

# Energy efficient coded random access for the wireless uplink

Suhas S Kowshik  
MIT  
suhask@mit.edu

Kirill Andreev  
SkolTech (Moscow)  
k.andreev@skoltech.ru

Alexey Frolov  
SkolTech (Moscow)  
al.frolov@skoltech.ru

Yury Polyanskiy  
MIT  
yp@mit.edu

## Abstract

We discuss the problem of designing channel access architectures for enabling fast, low-latency, grant-free and uncoordinated uplink for densely packed wireless nodes. Specifically, we study random-access codes, previously introduced for the AWGN multiple-access channel (MAC) by Polyanskiy'2017, in the practically more relevant case of users subject to Rayleigh fading, when channel gains are unknown to the decoder. We propose a random coding achievability bound, which we analyze both non-asymptotically (at finite blocklength) and asymptotically. As a candidate practical solution, we propose an explicit sparse-graph based coding scheme together with an alternating belief-propagation decoder. The latter's performance is found to be surprisingly close to the finite-blocklength bounds. Our main findings are twofold. First, just like in the AWGN MAC we see that jointly decoding large number of users leads to a surprising phase transition effect, where at spectral efficiencies below a critical threshold (5-15 bps/Hz depending on reliability) a perfect multi-user interference cancellation is possible. Second, while the presence of Rayleigh fading significantly increases the minimal required energy-per-bit  $E_b/N_0$  (from about 0-2 dB to about 8-11 dB), the inherent randomization introduced by the channel makes it much easier to attain the optimal performance via iterative schemes.

In all, it is hoped that a principled definition of the random-access model together with our information-theoretic analysis will open the road towards unified benchmarking and comparison performance of various random-access solutions, such as the currently discussed candidates (MUSA, SCMA, RSMA) for the 5G/6G.

## I. INTRODUCTION

Presently, wireless networks are starting to see a new type of load (a so-called mMTC or machine-type communication), in which hundreds of thousands of devices are serviced by a single base station, each communicating very small and infrequent data payloads. In the interest of reducing hardware complexity, reducing latency and improving energy consumption, the conceptual paradigm shift is to move to the *grant-free* access management, in which uplink communication is not orthogonalized by the base-station (as is done in today's systems). This requires new kinds of codes that can be decoded from uncoordinated and colliding transmissions.

In this work, we aim to understand the fundamental tradeoffs of these dense random access systems, and provide coding solutions that are close to achieving these fundamental limits. Specifically, we consider a problem of a large number of nodes (potentially unbounded) with any  $K_a$  of them communicating to a single access point or base station (BS) over a frame synchronous multiple access channel (MAC) with frame length  $n$ .

An information-theoretic formulation of this problem was done in [1] where the author considered an additive white Gaussian noise (AWGN) random access channel (RAC) model. In this formulation the random access is modeled as follows: each of  $K_a$  active users encodes his  $k$ -bit message into an  $n$ -symbol codeword. The receiver observes superposition of  $K_a$  codewords corrupted by the AWGN. There are a number of challenges in this model: finite blocklength (FBL) effects due to small payload size, massive number of users (comparable to blocklength), sparsity due to random access and incorporating accurate channel models. However, the most crucial departure from canonical MAC is that the users are required to share the same codebook (i.e. they are unidentifiable, unless they desire to put their identity as part of the  $k$ -bit payload), and the decoder is only required to provide an unordered list of user messages. In the follow-up works, this problem has also been called *unsourced random access* [2–4]. Another important aspect of this new formulation is the notion of per-user probability of error (PUPE) which is defined as the average (over the active users) fraction of the transmitted messages that are misdecoded. (Recall that classical definition declares error even if any one of the messages decoded incorrectly.)

In a quest towards low-complexity schemes achieving FBL bounds above, a scheme based on concatenated codes (with an inner binary linear code and an outer BCH codes) in conjunction with a protocol called  $T$ -fold ALOHA was considered in [5].  $T$ -fold ALOHA is a modification of the standard slotted aloha protocol, in that up to  $T$  collisions can be decoded in a slot. So, slotted ALOHA corresponds to  $T = 1$ . The idea of  $T$ -fold ALOHA itself is not new as the idea of employing multi-packet receivers to resolve small order collisions has reappeared periodically [6] and more recently [7, Appendix A]. The

This material is based upon work supported by the National Science Foundation under Grant No CCF-17-17842, CAREER award under grant agreement CCF-12-53205, the Hong Kong Innovation and Technology Fund (ITS/066/17FP) under the HKUST-MIT Research Alliance Consortium and a grant from Skoltech-MIT Joint Next Generation Program (NGP).

The research of A. Frolov and K. Andreev was supported by the Russian Science Foundation (project no. 18-19-00673).

gap between this low complexity scheme in [5] and the FBL bound [1] was reduced in [2] by employing a serial interference cancellation scheme on top of an interleaved LDPC code. Achievability bounds for serial interference cancellation scheme (also known as irregular repetition slotted ALOHA [7]) were further improved in [8], where density evolution method [7] and a finite length random coding bound for the Gaussian MAC [1] were combined. In [9] the LDPC portion of [2] was improved by optimizing the protograph of LDPC code for Gaussian MAC using generalized EXIT charts. Further improvements were obtained in [3] by developing a compressive sensing based algorithm. In [10] the idea of sparsifying collisions, inherent in T-fold ALOHA, was modified by randomizing (sparse) locations of the LDPC codeword symbols and by optimizing degree distributions via a suitable approximation of a density evolution.

Finally, we mention that there is another promising idea, proposed in 2001 by Muller and Caire [11], that uses non-orthogonal CDMA spreading coupled with an outer code. The key idea is to demodulate CDMA by leveraging the soft information from the outer decoder (and alternate between the two). In [11] authors observed a perfect multi-user cancellation effect, shown to exist also for the fundamental limit in [12]. It remains to explore whether this method is competitive for practically relevant blocklengths.

Another set of works considers the problem of sending a (distributedly detected) alarm signal with high-reliability on top of the regular low-rate update traffic, cf. [13].

All of the references above focused on the AWGN RAC (or, equivalently, assumed perfect power control of the users' transmissions equalizing received powers). In the presence of fading and MIMO, there have been various works on algorithms for on/off activity detection [14–16] that use compressive sensing ideas along with approximate message passing algorithm. (We note that the random-access problem can be seen as on/off activity detection within a population of  $2^k$  users, where  $k$  is the message length. However, already a moderate value of  $k = 100$  precludes the straightforward usage of activity detection protocols.) In [17], scaling laws were derived for activity detection in a massive MIMO scenario. This and the ideas from [3] have been used to develop a low complexity coding scheme in [4]. We also note here that our problem can be understood as a sparse support recovery in the compressed sensing literature [18–21]. Theoretical investigations in that literature predominantly consider iid Gaussian codebooks. In particular, in [19], the authors analyze various estimators like maximum likelihood (ML) and linear estimators like matched filter (MF) and linear minimum mean squared error (LMMSE) but in an asymptotic setting similar to a many-user MAC [1, 12, 22–24] where the number of active users scales linearly in blocklength.

