (d+\lambda) \geq 2\sqrt{(d+2\lambda)x} + 2x\right] \leq e^{-x} \\
&\mathbb{P}\left[\chi - (d+\lambda) \leq -2\sqrt{(d+2\lambda)x}\right] \leq e^{-x}
\end{aligned} \tag{65}$$

Hence, for $x > 0$, we have

$$\mathbb{P}\left[\chi - (d+\lambda) \geq x\right] \leq e^{-\frac{1}{2}\left(x + d + 2\lambda - \sqrt{d+2\lambda}\sqrt{2x+d+2\lambda}\right)}. \tag{66}$$

and for $x < (d+\lambda)$, we have

$$\mathbb{P}\left[\chi \leq x\right] \leq e^{-\frac{1}{4}\frac{(d+\lambda-x)^2}{d+2\lambda}}. \tag{67}$$

Observe that, in (66), the exponent is always negative for $x > 0$ and finite λ due to AM-GM inequality. When $\lambda = 0$, we can get a better bound for the lower tail in (67) by using [24, Lemma 25].

Lemma IV.3 ([24])**.** *Let $\chi = \chi_2(d)$ be a chi-squared distributed variable with d degrees of freedom. Then $\forall x > 1$*

$$\mathbb{P}\left[\chi \leq \frac{d}{x}\right] \leq e^{-\frac{d}{2}\left(\ln x + \frac{1}{x} - 1\right)} \tag{68}$$

Therefore, from (55), (63), (64) and (66), we have

$$\begin{aligned}
p_2 \leq &\sum_{t,S,K_1} \mathbb{E}\left[e^{-n' f_n(U^1)} 1[U^1 > \delta_1]\right] + \\
&\mathbb{P}\Bigg[\bigcup_{t,S,K_1} \left\{U^1 \leq \delta_1\right\}\Bigg]
\end{aligned} \tag{69}$$

where f_n is given by

$$\begin{aligned}
&f_n(x) = x + 1 + \frac{2V_{n,t}}{1-V_{n,t}}(1+x) \\
&- \sqrt{1 + \frac{2V_{n,t}}{1-V_{n,t}}(1+x)}\sqrt{2x + 1 + \frac{2V_{n,t}}{1-V_{n,t}}(1+x)}.
\end{aligned} \tag{70}$$

Next, we have the following claim.

Claim 4. *For $0 < V_{n,t} < 1$ and $x > 0$, $f_n(x)$ is a monotonically increasing function of x.*

Proof. See appendix B. □

From this claim, we get

$$p_2 \leq \sum_{t,S,K_1} e^{-n' f_n(\delta_1)} + p_3 \tag{71}$$

where $p_3 = \mathbb{P}\left[E_2^c\right]$.

Now, if, for each t, δ_1 is chosen such that $f_n(\delta_1) > \frac{Kh(\frac{K-K_1+t}{K})}{n-K_1+t}$, then the first term in (102) goes to 0 as $n \to \infty$. Therefore, for $t = \theta K_1$, setting c_θ and q_θ as in (28) and (29) respectively, and choosing δ_1 such that

$$\delta_1 > \delta_1^* \tag{72}$$

with δ_1^* given by (30), will ensure that the first term in (71) goes to 0 as $n \to \infty$.

Note that

$$\begin{aligned} p_3 &= \mathbb{P}\left[E_2^c\right] && (73)\\ &= \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Big\{V_{n,t}W_S - 1 \leq \delta_1(1-V_{n,t})\Big\}\Bigg]. && (74)\end{aligned}$$

Conditional on $H_{[K]}$,

$$\left\|P_{A_1}^{\perp}\sum_{i\in S_2^*} H_i c_i\right\|^2 \sim \frac{1}{2}P'\sum_{i\in S_2^*}|H_i|^2\chi_2^{S_2^*}(2n')$$

where $\chi_2(2n')$ is a chi-squared distributed random variable with $2n'$ degrees of freedom (here the superscript S_2^* denotes the fact that this random variable depends on the codewords corresponding to S_2^*). For $1 > \delta_2 > 0$, consider the event $E_4 = \bigcap_{t,S,K_1}\left\{\frac{\chi_2^{S_2^*}(2n')}{2n'} > 1-\delta_2\right\}$. Using (68) , we can bound p_3 as

$$p_3 \leq \sum_t \binom{K}{K-K_1+t} e^{-n'(-\ln(1-\delta_2)-\delta_2)} + p_4 \tag{75}$$

where

$$\begin{aligned} p_4 &= \mathbb{P}\left[E_4^c\right]\\ &= \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Bigg\{V_{n,t}\Bigg(1+\frac{P'\sum_{i\in S_2^*}|H_i|^2(1-\delta_2)}{\Big(1+P'\sum_{i\in S\setminus S_2^*}|H_i|^2\Big)}\Bigg)\\ &\qquad \leq 1+\delta_1(1-V_{n,t})\Bigg\}\Bigg]. && (76)\end{aligned}$$

Again, it is enough to choose δ_2 such that

$$\delta_2 > \delta_2^* \tag{77}$$

with δ_2^* given by (31), to make sure that the first term in (75) goes to 0 as $n \to \infty$.

Note that the union bound over S is the minimum over S, and this minimizing S should be contiguous amongst the indices arranged according the decreasing order of fading powers. Further, S_2^* is chosen to be corresponding to the top t fading powers in S. Hence, we get

$$\begin{aligned} p_4 = \mathbb{P}\Bigg[\bigcup_t\Bigg\{&\min_{0\leq j\leq K_1-t}\Bigg(\frac{P'\sum_{i=j+1}^{j+t}|H_{(i)}|^2(1-\delta_2)}{1+P'\sum_{i=j+t+1}^{j+t+K-K_1}|H_{(i)}|^2}\Bigg)\\ &\leq \frac{1+\delta_1(1-V_{n,t})}{V_{n,t}}-1\Bigg\}\Bigg]. && (78)\end{aligned}$$

We make the following claim

Claim 5.

$$\begin{aligned}&\limsup_{n\to\infty} p_4 \leq 1\Bigg[\bigcup_{\theta\in(\frac{\epsilon'}{\nu},1]\cap\mathbb{Q}}\\ &\Bigg\{\inf_{\xi\in[0,\nu(1-\theta)]}\Bigg(\frac{(1-\delta_2)P'_{tot}\alpha(\xi,\xi+\nu\theta)}{1+P'_{tot}\alpha(\xi+\nu\theta,\xi+1-\nu(1-\theta))}\Bigg)\\ &\leq \frac{1+\delta_1(1-V_\theta)}{V_\theta}-1\Bigg\}\Bigg] && (79)\end{aligned}$$

where $\alpha(a,b)$ is given by (32).

Proof. We have $|H_1|^2,...,|H_K|^2$ with CDF $F(x) = (1-e^{-x})1[x>=0]$. Let $\tilde{F}_K(x) = \frac{1}{K}\sum_{i=1}^K 1[|H_i|^2 \leq x]$ be the empirical CDF (ECDF). Then standard Chernoff bound gives, for $0<r<1$,

$$\mathbb{P}\left[|\tilde{F}_K(x)-F(x)| > rF(x)\right] \leq 2e^{-KcF(x)r^2} \tag{80}$$

where c is some constant.

From [46], we have the following representation. Let $0<\gamma<1$. Then

$$\begin{aligned}&|H_{(\lceil n\gamma\rceil)}|^2 =\\ &F^{-1}(1-\gamma) - \frac{\tilde{F}_K(F^{-1}(1-\gamma))-(1-\gamma)}{f\left(F^{-1}(1-\gamma)\right)} + R_K && (81)\end{aligned}$$

where f is the pdf corresponding to F, and with probability 1, we have $R_K = O(n^{-3/4}\log(n))$ as $n\to\infty$.

Let $\tau>0$. Then using (80) and (81), we have

$$\left||H_{(\lceil n\gamma\rceil)}|^2 - F^{-1}(1-\gamma)\right| \leq O\left(\frac{1}{n^{\frac{1-\tau}{2}}}\right) \tag{82}$$

with probability atleast $1-e^{-O(n^\tau)}$.

Hence, for $0<\xi<\zeta<1$, we have, with probability $1-e^{-O(n^\tau)}$,

$$\begin{aligned}&\frac{1}{K}\sum_{i=\lceil\alpha K\rceil}^{\lceil\beta K\rceil}|H_{(i)}|^2 =\\ &\left[\frac{1}{K}\sum_{i=1}^K|H_i|^2 1\left[b\leq|H_i|^2\leq a\right]\right]+o(1) && (83)\end{aligned}$$

where $a = F^{-1}(1-\xi)$ and $b = F^{-1}(1-\zeta)$. Now, by law of large numbers (and Bernstein's inequality [47]), with overwhelming probability (exponentially close to 1), we have

$$\frac{1}{K}\sum_{i=1}^K|H_i|^2 1\left[b\leq|H_i|^2\leq a\right] = \int_b^a x\,dF(x)+o(1) \tag{84}$$

and $\int_b^a x\,dF(x) = \int_\xi^\zeta F^{-1}(1-\gamma)d\gamma = \alpha(\xi,\zeta)$.

Define the events

$$J_{n,\theta,\xi} = \left\{ \left(\frac{P' \sum_{i=\lceil \xi K \rceil +1}^{\lceil (\xi+\nu\theta)K \rceil} |H_{(i)}|^2 (1-\delta_2)}{1 + P' \sum_{i=\lceil (\xi+\nu\theta)K \rceil +1}^{\lceil (\xi+1-\nu(1-\theta))K \rceil} |H_{(i)}|^2} \right) \right.$$
$$\left. \leq \frac{1+\delta_1(1-V_{n,\lceil \theta\nu K \rceil})}{V_{n,\lceil \theta\nu K \rceil}} - 1 \right\} \tag{85}$$

$$I_{n,\theta,\xi} = \left\{ \left(\frac{(1-\delta_2) P'_{tot} \alpha(\xi, \xi+\nu\theta)}{1 + P'_{tot} \alpha(\xi+\nu\theta, \xi+1-\nu(1-\theta))} \right) \right.$$
$$\left. \leq \frac{1+\delta_1(1-V_{n,\lceil \theta\nu K \rceil})}{V_{n,\lceil \theta\nu K \rceil}} - 1 \right\} \tag{86}$$

$$I_{\theta,\xi} = \left\{ \left(\frac{(1-\delta_2) P'_{tot} \alpha(\xi, \xi+\nu\theta)}{1 + P'_{tot} \alpha(\xi+\nu\theta, \xi+1-\nu(1-\theta))} \right) \right.$$
$$\left. \leq \frac{1+\delta_1(1-V_\theta)}{V_\theta} - 1 \right\} \tag{87}$$

$$E_{n,\theta,\xi} = \left\{ \left| \frac{1}{K} \sum_{i=\lceil \xi K \rceil +1}^{\lceil (\xi+\nu\theta)K \rceil} |H_{(i)}|^2 - \alpha(\xi, \xi+\nu\theta) \leq o(1) \right| \right\}$$
$$\bigcap \left\{ \left| \frac{1}{K} \sum_{i=\lceil (\xi+\nu\theta)K \rceil +1}^{\lceil (\xi+1-\nu(1-\theta))K \rceil} |H_{(i)}|^2 - \right. \right.$$
$$\left. \left. \alpha(\xi+\nu\theta, \xi+1-\nu(1-\theta)) \leq o(1) \right| \right\} \tag{88}$$

$$E_n = \left(\bigcap_{\theta \in A_n} \bigcap_{\xi \in B_{K,\theta}} E_{n,\theta,\xi} \right) \tag{89}$$

where $A_n = \left(\frac{\epsilon'}{\nu}, 1\right] \cap \left\{\frac{i}{K_1} : i \in [K_1]\right\}$ and $B_{K,\theta} = [0, \nu(1-\theta)] \cap \left\{\frac{i}{K} : i \in [K]\right\}$. Note that, from (83) and (84), $\mathbb{P}\left[E^c_{n,\theta,\xi}\right]$ is exponentially small in n.

Then we have

$$\begin{aligned}
p_4 &= \mathbb{P}\left[\bigcup_{\theta \in A_n} \bigcup_{\xi \in B_{K,\theta}} J_{n,\theta,\xi} \right] \\
&\leq \mathbb{P}\left[\bigcup_{\theta \in A_n} \bigcup_{\xi \in B_{K,\theta}} J_{n,\theta,\xi} \cap E_{n,\theta,\xi} \right] \\
&\quad + \sum_{\theta \in A_n} \sum_{\xi \in B_{K,\theta}} \mathbb{P}\left[E^c_{n,\theta,\xi}\right] \\
&\leq 1\left[\bigcup_{\theta \in A_n} \bigcup_{\xi \in B_{K,\theta}} I_{n,\theta,xi} \right] + o(1) \\
&\leq 1\left[\bigcup_{\theta \in (\frac{\epsilon'}{\nu}, 1]} \bigcup_{\xi \in [0, \nu(1-\theta)]} I_{n,\theta,\xi} \right] + o(1).
\end{aligned} \tag{90}$$

Therefore

$$\limsup_{n\to\infty} p_4 \leq 1\left[\bigcup_{\theta \in (\frac{\epsilon'}{\nu}, 1]} \bigcup_{\xi \in [0, \nu(1-\theta)]} I_{\theta,\xi} \right] \tag{91}$$

This concludes the proof of claim 5.

□

The statement of the theorem follows by choosing P'_{tot} to make sure that $\limsup_{n\to\infty} p_4 = 0$.

Remark 4. *In retrospect, our analysis is rather similar to the one in [24]. We remind that the problem considered there can be seen (as argued in [4]) as a version of the many-MAC problem with random-access, cf. Section II-A for more.*

□

B. CSIR

In this subsection, we focus on the CSIR scenario. We could use projection decoding to decode a fraction of users where decoding set is a function of CSIR. But a better bound is obtained by directly using euclidean metric to decode, similar to [4]. Then have the following theorem.

Theorem IV.4. *Consider the channel* (6) *(with CSIR) with $K = \mu n$ where $\mu < 1$. Fix the spectral efficiency S and target probability of error (per-user) ϵ. Let $M = 2^{S/\mu}$ denote the size of the codebooks and $P_{tot} = KP$ be the total power. Fix $\nu \in (1-\epsilon, 1]$. Let $\epsilon' = \epsilon - (1-\nu)$. Then if $\mathcal{E} > \mathcal{E}^*_{CSIR} = \sup_{\frac{\epsilon'}{\nu} < \theta \leq 1} \inf_{0 \leq \rho \leq 1} \frac{P_{tot,\nu}(\theta,\rho)}{S}$, there exists a sequence of $(n, M, \epsilon_n, \mathcal{E}, K = \mu n)$ codes such that $\limsup_{n\to\infty} \epsilon_n \leq \epsilon$, where, for $\frac{\epsilon'}{\nu} < \theta \leq 1$,*

$$P_{tot,\nu}(\theta,\rho) = \frac{(1+\rho)\left(e^{\mu\nu\left(\frac{h(\theta)}{\rho} + \theta \ln M\right)} - 1\right)}{\alpha(\nu(1-\theta), \nu) - \left(e^{\mu\nu\left(\frac{h(\theta)}{\rho} + \theta \ln M\right)} - 1\right)\alpha(\nu, 1)(1+\rho)} \tag{92}$$

$$\alpha(a,b) = a\ln(a) - b\ln(b) + b - a. \tag{93}$$

Hence $\mathcal{E}^* \leq \mathcal{E}^*_{CSIR}$.

The proof idea is a combination of techniques similar to [4] and theorem IV.1

Proof. Let each user generate a Gaussian codebook of size M and power $P' < P$ independently such that $KP' = P'_{tot} < P_{tot}$. Let W_j denote the random (in $[M]$) message of user j. So, if $\mathcal{C}_j = \{c^j_i : i \in [M]\}$ is the codebook of user j, he/she transmits $X_j = c^j_{W_j} 1\left\{\left\|c^j_{W_j}\right\|^2 \leq nP\right\}$. For simplicity let $(c_1, c_2, ..., c_K)$ be the sent codewords. Fix $\nu \in (1-\epsilon, 1]$. Let $K_1 = \nu K$ be the number of users that are decoded. Fix a decoding set $D \subset [K]$, possibly depending on $H_{[K]}$ such that $|D| = K_1, a.s.$ Since the receives knows $H_{[K]}$, we can use the euclidean distance used in [4] as the decoding metric. Formally, the decoder output $g_D(Y) \in \prod_{i=1}^K \mathcal{C}_i$ is given by

$$(g_D(Y))_i = \begin{cases} f_i^{-1}(\hat{c}_i) & i \in D \\ ? & i \notin D \end{cases}$$
$$(\hat{c}_i)_{i\in D} = \arg \min_{(c_i \in \mathcal{C}_i)_{i\in D}} \left\| Y - \sum_{i\in D} H_i c_i \right\|^2.$$

The probability of error is given by

$$P_e = \frac{1}{K} \sum_{j=1}^K \mathbb{P}\left[W_j \neq \hat{W}_j\right] \tag{94}$$

where $\hat{W}_j = (g(Y))_j$ is the decoded message of user j. Similar to the no-CSI case, we perform a change of measure to $X_j = c^j_{W_j}$ by adding a total variation distance bounded by $p_0 = K\mathbb{P}\left[\frac{\chi_2(2n)}{2n} > \frac{P}{P'}\right] \to 0$ as $n \to \infty$.

Let $\epsilon' = \epsilon - (1-\nu)$. Now we have

$$\begin{aligned} P_e &= \mathbb{E}\left[\frac{1}{K}\sum_{j=1}^{K} 1\{W_j \neq \hat{W}_j\}\right] \\ &= \frac{K-K_1}{K} + \mathbb{E}\left[\frac{1}{K}\sum_{j\in D} 1\{W_j \neq \hat{W}_j\}\right] \\ &\leq (1-\nu) + \epsilon' + \nu\mathbb{P}\left[\frac{1}{K}\sum_{j\in D} 1\{W_j \neq \hat{W}_j\} \geq \epsilon'\right] \\ &= \epsilon + \nu p_1 \end{aligned} \tag{95}$$

where $p_1 = \mathbb{P}\left[\bigcup_{t=\epsilon' K}^{\nu K}\left\{\sum_{j\in D} 1\{W_j \neq \hat{W}_j\} = t\right\}\right]$.

From now on, we just write $\bigcup_t$ to denote $\bigcup_{t=\epsilon' K}^{\nu K}$, $\sum_t$ for $\sum_{t=\epsilon' K}^{\nu K}$, and $\sum_S$ for $\sum_{\substack{S\subset D \\ |S|=t}}$. Let $c_{[S]} \equiv \{c_i : i \in [S]\}$ and $H_{[K]} = \{H_i : i \in [K]\}$.

Let $F_t = \left\{\sum_{j\in D} 1\{W_j \neq \hat{W}_j\} = t\right\}$. Let $\rho \in [0,1]$. We bound $\mathbb{P}[F_t]$ using Gallager's rho trick similar to [4] as

$$\begin{aligned} &\mathbb{P}\left[F_t | Z, c_{[K]}, H_{[K]}\right] \\ &\leq \mathbb{P}\Bigg[\exists S \subset D : |S| = t, \exists\{c'_i \in \mathcal{C}_i : i \in S, c'_i \neq c_i\} : \\ &\quad \Big\|Y - \sum_{i\in S} H_i c'_i - \sum_{i\in D\setminus S} H_i c_i\Big\|^2 < \\ &\quad \Big\|Y - \sum_{i\in D} H_i c_i\Big\|^2 \Big| Z, c_{[K]}, H_{[K]}\Bigg] \\ &\leq \sum_S \mathbb{P}\Bigg[\bigcup_{\substack{c'_i\in\mathcal{C}_i : i\in S \\ c'_i\neq c_i}} \Big\{\Big\|Z_D + \sum_{i\in S} H_i c_i - \sum_{i\in S} H_i c'_i\Big\|^2 < \\ &\quad \Big\|Z_D\Big\|^2\Big\} \Big| Z, c_{[K]}, H_{[K]}\Bigg] \\ &\leq \sum_S M^{\rho t}\mathbb{P}\Bigg[\Big\|Z_D + \sum_{i\in S} H_i c_i - \sum_{i\in S} H_i c'_i\Big\|^2 < \\ &\quad \Big\|Z_D\Big\|^2 \Big| Z, c_{[K]}, H_{[K]}\Bigg]^{\rho} \end{aligned} \tag{96}$$

where $Z_D = Z + \sum_{i\in[K]\setminus D} H_i c_i$ and $c'_{[S]}$ in the last display denotes a generic set of unsent codewords corresponding to codebooks of users in set S.

We use the following simple lemma which is a trivial extension of a similar result used in [4] to compute the above probability.

Lemma IV.5. *Let $Z \sim \mathcal{CN}(0, I_n)$ and $u \in \mathbb{C}^n$. Let $D = diag(d_1, ..., d_n) \in \mathbb{C}^{n\times n}$ be a diagonal matrix. If $\gamma > \sup_{j\in[n]} -\frac{1}{|d_j|^2}$, then*

$$\mathbb{E}\left[e^{-\gamma\|DZ+u\|^2}\right] = \frac{1}{\prod_{j\in[n]}(1+\gamma|d_j|^2)} e^{-\gamma\sum_{j\in[n]}\frac{|u_j|^2}{1+\gamma|d_j|^2}}$$

Proof. Omitted. □

So, using the above lemma, we have, for $\lambda_1 > 0$,

$$\begin{aligned} &\mathbb{E}_{\{c'_S\}}\Bigg[\mathbb{P}\Bigg[\Big\|Z_D + \sum_{i\in S} H_i c_i - \sum_{i\in S} H_i c'_i\Big\|^2 < \\ &\quad \|Z_D\|^2 \Big| Z, c_{[K]}, H_{[K]}\Bigg]^{\rho}\Bigg] \\ &= \mathbb{E}_{\{c'_S\}}\Bigg[\mathbb{P}\Bigg[\exp\left(-\lambda_1\Big\|Z_D + \sum_{i\in S} H_i c_i - \sum_{i\in S} H_i c'_i\Big\|^2\right) > \\ &\quad \exp\left(-\lambda_1\|Z_D\|^2\right)\Big| Z, c_{[K]}, H_{[K]}\Bigg]^{\rho}\Bigg] \\ &\leq \frac{e^{\rho\lambda_1\|Z_D\|^2}}{\left(1+\lambda_1 P'\sum_{i\in S}|H_i|^2\right)^{\rho n}} e^{\frac{-\rho\lambda_1\|Z_D+\sum_{i\in S} H_i c_i\|^2}{1+\lambda_1 P'\sum_{i\in S}|H_i|^2}} \end{aligned} \tag{97}$$

where $\mathbb{E}_{c'_S}$ denotes taking expectation with respect to $\{c'_i : i \in S\}$ alone, and $1 + \lambda_1 P'\sum_{i\in S}|H_i|^2 > 0$.

Let $\lambda_2 = \frac{\rho\lambda_1}{1+\lambda_1 P'\sum_{i\in S}|H_i|^2}$. Note that λ_2 is a function of H_S. Now using lemma IV.5 again to take expectation over c_S, we get

$$\begin{aligned} &\mathbb{E}_{c_S}\left[\frac{e^{\rho\lambda_1\|Z_D\|^2}}{\left(1+\lambda_1 P'\sum_{i\in S}|H_i|^2\right)^{\rho n}} e^{\frac{-\rho\lambda_1\|Z_D+\sum_{i\in S} H_i c_i\|^2}{1+\lambda_1 P'\sum_{i\in S}|H_i|^2}}\right] \\ &\leq \frac{1}{\left(1+\lambda_1 P'\sum_{i\in S}|H_i|^2\right)^{\rho n}} \frac{1}{\left(1+\lambda_2 P'\sum_{i\in S}|H_i|^2\right)^{n}} \cdot \\ &\quad e^{\left(\rho\lambda_1 - \frac{\lambda_2}{1+\lambda_2 P'\sum_{i\in S}|H_i|^2}\right)\|Z_D\|^2} \end{aligned} \tag{98}$$

with $1 + \lambda_2 P'\sum_{i\in S}|H_i|^2 > 0$. Finally, taking expectation over Z, we get

$$\mathbb{P}\left[F_t | H_{[K]}\right] \leq \sum_S M^{\rho t} e^{-nE_0(\lambda_1;\rho,H_{[K]},S)} \tag{99}$$

where

$$\begin{aligned} &E_0(\lambda_1; \rho, H_{[K]}, S) \\ &= \rho\ln\Big(1+\lambda_1 P'\sum_{i\in S}|H_i|^2\Big) + \ln\Big(1+\lambda_2 P'\sum_{i\in S}|H_i|^2\Big) + \\ &\quad \ln\Bigg(1 - \Big(1 + P'\sum_{i\in D^c}|H_i|^2\Big) \cdot \\ &\quad \left(\rho\lambda_1 - \frac{\lambda_2}{1+\lambda_2 P'\sum_{i\in S}|H_i|^2}\right)\Bigg) \end{aligned} \tag{100}$$

with

$$1 > \left(1 + P'\sum_{i\in D^c}|H_i|^2\right)\left(\rho\lambda_1 - \frac{\lambda_2}{1+\lambda_2 P'\sum_{i\in S}|H_i|^2}\right).$$

It is easy to see that the optimum value of λ_1 that maximizes E_0 is given by

$$\lambda_1^* = \frac{1}{\left(1 + P' \sum_{i \in D^c} |H_i|^2\right)(1+\rho)} \tag{101}$$

and hence the maximum value of the exponent

$$E_0(\rho, H_{[K]}, S) = E_0(\lambda_1^*; \rho, H_{[K]}, S)$$

is given by

$$E_0(\rho, H_{[K]}, S) = \rho \ln \left(1 + \frac{P' \sum_{i \in S} |H_i|^2}{(1+\rho)\left(1 + P' \sum_{i \in D^c} |H_i|^2\right)}\right).$$

Therefore, we have

$$p_1 \leq \mathbb{E}\left[\sum_t \sum_S e^{\rho t \ln M} e^{-nE_0(\rho, H_{[K]}, S)}\right]. \tag{102}$$

Since we want an upper bound for (102), we would like to take minimum over $S \subset D : |S| = t$. For a given choice of D, this corresponds to minimizing $P' \sum_{i \in S} |H_i|^2$ which mean we take S to contain indices in D which correspond to t smallest fading coefficients (within D). Then, the best such bound is obtained by choosing D that maximizes $\frac{P' \sum_{i \in S} |H_i|^2}{\left(1 + P' \sum_{i \in D^c} |H_i|^2\right)}$. Clearly this corresponds to choosing D to contain indices corresponding to top K_1 fading coefficients.

Therefore, we get

$$p_1 \leq \mathbb{E}\left[\sum_t \binom{K_1}{t} e^{\rho t \ln M} e^{-n\rho \ln\left(1 + \frac{P' \sum_{i=K_1-t+1}^{K_1} |H_{(i)}|^2}{(1+\rho)\left(1 + P' \sum_{i=K_1+1}^{K} |H_{(i)}|^2\right)}\right)}\right].$$

Let $A_n = [\frac{\epsilon'}{\nu}, 1] \cap \left\{\frac{i}{K_1} : i \in [K_1]\right\}$. For $\theta \in A_n$ and $t = \theta K_1$, using [44, Ex. 5.8] again, we have

$$\binom{K_1}{t} \leq \sqrt{\frac{K_1}{2\pi t(K_1 - t)}} e^{K_1 h(\frac{t}{K_1})} = O\left(\frac{1}{\sqrt{n}}\right) e^{n\mu\nu h(\theta)}. \tag{103}$$

The choice of ρ was arbitrary, and hence,

$$\begin{aligned} p_1 \leq \mathbb{E}\Bigg[\min\Bigg\{1, \sum_{\theta \in A_n} \exp\Bigg(-n \sup_{\rho \in [0,1]}\Bigg(&\rho \ln\left(1 + \frac{P' \sum_{i=\nu(1-\theta)K+1}^{\nu K} |H_{(i)}|^2}{(1+\rho)(1 + P' \sum_{i=\nu K+1}^{K} |H_{(i)}|^2)}\right) \\ &- \mu\nu h(\theta) - \mu\nu\theta \ln M\Bigg)\Bigg)\Bigg\}\Bigg] \\ \leq \mathbb{E}\Bigg[\min\Bigg\{1, |A_n| \exp\Bigg(-n \inf_{\theta \in A_n} \sup_{\rho \in [0,1]} &\left(\rho \ln\left(1 + \frac{P' \sum_{i=\nu(1-\theta)K+1}^{\nu K} |H_{(i)}|^2}{(1+\rho)(1 + P' \sum_{i=\nu K+1}^{K} |H_{(i)}|^2)}\right)\right. \\ &\left.- \mu\nu h(\theta) - \mu\nu\theta \ln M\right)\Bigg)\Bigg\}\Bigg] \end{aligned} \tag{104}$$

where we have used min since $p_1 \leq 1$. Now, using similar arguments as in the proof of claim 5 and taking limits, we can see that

$$\begin{aligned} &\inf_{\theta \in A_n} \sup_{\rho \in [0,1]} \left(\rho \ln\left(1 + \frac{P' \sum_{i=\nu(1-\theta)K+1}^{\nu K} |H_{(i)}|^2}{(1+\rho)(1 + P' \sum_{i=\nu K+1}^{K} |H_{(i)}|^2)}\right) - \mu\nu h(\theta) - \mu\nu\theta \ln M\right) = \\ &\inf_{\theta \in A_n} \sup_{\rho \in [0,1]} \left(\rho \ln\left(1 + \frac{P'_{tot}\alpha(\nu(1-\theta), \nu)}{(1+\rho)(1 + P'_{tot}\alpha(\nu, 1))}\right) - \mu\nu h(\theta) - \mu\nu\theta \ln M\right) + o(1) \end{aligned} \tag{105}$$

with exponentially high probability. Hence,

$$\begin{aligned} p_1 &\leq \mathbb{E}\Bigg[|A_n| \exp\Bigg(o(n) - n \inf_{\theta \in A_n} \sup_{\rho \in [0,1]} \left(\rho \ln\left(1 + \frac{P'_{tot}\alpha(\nu(1-\theta), \nu)}{(1+\rho)(1 + P'_{tot}\alpha(\nu, 1))}\right) - \mu\nu h(\theta) - \mu\nu\theta \ln M\right)\Bigg)\Bigg] + o(1) \\ &\leq \mathbb{E}\Bigg[|A_n| \exp\Bigg(o(n) - n \inf_{\theta \in A} \sup_{\rho \in [0,1]} \left(\rho \ln\left(1 + \frac{P'_{tot}\alpha(\nu(1-\theta), \nu)}{(1+\rho)(1 + P'_{tot}\alpha(\nu, 1))}\right) - \mu\nu h(\theta) - \mu\nu\theta \ln M\right)\Bigg)\Bigg] + o(1) \end{aligned} \tag{106}$$

where $A = [\frac{\epsilon'}{\nu}, 1]$.

Therefore, choosing $P'_{tot} > \sup_{\theta \in A} \inf_{\rho \in [0,1]} P_{tot}(\theta, \rho)$ will ensure that $\limsup_{n \to \infty} p_1 = 0$.

□

Remark 5. *Note that the analysis of the CSIR case in this paper and the AWGN case in [4] are similar, in particular both analyze a (suboptimal for PUPE) maximum likelihood decoder. However, there are two new subtleties, compared to [4]. First, [4] applies Gallager's ρ-trick twice, where the second application (with parameter ρ_1 in the notation of [4]) is applied just before taking the expectation over Z in* (99). *In the CSIR case, the summands of $\sum_S$ actually depend on the subset S through the fading gains, which makes the ρ-trick less appealing, and that is why we omitted it here. Secondly, because the summands depend on S, we upper bound each by taking the maximum over S, and this requires analysis of order statistics which is, of course, not present in the AWGN case.*

C. Achievability bound via scalar AMP

In this section, we will given an achievability bound on E_b/N_0 for the no-CSI case by the asymptotic analysis of the scalar AMP algorithm [24, 26, 27, 41]. Here, we recall the compressed sensing view of our model (11) where U is block sparse. As discussed in section II-A, a better algorithm to use in this case would be the vector or block version of AMP, whose analysis is also well studied, e.g. [41]. However, as we discussed in Section II-A evaluation of performance of this block-AMP requires computing $M = 2^{100}$ dimensional integrals, and thus does not result in computable bounds. Instead, here we take a different approach by analyzing the scalar AMP algorithm, whose asymptotic analysis in [27] in fact only requires that the empirical distribution of entries of U be convergent – a fact emphasized in [24]. Let us restate the signal model we have:

$$Y = AU+Z, \qquad A_{i,j} \overset{iid}{\sim} \mathcal{CN}(0, P_{tot}/\mu), i \in [n], j \in [KM], \tag{107}$$

where $U \in \mathbb{C}^{KM}$ is block sparse with $K = \mu n$ blocks each of length M, with a single non-zero entry U_j in each block with $U_j \sim \mathcal{CN}(0,1)$ (Rayleigh fading), and $Z \sim \mathcal{CN}(0, I_n)$. The support of U, denoted by $S \in \{0,1\}^{KM}$, is sampled uniformly from all such block sparse supports (there are M^K of them). The goal is to get an estimate $\hat{S} = \hat{S}(Y, A)$ of S where our figure of merit is the following:

$$\mathrm{PUPE}(\hat{S}) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{P}\left[S_{1+(k-1)M}^{kM} \neq \hat{S}_{1+(k-1)M}^{kM}\right], \tag{108}$$

which is also known as section error rate (SER) in the SPARC literature [38].

The AMP-based algorithm operates as follows. First we estimate U iteratively, then after estimating U we threshold its values to obtain an estimator for S.

To describe scalar AMP we first introduce the following scalar problem. For each $\sigma > 0$ define $\mu^{(\sigma)} = P_{X,V}$ to be the joint distribution of variables X and V:

$$V = X + \sigma W, \qquad X \perp W \sim \mathcal{CN}(0,1) \tag{109}$$

and

$$X \sim \mathrm{BG}(1, 1/M) = \begin{cases} \mathcal{CN}(0,1) & \text{w.p. } \frac{1}{M} \\ 0 & \text{w.p. } 1 - \frac{1}{M} \end{cases} \tag{110}$$

We also define

$$\eta(z, \sigma^2) \triangleq \mathbb{E}[X|V=z], \mathrm{mmse}(\sigma^2) \triangleq \mathbb{E}[(X - \mathbb{E}[X|V])^2]. \tag{111}$$

Next, start with $U^{(0)} = 0 \in \mathbb{C}^{KM}$, $R^{(0)} = Y$, $\hat{\sigma}_0^2 = \frac{\mu}{P_{tot}} + \mu$. Then for $t = 1, 2, \cdots$ we have the following iterations

$$U^{(t)} = \eta\left(A^* R^{(t-1)} + U^{(t-1)}, \hat{\sigma}_{t-1}^2\right) \tag{112}$$

$$R^{(t)} = Y - AU^{(t)} + \mu M R^{(t-1)} \frac{1}{KM} \sum_{i=1}^{KM} \eta'\left(\left(A^* R^{(t-1)} + U^{(t-1)}\right)_i, \hat{\sigma}_{t-1}^2\right) \tag{113}$$

$$\hat{\sigma}_t^2 = \frac{1}{n}\left\|R^{(t)}\right\|^2 \tag{114}$$

where $\eta'(x+iy, \sigma^2)$ denotes $\frac{1}{2}\left(\frac{\partial \eta(x+iy,\sigma^2)}{\partial x} - i\frac{\partial \eta(x+iy,\sigma^2)}{\partial y}\right)$ and $i = \sqrt{-1}$ is the imaginary unit (see [48, 49] for a more general derivation of complex AMP). The estimate of U after t steps is given by (see [24] for more details)

$$\hat{U}^{(t)} = A^* R^{(t)} + U^{(t)} \tag{115}$$

To convert $\hat{U}^{(t)}$ into $\hat{S}^{(t)}$ we perform a simple thresholding:

$$\hat{S}^{(t)}(\theta) = \{i \in [KM] : |\hat{U}_i^{(t)}|^2 > \theta\}. \tag{116}$$

Theorem IV.6 (Scalar AMP achievability)**.** *Fix any $\mu > 0$, $P_{tot} > 0$ and $M \geq 1$. Then for every $\mathcal{E} > \frac{P_{tot}}{\mu \log_2 M}$ there exist a sequence of $(n, M, \epsilon_n, \mathcal{E}, K = \mu n)$ codes (noCSI) such that AMP decoder* (116) *(with a carefully chosen $\theta = \theta(\mathcal{E}, M, \mu)$ and sufficiently large t) achieves*

$$\limsup_{n\to\infty} \epsilon_n \leq \pi^*(\sigma_\infty^2, M),$$

where $\pi^*(\tau, M) = 1 - \frac{1}{1+\tau}\left((M-1)\left(\frac{1}{\tau}+1\right)\right)^{-\tau}$ and σ_∞^2 is found from

$$\sigma_\infty^2 \equiv \sigma_\infty^2(\mu, P_{tot}, M) = \sup\left\{\tau \geq 0 : \tau = \frac{\mu}{P_{tot}} + \mu M \,\mathrm{mmse}(\tau)\right\} \tag{117}$$

Proof. Denote the Hamming distance

$$d_H(S, \hat{S}) = \frac{1}{KM}\sum_{i=1}^{KM} 1[S_i \neq \hat{S}_i] \tag{118}$$

Note that according to the definition (108) we have a bound

$$\mathrm{PUPE}(\hat{S}^{(t)}(\theta)) \leq M\mathbb{E}\left[d_H(S, \hat{S}^{(t)}(\theta))\right] \tag{119}$$

Indeed, this is a simple consequence of upper bounding each probability in (108) by the union bound.

The key result of [27] shows the following. Let the empirical joint distribution of entries in $(U, \hat{U}^{(t)})$ be denoted by

$$\hat{\mu}_{U,\hat{U}^{(t)}} \triangleq \frac{1}{KM}\sum_{i=1}^{KM} \delta_{(U_i, \hat{U}_i^{(t)})},$$

where δ_x is the Dirac measure at x. Then as $n \to \infty$ this (random) distribution on $\mathbb{C}^2$ converges weakly to a deterministic limit $\mu_{X,V}$ almost surely. More precisely, from [27, Lemma 1(b)] and proof of [24, Theorem 5] for any bounded Lipschitz continuous function $f : \mathbb{C}^2 \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \int f\, d\hat{\mu}_{U,\hat{U}^{(t)}} = \int f\, d\mu^{(\sigma_t)} \text{ a.s.} \tag{120}$$

where $\mu^{(\sigma_t)}$ is the joint distribution of $(X, V = X + \sigma_t W)$ defined in (109), and σ_t can further be determined from the so called state evolution sequence: Set $\sigma_0^2 = \frac{\mu}{P_{tot}} + \mu$ and then

$$\sigma_t^2 = \frac{\mu}{P_{tot}} + \mu M \,\mathrm{mmse}(\sigma_{t-1}^2) \tag{121}$$

where mmse is defined in (111).

Note that the assumptions on U A and Z in [27, Lemma 1(b)] hold in our case. In particular, since the support of U is sampled uniformly from all block sparse supports of size K and the entries in the support are iid $\mathcal{CN}(0,1)$ random variables, we have that the empirical distribution of entries of U converge weakly almost surely to the distribution P_X of X defined in (110). Further the moment conditions in [27, Theorem 2] are also satisfied. We note here that although [24, 27] consider only real valued signals, the results there also hold for the complex case (see [50, Theorem III.15], [51, Chapter 7]).