The structure and main contributions of this paper are:

- In Section II we formally define the problem of unsourced frame synchronized single antenna quasi-static Rayleigh fading RAC under per-user error (PUPE). We assume that the channel realizations are not known to the receiver or the transmitters.
- A  $T$ -fold ALOHA access method from [5] is reviewed in Section III. There are two ways we apply  $T$ -fold ALOHA in this paper. One is to get a random-coding (non-constructive) achievability bounds, this is done in Appendix A. Another is to use it as part of the explicit construction, which we do in Section V.
- A converse (lower) bound on energy-per-bit required for any random-access codes is developed in Section IV.
- The random coding achievability and converse bounds are evaluated in the asymptotic setting in Section VII.
- In Section V we develop a low-complexity iterative multi-user decoding scheme based on LDPC codes [25–27] and a belief propagation decoder on a joint Tanner graph.
- In Section VI we numerically compare various bounds in the finite-blocklength setting. It is found that our practical scheme is rather competitive compared to both our own finite-blocklength bounds and asymptotic benchmarks.
- Section VIII finishes with some future directions.

## A. Notation

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  $[n] = \{1, 2, \dots, n\}$ . Lastly,  $\|\cdot\|$  represents the standard euclidean norm.

## II. SYSTEM MODEL

We follow the definition of a code from [1]. Fix an integer  $K_a \geq 1$  – the number of active users. Let  $\{P_{Y^n|X^n} = P_{Y^n|X_1^n, X_2^n, \dots, X_{K_a}^n} : \times_{i=1}^{K_a} \mathcal{X}_i^n \rightarrow \mathcal{Y}^n\}_{n=1}^\infty$  be a multiple access channel (MAC), which is also permutation invariant: for any permutation  $\pi$  on  $[K_a]$ , the distribution  $P_{Y^n|X_1^n, \dots, X_{K_a}^n}(\cdot|x_1^n, \dots, x_{K_a}^n)$  coincides with  $P_{Y^n|X_{\pi(1)}^n, \dots, X_{\pi(K_a)}^n}(\cdot|x_{\pi(1)}^n, \dots, x_{\pi(K_a)}^n)$ . We also call this a random access channel (RAC).

**Definition 1** ([1]). An  $(M, n, \epsilon)$  random-access code for the  $K_a$  user MAC  $P_{Y^n|X^n}$  is a pair of (possibly randomized) maps  $f : [M] \rightarrow \mathcal{X}^n$  (the encoder) and  $g : \mathcal{Y}^n \rightarrow \binom{[M]}{K_a}$  such that if  $W_1, \dots, W_{K_a}$  are chosen independently and uniformly from  $[M]$  and  $X_j = f(W_j)$  then the average (per-user) probability of error satisfies

$$P_e = \frac{1}{K_a} \sum_{j=1}^{K_a} \mathbb{P}[E_j] \leq \epsilon \quad (1)$$

where  $E_j \triangleq \{W_j \notin g(Y^n)\} \cup \{W_j = W_i \text{ for some } i \neq j\}$  and  $Y^n$  is the channel output.

So, all users use the same codebook, and the receiver outputs a list of  $K_a$  codewords. Further, the probability of error is the average fraction of incorrectly decoded codewords. In the remainder of the paper we particularly focus on the single antenna quasi-static fading MAC:

$$Y^n = \sum_{i=1}^{K_a} H_i X_i^n + Z^n \quad (2)$$

where  $Z^n \sim \mathcal{CN}(0, I_n)$ , and  $H_i \stackrel{iid}{\sim} \mathcal{CN}(0, 1)$  are the fading coefficients which are independent of  $\{X_i^n\}$  and  $Z^n$ . Consequently, we require each codeword produced by the encoder  $f$  to satisfy a maximum power constraint:

$$\|f(w)\|^2 \leq nP, \quad \forall w \in [M]. \quad (3)$$

We emphasize that there can be potentially an unbounded number of users, but only  $K_a$  of them are active. If each user has a message of size  $k$  and transmits at power  $P$  per symbol, then the energy-per-bit is given by  $E_b/N_0 = \frac{nP}{k}$ .

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

### III. RANDOM-ACCESS VIA $T$ -FOLD ALOHA

In this section, we discuss our main achievability bound based on  $T$ -fold ALOHA protocol [5]. The idea is the following. Let  $T, n_1 \in \mathbb{N}$  such that  $T < K_a$  and  $n_1 < n$ . The time slot or frame of length  $n$  is partitioned into  $L = n/n_1$  subframes of length  $n_1$ . The common codebook is of blocklength  $n_1$  and thus may use a larger power  $LP$  per degree of freedom. Each user chooses a slot to send his message uniformly at random independently of other users. If there are  $r$  users placing their codewords in a particular  $n_1$ -slot, then the law of observations  $Y^{n_1}$  and messages  $W_1, \dots, W_r$  in this slot is given by

$$Y^{n_1} = \sum_{i=1}^r H_i f(W_i) + Z^{n_1}, \quad W_i \stackrel{iid}{\sim} \text{Unif}[M]. \quad (4)$$

Suppose there is a code such that if there are at most  $T$  users transmitting in a given block, then with good reliability decoder can estimate all  $\leq T$  messages, while if  $> T$  users were transmitting then no guarantees on the decoder performance are made. For  $T = 1$  this corresponds to the usual ‘‘collision model’’ prevalent in the analysis of the ALOHA. (Thus  $T > 1$  serves to partially address the more realistic physical layer behavior.) Intuitively, then, if the average number of users per slot, equal to  $K_a/L$ , is smaller than  $T$ , then with good probability all users will be properly decoded.

More specifically, for a given common codebook  $\mathcal{C} \subset B(n_1, \sqrt{n_1 LP})$  inside an  $\mathbb{C}^{n_1}$ -ball of radius  $\sqrt{n_1 LP}$  and size  $|\mathcal{C}| = M$  we let  $P_{e,\text{genie}}(\mathcal{C}, r)$  denote the following quantity:

$$P_{e,\text{genie}}(\mathcal{C}, r) = \frac{1}{r} \sum_{i=1}^r \mathbb{P}[W_i \notin \mathcal{L}(Y^{n_1}, r)],$$

where  $\mathcal{L}$  is the decoded list of messages. The subindex ‘‘genie’’ denotes the fact that the decoder is aware of the exact number of users active in a slot. Given this genie side-information we can show that the  $T$ -fold ALOHA access scheme then attains the overall PUPE for all of  $K_a$  users bounded by

$$\epsilon_{T,\text{genie}}(\mathcal{C}) \triangleq 1 - \sum_{r=1}^T (1 - P_{e,\text{genie}}(\mathcal{C}, r)) \binom{K_a - 1}{r - 1} \left(\frac{1}{L}\right)^{r-1} \left(1 - \frac{1}{L}\right)^{K_a - r} + \frac{K_a}{M}.$$

To get this bound, we first bound the probability that the  $i$ -th user’s message is in collision:  $\mathbb{P}[\exists j \neq i : W_j = W_i] \leq \frac{K_a - 1}{M}$ . Next, we note that the  $i$ -th user’s slot will have  $r - 1$  other users with probability  $\binom{K_a - 1}{r - 1} \left(\frac{1}{L}\right)^{r-1} \left(1 - \frac{1}{L}\right)^{K_a - r}$ . Note that the resulting bound is monotonically improving with increasing  $T$ .

**Remark 1.** We will use the genie bound for our random-coding constructions and upper bound  $P_{e,\text{genie}}$  via (33) in appendix A. Note that the genie assumption prevents the above from being a true achievability bound. In the AWGN (non-fading) setting

the number of users can be reliably estimated by simply measuring the total received energy in each  $n_1$ -long slot, cf. [5]. However, in the presence of fading this detector is a lot less reliable. Consequently, our genie-based bound strictly speaking is only an optimistic estimate of the performance achievable within a  $T$ -fold ALOHA scheme by the best possible component subcode.

To get the true (genie-free) bounds, we are going to use an explicit (LDPC-based) code inside each  $n_1$ -slot. Our decoder automatically detects the number of users in a slot and estimates the messages. To evaluate the performance we need to define two parameters corresponding to the  $n_1$ -code  $\mathcal{C}$ . Namely, we define  $P_e(\mathcal{C}, r)$  and  $Q_e(\mathcal{C}, r)$  as follows. Consider the setting of (4). Fix some decoder (unaware of the number  $r$ ) which outputs a variable-length list  $\mathcal{L} = \mathcal{L}(Y^{n_1}) \subset [M]$ . We define

$$P_e(\mathcal{C}, r) = \frac{1}{r} \sum_{i=1}^r \mathbb{P}[W_i \notin \mathcal{L}] ,$$

$$Q_e(\mathcal{C}, r) = \mathbb{P}[|\mathcal{L}| > r] .$$

With this definition we get the following bound on the overall PUPE (for all of  $K_a$  users):

$$\epsilon_T(\mathcal{C}) \triangleq 1 - \sum_{r=1}^T (1 - P_e(\mathcal{C}, r)) \binom{K_a - 1}{r - 1} \left(\frac{1}{L}\right)^{r-1} \left(1 - \frac{1}{L}\right)^{K_a - r} + \frac{K_a}{M} + q, \quad (5)$$

where

$$q = L \sum_{r=0}^{K_a} \binom{K_a}{r} L^{-r} \left(1 - \frac{1}{L}\right)^{K_a - r} Q_e(\mathcal{C}, r)$$

is an upper bound on  $\mathbb{P}[\cup_{j=1}^L F_j]$ , where  $F_j$  is the event that the  $j$ -th slot's decoded list has size strictly bigger than the number  $r$  of users active in that slot. Note that if the decoder never outputs a list of size  $> T$  then  $Q_e(\mathcal{C}, r) = 0$  for all  $r \geq T$ . In our simulations, we have  $Q_e(\mathcal{C}, r) \approx 0$  (within accuracy of the Monte Carlo) for all  $r \geq 0$ . In other words, our decoder does not ever overestimate the number of active users.

#### IV. CONVERSE BOUND

In this section we describe a simple converse bound based on results from [28] and the meta-converse from [29].

**Theorem IV.1.** *Let*

$$L_n = n \log(1 + PG) + \sum_{i=1}^n \left(1 - |\sqrt{PG}Z_i - \sqrt{1 + PG}|^2\right) \quad (6)$$

$$S_n = n \log(1 + PG) + \sum_{i=1}^n \left(1 - \frac{|\sqrt{PG}Z_i - 1|^2}{1 + PG}\right) \quad (7)$$

where  $G = |H|^2$  and  $Z_i \stackrel{iid}{\sim} \mathcal{CN}(0, 1)$ . Then for every  $n$  and  $0 < \epsilon < 1$ , any  $(M, n - 1, \epsilon)$  code for the quasi-static  $K_a$  MAC satisfies

$$\log(M) \leq \log(K_a) + \log \frac{1}{\mathbb{P}[L_n \geq n\gamma_n]} \quad (8)$$

where  $\gamma_n$  is the solution of

$$\mathbb{P}[S_n \leq n\gamma_n] = \epsilon. \quad (9)$$

*Proof:* Notice that the converse bound for the case where full CSI is available at the receiver (and/or transmitter) is a converse for the no-CSI case as well. Further, by symmetry, it is enough to get a lower bound on the probability that a particular user's message is not in the decoded list. Finally, we can assume that the decoder has the knowledge of the codewords of all other users. To formalize, let  $Y$  be the received vector and let  $L(Y)$  be the list of codewords output by the decoder (we use the list of codewords or messages interchangeably). The size of the list is  $|L(Y)| \leq K_a$ . Then we have the following implications:

$$\frac{1}{K_a} \sum_{t=1}^{K_a} \mathbb{P}[X_t \notin L(Y)] \geq 1 - \epsilon$$

$$\iff \mathbb{P}[X_1 \notin L(Y)] \geq 1 - \epsilon \quad (10)$$

$$\iff \mathbb{P}[X_1 \notin L(Y, H_1)] \geq 1 - \epsilon \quad (11)$$

$$\iff \mathbb{P}[X_1 \notin L(Y, H_{[K_a]}, X_{[K_a] \setminus \{1\}})] \geq 1 - \epsilon \quad (12)$$

where (11) and (12) represents the case when the decoder has access to the fading realization of user 1 and interference from all other users respectively.