We next consider the support recovery in the scalar model (109). Let $S_0 = 1[X \neq 0]$ denote the indicator of the event when X is non-zero. Let $\hat{S}_0 \equiv \hat{S}_0(\theta) = 1[|V|^2 > \theta]$ denote an estimator of S_0 using the observation $V = X + \sigma W$ in (109). Let

$$\psi(\sigma^2, \theta, M) = \mathbb{P}\left[S_0 \neq \hat{S}_0\right] \tag{122}$$

denote the probability of error in the scalar model (109) with σ dependence made explicit as an argument of ψ. The from the convergence of $\hat{\mu}_{U,\hat{U}^{(t)}}$ we conclude as in [24] that for any number t of steps of the AMP algorithm $\hat{S}^{(t)}(\theta)$ achieves

$$\lim_{n\to\infty} \text{PUPE}(\hat{S}^{(t)}(\theta)) \leq M\psi(\sigma_t^2, \theta, M)\,. \tag{123}$$

Since this holds for any t and any θ we can optimize both by taking $t \to \infty$ and $\inf_{\theta>0}$. From the proof of [24, Theorem 6] it follows that $\lim_{t\to\infty} \sigma_t^2 = \sigma_\infty^2$ exists and σ_∞ satisfies (117). The proof is completed by the application of the following Claim, which allows us to compute infimum over θ in closed form. □

Claim 6.

$$M \inf_\theta \psi(\tau, \theta, M) = 1 - \frac{1}{1+\tau} \frac{1}{\left((M-1)\left(\frac{1}{\tau}+1\right)\right)^\tau} \tag{124}$$

Proof. Let us define $\tau = \sigma^2$. We have

$$\begin{aligned}\psi(\tau, \theta, M) &= \mathbb{P}\left[S_0 \neq \hat{S}_0\right] \\ &= \frac{1}{M}\mathbb{P}\left[\hat{S}_0 = 0 | S_0 = 1\right] + \\ &\quad \left(1 - \frac{1}{M}\right)\mathbb{P}\left[\hat{S}_0 = 1 | S_0 = 0\right]\end{aligned}$$

Now conditioned on $S_0 = 1$, $|V|^2 \sim (1+\tau)\text{Exp}(1)$ and conditioned on $S_0 = 0$, $|V|^2 \sim \tau\text{Exp}(1)$ where $\text{Exp}(1)$ is the Exponential distribution with density function $p(x) = e^{-x}1[x \geq 0]$. Hence

$$\psi(\tau, \theta, M) = \frac{1}{M}\left(1 - e^{-\frac{\theta}{1+\tau}}\right) + \left(1 - \frac{1}{M}\right)e^{-\frac{\theta}{\tau}} \tag{125}$$

The claim follows by optimizing (125) over θ. The optimum occurs at

$$\theta^* = \tau(1+\tau)\ln\left(\frac{1+\tau}{\tau}(M-1)\right)$$

Substituting θ^* in (125) proves the claim. □

D. Converse

In this section we derive a converse for $\mathcal{E}^*$, based on the Fano inequality and the results from [28].

Theorem IV.7. *Let M be the codebook size. Given ϵ and μ, let $S = \mu \log M$. Then assuming that the distribution of $|H|^2$ has a density with $\mathbb{E}\left[|H|^2\right] = 1$ and $\mathbb{E}\left[|H|^4\right] < \infty$, $\mathcal{E}^*(M, \mu, \epsilon)$ satisfies the following two bounds*

1)

$$\mathcal{E}^*(M, \mu, \epsilon) \geq \inf \frac{P_{tot}}{S} \tag{126}$$

where infimum is taken over all $P_{tot} > 0$ that satisfies

$$\begin{aligned}&\theta S - \epsilon\mu \log\left(2^{S/\mu} - 1\right) - \mu h_2(\epsilon) \leq \\ &\log\left(1 + P_{tot}\alpha\left(1-\theta, 1\right)\right), \forall \theta \in [0,1]\end{aligned} \tag{127}$$

where $\alpha(a, b) = \int_a^b F_{|H|^2}^{-1}(1-\gamma)d\gamma$, and $F_{|H|^2}$ is the CDF of squared absolute value of the fading coefficients.

2)

$$\mathcal{E}^*(M, \mu, \epsilon) \geq \inf \frac{P_{tot}}{S} \tag{128}$$

where infimum is taken over all $P_{tot} > 0$ that satisfies

$$\epsilon \geq 1 - \mathbb{E}\left[Q\left(Q^{-1}\left(\frac{1}{M}\right) - \sqrt{\frac{2P_{tot}}{\mu}|H|^2}\right)\right] \tag{129}$$

where Q is the complementary CDF function of the standard normal distribution.

Proof. First, we use the Fano inequality.

Let $W = (W_1, ..., W_K)$, where $W_i \stackrel{iid}{\sim} Unif[M]$ denote the sent messages of K users. Let $X = (X_1, ..., X_K)$ where $X_i \in \mathbb{C}^n$ be the corresponding codewords, $Y \in \mathbb{C}^n$ be the received vector. Let $\hat{W} = \left(\hat{W}_1, ...\hat{W}_K\right)$ be the decoded messages. Then $W \to X \to Y \to \hat{W}$ forms a Markov chain. Then $\epsilon = P_e = \frac{1}{K}\sum_{i\in[K]} \mathbb{P}\left[W_i \neq \hat{W}_i\right]$.

Suppose a genie G reveal a set $S_1 \subset [K]$ for transmitted messages $W_{S_1} = \{W_i : i \in S_1\}$ and the corresponding fading coefficients H_{S_1} to the decoder. So, a converse bound in the Genie case is a converse bound for our problem (when there is no Genie). Further, the equivalent channel at the receiver is

$$Y_G = \sum_{i\in S_2} H_i X_i + Z \tag{130}$$

where $S_2 = [K] \setminus S_1$, and the decoder outputs a $[K]$ sized tuple. So, PUPE with Genie is given by

$$P_e^G = \frac{1}{K}\sum_{i\in[K]} \mathbb{P}\left[W_i \neq \hat{W}_i^G\right]. \tag{131}$$

Now, it can be seen that the optimal decoder must have the codewords revealed by the Genie in the corresponding locations in the output tuple, i.e., if $\hat{W}^G$ denotes the output tuple (in the Genie case), for $i \in S_1$, we must have that $W_i = \hat{W}_i^G$. Otherwise, PUPE can be strictly decreased by including these Genie revealed codewords.

So, letting $E_i = 1[W_i \neq \hat{W}_i^G]$ and $\epsilon_i^G = \mathbb{E}[E_i]$, we have that $\epsilon_i^G = 0$ for $i \in S_1$. For $i \in S_2$, a Fano type argument gives

$$I(W_i; \hat{W}_i^G) \geq \log M - \epsilon_i^G \log(M-1) - h_2(\epsilon_i^G). \tag{132}$$

So, using the fact that

$$\begin{aligned}\sum_{i\in S_2} I(W_i; \hat{W}_i^G) &\leq I(W_{S_2}; \hat{W}_{S_2}^G) \\ &\leq n\mathbb{E}\left[\log(1 + P\sum_{i\in S_2}|H_i|^2)\right]\end{aligned}$$

we have

$$|S_2|\log M-\sum_{i\in S_2}\epsilon_i^G\log(M-1)-\sum_{i\in S_2}h_2(\epsilon_i^G)$$
$$\leq n\mathbb{E}\left[\log(1+P\sum_{i\in S_2}|H_i|^2)\right]. \quad (133)$$

By concavity of h_2, we have

$$\frac{1}{K}\sum_{i\in S_2}h_2(\epsilon_i^G)=\frac{1}{K}\sum_{i\in[K]}h_2(\epsilon_i^G)\leq h_2(P_e^G). \quad (134)$$

Hence we get

$$\frac{|S_2|}{K}\log M-P_e^G\log(M-1)-h_2(P_e^G)$$
$$\leq\frac{n}{K}\mathbb{E}\left[\log(1+P\sum_{i\in S_2}|H_i|^2)\right]. \quad (135)$$

Next, notice that $P_e^G\leq P_e\leq 1-\frac{1}{M}$ and hence $P_e^G\log(M-1)+h_2(P_e^G)\leq P_e\log(M-1)+h_2(P_e)$. Further the inequality above hols for all $S_2\subset[K]$ (which can depend of $H_{[K]}$ as well). Hence, letting $|S_2|=\theta K$

$$\theta\log M-P_e\log(M-1)-h_2(P_e)$$
$$\leq\frac{1}{\mu}\mathbb{E}\left[\log\left(1+\inf_{S_2:|S_2|=\theta K}\frac{P_{tot}}{K}\sum_{i\in S_2}|H_i|^2\right)\right]. \quad (136)$$

Now, taking limit as $K\to\infty$ and using results on strong laws of order statistics [52, Theorem 2.1], we get that

$$\log\left(1+\inf_{S_2:|S_2|=\theta K}\frac{P_{tot}}{K}\sum_{i\in S_2}|H_i|^2\right)$$
$$\to\log\left(1+P_{tot}\alpha(1-\theta,1)\right). \quad (137)$$

For any $a,b\in[0,1]$ with $a<b$, let $S_K\equiv S_K(a,b)=\frac{1}{K}\sum_{i=aK}^{bK}|H_{(i)}|^2$. Note that $S_K\to\alpha(a,b)$ as $K\to\infty$. Then

$$\mathbb{E}\left[S_K^2\right]\leq\mathbb{E}\left[\left(\frac{1}{K}\sum_{i=1}^{K}|H_i|^2\right)^2\right]$$
$$=1+\frac{\mathbb{E}\left[|H|^4\right]-1}{K}\leq\mathbb{E}\left[|H|^4\right]. \quad (138)$$

Hence the family of random variables $\{S_K:K\in\mathbb{N}\}$ is *uniformly integrable*. Further

$$0\leq\log(1+P_{tot}S_K)\leq P_{tot}S_K.$$

Hence the family $\{\log(1+P_{tot}S_K):K\geq 1\}$ is also uniformly integrable. Then from theorem [53, Theorem 9.1.6],

$$\mathbb{E}\left[\log(1+P_{tot}S_K)\right]\to\log(1+P_{tot}\alpha(a,b)).$$

Using this in (136) with $a=1-\theta$ and $b=1$, we obtain (127).

Next we use the result from [28] to get another bound.

Using the fact that S/μ bits are needed to be transmitted under a per-user error of ϵ, we can get a converse on the minimum E_b/N_0 required by deriving the corresponding results for a single user quasi-static fading MAC. In [28], the authors gave the following non-asymptotic converse bound on the minimum energy required to send k bits for an AWGN channel. Consider the single user AWGN channel $Y=X+Z$, $Y,X\in\mathbb{R}^\infty$, $Z_i\overset{iid}{\sim}\mathcal{N}(0,1)$. Let $M^*(E,\epsilon)$ denote the largest M such that there exists a (E,M,ϵ) code for this channel: codewords $(c_1,...,c_M)$ with $\|c_i\|^2\leq E$ and a decoder such that probability of error is smaller than ϵ. The following is a converse bound from [28].

Lemma IV.8 ([28])**.** *Any (E,M,ϵ) code satisfies*

$$\frac{1}{M}\geq Q\left(\sqrt{2E}+Q^{-1}\left(1-\epsilon\right)\right) \quad (139)$$

Translating to our notations, for the channel $Y=HX+Z$, conditioned on H, if $\epsilon(H)$ denotes the probability of error for each realization of H, then we have

$$\frac{1}{M}\geq Q\left(\sqrt{\frac{2P_{tot}}{\mu}|H|^2}+Q^{-1}\left(1-\epsilon(H)\right)\right). \quad (140)$$

Further $\mathbb{E}\left[\epsilon(H)\right]=\epsilon$. Therefore we have

$$\epsilon\geq 1-\mathbb{E}\left[Q\left(Q^{-1}\left(\frac{1}{M}\right)-\sqrt{\frac{2P_{tot}}{\mu}|H|^2}\right)\right]. \quad (141)$$

Hence we have the required converse bound.

Remark 6. *We also get the following converse from [21, theorem 7] by taking the appropriate limits $P=\frac{P_{tot}}{\mu n}$ and $n\to\infty$.*

$$\log M\leq-\log\left(\mathbb{E}\left[\mathcal{Q}\left(\frac{c+\frac{P_{tot}|H|^2}{\mu}}{\sqrt{\frac{2P_{tot}|H|^2}{\mu}}}\right)\right]\right) \quad (142)$$

where c satisfies

$$\mathbb{E}\left[\mathcal{Q}\left(\frac{c-\frac{P_{tot}|H|^2}{\mu}}{\sqrt{\frac{2P_{tot}|H|^2}{\mu}}}\right)\right]=1-\epsilon. \quad (143)$$

But this is strictly weaker than (141)*. This is because, using lemma IV.8, we perform hypothesis testing (in the meta-converse) for each realization of H but in the bound used in [21], hypothesis testing is performed over the joint distribution (including the distribution of H). This is to say that if H is presumed to be constant (and known), then in* (142) *and* (143) *we can remove the expectation over H and this gives precisely the same bound as* (140)*.*

□

Bounds tighter than (127) can be obtained if further assumptions are made on the codebook. For instance, if we assume that each codebook consists of iid entries of the form $\frac{C}{K}$ where C is sampled from a distribution with zero mean and finite variance, then using ideas similar to [36, Theorem 3] we have the following converse bound.

Theorem IV.9. *Let M be the codebook size, and let μn users ($\mu<1$) generate their codebooks independently with each code symbol iid of the form $\frac{C}{K}$ where C is of zero mean and variance P_{tot}. Then in order for the iid codebook to achieve*

PUPE ϵ with high probability, the energy-per-bit $\mathcal{E}$ should satisfy

$$\mathcal{E} \geq \inf \frac{P_{tot}}{\mu \log M} \tag{144}$$

where infimum is taken over all $P_{tot} > 0$ that satisfies

$$\begin{aligned} &\ln M - \epsilon \ln(M-1) - h(\epsilon) \\ &\leq \left(M\mathcal{V}\left(\frac{1}{\mu M}, P_{tot}\right) - \mathcal{V}\left(\frac{1}{\mu}, P_{tot}\right)\right) \end{aligned} \tag{145}$$

where $\mathcal{V}$ is given by [36]

$$\begin{aligned} \mathcal{V}(r,\gamma) &= r \ln\left(1+\gamma - \mathcal{F}(r,\gamma)\right) + \ln\left(1 + r\gamma - \mathcal{F}(r,\gamma)\right) \\ &\quad - \frac{\mathcal{F}(r,\gamma)}{\gamma} \end{aligned} \tag{146}$$

$$\mathcal{F}(r,\gamma) = \frac{1}{4}\left(\sqrt{\gamma\left(\sqrt{r}+1\right)^2 + 1} - \sqrt{\gamma\left(\sqrt{r}-1\right)^2+1}\right)^2 \tag{147}$$

Proof sketch. The proof is almost the same as in [36, Theorem 3] (see [36, Remark 3] as well). We will highlight the major differences here. First, our communication system can be modeled as a support recovery problem as follows. Let A be the $n \times KM$ matrix consisting of $n \times M$ blocks of codewords of users. Let $\boldsymbol{H}$ be the $KM \times KM$ block diagonal matrix with block i being a diagonal $M \times M$ matrix with all diagonal entries being equal to H_i. Finally let $W \in \{0,1\}^{KM}$ with K blocks of size M each and within each M sized block, there is exactly one 1. So the position of 1 in block i of W denotes the message or codeword corresponding to the user i which is the corresponding column in block i of matrix A. Hence our channel can be represented as

$$Y = A\boldsymbol{H}W + Z \tag{148}$$

with the goal of recovering W.

Next the crucial step is bound $R^K(\epsilon, M)$ in (132) as

$$\begin{aligned} R^K(\epsilon, M) &\leq I(W; Y|A) \\ &= I(\boldsymbol{H}W; Y|A) - I(\boldsymbol{H}W; Y|A, W) \end{aligned} \tag{149}$$

where the equality in the above display follows from [36, equation (78)]. The first term in above display is bounded as

$$\begin{aligned} I(\boldsymbol{H}W; Y|A = A_1) &= I(\boldsymbol{H}W; A_1 HW + Z) \\ &\leq \sup_U I(U; A_1 U + Z) \end{aligned} \tag{150}$$

where A_1 is a realization of A and supremum is over random vectors $U \in \mathbb{C}^{KM}$ such that $\mathbb{E}[U] = 0$ and $\mathbb{E}[UU^*] = \mathbb{E}[(\boldsymbol{H}W)(\boldsymbol{H}W)^*] = \frac{\mathbb{E}[|H_1|^2]}{M} I_{KM\times KM}$. Now similar to [36], the supremum is achieved when

$$U \sim \mathcal{CN}\left(0, \frac{\mathbb{E}\left[|H_1|^2\right]}{M} I_{KM\times KM}\right).$$

Hence

$$I(\boldsymbol{H}W; Y|A = A_1) \leq \log\det\left(I_{n\times n} + \frac{1}{M}AA^*\right). \tag{151}$$

Next, for any realization A_1 and W_1 of A and W respectively, we have

$$\begin{aligned} &I(\boldsymbol{H}W; Y|A = A_1, W = W_1) \\ &= I(\boldsymbol{H}W_1; A_1\boldsymbol{H}W_1 + Z) \\ &= I(\tilde{\boldsymbol{H}}; (A_1)_{W_1}\tilde{\boldsymbol{H}} + Z) \\ &\geq I(\tilde{\boldsymbol{H}}; \tilde{\boldsymbol{H}} + (A_1)_{W_1}^{\dagger} Z) \end{aligned} \tag{152}$$

where $\tilde{\boldsymbol{H}} = [H_1, ..., H_K]^T$ and $(A_1)_{W_1}$ is the $n\times K$ submatrix of A_1 formed by columns corresponding to the support of W_1 and $\dagger$ denotes the Moore-Penrose inverse (pseudoinverse). The last equality in the above follows from the data processing inequality. Now, by standard mutual information of Gaussians, we have

$$\begin{aligned} &I(\tilde{\boldsymbol{H}}; \tilde{\boldsymbol{H}} + (A_1)_{W_1}^{\dagger} Z) \\ &= \log\det\left(I_{K\times K} + ((A_1)_{W_1})^* (A_1)_{W_1}\right). \end{aligned} \tag{153}$$

Hence

$$I(\boldsymbol{H}W; Y|A, W) = \mathbb{E}\left[\log\det\left(I_{K\times K} + A_W^* A_W\right)\right]. \tag{154}$$

Hereafter, the we can proceed similarly to the proof of [36, Theorem 3] using results from random-matrix theory [54, 55] to finish the proof. □

We remark here that for a general fading distribution, the term $I(\tilde{\boldsymbol{H}}; \tilde{\boldsymbol{H}} + (A_1)_{W_1}^{\dagger} Z)$ can be lower bounded similar to the proof of [36, Theorem 3] using EPI (and its generalization [56]) to get

$$\begin{aligned} &I(\tilde{\boldsymbol{H}}; \tilde{\boldsymbol{H}} + ((A_1)_{W_1})^{\dagger} Z) \\ &\geq K \log\left(1 + N_H \left(\det\left(((A_1)_{W_1})^* (A_1)_{W_1}\right)\right)^{\frac{1}{K}}\right) \end{aligned} \tag{155}$$

where $N_H = \frac{1}{\pi e}\exp(h(H))$ is the entropy power of fading distribution. Hence

$$\begin{aligned} &I(\boldsymbol{H}W; Y|A, W) \\ &\geq K\mathbb{E}\left[\log\left(1 + N_H\left(\det\left(A_W^* A_W\right)\right)^{\frac{1}{K}}\right)\right]. \end{aligned} \tag{156}$$

Again, we can use results from random-matrix theory [55] and proceed similarly to the proof of [36, Theorem 3] to get a converse bound with the second term in (145) replaced by $\mathcal{V}_{LB}\left(\frac{1}{\mu}, P_{tot}\right)$ and

$$\mathcal{V}_{LB}(r,\gamma) = \ln\left(1 + \gamma r\left(\frac{r}{r-1}\right)^{r-1}\frac{1}{e}\right) \tag{157}$$

We make a few observations regarding the preceding theorem. First and foremost, this hold only for the case of no-CSI because the term analogous to $I(\boldsymbol{H}W; Y|A, W)$ in the case of CSIR is $I(\boldsymbol{H}W; Y|A, \boldsymbol{H}, W)$ which is zero. Next, it assumes that the codebooks have iid entries with variance scaling $\Theta(1/n)$. This point is crucial to lower bounding $I(\boldsymbol{H}W; Y|A, W)$, and this is where a significant improvement comes when compared to (127). Indeed, EPI and results from random matrix theory give $O(n)$ lower bound for $I(\boldsymbol{H}W; Y|A, W)$. This once again brings to focus the the difference between classical regime and the scaling regime, where in the former, this term is negligible. Further this leaves open the question of whether we could improve performance in the high-density of users case by using non-iid codebooks.

Now, as to what types of codebooks give a $\Theta(n)$ lower bound for $I(\boldsymbol{H}W; Y|A, W)$, a partial answer can be given by carefully analyzing the full proof of the theorem. In particular, if $\mathcal{S} = \text{supp}W$ i.e, the support of W, then as seen from [36, equation (85)], any non zero lower bound on $\det(A_{\mathcal{S}}^* A_{\mathcal{S}})^{1/K}$ in the limit is enough. So if the matrix $A_{\mathcal{S}}^* A_{\mathcal{S}}$ possesses strong diagonal dominance then it is possible to have such a non zero lower bound on $\det(A_{\mathcal{S}}^* A_{\mathcal{S}})^{1/K}$ for every S [57]. These could be ensured by having codewords that are overwhelmingly close to orthogonal.

V. Numerical evaluation and discussion

In this section, we provide the results of numerical evaluation of the bounds in the paper. We focus on the trade-off of user density μ with the minimum energy-per-bit $\mathcal{E}^*$ for a given message size k and target probability of error P_e.

For $k = 100$ bits, we evaluate the trade-off from the bounds in this paper for $P_e = 0.1$ and $P_e = 0.001$ in figures 3 and 2 respectively. For TDMA, we split the frame of length n equally among K users, and compute the smallest P_{tot} the ensures the existence of a single user quasi-static AWGN code of rate S, blocklength $\frac{1}{\mu}$ and probability of error ϵ using the bound from [21]. The simulations of the single user bound is performed using codes from [58]. TIN is computed using a method similar to theorem IV.4. In particular, the codeword of user i is decoded as $\hat{c}_i = \arg\min_{c' \in \mathcal{C}_i} \|Y - H_i c_i\|^2$ where we assume that the decoder has the knowledge of CSI. The analysis proceeds in a similar way as theorem IV.4.

Achievability bounds. It can be seen that for small μ the scalar-AMP bound of Theorem IV.6 is better than the projection decoder bounds of Theorems IV.4 and IV.1. The latter bounds have another artifact. For example, the no-CSI bound on $\mathcal{E}^*$ from Theorem IV.1 increases sharply as $\mu \downarrow 0$, in fact one can show that the said bound behaves as $\mathcal{E} = \Omega(\sqrt{-\ln \mu})$.

Engineering insights. From these figures, we clearly observe the *perfect MUI cancellation* effect mentioned in the introduction and previously observed for the non-fading model [4, 7]. Namely, as μ increases from 0, the $\mathcal{E}^*$ is almost a constant, $\mathcal{E}^*(\mu, \epsilon, k) \approx \mathcal{E}_{\text{s.u.}}(\epsilon, k)$ for $0 < \mu < \mu_{\text{s.u.}}$. As μ increases beyond $\mu_{\text{s.u.}}$ the tradeoff undergoes a "phase transition" and the energy-per-bit $\mathcal{E}^*$ exhibits a more familiar increase with μ. Further, standard schemes for multiple-access like TDMA and TIN do not have this behavior. Moreover, although these suboptimal schemes have an optimal trade-off at $\mu \to 0$ they show a significant suboptimality at higher μ. We note again that this perfect MUI cancellation which was observed in standard GMAC [4, 7] is also present in the more practically relevant quasi-static fading model. So, we suspect that this effect is a general characteristic of the many-user MAC.

Suboptimality of orthogonalization. The fact that orthogonalization is not optimal is one of the key practical implications of our work. It was observed before in the GMAC and here we again witness it in the more relevant QS-MAC. How to understand this suboptimality? First, in the fading case we have already seen this effect even in the classical regime (but under PUPE) – see (21). To give another intuition we consider a $K = \mu n$ user binary adder MAC

$$Y = \sum_{i=1}^{K} X_i \tag{158}$$

where $X_i \in \{0, 1\}$ and addition is over $\mathbb{Z}$. Now, using TDMA on this channel, each user can send at most $n/K = 1/\mu$ bits. Hence the message size is bounded by

$$\log M \leq \frac{1}{\mu}. \tag{159}$$

Next, let us consider TIN. Assume $X_i \sim \text{Ber}(1/2)$. For user 1, we can treat $\sum_{i=2}^{\mu n} X_i$ as noise. By central limit theorem, this noise can be approximated as $\sqrt{\frac{1}{4}\mu n Z}$ where $Z \sim N(0, 1)$. Thus we have a binary input AWGN (BIAWGN) channel

$$Y = X_1 + \sqrt{\frac{1}{4}\mu n} Z. \tag{160}$$

Therefore, the message size is bounded as

$$\begin{aligned} \log M &\leq n C_{\text{BIAWGN}} \left(1 + \frac{4}{\mu n}\right) \\ &\leq \frac{n}{2} \log\left(1 + \frac{4}{\mu n}\right) \to \frac{2}{\mu \ln 2} \end{aligned} \tag{161}$$

where C_{BIAWGN} is the capacity of the BIAWGN channel. Note that in both the above schemes the achievable message size is a constant as $n \to \infty$.

On the other hand, the true sum-capacity of the K-user adder MAC is given by

$$C_{sum} = \max_{X_1, \cdots, X_K} H(X_1 + \cdots + X_K).$$

As shown in [59] this maximum as achieved at $X_i \overset{iid}{\sim} \text{Ber}(1/2)$. Since the the entropy of binomial distributions [60] can be computed easily, we obtain

$$C_{sum} = \frac{1}{2} \log K + o(\log K).$$

In particular, for our many-user MAC setting we obtain from the Fano inequality (and assuming PUPE is small)

$$\log M \lessapprox \frac{\log(\mu n)}{2\mu}.$$

Surprisingly, there exist explicit codes that achieve this limit and with a very low-complexity (each message bit is sent separately),– a construction rediscovered several times [61–63]. Hence the optimal achievable message size is

$$\log M \approx \frac{\log n}{2\mu} \to \infty \tag{162}$$

as $n \to \infty$. And again, we see that TDMA and TIN are severely suboptimal for the many-user adder MAC as well.

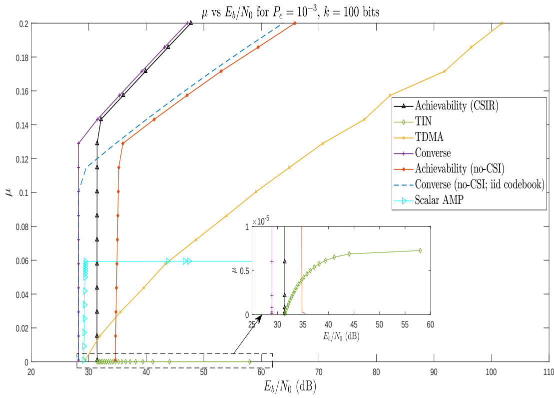

Fig. 2: μ vs E_b/N_0 for $\epsilon \le 10^{-3}$, $k = 100$

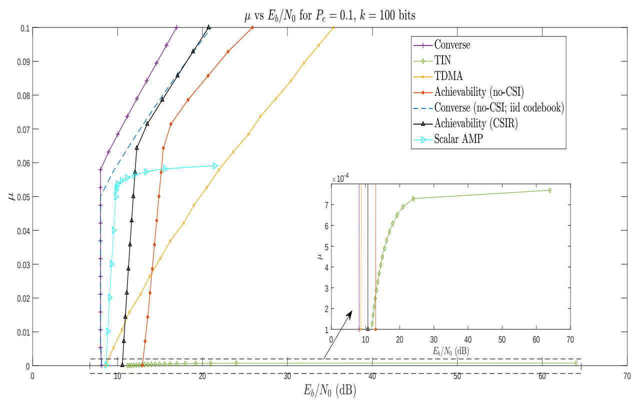

Fig. 3: μ vs E_b/N_0 for $\epsilon \le 10^{-1}$, $k = 100$

A. The "curious behavior" in phase transition

As we emphasized in the Introduction, the most exciting conclusion of our work is the existence of the almost vertical part on the μ vs $\frac{E_b}{N_0}$ plots of Fig. 3 and 2. In this section we want to explain how this effect arises, why it can be called the "almost perfect MUI cancellation" and how it relates (but is not equivalent) to well-known phase transitions in compressed sensing.

To make things easier to evaluate, however, we depart from the model in the previous sections and do two relaxations. First, we consider a non-fading AWGN. Second, we endow all users with the same codebook. The second assumption simply means that the decoding from now on is only considered up to permutation of messages, see [4] for more on this. Technically, these two assumptions mean that we are considering a model (11) with U vector that is K-sparse (as opposed to block-sparse) and that all non-zero entries of U are equal to 1. Finally, we will consider the real-valued channel. In all, we get the following signal model [5, Section IV]:

$$Y^n = AU^p + Z^n\,, \qquad Z^n \sim \mathcal{N}(0, I_n)\,, \tag{163}$$

with $A_{i,j} \overset{iid}{\sim} \mathcal{N}(0, b^2/n)$, $(i,j) \in [n] \times [p]$, $U_i \overset{iid}{\sim} \mathrm{Ber}(K/p)$, so that $\mathbb{E}[\|U^p\|_0] = K$. We take the proportional scaling limit with $K = \mu n$ and $p = KM$. Interpretation of these parameters in the context of communication problem are:

- M as the number of messages that each user wants to communicate
- μ is the user density per (real) degree of freedom
- $P_{tot} \triangleq b^2\mu$ is the total received power from all K users at the receiver.
- We also introduce *the spectral efficiency (in nats)* as
$$S_e \triangleq \mu M h\left(\frac{1}{M}\right) = \mu(\ln M - 1) + O(\frac{\mu}{M}) \tag{164}$$
where $h(x) = -x\ln x - (1-x)\ln(1-x)$ is the binary entropy function in nats. Note that $S_e = \frac{1}{n}H(U^p)$.
- Consequently, we may define energy-per-bit as[3]
$$\frac{E_b}{N_0} \triangleq \frac{P_{tot}\ln 2}{2S_e}\,.$$

Given (Y^n, A), the decoder outputs an estimate $\hat{U}^p \in \{0,1\}^p$ with $\mathbb{E}\left[\|\hat{U}^p\|_0\right] = K$ and we are interested in the minimal achievable PUPE, or

$$\begin{aligned} &P_e^*(\mu, M, b) \\ &\triangleq \limsup_{n\to\infty} \min_{\hat{U}^p : \mathbb{E}[\|\hat{U}\|_0] = K} \frac{1}{K}\sum_{i\in[p]} \mathbb{P}\left[U_i = 1, \hat{U}_i = 0\right] \end{aligned} \tag{165}$$

To discuss performance of the optimal decoder, we need to return to the scalar channel (109) with the following modifications: $X \sim \mathrm{Ber}(1/M)$, $W \sim \mathcal{N}(0,1)$. Now, for every value of σ in (109) we may ask for the smallest possible error $\epsilon^*(\sigma, M) = \min \mathbb{P}\left[\hat{X} \neq X\right]$ where minimum is taken over all estimators $\hat{X} = \hat{X}(V)$ such that $\mathbb{P}\left[\hat{X} = 1\right] = \frac{1}{M}$. As discussed in [5, Section IV.B], this minimal $\epsilon^*(\sigma, M)$ satisfies [5] is found from solving:

$$\frac{1}{\sigma} = \mathcal{Q}^{-1}\left(\frac{\epsilon^*}{M-1}\right) + \mathcal{Q}^{-1}(\epsilon^*) \tag{166}$$

where $\mathcal{Q}(\cdot)$ is the complementary CDF of the standard normal distribution.

Now the limit P_e^* in (165) can be computed via the replica method.[4] Namely, replica predictions tell us that

$$P_e^*(b, \mu) = \epsilon^*(\sigma, M)\,,$$

where $\sigma^2 = \frac{1}{\eta^* b^2}$ and the *multiuser efficiency*

$$\eta^* = \eta^*(M, b, \mu)$$

is given by [5, 35, 66]

$$\eta^* \equiv \eta_M^* \in \arg\min_{\eta\in[0,1]} \tilde{\mathcal{F}}_M(\eta; b^2, \mu) \tag{167}$$

[3]This definition is slightly different, but appears more reasonable than the one in [5]

[4]Note that in [64, 65] it was shown that the replica-method prediction is correct for estimating $I(U_i; Y^n, A)$ and $\mathrm{Var}[U_i|Y^n, A]$, but what we need for computing the P_e is asymptotic distribution of a random variable $\mathbb{P}[U_i = 1|Y^n, A]$. First, it is known that AMP initialized at the true value U converges to an asymptotically MMSE-optimal estimate. Second, distribution of the AMP estimates are known to belong to a $P_{X|V}$ in (109) (with σ identified from the replica method). Finally, any asymptotically MMSE-optimal estimator $\hat{U}$ should satisfy $\hat{U}_i \overset{(d)}{\to} \mathbb{E}[U_i|Y^n, A] = \mathbb{P}[U_i = 1|Y^n, A]$, and thus $\mathbb{P}[U_i = 1|Y^n, A]$ should match the replica-method predicted one.

where

$$\tilde{\mathcal{F}}_M(\eta; b^2, \mu) = \frac{I_M\left(\frac{\mu}{\eta P_{tot}}\right)}{h(1/M)} + \frac{1}{2S_e}\phi(\eta) \tag{168}$$

where $\phi(x) = x - 1 - \ln x$ and $I_M(\sigma^2) = I(X; X + \sigma W)$ is the mutual information between the signal $X \sim \mathrm{Ber}(1/M)$ and observation in the scalar channel (109) in nats.

In the figure 4 we have shown the plots of optimal PUPE P_e for the model (163) versus E_b/N_0 for various values of μ when $M = 2^{100}$, computed via replica predictions. What is traditionally referred to as the phase transition in compressed sensing is the step-function drop from $P_e \approx 1$ to $P_e \ll 1$. However, there is a *second effect* here as well. Namely that all the curves with different μ seem to have a *common envelope*. The former has not only been observed in compressed sensing, e.g. [24, 36, Fig.1] and [67, Fig.4] among others, but also in a number of other inference problems: randomly-spread CDMA [35], LDPC codes [68] and random SAT [69]. However, the second effect appears to be a rather different phenomenon, and in fact it is exactly the one that corresponds to the existence of the vertical part of the curves on Fig. 2-3.

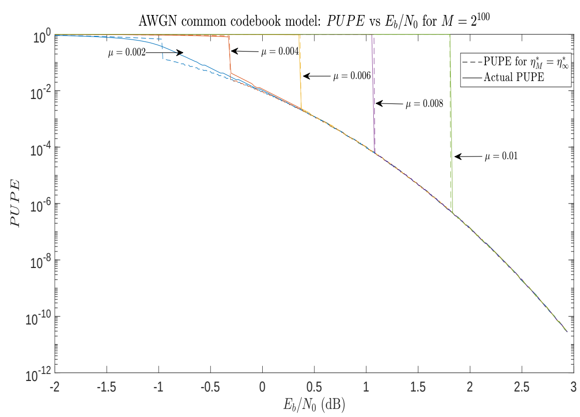

Fig. 4: AWGN same codebook model: P_e vs E_b/N_0 for $M = 2^{100}$. The solid lines correspond to using $\eta^* \equiv \eta^*_M$ from (167) in (166). The dashed lines correspond to η^*_∞ defined in (171).

Let us, for the moment, assume that the envelope is actually exactly the same for all μ. Fix a value of PUPE $P_e = 10^{-3}$ (say) and consider how the intercept of the horizontal line at $P_e = 10^{-3}$ on Fig. 4 changes with μ. It is easy to see that as long as the value of μ is small enough the intercept will not be moving (corresponding to constancy of the E_b/N_0 as a function of μ). However, once the value of μ exceeds a value (dependent on the fixed value of P_e) the intercept starts moving to the right together with the step-drop portion of the curves. From this we conclude that indeed, existence of the (almost) common envelope on Fig. 4 results in the (almost) vertical part on Fig. 2-3. (As a side note, we also note that since the slanted portions of the tradeoff curves on those figures correspond to the vertical drop on the Fig. 4 and hence the slanted portion is virtually independent of the fixed value of P_e – as predicted by (4).)

How can the curves have common envelope? Notice that in the expression for P_e^* only η^* is a function of μ. Thus, we conclude that for small μ we must have $\eta^*(\mu) \approx \text{const}$. But as $\mu \to 0$ we should get a $\eta^* \to 1$. Thus we see that common envelopes are only possible if to the right of the step-drops on Fig. 4 we get $\eta^* \approx 1$. Is that indeed so? Fig. 5 provides an affirmative answer.

In reality, the "vertical part" is not truly vertical and the common envelope is not exactly common. In truth the right portions of the curves on Fig. 4 (following the drop) are all very slightly different, but this difference is imperciptible to the eye (and irrelevant to an engineer). What makes them so close is the incredible degree of sparsity $\frac{1}{M} = 2^{-100}$. Indeed, as Fig. 6 demonstrates that as $M \to \infty$ the value of η^* to the right of the step transition approaches 1.

To summarize, we conclude that what determines our "curious behavior" is not a sudden change in the estimation performance (typically credited as "phase transition" in compressed sensing), but rather a more subtle effect arising in the super-low sparsity limit: the step-transition of the parameter η^* from a moderate value in the interior of $(0, 1)$ to a value close to 1. The fact that only the incredibly low sparsity values $\frac{1}{M}$ are relevant for the many-MAC problems makes this new effect practically interesting.

To study η^* in the limit of large M, we apply a clever observation of [70, 71]: in the limit of $M \to \infty$ (low sparsity), instead of using parameters (M, μ, b^2) it is better to use (M, P_{tot}, S_e). Holding P_{tot} and S_e constant implies that

$$\mu = \frac{S_e}{Mh(1/M)} = \frac{S_e}{\ln M} \cdot (1 + o(1))$$
$$b^2 = \frac{P_{tot} M h(1/M)}{S_e} = \frac{P_{tot}}{S_e} \ln M \cdot (1 + o(1))$$

as $M \to \infty$. Under this scaling of μ and b^2 we get [71, Section 3], for every $\eta \in (0, 1]$ we have

$$\lim_{M\to\infty} \tilde{\mathcal{F}}_M(\eta) = \tilde{\mathcal{F}}_\infty(\eta) \tag{169}$$

$$\tilde{\mathcal{F}}_\infty(\eta) \triangleq \min\left(1, \frac{P_{tot}}{2S_e}\eta\right) + \frac{1}{2S_e}\phi(\eta) \tag{170}$$

Furthermore, for $S_e \neq \frac{1}{2}\ln(1 + P_{tot})$ we have

$$\eta^*_M \to \eta^*_\infty = \begin{cases} \frac{1}{1+P_{tot}}, & \frac{1}{2}\ln(1 + P_{tot}) < S_e \\ 1, & \frac{1}{2}\ln(1 + P_{tot}) > S_e \end{cases} \tag{171}$$

Thus, we see that (at least in this regime of μ and b^2) indeed the η^*_M jumps from < 1 to ≈ 1 the energy-per-bit crosses over a threshold. Furthermore, in figures 4, 5 and 6 we show that $\eta^*_\infty(P_{tot}, S_e)$ gives an excellent agreement with $\eta^*(M, \mu, b^2)$.

Incidentally, under this scaling of $\mu \to 0$ and $b^2 \to \infty$ with a fixed P_{tot} and S_e, it turns out that PUPE experiences a step-transition from 0 to 1 upon crossing the "slanted part" of the S_e vs E_b/N_0 diagram. This step transition is dubbed all-or-nothing phenomenon and was indeed the subject of [70].