Now, given  $H_{[K_a]}$  and  $X_{[K_a]\setminus\{1\}}$  at the receiver, the channel is equivalent to

$$Y_1 = H_1 X_1 + Z$$

where  $H_1$  and  $Z$  are same as before, the decoder outputs a list of messages  $\hat{W} = L(Y_1, H_1)$  of size at most  $K_a$  and the probability of error is  $\mathbb{P}[W_1 \notin \hat{W}]$  where  $W_1 \sim \text{unif}[M]$  is the users message. Observe that this is similar to the case dealt in [28], but the decoder is performing list decoding. Using the meta converse variation for list decoding, e.g. [30, Proposition 3], we can modify the converse bound in [28] that results in replacement of  $\log M$  with  $\log(M/K_a)$ . We note that [31, Lemma 39] holds here (this is used in the the converse bound of [28]). Combining theses with implications (10), (11) and (12) we have the theorem. ■

## V. LOW-COMPLEXITY ITERATIVE CODING SCHEME

In this section, we present a low-complexity iterative coding scheme based on LDPC codes, which allows one to decode user messages in a slot.

Recall that the users utilize the same codebook. Let us denote it by  $\mathcal{C}$  and explain how to construct it. We start with a binary  $[n, k]$  LDPC codebook and replace each 0 with  $+\sqrt{P}$  and each 1 with  $-\sqrt{P}$ . Let us show the bit-wise MAP decoding rule for the  $j$ -th bit of the  $i$ -th user below

$$\hat{X}_{i,j} = \arg \max_{X_{i,j} \in \pm\sqrt{P}} \mathbb{E} \left[ \sum_{\sim X_{i,j}} p_{Y|X} \left( Y \mid \sum_{k=1}^T H_k X_k \right) \prod_{k=1}^T \mathbb{1}_{X_k \in \mathcal{C}} \right] \quad (13)$$

where the expectation is taken over  $H_1, H_2, \dots, H_T$ . Following [27], the summation “ $\sim X_{i,j}$ ” means that we sum over all positions in all user codewords, except  $X_{i,j}$ .

### A. Alternating BP-decoder general description

The decoder aims to recover all the codewords based on the received vector  $Y$ . The decoder employs a low-complexity iterative belief propagation (BP) decoder that deals with a received soft information presented in a log-likelihood ratio (LLR) form. The decoding system can be represented as a graph (factor graph, [32]), which is shown in Fig. 1.

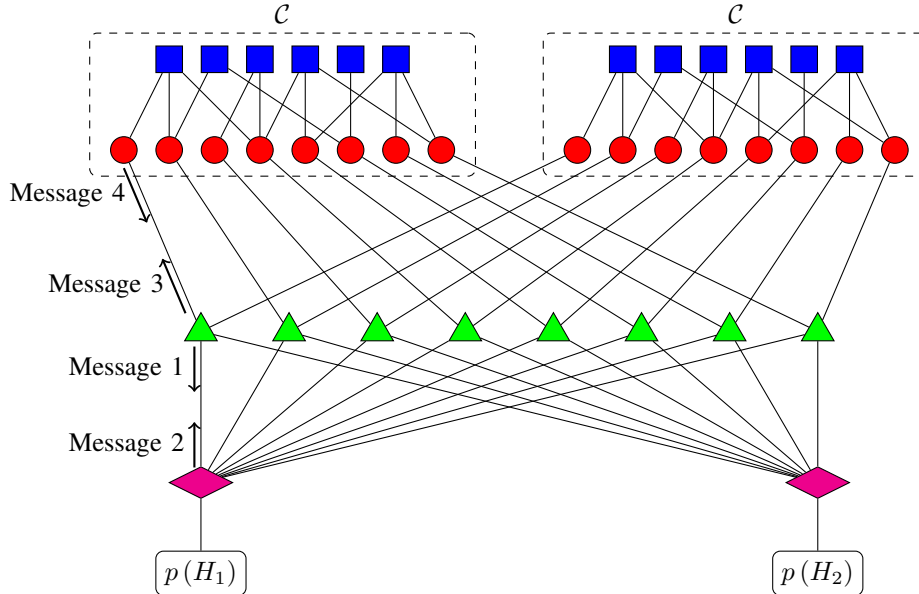


Fig. 1: Iterative joint decoding algorithm (alternating BP-decoder), factor graph

There are four types of nodes in the graph. User LDPC codes are presented with the use of Tanner graphs with variable (red color) and check nodes (blue color). At the same time, there is a third kind of nodes in the figure – functional nodes (green color). These nodes correspond to the elements of the received vector  $Y$ . The fourth kind of nodes (magenta nodes) corresponds to fading coefficients. We note that the decoder also performs an estimation of fading coefficients (latent variables).

The decoding algorithm is based on the iterative message passing procedure. There are two types of iterations in our system: inner iterations, which are used for LDPC code decoding and outer iterations used for fading coefficients estimation. In what follows we mean an outer iteration in all the cases where the type of iteration is not specified. The user codewords are decoded in a sequential manner. Let us consider a single user decoding. This process consists of the calculation and passing of four message types (see Fig. 1). We note that both fading coefficients and LLRs for other users remain fixed during this process. Every message is described in details below:

*a) Message type 1 (from functional nodes to fading nodes):* Without loss of generality let us consider the first functional node. Assume we received a symbol  $y$ . By  $x_i = X_{i,1} \in \{+\sqrt{P}, -\sqrt{P}\}$ ,  $i = 1, \dots, T$ , we denote symbols sent by the users. Let us show how to calculate a posterior probability density function (pdf) of  $H_1$  from the first functional node. We denote this message by  $R_1^{(1)}$  and calculate it as follows

$$R_1^{(1)}(h_1) \propto \mathbb{E} \left[ \sum_{x_1, x_2, \dots, x_T} p(y | \sum_{j=1}^T H_j x_j) \prod_{j=2}^T \Pr(x_j) \right], \quad (14)$$

where the expectations are taken over  $H_2, \dots, H_T$ . Such updates are calculated at every functional node and denoted by  $R_1^{(i)}$ ,  $i = 1, \dots, n$ .

*b) Message type 2 (from fading nodes to functional nodes):* We denote the message from  $j$ -th fading node to  $i$ -th functional node by  $Q_j^{(i)}$ , this message is a pdf. To find it we need to calculate the product of incoming messages. Let us consider a message from the first fading to the first functional node, we have

$$Q_1^{(1)}(h_1) = \prod_{i=1}^n R_1^{(i)}(h_1), \quad (15)$$

**Remark 2.** In a conventional message passing algorithm, the outgoing message is calculated based on messages which come through all the edges except its own edge. But here to reduce the complexity we approximate the complicated message update at fading nodes via the product of a few randomly selected incoming messages.