To see that notice that (166) corresponds to

$$\sqrt{2\eta^*_M M h(1/M)\frac{E_b/N_0}{\ln 2}} = \mathcal{Q}^{-1}\left(\frac{\epsilon^*}{M-1}\right) + \mathcal{Q}^{-1}(\epsilon^*) \tag{172}$$

Next, using the approximation

$$\mathcal{Q}^{-1}(\delta) \approx \sqrt{2\ln\frac{1}{\delta} - \ln\left(4\pi\ln\frac{1}{\delta}\right)} \qquad (173)$$

one can check that

$$\lim_{M\to\infty} \epsilon^* \to \begin{cases} 0, & \eta^*_\infty E_b/N_0 > \ln 2 \\ 1, & \eta^*_\infty E_b/N_0 < \ln 2 \end{cases},$$

which incidentally corresponds to the two cases $\eta^*_\infty \frac{E_b}{N_0} \lessgtr -1.59\ dB$. From (171) we get

$$\lim_{M\to\infty} \epsilon^* \to \begin{cases} 0, & \frac{1}{2}\ln(1+P_{tot}) > S_e \\ 1, & \frac{1}{2}\ln(1+P_{tot}) < S_e \end{cases},$$

In all, we see that the "mother" effect is that of η^* jumping from < 1 to ≈ 1 in the regime of low sparsity (large M). It causes both the "curious behavior" for us, and the all-or-nothing for sparse regression.

Finally we note that when $\eta^*_M \approx 1$, then from (172) the asymptotic expansion of $\ln(M-1)$ as $\frac{P_{tot}}{\mu} \to \infty$ (and fixed ϵ) matches with [28, Theorem 3] up to the $O(1)$ term, thus showing that the replica prediction for the location of the vertical part matches with the finite-blocklength prediction in (5).

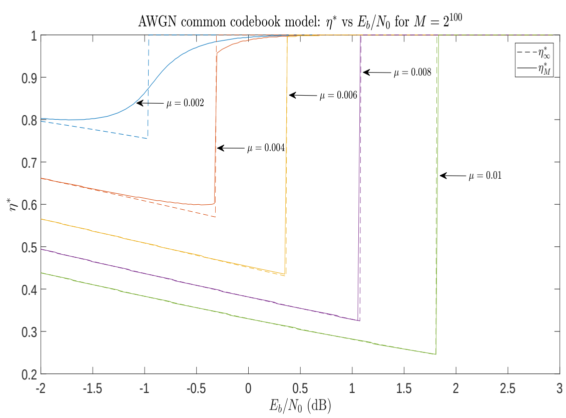

Fig. 5: AWGN same codebook model: η^* vs E_b/N_0 for $M = 2^{100}$. The solid lines correspond to $\eta^* \equiv \eta^*_M$ from (167). The dashed lines correspond to η^*_∞ defined in (171).

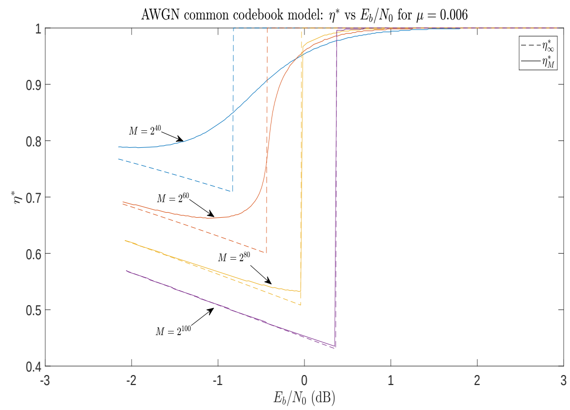

Fig. 6: AWGN same codebook model: η^* vs E_b/N_0 for $\mu = 0.006$. The solid lines correspond to $\eta^* \equiv \eta^*_M$ from (167). The dashed lines correspond to η^*_∞ defined in (171).

B. Future work

There are a lot of interesting directions for future work. A natural extension, already undertaken for the setting of random access [72], would be to analyze the many-user massive MIMO fading channel with receiver having $N > 1$ antennas under different fading scenarios (block-fading and quasi-static fading would probably be most relevant). Further, different asymptotic scalings of N and n may lead to radically different tradeoffs. Another interesting direction is to see how much improvements would the (so far incomputable) vector-AMP (and replica-method) yield over the bounds presented in this work. Note that for practical systems the asymptotic limit $n \to \infty$ is less relevant than finite-n bounds. However, AMP based bounds are inherently asymptotic. In this regard, it would be interesting to derive finite-blocklength versions of achievability bounds. Theorems IV.1 and IV.4 should be possible to "de-asymptotize". Similarly, recent bounds based on Gaussian processes [5] should be extendable to both the quasi-static fading and finite blocklength. From a practical standpoint, the most pressing issue is finding *any* architecture for the many-MAC setting that would have an approximately perfect MUI cancellation (i.e. have vertical part in the $\mathcal{E}$ vs μ tradeoff). One imagines such a system should be comprised of a message-passing decoding alternating between interference cancellation and signal re-estimation – as for example proposed in [73].

APPENDIX A
PROOFS OF SECTION III

A. Joint error

Proof of theorem III.1. Let

$$R = (R_1, ..., R_K) \in C_{\epsilon,J}.$$

We need to show that there exists a sequence of

$$\Big((M_1^{(n)}, M_2^{(n)}, ..., M_K^{(n)}), n, \epsilon_n\Big)_J$$

codes with projection decoding, such that

$$\liminf_{n\to\infty} \frac{1}{n} \log\left(M_i^{(n)}\right) \geq R_i, \forall i \in [K] \tag{174}$$

$$\limsup_{n\to\infty} \epsilon_n \leq \epsilon. \tag{175}$$

Let $\eta_i > 0, i \in [K]$. Choose

$$M_i^{(n)} = \lfloor 2^{n(R_i-\eta_i)} \rfloor, \forall i \in [K].$$

We use random coding: user j, independently generates $M_j^{(n)}$ vectors, each independently and uniformly distributed on the $\sqrt{nP}$–complex sphere. That is $X_i \stackrel{iid}{\sim} \text{Unif}\left(\sqrt{nP}(\mathcal{CS})^{n-1}\right)$. Hence the channel inputs are given by

$$X_i^{(n)} \stackrel{iid}{\sim} \left(\sqrt{nP}(\mathcal{CS})^{n-1}\right).$$

We will drop the superscript n for brevity.

Suppose the codewords $(c_1, c_2, ..., c_K) \in \prod_{i=1}^K \mathcal{C}_i$ were actually sent. Then by (14), error occurs iff $\exists (c_1', c_2', ..., c_K') \in \mathcal{C}_1 \times \mathcal{C}_2 ... \times \mathcal{C}_K$ such that $(c_1', c_2', ..., c_K') \neq (c_1, c_2, ..., c_K)$ and

$$\left\|P_{c_1',...,c_K'} Y\right\|^2 \geq \left\|P_{c_1,...,c_K} Y\right\|^2. \tag{176}$$

This can be equivalently written as follows. Let $S \subset [K]$ be such that

$$i \in [S] \iff \hat{c}_i \neq c_i \tag{177}$$

where $(\hat{c}_i)_{i=1}^K$ denote the decoded codewords.

Let $c_{[S]} \equiv \{c_i : i \in [S]\}$. Then, error occurs iff $\exists S \subset [K]$ and $S \neq \emptyset$, and $\exists \{c_i' : i \in [S], c_i' \neq c_i\}$ such that

$$\left\|P_{c'_{[S]}, c_{[S^c]}} Y\right\|^2 \geq \left\|P_{c_{[[K]]}} Y\right\|^2. \tag{178}$$

Let $B_S = \left\{\left\|P_{c'_{[S]}, c_{[S^c]}} Y\right\|^2 \geq \left\|P_{c_{[[K]]}} Y\right\|^2\right\}$ (here primes denote unsent codewords i.e., c_i' here means that it is independent of the channel inputs/output and distributed with the same law as c_i). Note that, for the sake of brevity, we are suppressing the dependence on c'.

So, the average probability of error is given by

$$\begin{aligned}\epsilon_n &= \mathbb{P}\Bigg[\bigcup_{\substack{S\subset[K]\\S\neq\emptyset}} \bigcup_{\substack{\{c_i'\in\mathcal{C}_i:\\ i\in S, c_i'\neq c_i\}}} B_S\Bigg] \\ &= \mathbb{P}\Bigg[\bigcup_{t\in[K]} \bigcup_{\substack{S\subset[K]\\|S|=t}} \bigcup_{\substack{\{c_i'\in\mathcal{C}_i:\\ i\in S, c_i'\neq c_i\}}} B_S\Bigg] \end{aligned}\tag{179}$$

Using ideas similar to the Random Coding Union (RCU) bound [74], we have

$$\begin{aligned}\epsilon_n \leq \mathbb{E}\Bigg[\min\Bigg\{1, \sum_{t\in[K]} \sum_{S\subset[K]:|S|=t} \left(\prod_{j\in S}(M_j-1)\right) \cdot \\ \mathbb{P}\Big[B_S \mid c_{[K]}, H_{[K]}, Z\Big]\Bigg\}\Bigg]\end{aligned} \tag{180}$$

where $H_{[K]} = \{H_i : i \in [K]\}$.

From now on we denote

$$\bigcup_{t\in[K]} \bigcup_{\substack{S\subset[K]\\|S|=t}} \equiv \bigcup_{t,S}$$

$$\sum_{t\in[K]} \sum_{S\subset[K]:|S|=t} \equiv \sum_{t,S}$$

$$\bigcap_{t\in[K]} \bigcap_{\substack{S\subset[K]\\|S|=t}} \equiv \bigcap_{t,S}.$$

Claim 7. *For $t \in [K]$ and $S \subset [K]$ with $|S| = t$,*

$$\begin{aligned}&\mathbb{P}\left[\left\|P_{c'_{[S]}, c_{[S^c]}} Y\right\|^2 > \left\|P_{c_{[[K]]}} Y\right\|^2 \middle| c_{[K]}, H_{[K]}, Z\right] \\ &= F\left(\frac{\|Y\|^2 - \left\|P_{c_{[[K]]}} Y\right\|^2}{\|Y\|^2 - \left\|P_{c_{[S^c]}} Y\right\|^2}; n-K, t\right)\end{aligned} \tag{181}$$

where $F(x; a, b)$ is the cdf of beta distribution $Beta(a, b)$. Further, from [21], we have

$$F(x; n-K, t) \leq (n-K+t-1)^{t-1} x^{n-K} \tag{182}$$

Proof. See proof of claim 1. ☐

Letting $G_S \equiv g(Y, c_{[K]}, S) = \frac{\|Y\|^2 - \left\|P_{c_{[[K]]}} Y\right\|^2}{\|Y\|^2 - \left\|P_{c_{[S^c]}} Y\right\|^2}$, $M_S = \prod_{j\in S}(M_j - 1)$, $s_t = (t-1)\frac{\ln(n-K+t-1)}{(n-K)}$ and $r_t = s_t + \frac{\ln M_S}{(n-K)}$, we have the following from (180), (181) and (182)

$$\epsilon_n \leq \mathbb{E}\left[\min\left\{1, \sum_{t,S} \exp\left(-(n-K)\left[-r_t - \ln G_S\right]\right)\right\}\right] \tag{183}$$

Let $\delta > 0$ and let E_1 be the following event

$$E_1 = \bigcap_{t,S} \{-\ln G_S - r_t > \delta\} \tag{184}$$

$$= \bigcap_{t,S} \left\{-\ln G_S > \tilde{V}_{n,S}\right\}$$

$$= \bigcap_{t,S} \{G_S < V_{n,S}\} \tag{185}$$

where $\tilde{V}_{n,S} = r_t + \delta$ and $V_{n,S} = e^{-\tilde{V}_{n,S}}$. Note that $V_{n,S}$ depends on S and t.

Then, from (183) we have the following

Lemma A.1. *For the K-user MAC defined above, with the projection decoder, the average probability of error is upper bounded as*

$$\epsilon_n \leq \sum_{t,S} e^{-(n-K)\delta} + \mathbb{P}\left[\bigcup_{t,S} G_S \geq V_{n,S}\right] \tag{186}$$

Proof. By (183),

$$\begin{aligned}&\epsilon_n\\ &\leq \mathbb{E}\left[\min\left\{1, \sum_{t,S} e^{(-(n-K)[-r_t - \ln G_S])}\right\}(1[E_1] + 1[E_1^c])\right]\\ &\leq \sum_{t,S} e^{-(n-K)\delta} + \mathbb{P}[E_1^c] \end{aligned}\tag{187}$$

□

Hence, as $n \to \infty$, it is the second term in the above expression that potentially dominates.

Claim 8. *For $t \in [K]$, $S \subset [K]$ with $|S| = t$, we have*

$$\begin{aligned}&\mathbb{P}\left[G_S \geq V_{n,S}\right]\\ &\leq \mathbb{P}\left[\left\|(1-V_{n,S})P^{\perp}_{c_{[S^c]}} Z - V_{n,S} P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2\right.\\ &\qquad \left.\geq V_{n,S}\left\|P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2\right]\end{aligned}\tag{188}$$

where $P^{\perp}_{c_{[S^c]}}$ represents the orthogonal projection onto the orthogonal complement of the space spanned by $c_{[S^c]}$.

Proof. See proof of claim 2 □

To evaluate the above probability, we condition on $c_{[K]}$ and $H_{[K]}$. For ease of notation, we will not explicitly write the conditioning.

Let $\chi_2'(\lambda, d)$ denote the non-central chi-squared distributed random variable with non-centrality λ and degrees of freedom d. That is, if $Z_i \sim \mathcal{N}(\mu_i, 1), i \in [d]$ and $\lambda = \sum_{i\in[d]} \mu_i^2$, then $\chi_2'(\lambda, d)$ has the same distribution as that of $\sum_{i\in[d]} Z_i^2$.

Since $Z \sim \mathcal{CN}(0, I_n)$, we have

$$Z - \frac{V_{n,S}}{1-V_{n,S}} \sum_{i\in S} H_i c_i \sim \mathcal{CN}(-\frac{V_{n,S}}{1-V_{n,S}} \sum_{i\in S} H_i c_i, I_n).$$

Hence

$$\begin{aligned}&P^{\perp}_{c_{[S^c]}}\left(Z - \frac{V_{n,S}}{1-V_{n,S}} \sum_{i\in S} H_i c_i\right)\\ &\sim \mathcal{CN}(-\frac{V_{n,S}}{1-V_{n,S}} P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i, P^{\perp}_{c_{[S^c]}}).\end{aligned}$$

Now using the fact that if $W = P + iQ \sim \mathcal{CN}(\mu, \Gamma, 0)$ then

$$\begin{bmatrix} P \\ Q \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} Re(\mu) \\ Im(\mu) \end{bmatrix}, \frac{1}{2}\begin{bmatrix} Re(\Gamma) & -Im(\Gamma) \\ Im(\Gamma) & Re(\Gamma) \end{bmatrix}\right) \tag{189}$$

we can show the following

Lemma A.2. *Let $F = \left\|\frac{V_{n,S}}{1-V_{n,S}} P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2$ and $n' = n - K + t$. Conditioned on $H_{[K]}$ and $c_{[K]}$, we have*

$$\left\|P^{\perp}_{c_{[S^c]}}\left(Z - \frac{V_{n,S}}{1-V_{n,S}} \sum_{i\in S} H_i c_i\right)\right\|^2 \sim \frac{1}{2}\chi_2'\left(2F, 2n'\right) \tag{190}$$

Hence its conditional expectation is

$$\mu = n' + F. \tag{191}$$

Proof. See proof of claim 3. □

Let $U = \frac{V_{n,S}}{(1-V_{n,S})}\left\|P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2 - n'$. Hence $F = \frac{V_{n,S}}{1-V_{n,S}}(U + n')$. Note that $V_{n,S}$, U ,λ all depend on t and S.

Letting $T = \frac{1}{2}\chi_2'(2F, 2n') - (F + n')$, we have,

$$\begin{aligned}&\mathbb{P}\left[\bigcup_{t,S}\left\|P^{\perp}_{c_{[S^c]}}\left(Z - \frac{V_{n,S}}{1-V_{n,S}} \sum_{i\in S} H_i c_i\right)\right\|^2\right.\\ &\quad\left.\geq \frac{V_{n,S}}{(1-V_{n,S})^2}\left\|P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2\right]\\ &= \mathbb{P}\left[\bigcup_{t,S}\left\|P^{\perp}_{c_{[S^c]}}\left(Z - \frac{V_{n,S}}{1-V_{n,S}} \sum_{i\in S} H_i c_i\right)\right\|^2 - \mu \geq U\right]\\ &= \mathbb{E}\left[\mathbb{P}\left[\left\{\bigcup_{t,S}\{T \geq U\}\right\}\middle| c_{[K]}, H_{[K]}\right]\right].\end{aligned}\tag{192}$$

Next we use lemma IV.2 to bound (192).

First, note that

$$\begin{aligned}U &= \frac{V_{n,S}}{(1-V_{n,S})}\left\|P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2 - n'\\ &= \frac{n'}{1-V_{n,S}}\left(V_{n,S}\left(1 + \frac{\left\|P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2}{n'}\right) - 1\right)\\ &= n' U^1\end{aligned}\tag{193}$$

where

$$U^1 = \frac{1}{1-V_{n,S}}\left(V_{n,S} W_S - 1\right)$$

and

$$W_S = \left(1 + \frac{\left\|P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2}{n'}\right).$$

Hence

$$F = \frac{V_{n,S}}{1-V_{n,S}}(U + n') = n' \frac{V_{n,S}}{1-V_{n,S}}(U^1 + 1). \tag{194}$$

Let $\delta_1 > 0$. Let $E_{11} = \bigcap_{S,t}\{U^1 > \delta_1\}$. From (192) we have

$$\begin{aligned}&\mathbb{P}\left[\bigcup_{t,S} T \geq U\right]\\ &\leq \sum_{t,S} \mathbb{E}\left[\mathbb{P}\left[\{T \geq U\}\middle| c_{[K]}, H_{[K]}\right] 1[E_{11}]\right]\\ &\quad + \mathbb{P}\left[E_{11}^c\right]\end{aligned}$$

$$\leq \sum_{t,S} \mathbb{E}\left[\mathbb{P}\left[\{T \geq U\}\middle| c_{[K]}, H_{[K]}\right] 1[U^1 > \delta_1]\right] + \mathbb{P}\left[E_{11}^c\right]$$
$$\leq \sum_{t,S} \mathbb{E}\left[e^{-n' f_n(U^1)} 1[U^1 > \delta_1]\right] + \mathbb{P}\left[E_{11}^c\right] \quad (195)$$

where the last inequality follows from (66), and

$$f_n(x) = x + 1 + \frac{2V_{n,S}}{1 - V_{n,S}}(1+x) - \sqrt{1 + \frac{2V_{n,S}}{1-V_{n,S}}(1+x)}\sqrt{2x + 1 + \frac{2V_{n,S}}{1-V_{n,S}}(1+x)} \quad (196)$$

Now, from claim 4, we have that for $0 < V_{n,S} < 1$ and $x > 0$, $f_n(x)$ is a monotonically increasing function of x

Hence we have

$$\mathbb{P}\left[\bigcup_{t,S}\{T \geq U\}\right] \leq \sum_{t,S} e^{-n' f_n(\delta_1)} + \mathbb{P}\left[E_{11}^c\right]. \quad (197)$$

So, we have the following claim

Claim 9. *Let $A_S = \{V_{n,S} W_S - 1 \leq \delta_1\}$ and $E_{12} = \bigcup_{t,S} A_S$. If $0 < V_{n,S} < 1$ for all $t \in [K]$, $S \subset [K]$ with $|S| = t$ then we have*

$$\mathbb{P}\left[\bigcup_{t,S}\{G_S \geq V_{n,S}\}\right] \leq \sum_{t,S} e^{-n' f_n(\delta_1)} + \mathbb{P}\left[E_{12}\right]. \quad (198)$$

Proof.

$$\mathbb{P}\left[\bigcup_{t,S}\{G_S \geq V_{n,S}\}\right] \leq \sum_{t,S} e^{-n' f_n(\delta_1)} + \mathbb{P}\left[E_{11}^c\right] \leq \sum_{t,S} e^{-n' f_n(\delta_1)} + \mathbb{P}\left[E_{12}\right]. \quad (199)$$

□

Now, we need to upper bound $\mathbb{P}\left[E_{12}\right]$.