*c) Message type 3 (from functional nodes to LDPC codes):* Let us note, that a posterior LLR for  $x_1$  can be calculated as follows

$$L(x_1) = \log \frac{\mathbb{E} \left[ \sum_{x_1 = +\sqrt{P}, x_2, \dots, x_T} p(y | \sum_{j=1}^T H_j x_j) \prod_{j=2}^T \Pr(x_j) \right]}{\mathbb{E} \left[ \sum_{x_1 = -\sqrt{P}, x_2, \dots, x_T} p(y | \sum_{j=1}^T H_j x_j) \prod_{j=2}^T \Pr(x_j) \right]}, \quad (16)$$

where the expectations are taken over  $H_1, H_2, \dots, H_T$  and  $p(y|a) = \frac{1}{\pi} \exp(-(y-a)^2)$ . Note that for practical implementation the Monte-Carlo sampling method can be used for expectations.

*d) Message type 4 (LDPC decoding):* After functional nodes decoding one needs to update the LLR for a given user with LDPC iterative decoder. Each user utilizes a standard BP decoding algorithm (Sum-Product or Min-Sum, [27]) to decode an LDPC code.

Now, let us present the final message passing decoding algorithm (see Algorithm 1).

---

**Algorithm 1** Iterative decoding algorithm (alternating BP-decoder)

---

- 1: initialize the LLR values of variable nodes for each user code with zero values assuming equal probability for  $\sqrt{P}$  and  $-\sqrt{P}$  values
  - 2: initialize pdf of  $H_i$ ,  $i = 1, \dots, T$ . For each coefficient we have pdf for both real and imaginary parts with prior distribution  $\mathcal{N}(0, 1/2)$  corresponding to Rayleigh fading.
  - 3: **for**  $i_O = 1, \dots, I_O$  **do** ▷ perform  $I_O$  outer iterations
  - 4:   **for**  $u = 1, \dots, T$  **do** ▷ decode users sequentially
  - 5:     Propagate message type 1, eq. (14) ▷ from functional nodes to fading nodes
  - 6:     Propagate message type 2, eq. (15) ▷ from fading nodes to functional nodes
  - 7:     Sample fading coefficients for further expectation estimation at (16) from the fading coefficients pdfs
  - 8:     Propagate message type 3 using sampled fading coefficients, eq. (16) ▷ from functional nodes to LDPC codes
  - 9:     Propagate message type 4 ▷ perform  $I_I$  inner iterations of BP decoder for  $u$ -th user LDPC code.
  - 10:   **end for**
  - 11: **end for**
- 

Below the efficient implementation with Gaussian mixtures (GM) approximating the pdf of fading coefficients is discussed.

### B. Alternating BP-decoder implementation with Gaussian mixtures

Alternating BP-decoder is based on a successive update of LLRs for every codeword and a successive update of the pdfs of fading coefficients  $H_j$ ,  $j = 1, 2, \dots, T$ . To construct a practical implementation of this algorithm, one needs a tractable representation of probability density functions that

- can be easily manipulated during convolution and multiplication procedures,
- retain their form after such kind of transformations through multiple iterations.

The simplest form of pdf approximation that satisfies the listed above requirements is the GM model:

$$\pi(\cdot) = \sum_{l=1}^{\nu} \omega_l \mathcal{N}(\mu_l, \sigma_l^2), \quad \sum_{l=1}^{\nu} \omega_l = 1, \quad (17)$$

where  $\pi$  is the pdf approximation and  $\mathcal{N}(\mu, \sigma^2)$  is the Gaussian pdf with the mean  $\mu$  and variance  $\sigma^2$ . The sum of two random variables having the pdf in the form of (17) remains a GM. We also note that GM is a conjugate prior with respect to itself, which helps to construct the pdf of  $H_j$  given the pdf of (14) at each functional node.

Now let us specify how the Algorithm 1 can be implemented with the use of GMs and describe the steps of every outer iteration. Without loss of generality suppose that the user 1 is being decoded.

The first step of outer iteration is to update the fading coefficient for a given user at every functional node (see eq. (14)). This can be done as follows. Rewrite eq. (14) via GMs. Let us consider the  $i$ -th functional node. Note that

$$H_1^{(i)} x_1^i = y_i - \left( \sum_{j=2}^T H_j^{(i)} x_j^{(i)} + z \right). \quad (18)$$

Given the LLR of some bit  $x_j^{(i)}$  and the GM representing the coefficient

$$H_j^{(i)} \sim \sum_{l=1}^{\nu} \omega_l \mathcal{N}(\mu_l, \sigma_l^2),$$

the variable  $H_j^{(i)} x_j^{(i)}$  will also be a GM in the following form

$$H_j^{(i)} x_j^{(i)} \sim \sum_{l=1}^{\nu} \omega_l P(x_j^{(i)} = +\sqrt{P}) \mathcal{N}(\sqrt{P}\mu_l, P\sigma_l^2) + \sum_{l=1}^{\nu} \omega_l P(x_j^{(i)} = -\sqrt{P}) \mathcal{N}(-\sqrt{P}\mu_l, P\sigma_l^2) \quad (19)$$

As soon as the sum of random variables has the pdf that equals to the convolution of every single pdf, the right-hand side of equation (18) is a convolution of GMs. This procedure is straightforward, but the resulting GM component count grows as a product of component counts of every GM under the convolution. One can see, that the  $y_i - \sum_{j=2}^T H_j x_j^{(i)}$  is also a GM as  $y_i$  is a constant. Also, note that this procedure is performed separately for both real and imaginary parts of the signal.

The final step is to construct the  $H_1^{(i)}$  pdf given the GM on the right-hand side of eq. (18) and the LLR for  $x_1^{(i)}$ . The coefficient  $H_1^{(i)}$  has a GM pdf that can be calculated in exactly the same manner as in equation (19). Suppose the RHS of (18) has a pdf

$$H_1^{(i)} x_1^i \sim \sum_{l=1}^{\nu} \omega_l \mathcal{N}(\mu_l, \sigma_l^2).$$

then the pdf of the coefficient  $H_1^{(i)}$  can be calculated as follows

$$H_1^{(i)} \sim \sum_{l=1}^{\nu} \omega_l P(x_1^{(i)} = +\sqrt{P}) \mathcal{N}\left(\frac{\mu_l}{\sqrt{P}}, \frac{\sigma_l^2}{P}\right) + \sum_{l=1}^{\nu} \omega_l P(x_1^{(i)} = -\sqrt{P}) \mathcal{N}\left(-\frac{\mu_l}{\sqrt{P}}, \frac{\sigma_l^2}{P}\right). \quad (20)$$

The second step of the outer iteration is to derive the fading coefficient estimate  $H_1$  given the messages  $H_1^{(i)}$  from every functional node (20). This can be done by multiplying the corresponding pdfs (see eq. (15)). Note that as in the case of convolution, the product of two GMs is also a GM with the number of components equal to the product of the number of components in the multipliers.