We have

$$\left\|P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\right\|^2 = \sum_{i \in S} |H_i|^2 \left\|P^{\perp}_{c_{[S^c]}} c_i\right\|^2 + 2\sum_{i<j: i,j\in S} Re\left(\left\langle P^{\perp}_{c_{[S^c]}} c_i, P^{\perp}_{c_{[S^c]}} c_j\right\rangle H_i \bar{H}_j\right). \quad (200)$$

Further,

$$\left\langle P^{\perp}_{c_{[S^c]}} c_i, P^{\perp}_{c_{[S^c]}} c_j\right\rangle = \langle c_i, c_j\rangle - \left\langle P_{c_{[S^c]}} c_i, P_{c_{[S^c]}} c_j\right\rangle \quad (201)$$

Hence we have

$$\left|Re\left\langle P^{\perp}_{c_{[S^c]}} c_i, P^{\perp}_{c_{[S^c]}} c_j\right\rangle\right| \leq \left|\left\langle P^{\perp}_{c_{[S^c]}} c_i, P^{\perp}_{c_{[S^c]}} c_j\right\rangle\right| \leq |\langle c_i, c_j\rangle| + \left|\left\langle P_{c_{[S^c]}} c_i, P_{c_{[S^c]}} c_j\right\rangle\right|$$
$$\leq |\langle c_i, c_j\rangle| + \left\|P_{c_{[S^c]}} c_j\right\| \left\|P_{c_{[S^c]}} c_i\right\| = nP\left(|\langle \hat{c}_i, \hat{c}_j\rangle| + \left\|P_{c_{[S^c]}} \hat{c}_i\right\| \left\|P_{c_{[S^c]}} \hat{c}_j\right\|\right) \quad (202)$$

where hats denote corresponding normalized vectors. Since these unit vectors are high dimensional, their dot products and projection onto a smaller, fixed dimension surface is very small. Indeed, we have the following two lemmas.

Lemma A.3. *If $e_1, e_2 \overset{iid}{\sim} Unif((\mathcal{CS})^{n-1})$, then for any $\delta_2 > 0$, we have*

$$\mathbb{P}\left[|\langle e_1, e_2\rangle| > \delta_2\right] \leq 4e^{-\frac{n\delta_2^2}{2}} \quad (203)$$

Proof. First, lets take $e_1, e_2 \overset{iid}{\sim} S^{n-1}$. Let x be a fixed unit vector in $\mathbb{R}^n$. Due to symmetry, we have $\mathbb{P}\left[\langle e_1, x\rangle \geq 0\right] = 1/2$. Hence, by Levy's Isoperimetric inequality on the sphere [75], we have

$$\mathbb{P}\left[\langle e_1, x\rangle > \delta_2\right] \leq e^{-n\delta_2^2/2}. \quad (204)$$

Again by symmetry, and then taking x as e_2, we have

$$\mathbb{P}\left[|\langle e_1, e_2\rangle| > \delta_2\right] \leq 2e^{-n\delta_2^2/2}. \quad (205)$$

Now uniform distribution on $(\mathcal{CS})^{n-1}$ is same as the uniform distribution on S^{2n-1}, and for complex vectors $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ we have $Re\langle z_1, z_2\rangle = x_1^T x_2 + y_1^T y_2 = (x_1, y_1)^T (x_2, y_2)$. Hence if $e_1, e_2 \overset{iid}{\sim} (\mathcal{CS})^{n-1}$, and $u_1, u_2 \overset{iid}{\sim} S^{2n-1}$ then $Re\langle e_1, e_2\rangle$ has same law as $\langle u_1, u_2\rangle$. Hence we have

$$\mathbb{P}\left[|Re\langle e_1, e_2\rangle| > \delta_2\right] \leq 2e^{-\frac{2n\delta_2^2}{2}}. \quad (206)$$

Also, $Im\langle z_1, z_2\rangle = x_1^T y_2 - y_1^T x_2$. Hence $Im\langle e_1, e_2\rangle$ has the same law as $Re\langle e_1, e_2\rangle$. Hence we have

$$\mathbb{P}\left[|\langle e_1, e_2\rangle| > \delta_2\right] = \mathbb{P}\left[|\langle e_1, e_2\rangle|^2 > \delta_2^2\right] = \mathbb{P}\left[|Re\langle e_1, e_2\rangle|^2 + |Im\langle e_1, e_2\rangle|^2 > \delta_2^2\right]$$
$$\leq \mathbb{P}\left[|Re\langle e_1, e_2\rangle| > \frac{\delta_2}{\sqrt{2}}\right] + \mathbb{P}\left[|Im\langle e_1, e_2\rangle| > \frac{\delta_2}{\sqrt{2}}\right] \leq 4e^{-\frac{n\delta_2^2}{2}}. \quad (207)$$

□

Next we have a similar lemma for low dimensional projections from [76, Lemma 5.3.2]

Lemma A.4 ([76]). *Let $x \sim Unif(S^{n-1})$ and P be a projection to an m dimensional subspace of $\mathbb{R}^n$. Then for any $\delta_3 > 0$, we have*

$$\mathbb{P}\left[\left|\|Px\| - \sqrt{\frac{m}{n}}\right| > \delta_3\right] \leq 2e^{-cn\delta_3^2} \quad (208)$$

where c is some absolute constant. Hence, by symmetry, the result remains true if P is a uniform random projection, independent of x.

Now we need to prove that a similar result holds for the complex variable case as well. We have the following lemma

Lemma A.5. *Let $z \sim Unif(\mathcal{CS})^{n-1}$ and P be a projection to an m dimensional subspace V of $\mathbb{C}^n$. Then for any $\delta_3 > 0$, we have*

$$\mathbb{P}\left[\left|\|Pz\| - \sqrt{\frac{m}{n}}\right| > \delta_3\right] \leq 2e^{-2cn\delta_3^2} \tag{209}$$

where c is some absolute constant. Hence, by symmetry, the result remains true if P is a uniform random projection, independent of z.

Proof. Consider $\|Pz\|$. Let U be the unitary change of basis matrix which converts V to first m coordinates. Hence $\|Pz\| = \|UPz\|$. Therefore we can just consider the orthogonal projection onto first m coordinates. Hence the projection matrix P is real. Let $e_1, ..., e_m$ be the standard basis corresponding to the first m coordinates. Let A be the $n \times m$ matrix whose columns are $e_1, ..., e_m$. Then $P = AA^*$ ($*$ denotes conjugate transpose). Since A is real, we have $Re(Pz) = AA^* Re(z)$ and $Im(Pz) = AA^* Im(z)$. Now, if $z \sim Unif((\mathcal{CS})^{n-1})$ then $Re(z)$ has same law as $Im(z)$. Hence $Re(Pz)$ has same law as $Im(Pz)$. Further $A^* = A^T$. Also note that, if $z = x + iy$ then $\|Pz\|^2 = z^* AA^* Z = x^T AA^T x + y^T AA^T y = \begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} AA^T & 0 \\ 0 & AA^T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \left\|\hat{P}\begin{bmatrix} x \\ y \end{bmatrix}\right\|$ where $\hat{P}$ denotes the orthogonal projection from $\mathbb{R}^{2n}$ to a $2m$ dimensional subspace. Hence $\|Pz\|^2$ has the same law as that of the projection of a uniform random vector on S^{2n-1} to a $2m$ dimensional subspace. Hence using lemma A.4, we have

$$\mathbb{P}\left[\left|\|Pz\| - \sqrt{\frac{m}{n}}\right| > \delta_3\right] \leq 2e^{-2cn\delta_3^2} \tag{210}$$

□

Since $H_i \sim \mathcal{CN}(0, 1)$, we have $|H_i|^2 \sim \frac{1}{2}\chi_2(2) = \exp(1)$ where $\chi_2(d)$ denotes the chi-squared distribution with d degrees of freedom and $\exp(1)$ represents an exponentially distributed random variable with rate 1. Therefore, for $\nu \geq 0$,

$$\mathbb{P}\left[|H_i|^2 \geq \nu\right] = e^{-\nu} \tag{211}$$

Now, we are in a position to bound $\mathbb{P}[E_{12}]$.

For $S \subset [K]$ with $|S| = t$, define the events E_2, E_3 and E_4 as follows:

$$E_2 = \bigcap_{i<j: i,j \in [K]} \{|\langle \hat{c}_i, \hat{c}_j \rangle| \leq \delta_2\} \tag{211a}$$

$$E_3(S,t) = \bigcap_{i \in S} \left\{\left|\left\|P_{c_{[S^c]}} \hat{c}_i\right\| - \sqrt{\frac{K-t}{n}}\right| \leq \delta_3\right\} \tag{212a}$$

$$E_4 = \bigcap_{i \in [K]} \{|H_i|^2 \leq \nu\} \tag{212b}$$

where we choose $\delta_2 = n^{-\frac{1}{3}} = \delta_3$ and $\nu = n^{\frac{1}{4}}$. Hence we have

$$\mathbb{P}[E_{12}] \leq \mathbb{P}\left[\bigcup_{t,S} (A_S \cap E_2 \cap E_3 \cap E_4)\right] + \mathbb{P}[E_2^c] + \mathbb{P}[E_4^c] + \sum_{t,S} \mathbb{P}[E_3^c(S,t)]. \tag{213}$$

Using lemmas A.3 and A.5 and eq. (211), we have

$$\mathbb{P}[E_2^c] + \mathbb{P}[E_4^c] + \sum_{t,S} \mathbb{P}[E_3^c(S,t)] \leq 2K(K-1)e^{-\frac{n\delta_2^2}{2}} + Ke^{-\nu} + \sum_{t,S} 2te^{-cn\delta_3^2}. \tag{214}$$

Note that the above quantity goes to 0 as $n \to \infty$ due to the choice of δ_2, δ_3 and ν. Also, the choice of parameters is not the optimum. Nevertheless, this is enough to prove the result.

Let

$$\delta_{1,t} = \left(\delta_2 + \left(\delta_3 + \sqrt{\frac{K-t}{n}}\right)^2\right)$$

$$\delta_{2,t} = \left(\delta_3 + \sqrt{\frac{K-t}{n}}\right)^2$$

Observe that on the sets E_2, E_3 and E_4, we have from (202)

$$\left|Re\left\langle P_{c_{[S^c]}}^{\perp} c_i, P_{c_{[S^c]}}^{\perp} c_j\right\rangle H_i \bar{H}_j\right| \leq \nu\delta_{1,t} = O(n^{-\frac{1}{12}}) \tag{215a}$$

$$|H_i|^2 \left\|P_{c_{[S^c]}} \hat{c}_i\right\|^2 \leq \nu\delta_{2,t} = O(n^{-\frac{5}{12}}) \tag{215b}$$

So we have

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{t,S} (A_S \cap E_2 \cap E_3 \cap E_4)\right] \\
&\leq \mathbb{P}\left[\bigcup_{t,S}\left\{V_{n,S}\left[1 + \frac{nP}{n'}\left\{\sum_{i \in S} |H_i|^2 \|\hat{c}_i\|^2 - t\nu\delta_{2,t} - t(t-1)\delta_{1,t}\right\}\right] - 1 \leq \delta_1\right\}\right] \\
&\leq \mathbb{P}\left[\bigcup_{t,S}\left\{V_{n,S}\left[1 + \frac{nP}{n'}\sum_{i \in S} |H_i|^2\right] \leq 1 + \delta_1 + O(n^{-\frac{1}{12}})\right\}\right] \\
&\leq \mathbb{P}\left[\bigcup_{t,S}\left\{\ln\left[1 + P\sum_{i \in S} |H_i|^2\right] \leq V'_{n,S}\right\}\right]
\end{aligned} \tag{216}$$

where $V'_{n,S} = \tilde{V}_{n,S} + \ln(1 + \delta_1 + O(n^{-1/12}))$, and O depends on K and t.

Let $\delta_n = \ln(1 + \delta_1 + O(n^{-1/12}))$. We have $\frac{\log M_S}{n-K} = \left(\sum_{i \in S}(R_i - \eta_i)\right)(1 + o(1))$ and $s_t = O\left(\frac{\log n}{n}\right)$.

By the choice of $M_i^{(n)}$, for sufficiently large n, sufficiently small δ and δ_1, we have

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{t,S}\left\{\ln\left[1 + P\sum_{i \in S} |H_i|^2\right] \leq V'_{n,S}\right\}\right] \\
&= \mathbb{P}\left[\bigcup_{t,S} \ln\left[1 + P\sum_{i \in S} |H_i|^2\right] \leq s_t + \frac{\ln M_S}{n-K} + \delta + \delta_n\right] \\
&= \mathbb{P}\left[\bigcup_{t,S}\left\{\log\left[1 + P\sum_{i \in S} |H_i|^2\right] \leq s_t \log_2(e) + \left(\sum_{i \in S}(R_i - \eta_i)\right)(1 + o(1)) + (\delta + \delta_n)\log_2(e)\right\}\right]
\end{aligned}$$

$$\leq \mathbb{P}\left[\bigcup_{t,S}\left\{\log\left[1+P\sum_{i\in S}|H_i|^2\right]\leq\left(\sum_{i\in S}R_i\right)\right\}\right] \tag{217}$$

Finally combining everything, we have

$$\begin{aligned}&\epsilon_n\\ &\leq \mathbb{P}\left[\bigcup_{t,S}\left\{\log\left[1+P\sum_{i\in S}|H_i|^2\right]\leq s_t\log_2(e)\right.\right.\\ &\quad\left.\left.+\left(\sum_{i\in S}(R_i-\eta_i)\right)(1+o(1))+(\delta+\delta_n)\log_2(e)\right\}\right]\\ &\quad+2K(K-1)e^{-\frac{n^{1/3}}{2}}+Ke^{-n^{1/4}}\\ &\quad+\sum_{t,S}\left[2te^{-cn^{1/3}}+e^{-\delta(n-K)}+e^{-nf_n(\delta_1)}\right].\end{aligned} \tag{218}$$

Therefore for this choice of $\left(M_i^{(n)}\right)$, from (217) we have

$$\begin{aligned}&\limsup_{n\to\infty}\epsilon_n\\ &\leq \mathbb{P}\left[\bigcup_{t,S}\left\{\log\left[1+P\sum_{i\in S}|H_i|^2\right]\leq\left(\sum_{i\in S}R_i\right)\right\}\right]\\ &\leq\epsilon\end{aligned} \tag{219}$$

Since $\eta_i>0$ were arbitrary, we are done. That is (175) is also satisfied.

$\square$

B. Per-user error

Proof of theorem III.2. We need to show that there exists a sequence of $\left((M_1^{(n)},M_2^{(n)},...,M_K^{(n)}),n,\epsilon_n\right)_{PU}$ codes with the decoder given by (17) and (18) such that

$$\liminf_{n\to\infty}\frac{1}{n}\log\left(M_i^{(n)}\right)\geq R_i,\forall i\in[K] \tag{220}$$

$$\limsup_{n\to\infty}\epsilon_n\leq\epsilon. \tag{221}$$

Let $P_e^S(R)<\epsilon$ and $\eta_i>0, i\in[K]$. Choose $M_i^{(n)}=\lceil e^{n(R_i-\eta_i)}\rceil,\forall i\in[K]$. We use random coding with Gaussian codebooks: user i generates M_i codewords $\{c_j^i : j\in[M_i]\}\overset{iid}{\sim}\mathcal{CN}(0,P_n'I_n)$ independent of other users, where $P_n'=\frac{P}{1+n^{-\frac{1}{3}}}$. Here $\mathcal{CN}(\mu,\Sigma)$ denotes the complex normal distribution with mean μ, covariance Σ and pseudo-covariance 0. For the (random) message $W_i\in[M_i]$, user i transmits $X_i=c_{W_i}^i 1\{\left\|c_{W_i}^i\right\|^2>nP\}$. The channel model is given in (6) and the decoder is given by (17) and (18). The per-user probability of error is given by (9)

$$P_e=\mathbb{E}\left[\frac{1}{K}\sum_{j=1}^K 1\left\{W_j\neq(g_D(Y))_j\right\}\right]. \tag{222}$$

Similar to the proof of [4, Theorem 1], we change the measure over which $\mathbb{E}$ is taken in (222) to the one where $X_i=c_{W_i}^i$ at the cost of adding a total variation distance. Hence the probability of error under this change of measure becomes

$$P_e\leq p_1+p_0$$

with

$$p_0=K\mathbb{P}\left[\|w\|^2>n\frac{P}{P_n'}\right] \tag{223}$$

$$p_1=\mathbb{E}\left[\frac{1}{K}\sum_{j=1}^K 1\left\{W_j\neq(g_D(Y))_j\right\}\right] \tag{224}$$

where $w\sim\mathcal{CN}(0,I_n)$ and, with abuse of notation, $\mathbb{E}$ in p_1 is taken over the new measure. It can be easily seen that by the choice of P_n' and lemma IV.2, $p_0\to 0$ as $n\to\infty$. From now on, we exclusively focus on bounding p_1.

p_1 can also be written as

$$\begin{aligned}p_1&=\frac{1}{K}\mathbb{E}\left[\sum_{i\in D}1\left\{W_j\neq(g_D(Y))_j\right\}+|D^c|\right]\\ &=1-\frac{\mathbb{E}[D]}{K}+\frac{1}{K}\mathbb{E}\left[\sum_{i\in D}1\left\{W_j\neq(g_D(Y))_j\right\}\right]\end{aligned} \tag{225}$$

because, for $i\in D^c$, $1\left\{W_j\neq(g_D(Y))_j\right\}=1,\ a.s.$ Define p_2 as

$$p_2=\mathbb{P}\left[\sum_{i\in D}1\left\{W_j\neq(g_D(Y))_j\right\}>0\right]. \tag{226}$$

So, it's enough to show that $p_2\to 0$ as $n\to\infty$. This is because, if $p_2\to 0$, then the non-negative random variables $A_n=\sum_{i\in D}1\left\{W_j\neq(g_D(Y))_j\right\}$ converge to 0 in probability. Since $A_n\leq K,\ a.s$, we have, by dominated convergence, $\mathbb{E}[A_n]=\mathbb{E}\left[\sum_{i\in D}1\left\{W_j\neq(g_D(Y))_j\right\}\right]\to 0$. To this end, we upper bound p_2.

Let $c=(c_1\in\mathcal{C}_1,...,c_K\in\mathcal{C}_K)$ be the tuple of sent codewords. Let $K_1=|D|$. Let $c_{(D)}$ denote the ordered tuple corresponding to indices in D. That is, if $i_1<i_2<...<i_{K_1}$ are the elements of D, then $(c_{(D)})_j=c_{i_j},\forall j\in[K_1]$. Let $B_S=\left\{\left\|P_{c'_{[S]},c_{[S^c]}}Y\right\|^2>\left\|P_{c_{[D]}}Y\right\|^2\right\}$. Then p_2 can also be written as

$$\begin{aligned}p_2&=\mathbb{P}\left[\sum_{i\in D}1\left\{W_j\neq(g_D(Y))_j\right\}>0\right] &(227)\\ &=\mathbb{P}\left[\exists S\subset D,S\neq\emptyset:\forall i\in S,(g_D(Y))_i\neq W_i\right] &(228)\\ &=\mathbb{P}\left[\exists c'_{(D)}\neq c_{(D)}:\left\|P_{c'_{[D]}}Y\right\|^2>\left\|P_{c_{[D]}}Y\right\|^2\right] &(229)\\ &=\mathbb{P}\left[\bigcup_{t\in[K_1]}\bigcup_{\substack{S\subset D\\|S|=t}}\bigcup_{\substack{c_i'\in\mathcal{C}_i\setminus\{c_i\}\\i\in S}}B_S\right]. &(230)\end{aligned}$$

Let $\delta>0$, $G_S\equiv g(Y,c_{[K]},S,D)=\frac{\|Y\|^2-\left\|P_{c_{[D]}}Y\right\|^2}{\|Y\|^2-\left\|P_{c_{[S^c]}}Y\right\|^2}$, $M_S=\prod_{j\in S}(M_j-1)$, $s_t=(t-1)\frac{\ln(n-K_1+t-1)}{n-K_1}$, $r_t=s_t+\frac{\ln M_S}{n-K_1}$, $\tilde{V}_{n,S}=r_t+\delta$ and $V_{n,S}=e^{-\tilde{V}_{n,S}}$. Denote $\bigcup_{t\in[K_1]}\bigcup_{\substack{S\subset D\\|S|=t}}$ as $\bigcup_{t,S,K_1}$, similarly for $\bigcap$ and $\sum$. Further, denote $\bigcup_{t\in[K]}\bigcup_{\substack{S\subset[K]\\|S|=t}}$ as $\bigcup_{t,S}$, again, similarly for $\bigcap$ and $\sum$.