The next two steps in the outer iteration are sampling from GM and functional nodes decoding procedure (see eq. (16)).

As it was mentioned before, the final step of the outer iteration is a simple iterative decoding algorithm, that just updates the user's codeword LLRs. The outer iterations are performed over every user successively until the maximum iteration count per user is reached.

### C. Gaussian mixture pruning and components merging

The decoding algorithm performs the convolution and multiplication of multiple GMs at every iteration. In this subsection, these procedures are described in more detail as well as the approach to limiting the ever-growing number of components in the final GM is presented.

The convolution of  $GM_1 \otimes GM_2$  with

$$GM_1 = \sum_{l_1=1}^{\nu_1} \omega_{l_1} \mathcal{N}(\mu_{l_1}, \sigma_{l_1}^2), \quad GM_2 = \sum_{l_2=1}^{\nu_2} \omega_{l_2} \mathcal{N}(\mu_{l_2}, \sigma_{l_2}^2).$$

is the GM that has  $\nu_1 \times \nu_2$  components:

$$GM_1 \otimes GM_2 = \sum_{l_1=1}^{\nu_1} \sum_{l_2=1}^{\nu_2} \omega_{l_1} \omega_{l_2} \mathcal{N}(\mu_{l_1} + \mu_{l_2}, \sigma_{l_1}^2 + \sigma_{l_2}^2).$$

The GM product is given by (21). The result has also  $\nu_1 \times \nu_2$  components.

$$GM_1 \times GM_2 = \sum_{l_1=1}^{\nu_1} \sum_{l_2=1}^{\nu_2} \frac{\omega_{l_1} \omega_{l_2}}{\sqrt{2\pi(\sigma_{l_1}^2 + \sigma_{l_2}^2)}} \exp \left\{ -\frac{(\mu_{l_1} - \mu_{l_2})^2}{2(\sigma_{l_1}^2 + \sigma_{l_2}^2)} \right\} \mathcal{N} \left( \frac{\sigma_{l_1}^2 \sigma_{l_2}^2}{\sigma_{l_1}^2 + \sigma_{l_2}^2} \left( \frac{\mu_{l_1}}{\sigma_{l_1}^2} + \frac{\mu_{l_2}}{\sigma_{l_2}^2} \right), \frac{\sigma_{l_1}^2 \sigma_{l_2}^2}{\sigma_{l_1}^2 + \sigma_{l_2}^2} \right) \quad (21)$$

Note, that in practical implementation it is better to manipulate with the logarithm of the Gaussian component weight for numerical stability. The component count optimization procedures are described below and include merge and prune steps.

1) *Gaussian mixtures pruning*: One can see that both GM convolution and product significantly increase the number of components. For practical implementation, one needs to limit the number of Gaussian components. The first step consists of removing the components with low weights (pruning). This can be easily done by sorting the weights in ascent order and removing several first components whose cumulative weight is less than some threshold.

2) *Gaussian mixtures components merge*: The GM components which are “close” to each other (with the distance measure specified below) must be merged. This approach is described in details in [33]. The procedure starts from the “heaviest” component. All other components that have the distance less than some threshold form a merge-list. This distance can be calculated as follows

$$d = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2} \leq d_{min},$$

where component 1 has a higher weight than component 2. After the merge list of length  $\nu_0$  has been constructed, all the components from this list are replaced by a new component  $\omega \mathcal{N}(\mu, \sigma^2)$  with the following parameters:

$$\omega = \sum_{l=1}^{\nu_0} \omega_l, \quad \mu = \sum_{l=1}^{\nu_0} c_l \mu_l, \quad \sigma^2 = \frac{1}{\omega} \sum_{l=1}^{\nu_0} \mu_l \left( \sigma_l^2 + (\mu_l - \mu)^2 \right)$$

Note that each component can be merged with any other only once during the GM-merge procedure.

The final step of GM pruning is to apply a hard limit on the maximum components count. This is done for performance stability and helps to control the maximum GM length.

### D. Blind detection and error floor

As soon as the iterative decoder operates as an optimization task and this optimization procedure is split between two groups of variables (user LLRs and fading coefficients), one can expect this algorithm to converge to some local maximum of (13). Convergence to a local maximum can be a source of the error floor. To overcome the error floor problem one can start the decoding algorithm multiple times and handle functional nodes in random order at every decoding iteration. As soon as GMs are merged and pruned, this provides some source of randomness and pushes the decoding procedure to possibly different local maximums. This approach has eliminated the error floor problem and allowed another opportunity – a blind detection. Given the multiple decoding attempts, one can select a set of unique codewords that were successfully decoded. Every attempt can detect different codewords. The final output of the decoder is the union of such sets. Without loss of generality, this approach can be applied to the case of unknown user count. As further numerical experiments (see appendix D) show, this approach is a promising one.

**Remark 3.** *Even though the number of users in a slot is unknown we never faced with a false alarm problem in our simulations. By false alarm, we mean a situation in which the output list contains codewords that were not transmitted. To explain this fact we note that LDPC codes have a large area of inputs for which they report a failure (the decoder cannot converge to a codeword). Thus, we mention once again that  $Q_e(\mathcal{C}, r) \approx 0$  (within the accuracy of the Monte Carlo) for all  $r \geq 0$ .*



**Remark 4.** The approach presented in this paper is similar to the approach from [34]. Nevertheless, the main differences are: a) we consider same codebook case and changed the parallel schedule with serial schedule in order to break symmetry, b) we show that this approach allows to efficiently perform blind user decoding, i.e. determine the number of active users in a slot and recover their messages, c) we suggest an approach how to deal with the error floor caused by the inaccuracy in the estimation of fading coefficients ( $H_i$ ,  $i = 1, \dots, T$ ).

## VI. NUMERICAL RESULTS AND DISCUSSION

In this section we present the plots of the minimum energy per bit required to achieve a probability of error  $\epsilon = 0.1$  as a function of  $K_a$  for the channel (2). Figure 2 shows plots of various schemes. The parameters used for evaluation are blocklength  $n = 30000$  and message size  $k = 100$  bits. Next we describe how each of these curves was obtained.

For  $T$ -fold ALOHA using FBL bound, we use the bound for  $p_t$  given in (33). For each  $K_a$  we find the optimum  $L$  (as an optimization over both  $L$  and  $P$ ) so that we minimize  $E_b/N_0$  such that the probability of error in (5) is less than 0.1. Since directly optimizing the bound is not easy, we approximate PUPE for the fading channel as [31]

$$P_e(M, n_1, r, LP) \approx \mathbb{E} \left[ \mathcal{Q} \left( \frac{n_1 C_{AWGN}(LP \sum_{i=1}^r |H_i|^2) - \log_2 M}{\sqrt{n V_{AWGN}(LP \sum_{i=1}^r |H_i|^2)}} \right) \right] \quad (22)$$

where  $C_{AWGN}(x) = \log(1 + x)$  and  $V_{AWGN}(x) = 1 - \frac{1}{(1+x)^2}$  are the capacity and dispersion of a (complex) AWGN channel, respectively. We choose  $L$  by using (22) in (5). Then we use the spherical codebook, i.e. codewords uniformly and independently sampled from the (complex) power shell in dimension  $n_1 = \lfloor n/L \rfloor$  to compute the probability of error according to (5) where  $P_e(M, n_1, r, LP)$  is computed using brute-force Monte-Carlo simulation of (33) with the choice  $K_1 = K_2 = r$ . Since  $r \leq T$  is small it would not make sense to drop a user. To this end, we produce 2000 samples, from which we construct the kernel density approximation of the cumulative distributive function (CDF) of the statistic  $\max_{|S_0| \leq t} G(Y, S_0, c_{S_0}, t)$  (given in (34)) for each  $t \leq r$ . Then this smooth approximation is used to optimize over  $\delta$  in (5).

For  $T$ -fold ALOHA using the iterative coding scheme, we have used  $(n_1, k)$  LDPC codes with  $k = 100$  and blocklength  $n_1 \in \{200, 400\}$ . We note, that two codes are enough to cover the interval  $1 \leq K_a \leq 250$ . For each of these codes, we get PUPE vs  $E_b/N_0$  curves and choose the best code (the best code requires the smallest  $E_b/N_0$  in order to achieve  $\text{PUPE} \leq \epsilon = 0.1$ ) for each value of  $K_a$ . The best waterfall curves for the different number of users are presented in Fig. 3. Iterative decoder used the multiple component Gaussian mixture model with parameters listed in Table I. Note again, that in LDPC-based scheme we perform honest blind slot decoding (without assuming the knowledge of user count in a slot).

It can be seen from Fig. 2 that the performance of  $T$ -fold ALOHA for iterative decoding scheme is very close to that of  $T$ -fold ALOHA with random coding bounds for small  $K_a$ . The gap increases with  $K_a$  because of our limited choices of LDPC codes, i.e. due to BPSK modulation, we are constrained by  $n_1 \geq k$ . We refer to remark 1 again to emphasize that the  $T$ -fold ALOHA with the FBL bound is not a true achievability bound since it assumes that the decoder has knowledge of the number of users in each slot or subframe.

We have also plotted the result of treat interference as noise (TIN) decoding. Here we have used optimistic capacity approximation for PUPE.

$$\epsilon \approx \mathbb{E} \left[ \mathcal{Q} \left( \frac{n C_{AWGN} \left( \frac{P |H_1|^2}{1 + P \sum_{i=2}^T |H_i|^2} \right) - k}{\sqrt{n V_{AWGN} \left( \frac{P |H_1|^2}{1 + P \sum_{i=2}^T |H_i|^2} \right)}} \right) \right] \quad (23)$$

It is easy to get an actual random coding bound for TIN similar to theorem A.1, but we don't expect it to be better than (23).

Also plotted for reference is the Shamai-Bettesh capacity bound from [35]. It is an asymptotic bound ( $n \rightarrow \infty$ ) for the probability of error per-user in the case of symmetric rate and large  $K_a$ . But, it doesn't assume same codebook. The idea is the following. The joint decoder knows the realization of fading coefficients and users are ranked according the strength of their fading coefficients. It first tries to decode all users. If it fails (i.e., the rate vector is not inside the instantaneous full capacity region), it drops the user with least fading coefficient and tries to decode the remaining  $K_a - 1$  users. The dropped user forms part of the noise. This process continues iteratively, and the fraction of users that were not decoded is precisely the outage/probability of error per-user. Since the case under discussion is for large  $K_a$ , the order statistics of the absolute value of fading coefficients crystallize (i.e., become almost non-random) and hence analytical expressions can be derived for outage in terms of spectral efficiency ( $kK_a/n$ ) and total power. So for each  $K_a$ , we know our operating spectral efficiency and total power, and hence we can use the asymptotic bound to find the probability of error. Most importantly observe that even at  $K_a = 100$ , the random coding based 4-fold ALOHA performance is off from the capacity bound of [35] by just 3 dB.

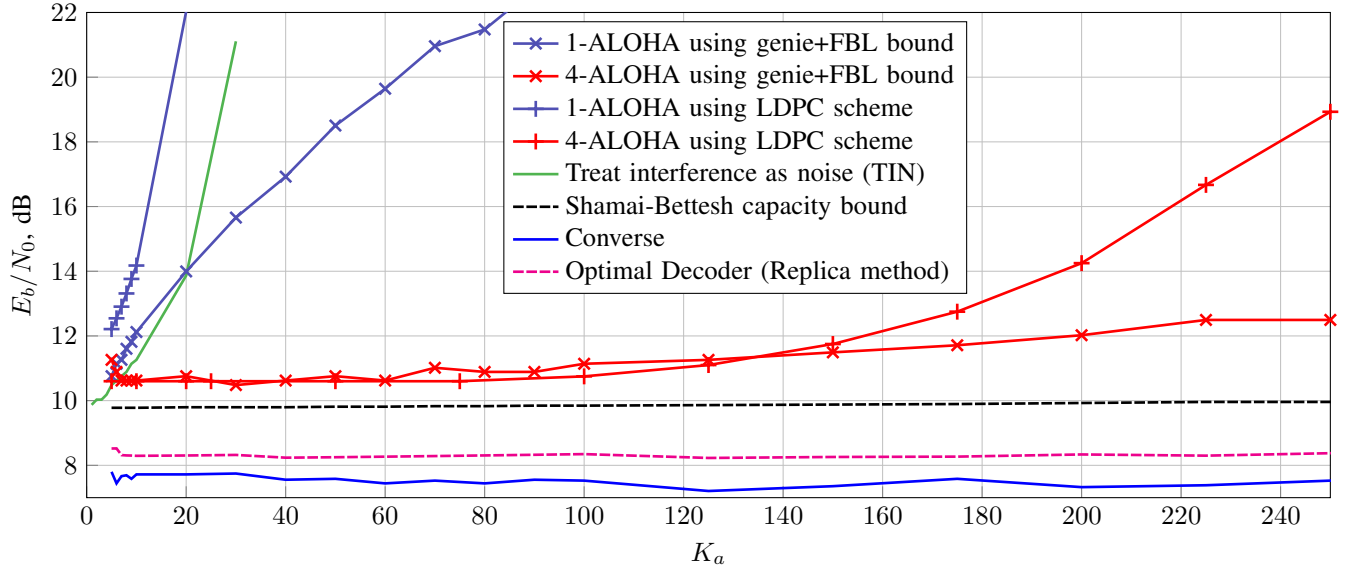


Fig. 2:  $K_a$  vs  $E_b/N_0$  for  $\epsilon \leq 0.1$ ,  $n = 30000$ ,  $k = 100$  bits. Dashed lines correspond to asymptotic approximation obtained by taking  $n \rightarrow \infty$  and are shown only for reference.