Note that, since D is random, both M_S and $V_{n,S}$ are random. But in the symmetric case only M_S is not random.

Now, following steps similar to (180), (181), (183) and (186), we have

$$p_2 \leq \mathbb{E}\bigg[\sum_{t,S,K_1} e^{-(n-K_1)\delta}\bigg] + \mathbb{P}\bigg[\bigcup_{t,S,K_1} G_S \geq V_{n,S}\bigg] \tag{231}$$

$$\leq \sum_{t,K} e^{-(n-K)\delta} + \mathbb{P}\bigg[\bigcup_{t,S,K_1} G_S \geq V_{n,S}\bigg]. \tag{232}$$

So, the first term goes to 0 as $n \to \infty$.

Let $Z_D = Z + \sum_{i \in D^c} H_i c_i$. It can be easily seen that, similar to (188), we have

$$\begin{aligned}&\mathbb{P}\Big[G_S \geq V_{n,S}\Big]\\ &\leq \mathbb{P}\bigg[\bigg\|(1-V_{n,S})P^{\perp}_{c_{[S^c]}} Z_D - V_{n,S} P^{\perp}_{c_{[S^c]}} \sum_{i\in S} H_i c_i\bigg\|^2 \\ &\qquad \geq V_{n,S}\bigg\|P^{\perp}_{c_{[S^c]}}\sum_{i\in S} H_i c_i\bigg\|^2\bigg].\end{aligned} \tag{233}$$

Now, conditional of $H_{[K]}$ and $c_{[D]}$, $Z_D \sim \mathcal{CN}(0, (1 + P'_n \sum_{i\in D^c}|H_i|^2))$. Hence

$$P^{\perp}_{c_{[S^c]}}\bigg(Z_D - \frac{V_{n,S}}{1-V_{n,S}}\sum_{i\in S} H_i c_i\bigg) \sim$$
$$\mathcal{CN}\bigg(-\frac{V_{n,S}}{1-V_{n,S}} P^{\perp}_{c_{[S^c]}}\sum_{i\in S} H_i c_i, (1 + P'_n \sum_{i\in D^c}|H_i|^2) P^{\perp}_{c_{[S^c]}}\bigg)$$

Therefore

$$\begin{aligned}&\bigg\|P^{\perp}_{c_{[S^c]}}\bigg(Z_D - \frac{V_{n,S}}{1-V_{n,S}}\sum_{i\in S} H_i c_i\bigg)\bigg\|^2\\ &\sim \bigg(1 + P'_n\sum_{i\in D^c}|H_i|^2\bigg)\frac{1}{2}\chi'_2\left(2F, 2n'\right)\end{aligned} \tag{234}$$

where

$$F = \frac{\Big\|\frac{V_{n,S}}{1-V_{n,S}} P^{\perp}_{c_{[S^c]}}\sum_{i\in S} H_i c_i\Big\|^2}{\left(1 + P'_n\sum_{i\in D^c}|H_i|^2\right)} \tag{235}$$

$$n' = n - K_1 + t. \tag{236}$$

Let

$$U = \frac{V_{n,S}}{(1-V_{n,S})}\frac{\Big\|P^{\perp}_{c_{[S^c]}}\sum_{i\in S} H_i c_i\Big\|^2}{\left(1 + P'_n\sum_{i\in D^c}|H_i|^2\right)} - n' \tag{237}$$

$$U^1 = \frac{1}{1-V_{n,S}}\left(V_{n,S}W_S - 1\right) \tag{238}$$

where $W_S = \left(1 + \frac{\left\|P^{\perp}_{c_{[S^c]}}\sum_{i\in S} H_i c_i\right\|^2}{n'\left(1+P'_n\sum_{i\in D^c}|H_i|^2\right)}\right)$

Hence $U = n'U^1$ and $F = \frac{V_{n,S}}{1-V_{n,S}} n'(1+U^1)$. So, similar to (192), we have

$$\mathbb{P}\left[\bigcup_{t,S,K_1}\{G_S \geq V_{n,S}\}\right] \leq \mathbb{P}\left[\bigcup_{t,S,K_1}\{T \geq U\}\right] \tag{239}$$

where $T = \frac{1}{2}\chi'_2(2F, 2n') - (F + n')$.

Let $\delta_1 > 0$ and $E_{11} = \bigcap_{t,S,K_1}\{U^1 > \delta_1\} \in \sigma(H_{[K]}, c_{[D]})$. Now, similar to (197), we have

$$\begin{aligned}&\mathbb{P}\bigg[\bigcup_{t,S,K_1}\{T \geq U\}\bigg]\\ &\leq \mathbb{E}\bigg[\sum_{t,S,K_1} e^{-n' f_n(\delta_1)}\bigg] + \mathbb{P}\left[E_{11}^c\right].\end{aligned} \tag{240}$$

where f_n (now a random function) was defined in (196). So, again by claim 4 and dominated convergence, the first term in (240) converges to 0 as $n \to \infty$. Next, we upper bound the second term $\mathbb{P}\left[E_{11}^c\right]$.

Let $A_S = \{V_{n,S}W_S - 1 \leq \delta_1\}$ and $E_{12} = \bigcup_{t,S,K_1} A_S$. Similar to (199), we have

$$\mathbb{P}\left[E_{11}^c\right] = \mathbb{P}\left[\bigcup_{t,S,K_1}\{U^1 \leq \delta_1\}\right] \leq \mathbb{P}\left[E_{12}\right]. \tag{241}$$

Let $\hat{c}_i = c_i/\|c_i\|$. Let $\delta_2 > 0$, $\delta_3 > 0$, $\delta_4 > 0$ and $\nu > 1$. Define the events

$$E_2 = \bigcap_{i<j: i,j\in[K]}\{|\langle \hat{c}_i, \hat{c}_j\rangle| \leq \delta_2\} \tag{242a}$$

$$E_3(S,t) = \bigcap_{i\in S}\left\{\left|\left\|P_{c_{[S^c]}}\hat{c}_i\right\| - \sqrt{\frac{K_1 - t}{n}}\right| \leq \delta_3\right\} \tag{242b}$$

$$E_4 = \bigcap_{i\in[K]}\left\{|H_i|^2 \leq \nu\right\} \tag{242c}$$

$$E_5 = \bigcap_{i\in[K]}\left\{|\,\|c_i\| - \sqrt{nP'_n} \leq \delta_4\sqrt{nP'}|\right\} \tag{242d}$$

and choose $\delta_2 = O(n^{-\frac{1}{3}}) = \delta_3 = \delta_4$ and $\nu = O(n^{1/4})$.

Using these events we can bound $\mathbb{P}\left[E_{11}^c\right]$ as

$$\begin{aligned}\mathbb{P}\left[E_{11}^c\right] &\leq \mathbb{P}\bigg[\bigcup_{t,S,K_1}\Big(A_S \cap E_2 \cap E_3(S,t) \cap E_4 \cap E_5\Big)\bigg]\\ &\quad + \mathbb{P}\left[E_2^c\right] + \mathbb{P}\left[E_4^c\right] + \mathbb{P}\left[E_5^c\right]\\ &\quad + \mathbb{E}\bigg[\sum_{t,S,K_1}\mathbb{P}\left[E_3^c(S,t)|H_{[K]}\right]\bigg].\end{aligned} \tag{243}$$

From [76, Theorem 3.1.1], we have

$$\mathbb{P}\left[E_5^c\right] \leq 2Ke^{-c_1 n\delta_4^2} \tag{244}$$

for some constant $c_1 > 0$. So, from lemma A.3, lemma A.5, (211) and (244), we have

$$\begin{aligned}\mathbb{P}\left[E_{11}^c\right] &\leq \mathbb{P}\left[\bigcup_{t,S,K_1}\left(A_S \cap E_2 \cap E_3(S,t) \cap E_4 \cap E_5\right)\right] +\\ &\quad 2K(K-1)e^{-\frac{n\delta_2^2}{2}} + +Ke^{-\nu} + 2Ke^{-c_1 n\delta_4^2}\\ &\quad + \sum_{t,S} 2te^{-cn\delta_3^2}.\end{aligned} \tag{245}$$

So, by the chose of $\delta_i, i \in \{2,3,4\}$ and ν, the exponential terms in the last expression go to 0 as $n \to \infty$.

Let $N = \left(1 + P'_n \sum_{i \in D^c} |H_i|^2\right)$,

$$\delta_{1,t} = \left(\delta_2 + \left(\delta_3 + \sqrt{\frac{K_1 - t}{n}}\right)^2\right),$$

$$\delta_{2,t} = \left(\delta_3 + \sqrt{\frac{K_1 - t}{n}}\right)^2.$$

Let $SINR_n = \frac{P'_n \sum_{i \in S} |H_i|^2}{N}$. Now, arguing similar to (216), we get

$$\begin{aligned}
&\mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Big(A_S \cap E_2 \cap E_3(S,t) \cap E_4 \cap E_5\Big)\Bigg] \\
&\leq \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Bigg\{V_{n,S}\Bigg[1+\Bigg\{\frac{n(1-\delta_4)^2}{n'}SINR_n \\
&\quad - (1+\delta_4)^2\left(\frac{nP'_n}{n-K}t\nu\delta_{2,t} + \frac{nP'_n t(t-1)}{n-K}\delta_{1,t}\right)\Bigg\}\Bigg] \\
&\quad - 1 \leq \delta_1\Bigg\}\Bigg] \qquad (246)\\
&\leq \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Big\{\ln\left[1 + SINR_n\right] \leq V'_{n,S}\Big\}\Bigg] \qquad (247)
\end{aligned}$$

where $V'_{n,S} = \tilde{V}_{n,S} + \ln(1 + \delta_1 + O(n^{-1/12}))$.

Let $\delta_n = \ln(1 + \delta_1 + O(n^{-1/12}))$. We have $\frac{\log M_S}{n-K} = \left(\sum_{i \in S}(R_i - \eta_i)\right)(1 + o(1))$. There for sufficiently large n and sufficiently small δ and δ_1, we have $V'_{n,S} \leq \sum_{i \in S} R_i$ *a.s.* Hence

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{t,S,K_1}\log\left[1 + SINR_n\right] \leq \log_2(e)V'_{n,S}\right] \\
&\quad \leq \mathbb{P}\left[\bigcup_{t,S,K_1}\log\left[1 + SINR_n\right] \leq \sum_{i \in S} R_i\right]. \qquad (248)
\end{aligned}$$

But we know that $P'_n \to P$, and on D, from (17) we have

$$\sum_{i \in S} R_i < \log\left(1 + \frac{P\sum_{i \in S}|H_i|^2}{1 + P\sum_{i \in D^c}|H_i|^2}\right) \ a.s. \qquad (249)$$

Hence the probability in (248) goes to 0 as $n \to \infty$.

So combining everything from (232), (241), (245), (246), (247), (248) and (249), we get $p_2 \to 0$ as $n \to \infty$. Therefor $p_1 \to 1 - \frac{\mathbb{E}[D]}{K}$ as $n \to \infty$. Hence we have

$$\epsilon_n = P_e \to 1 - \frac{\mathbb{E}\left[D\right]}{K} < \epsilon. \qquad (250)$$

Hence $\limsup_{n\to\infty} \epsilon_n \leq \epsilon$. Further, since $\eta_i > 0$ were arbitrary, we can ensure $\liminf_{n\to\infty} \frac{1}{n}\log M_i^{(n)} \geq R_i, \forall i \in [K]$.

□

C. Proof of proposition 1

Proof. We prove the second upper bound in (19). This is based on a single-user converse using the genie argument. Formally, since we consider per-user error, it is enough to look at the event that a particular user is not decoded. Let $W_i \overset{iid}{\sim} unif[M]$ be the message of user i. The channel (6) can be written as $Y = H_1X_1 + \hat{Z} + Z$ where $\hat{Z} = \sum_{i=2}^K H_iX_i$ denotes the interference. Let $L(Y)$ be the decoder output. Also, let $L(Y, \hat{Z})$ be the decoder output when it has knowledge of $\hat{Z}$. Hence a converse bound $\mathbb{P}\left[W_1 \neq (L(Y))_1\right] \geq \epsilon$ is implied by $\mathbb{P}\left[W_1 \neq \left(L(Y,\hat{Z})\right)_1\right] \geq \epsilon$ for all $L(\cdot,\cdot)$. Since $Y - \hat{Z}$ is a sufficient statistic of $(Y, \hat{Z})$ for W_1, we have, equivalently, $\mathbb{P}\left[W_1 \neq \left(L(Y - \hat{Z})\right)_1\right] \geq \epsilon$ for all $L(\cdot)$. Letting $\hat{Y} = Y - \hat{Z}$, this is equivalent to a converse for the channel $\hat{Y} = H_1X_1 + Z$: $\mathbb{P}\left[W_1 \neq \left(L(\hat{Y})\right)_1\right] \geq \epsilon$ for all $L(\cdot)$. This is just the usual single user converse, and hence the bound is given by $R \leq C_\epsilon = \sup\{\xi : \mathbb{P}\left[\log_2(1 + P|H_1|^2) \leq \xi\right] \leq \epsilon\} = \log_2(1 - P\ln(1-\epsilon))$[21].

□

APPENDIX B
PROOFS OF CERTAIN CLAIMS

Proof of claim 2. We have $\|Y\|^2 - \|P_{A_0}Y\|^2 = \left\|P_{A_0}^\perp \hat{Z}\right\|^2 \leq \left\|\hat{Z}\right\|^2 - \left\|P_{A_1}\hat{Z}\right\|^2 = \left\|P_{A_1}^\perp \hat{Z}\right\|^2$.

Also, $\left\|P_{A_1}^\perp Y\right\|^2 = \left\|P_{A_1}^\perp \sum_{i \in S_2^*} H_ic_i + P_{A_1}^\perp \hat{Z}\right\|^2$. Hence we have

$$\begin{aligned}
p_2 &= \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\left\{\frac{\|Y\|^2 - \|P_{A_0}Y\|^2}{\|Y\|^2 - \|P_{A_1}Y\|^2} \geq V_{n,t}\right\}\Bigg] \\
&= \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Bigg\{\left\|\hat{Z}\right\|^2 - \left\|P_{A_0}\hat{Z}\right\|^2 \\
&\qquad \geq V_{n,t}\left\|P_{A_1}^\perp \sum_{i \in S_2^*} H_ic_i + P_{A_1}^\perp \hat{Z}\right\|^2\Bigg\}\Bigg] \\
&\leq \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Bigg\{\left\|P_{A_1}^\perp \hat{Z}\right\|^2 \\
&\qquad \geq V_{n,t}\left\|P_{A_1}^\perp \sum_{i \in S_2^*} H_ic_i + P_{A_1}^\perp \hat{Z}\right\|^2\Bigg\}\Bigg] \\
&= \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Bigg\{(1 - V_{n,t})\left\|P_{A_1}^\perp \hat{Z}\right\|^2 \\
&\qquad - 2V_{n,t}Re\left\langle P_{A_1}^\perp \hat{Z}, P_{A_1}^\perp \sum_{i \in S_2^*} H_ic_i\right\rangle \\
&\qquad \geq V_{n,t}\left\|P_{A_1}^\perp \sum_{i \in S_2^*} H_ic_i\right\|^2\Bigg\}\Bigg] \\
&= \mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Bigg\{(1 - V_{n,t})^2\left\|P_{A_1}^\perp \hat{Z}\right\|^2 \\
&\qquad - 2V_{n,t}(1 - V_{n,t})Re\left\langle P_{A_1}^\perp \hat{Z}, P_{A_1}^\perp \sum_{i \in S_2^*} H_ic_i\right\rangle
\end{aligned}$$

$$\geq V_{n,t}(1-V_{n,t})\left\|P_{A_1}^{\perp}\sum_{i\in S_2^*}H_ic_i\right\|^2\Bigg\}\Bigg]$$

$$=\mathbb{P}\Bigg[\bigcup_{t,S,K_1}\Bigg\{\left\|(1-V_{n,t})P_{A_1}^{\perp}\hat{Z}-V_{n,t}P_{A_1}^{\perp}\sum_{i\in S_2^*}H_ic_i\right\|^2 \geq V_{n,t}\left\|P_{A_1}^{\perp}\sum_{i\in S_2^*}H_ic_i\right\|^2\Bigg\}\Bigg] \tag{251}$$

□

Proof of claim 3. Conditional of $H_{[K]}$ and A_0,

$$\hat{Z}\sim\mathcal{CN}\Bigg(0,\Bigg(1+P'\sum_{i\in S\setminus S_2^*}|H_i|^2\Bigg)\Bigg).$$

Hence

$$P_{A_1}^{\perp}\Bigg(\hat{Z}-\frac{V_{n,t}}{1-V_{n,t}}\sum_{i\in S_2^*}H_ic_i\Bigg)\sim \mathcal{CN}\Bigg(-\frac{V_{n,t}}{1-V_{n,t}}P_{A_1}^{\perp}\sum_{i\in S_2^*}H_ic_i,\Bigg(1+P'\sum_{i\in S\setminus S_2^*}|H_i|^2\Bigg)P_{A_1}^{\perp}\Bigg) \tag{252}$$

Now, the rank of $P_{A_1}^{\perp}$ is $n-K_1+t$ because the vectors in A_1 are linearly independent almost surely. Let $\mathcal{U}$ be a unitary change of basis matrix that rotates the range space of $P_{A_1}^{\perp}$ to the space corresponding to first $(n-K_1+t)$ coordinates. Then

$$\begin{aligned}
&\Bigg\|\mathcal{CN}\Bigg(-\frac{V_{n,t}}{1-V_{n,t}}P_{A_1}^{\perp}\sum_{i\in S_2^*}H_ic_i,\Bigg(1+P'\sum_{i\in S\setminus S_2^*}|H_i|^2\Bigg)P_{A_1}^{\perp}\Bigg)\Bigg\|^2\\
&=\Bigg\|\mathcal{U}\Bigg(\mathcal{CN}\Bigg(-\frac{V_{n,t}}{1-V_{n,t}}P_{A_1}^{\perp}\sum_{i\in S_2^*}H_ic_i,\Bigg(1+P'\sum_{i\in S\setminus S_2^*}|H_i|^2\Bigg)P_{A_1}^{\perp}\Bigg)\Bigg)\Bigg\|^2\\
&=\Bigg\|\mathcal{CN}\Bigg(-\frac{V_{n,t}}{1-V_{n,t}}\mathcal{U}P_{A_1}^{\perp}\sum_{i\in S_2^*}H_ic_i,\Bigg(1+P'\sum_{i\in S\setminus S_2^*}|H_i|^2\Bigg)\mathcal{U}P_{A_1}^{\perp}\mathcal{U}^*\Bigg)\Bigg\|^2.
\end{aligned} \tag{253}$$

Observe that $\mathcal{U}P_{A_1}^{\perp}\mathcal{U}^*$ is a diagonal matrix with first $(n-K_1+t)$ diagonal entries being ones and rest all 0. Also, if $W=P+iQ\sim\mathcal{CN}(\mu,\Gamma)$ (with pseudo-covariance being 0) then

$$\begin{bmatrix}P\\Q\end{bmatrix}\sim\mathcal{N}\left(\begin{bmatrix}Re(\mu)\\Im(\mu)\end{bmatrix},\frac{1}{2}\begin{bmatrix}Re(\Gamma) & -Im(\Gamma)\\ Im(\Gamma) & Re(\Gamma)\end{bmatrix}\right). \tag{254}$$

Using this and the definition of non-central chi-squared distribution the claim follows.

□

Proof of Claim 4. We have

$$\begin{aligned}
f_n(x)&=x+1+\frac{2V_{n,t}}{1-V_{n,t}}(1+x)\\
&\quad-\sqrt{1+\frac{2V_{n,t}}{1-V_{n,t}}(1+x)}\sqrt{2x+1+\frac{2V_{n,t}}{1-V_{n,t}}(1+x)}\\
&=\frac{1}{1-V_{n,t}}\Bigg[(1+V_{n,t})(x+1)\\
&\quad-2\sqrt{V_{n,t}}\sqrt{\left(x+\frac{(1+V_{n,t})^2}{4V_{n,t}}\right)^2-\frac{(1-V_{n,t}^2)^2}{16V_{n,t}^2}}\Bigg]
\end{aligned} \tag{255}$$

Hence

$$\begin{aligned}
f'(x)&=\frac{1}{1-V_{n,t}}\left[1+V_{n,t}-2\sqrt{V_{n,t}}\frac{a}{\sqrt{a^2-b^2}}\right]\\
&=\frac{1}{1-V_{n,t}}\left(\sqrt{V_{n,t}}-\sqrt{\frac{a+b}{a-b}}\right)\cdot\\
&\qquad\left(\sqrt{V_{n,t}}-\sqrt{\frac{a-b}{a+b}}\right)
\end{aligned} \tag{256}$$

where $a=\left(x+\frac{(1+V_{n,t})^2}{4V_{n,t}}\right)$ and $b=\frac{1-V_{n,t}^2}{4V_{n,t}}$. Also $a>0$ and $b>0$. Further $a+b>a-b$ and

$$\begin{aligned}
&\sqrt{V_{n,t}}<\sqrt{\frac{a-b}{a+b}}=\sqrt{\frac{V_{n,t}(1+V_{n,t}+2x)}{1+V_{n,t}+2V_{n,t}x}}\\
&\iff 2V_{n,t}x+1+V_{n,t}<2x+1+V_{n,t}\\
&\iff 0<V_{n,t}<1
\end{aligned}$$

which is true. Hence both the factors in (256) are negative. Therefore $f'(x)>0$. □

Appendix C
Maximal per-user error

In this section we briefly describe relations between maximal per-user error (PUPE-max) defined in (10) and PUPE. First, we represent our system as in (148)

$$Y=AHW+Z. \tag{257}$$

Let $P_{e,i}(A)=\mathbb{P}\left[W_i\neq\hat{W}_i\right]$. We are interested in bounding the variance of $P_{e,i}(A)$ so that

$$\mathbb{E}\left[P_{e,u}^{\max}(A)\right]=\mathbb{E}\left[\max\{P_{e,i}(A):i\in[K]\}\right]$$

can be related to $\mathbb{E}\left[P_{e,i}(A)\right]=\mathbb{E}\left[P_{e,u}\right]$ due to symmetry on users by random codebook generation. Consider two coupled systems

$$Y=AHW+Z \tag{258}$$

$$Y'=A'HW+Z \tag{259}$$

where A and A' are fixed so that the channels are dependent on these.

Now we have

$$\begin{aligned}
|P_{e,i}(A)-P_{e,i}(A')|&\leq d_{TV}\left(P_{Y,H,W},P_{Y',H,W}\right)\\
&\leq\sqrt{\frac{1}{2}D\left(P_{Y,H,W}||P_{Y',H,W}\right)}
\end{aligned} \tag{260}$$

where $d_{TV}(P,Q) = \sup\{|P(A) - Q(A)| : A \, \text{is measurable}\}$ is the total variation distance between measures P and Q, $D(P||Q) = \mathbb{E}_P\left[\ln \frac{dP}{dQ}\right]$ is the Kullback-Leibler divergence (in nats) and the last inequality is the Pinsker's inequality (see [77]). Now using properties of D (see [78, Theorem 2.2])

$$\begin{aligned} & D\left(P_{Y,H,W}||P_{Y',H,W}\right) \\ &= D(P_{Y|H,W}||P_{Y'|H,W}|P_{H,W}) \\ &= \int_{H,W} D(P_{Y|H=h,W=w}||P_{Y'|H=h,W=w}) dP_{H,W}(h,w) \end{aligned} \tag{261}$$

Now note that conditioned on $H = h, W = w$, we have $Y \sim \mathcal{CN}(Ahw, I_n)$ and $Y' \sim \mathcal{CN}(A'hw, I_n)$. Hence a simple computation shows that $D(P_{Y|H=h,W=w}||P_{Y'|H=h,W=w}) = \|Ahw - A'hw\|^2$. Therefore we have

$$D\left(P_{Y,H,W}||P_{Y',H,W}\right) = \mathbb{E}\left[\|(A - A')HW\|^2\right]. \tag{262}$$

Now let $B = A - A'$ and $X = HW$. Then

$$\mathbb{E}\left[\|BX\|^2\right] = \sum_{i\in[n]} \mathbb{E}\left[\sum_{j,k\in[KM]} B_{i,j}\bar{B}_{i,k} X_i \bar{X}_k\right] \tag{263}$$

Note that $\mathbb{E}\left[X_j \bar{X}_k\right]$ is zero if $j \neq k$ and it is $1/M$ otherwise. Hence

$$\mathbb{E}\left[\|BX\|^2\right] = \frac{1}{M}\sum_{i\in[n]}\sum_{j\in[KM]} B_{i,j}\bar{B}_{i,j} = \frac{1}{M}\|B\|_F^2. \tag{264}$$

Therefore

$$D\left(P_{Y,H,W}||P_{Y',H,W}\right) = \frac{1}{M}\|A - A'\|_F^2. \tag{265}$$

So combining this with (260), we obtain

$$|P_{e,i}(A) - P_{e,i}(A')| \leq \sqrt{\frac{1}{2M}}\|A - A'\|_F. \tag{266}$$

Now let each entry of A and A' to be distributed iid as $\mathcal{CN}(0,P)$ where $P = P_{tot}/K$. Further, let $\tilde{A} = \sqrt{\frac{K}{P_{tot}}}A$ and $\tilde{A}' = \sqrt{\frac{K}{P_{tot}}}A'$. So the entries of $\tilde{A}$ and $\tilde{A}'$ are iid $\mathcal{CN}(0,1)$. Therefore, with slight abuse of notation, we can rewrite (266) as

$$|P_{e,i}(\tilde{A}) - P_{e,i}(\tilde{A}')| \leq \sqrt{\frac{P_{tot}}{2MK}}\left\|\tilde{A} - \tilde{A}'\right\|_F. \tag{267}$$

Hence the function $P_{e,i}$ is Lipschitz with Lipschitz constant $L = \sqrt{\frac{P_{tot}}{2MK}}$. By concentration of Lipschitz functions of Gaussian random vectors [47, Theorems 5.5, 5.6], we have that $P_{e,i}(\tilde{A})$ is sub-Gaussian with

$$\mathrm{Var}(P_{e,i}(A)) \leq 4L^2 = \frac{2P_{tot}}{KM}. \tag{268}$$

Hence, using bounds on the expected maximum of sub-Gaussian random variables (see [47, Section 2.5]), we obtain

$$\begin{aligned} & \mathbb{E}\left[\max_{i\in[K]} P_{e,i}(A)\right] \\ &\leq \mathbb{E}\left[P_{e,u}\right] + \sqrt{\mathrm{Var}(P_{e,i}(A))\ln K} \\ &= \mathbb{E}\left[P_{e,u}\right] + \sqrt{\frac{2P_{tot}}{M}\frac{\ln K}{K}} \xrightarrow{K\to\infty} \mathbb{E}\left[P_{e,u}\right]. \end{aligned} \tag{269}$$

Therefore, a random coding argument along with (269) shows that PUPE-max has same asymptotics as PUPE in the linear scaling regime. For FBL performance, if each user sends $k = 100$ bits then $M = 2^k$ and hence $\mathbb{E}\left[P_{e,u}^{\max}\right] \approx \mathbb{E}\left[P_{e,u}\right]$

Acknowledgements

We thank the anonymous reviewer for suggesting to apply the scalar AMP to derive explicit achievability bounds in the many MAC case (resulting in Section IV-C) and pointing out to use (171) to approximate the η^*. We also thank Prof. A. Montanari for the footnote in Section V-A regarding rigorization of the replica-method prediction for PUPE.

References

[1] S. S. Kowshik and Y. Polyanskiy, "Quasi-static fading MAC with many users and finite payload," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 440–444.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[3] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian Many-Access Channels." *IEEE Trans. Information Theory*, vol. 63, no. 6, pp. 3516–3539, 2017.

[4] Y. Polyanskiy, "A perspective on massive random-access," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 2523–2527.

[5] I. Zadik, Y. Polyanskiy, and C. Thrampoulidis, "Improved bounds on Gaussian MAC and sparse regression via Gaussian inequalities," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019.

[6] F. Wei, Y. Wu, W. Chen, W. Yang, and G. Caire, "On the Fundamental Limits of MIMO Massive Multiple Access Channels," *arXiv preprint arXiv:1807.05553*, 2018.

[7] Y. Polyanskiy, "Information theoretic perspective on massive multiple-access," in *Short Course (slides)*, Skoltech Inst. of Tech., Moscow, Russia, Jul. 2018. [Online]. Available: https://people.lids.mit.edu/yp/homepage/data/SkolTech18-MAC-lectures.pdf

[8] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.

[9] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits with and without feedback," in *2010 IEEE International Symposium on Information Theory*. IEEE, 2010, pp. 221–225.

[10] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading MAC," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019.

[11] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access gaussian channel," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 2528–2532.

[12] A. Vem, K. R. Narayanan, J. Cheng, and J.-F. Chamberland, "A user-independent serial interference cancellation based coding scheme for the unsourced random access gaussian channel," in *Information Theory Workshop (ITW), 2017 IEEE*. IEEE, 2017, pp. 121–125.

[13] V. K. Amalladinne, A. Vem, D. K. Soma, K. R. Narayanan, and J.-F. Chamberland, "A coupled compressive sensing scheme for uncoordinated multiple access," *arXiv preprint arXiv:1809.04745*, 2018.

[14] A. Glebov, L. Medova, P. Rybin, and A. Frolov, "On LDPC Code Based Massive Random-Access Scheme for the Gaussian Multiple Access Channel," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer, 2018, pp. 162–171.

[15] A. Thompson and R. Calderbank, "Compressed neighbour discovery using sparse kerdock matrices," *arXiv preprint arXiv:1801.04537*, 2018.

[16] R. Calderbank and A. Thompson, "CHIRRUP: a practical algorithm for unsourced multiple access," *arXiv preprint arXiv:1811.00879*, 2018.

[17] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "Polar coding and random spreading for unsourced multiple access," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[18] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "Unsourced Random Access with Coded Compressed Sensing: Integrating AMP and Belief Propagation," *arXiv preprint arXiv:2010.04364*, 2020.

[19] A. Fengler, P. Jung, and G. Caire, "Sparcs for unsourced random access," *arXiv preprint arXiv:1901.06234*, 2019.

[20] K. Andreev, E. Marshakov, and A. Frolov, "A polar code based tin-sic scheme for the unsourced random access in the quasi-static fading mac," *arXiv preprint arXiv:2005.06899*, 2020.

[21] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4232–4265, 2014.

[22] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[23] G. Reeves and M. Gastpar, "A note on optimal support recovery in compressed sensing," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*. IEEE, 2009, pp. 1576–1580.

[24] ——, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3065–3092, 2012.

[25] I. Bettesh and S. Shamai, "Outages, expected rates and delays in multiple-users fading channels," in *Proceedings of the 2000 Conference on Information Science and Systems*, vol. 1, 2000.

[26] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[27] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.

[28] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits through the Gaussian channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4880–4902, 2011.

[29] D. Tuninetti and G. Caire, "The throughput of some wireless multiaccess systems," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2773–2785, 2002.

[30] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.

[31] Y. Jin, Y.-H. Kim, and B. D. Rao, "Limits on support recovery of sparse signals via multiple-access communication techniques," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7877–7892, 2011.

[32] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient coded random access for the wireless uplink," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4694–4708, 2020.

[33] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[34] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2773–2795, 1998.

[35] D. Guo and S. Verdú, "Randomly spread cdma: Asymptotics via statistical physics," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.

[36] G. Reeves and M. C. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3451–3465, 2013.

[37] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2541–2557, 2012.

[38] R. Venkataramanan, S. Tatikonda, and A. Barron, "Sparse regression codes," *Foundations and Trends® in Communications and Information Theory*, vol. 15, no. 1-2, pp. 1–195, 2019. [Online]. Available: http://dx.doi.org/10.1561/0100000092

[39] C. Rush and R. Venkataramanan, "The error probability of sparse superposition codes with approximate message passing decoding," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3278–3303, 2018.

[40] C. Rush, K. Hsieh, and R. Venkataramanan, “Capacity-achieving Spatially Coupled Sparse Superposition Codes with AMP Decoding,” *arXiv preprint arXiv:2002.07844*, 2020.

[41] J. Barbier and F. Krzakala, “Approximate message-passing decoder and capacity achieving sparse superposition codes,” *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 4894–4927, 2017.

[42] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[43] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, “Quasi-static SIMO fading channels at finite blocklength,” in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 1531–1535.

[44] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[45] L. Birgé, “An alternative point of view on Lepski’s method,” *Lecture Notes-Monograph Series*, pp. 113–133, 2001.

[46] R. R. Bahadur, “A note on quantiles in large samples,” *The Annals of Mathematical Statistics*, vol. 37, no. 3, pp. 577–580, 1966.

[47] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[48] X. Meng, S. Wu, L. Kuang, and J. Lu, “Concise derivation of complex bayesian approximate message passing via expectation propagation,” *arXiv preprint arXiv:1509.08658*, 2015.

[49] Q. Zou, H. Zhang, C.-K. Wen, S. Jin, and R. Yu, “Concise derivation for generalized approximate message passing using expectation propagation,” *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1835–1839, 2018.

[50] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, “Asymptotic analysis of complex lasso via complex approximate message passing (camp),” *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4290–4308, 2013.

[51] J. Barbier, “Statistical physics and approximate message-passing algorithms for sparse linear estimation problems in signal processing and coding theory,” *arXiv preprint arXiv:1511.01650*, 2015.

[52] W. Van Zwet, “A strong law for linear functions of order statistics,” *The Annals of Probability*, pp. 986–990, 1980.

[53] J. S. Rosenthal, *A first look at rigorous probability theory*. World Scientific Publishing Company, 2006.

[54] S. Verdú and S. Shamai, “Spectral efficiency of CDMA with random spreading,” *IEEE Transactions on Information theory*, vol. 45, no. 2, pp. 622–640, 1999.

[55] J. Salo, D. Seethaler, and A. Skupch, “On the asymptotic geometric mean of MIMO channel eigenvalues,” in *2006 IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 2109–2113.

[56] R. Zamir and M. Feder, “A generalization of the entropy power inequality with applications,” *IEEE transactions on information theory*, vol. 39, no. 5, pp. 1723–1728, 1993.

[57] W. Li and Y. Chen, “Some new two-sided bounds for determinants of diagonally dominant matrices,” *Journal of Inequalities and Applications*, vol. 2012, no. 1, p. 61, 2012.

[58] A. Collins, G. Durisi, T. Erseghe, V. Kostina, J. Östman, Y. Polyanskiy, I. Tal, and W. Yang, “SPECTRE: Short Packet Communication Toolbox,” https://github.com/yp-mit/spectre, 2016.

[59] L. A. Shepp and I. Olkin, “Entropy of the sum of independent bernoulli random variables and of the multinomial distribution,” in *Contributions to probability*. Elsevier, 1981, pp. 201–206.

[60] C. Knessl, “Integral representations and asymptotic expansions for Shannon and Renyi entropies,” *Applied mathematics letters*, vol. 11, no. 2, pp. 69–74, 1998.

[61] B. Lindström, “On a combinatorial problem in number theory,” *Canadian Mathematical Bulletin*, vol. 8, no. 4, pp. 477–490, 1965.

[62] D. G. Cantor and W. Mills, “Determination of a subset from certain combinatorial properties,” *Canadian Journal of Mathematics*, vol. 18, pp. 42–48, 1966.

[63] G. H. Khachatrian and S. S. Martirossian, “A new approach to the design of codes for synchronous-CDMA systems,” *IEEE Transactions on Information Theory*, vol. 41, no. 5, pp. 1503–1506, 1995.

[64] G. Reeves and H. D. Pfister, “The replica-symmetric prediction for random linear estimation with gaussian matrices is exact,” *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2252–2283, 2019.

[65] J. Barbier, M. Dia, N. Macris, and F. Krzakala, “The mutual information in random linear estimation,” in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2016, pp. 625–632.

[66] D. Guo, D. Baron, and S. Shamai, “A single-letter characterization of optimal noisy compressed sensing,” in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2009, pp. 52–59.

[67] J. Barbier, F. Krzakala, M. Mézard, and L. Zdeborová, “Compressed sensing of approximately-sparse signals: Phase transitions and optimal reconstruction,” in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2012, pp. 800–807.

[68] R. Vicente, D. Saad, and Y. Kabashima, “Low-density parity-check codes: statistical physics perspective,” in *Advances in imaging and electron physics*. Elsevier, 2003, vol. 125, pp. 231–353.

[69] S. Kirkpatrick and B. Selman, “Critical behavior in the satisfiability of random boolean expressions,” *Science*, vol. 264, no. 5163, pp. 1297–1301, 1994.

[70] G. Reeves, J. Xu, and I. Zadik, “All-or-nothing phenomena: From single-letter to high dimensions,” in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*.

IEEE, 2019, pp. 654–658.

[71] ——, "All-or-Nothing Phenomena: From Single-Letter to High Dimensions," *arXiv preprint arXiv:1912.13027*, 2019.

[72] A. Fengler, G. Caire, P. Jung, and S. Haghighatshoar, "Massive MIMO unsourced random access," *arXiv preprint arXiv:1901.00828*, 2019.

[73] G. Caire and R. Muller, "The optimal received power distribution for ic-based iterative multiuser joint decoders," in *PROCEEDINGS OF THE ANNUAL ALLERTON CONFERENCE ON COMMUNICATION CONTROL AND COMPUTING*, vol. 39, no. 2. The University; 1998, 2001, pp. 1132–1141.

[74] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits*. Princeton University, 2010.

[75] M. Ledoux, *The concentration of measure phenomenon*. American Mathematical Soc., 2001, no. 89.

[76] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018, vol. 47.

[77] S. Verdú, "Total variation distance and the distribution of relative information," in *2014 Information Theory and Applications Workshop (ITA)*. IEEE, 2014, pp. 1–3.

[78] Y. Polyanskiy and Y. Wu, "Lecture notes on Information Theory," 2017. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf

Suhas S Kowshik is a PhD student in the department of Electrical Engineering and Computer Science (EECS) at Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He obtained his S.M. degree from EECS, MIT in 2018, and B.Tech and M.Tech dual degrees from Indian Institute of Technology Madras, Chennai, India in 2016. His research interests include information theory, communication, machine learning and optimization.

Yury Polyanskiy Yury Polyanskiy (S'08-M'10-SM'14) is an Associate Professor of Electrical Engineering and Computer Science and a member of IDSS and LIDS at MIT. Yury received M.S. degree in applied mathematics and physics from the Moscow Institute of Physics and Technology, Moscow, Russia in 2005 and Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ in 2010. His research interests span information theory, statistical learning, error-correcting codes, wireless communication and fault tolerance. Dr. Polyanskiy won the 2020 IEEE Information Theory Society James Massey Award, 2013 NSF CAREER award and 2011 IEEE Information Theory Society Paper Award.