The converse from (8) and (9) is also plotted. This is in essence a single user based converse bound. We can also derive a Fano type converse, but for the range of parameters we work with, it is worse than the presented one. The converse presented here illustrates the fact the  $E_b/N_0$  requirements are necessarily higher compared to the AWGN channel in [1].

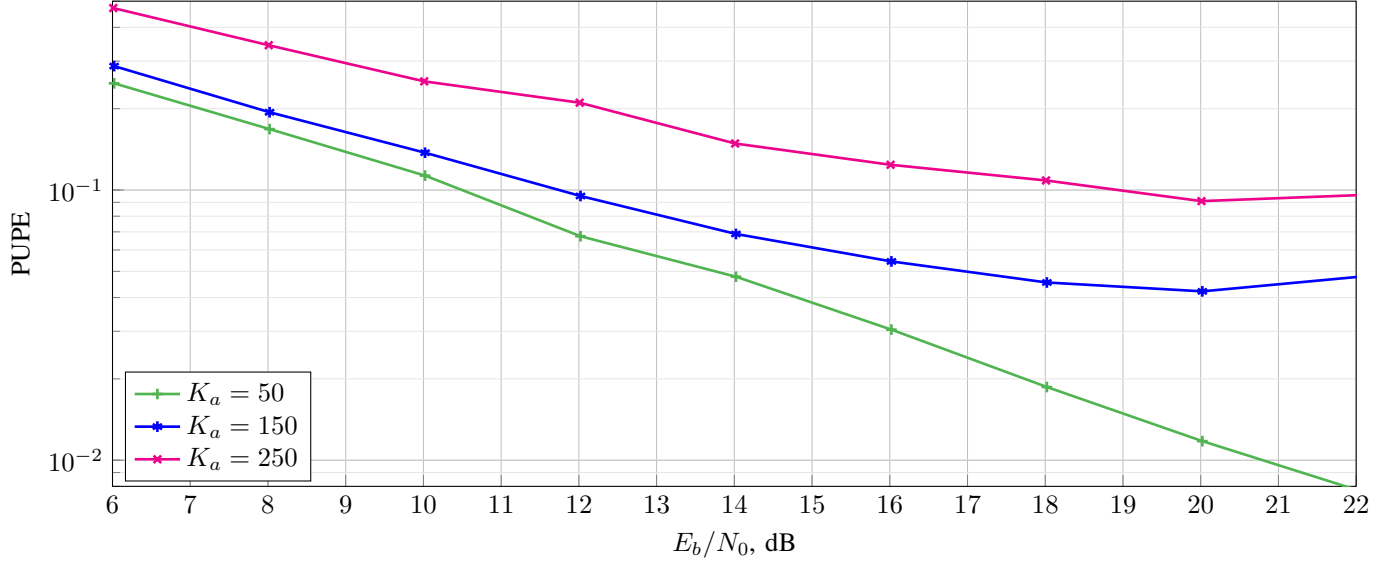


Fig. 3:  $E_b/N_0$  vs  $PUPE$  for  $n = 30000$ ,  $k = 100$  bits

## VII. ASYMPTOTICS OF RANDOM-ACCESS

In [1] the authors evaluated a random coding bound for AWGN RAC with  $n = 30000$  and  $K_a = 1, \dots, 300$ . The most interesting observation was that the bound on energy-per-bit was essentially constant up until about  $K_a = 150$  and only then started to increase with  $K_a$ . To explain this "phase transition" behavior a particular asymptotics was postulated in [12], which predicts the phase transition at roughly the same value of  $K_a = 150$ . It turned out that at low  $K_a$  the performance was essentially limited by the minimal energy required for a single user to send  $k$  bits over a fixed (but effectively infinite) blocklength. For larger number of  $K_a$  the performance is limited by the multi-user requirement: the total number of  $K_a \times k$  bits should not exceed the combined mutual information of  $n \log(1 + PK_a)$ .

In the present paper we adopt the very same asymptotics of [1, 12]. Again, we stress that the only ultimately relevant question is the one at finite blocklength. The asymptotic analysis here is only to get some insight into the possible regimes.

Specifically, we consider scaling of  $n \rightarrow \infty$  with  $K_a$ , the number of active users, scaling linearly with blocklength (similar to the *many-access* regime [1, 22, 23]) i.e.,  $K_a = \mu n$ . At the same time, the size of the common codebook is also scaling linearly:  $M = M_1 K_a$ . Since we operate in the same-codebook scenario this means that the common codebook size scales linearly with number of users:  $M = M_1 K_a$ . We think of  $M_1$  as the effective payload per user. We also modify the random-access model slightly by requiring that the messages of active users  $\{W_1, \dots, W_{K_a}\}$  are sampled uniformly from  $\binom{[M]}{K_a}$  i.e., user messages are sampled *uniformly without replacement* from  $[M]$ . (In reality, the user messages are distributed iid  $\text{Unif}[M]$  which leads to around  $\frac{\binom{K_a}{2}}{M}$  collisions but for finite length scenarios with  $M_1 = 2^{100}$ , this is essentially zero, hence we may ignore collisions in our asymptotic setup and simplify the analysis.) If  $P$  denotes the power (per symbol) of each user, then the energy-per-bit  $E_b/N_0$  is defined by

$$E_b/N_0 = \frac{nP}{\log M_1}. \quad (24)$$

Hence, for finite  $E_b/N_0$ , we need the total sum-power  $P_{tot} = K_a P$  to be constant. Therefore, the asymptotic energy-per-bit, denoted by  $\mathcal{E}$  is given by

$$\mathcal{E} = \frac{P_{tot}}{\mu \log M_1}. \quad (25)$$

We note that  $E_b/N_0$  is defined this way for the reason that  $\log \binom{M}{K_a} \approx K_a \log M$  for relevant finite-length values.

Lastly, the error metric is PUPE. We are interested in the trade-off of minimum  $\mathcal{E}$  required to achieve a target PUPE with the user density  $\mu$  as  $n \rightarrow \infty$ .

This setup is equivalent to the support recovery in compressed sensing considered in [19, 36]. Here, we provide a comparison of the fundamental trade-off of energy-per-bit with user density, for given PUPE and  $\rho$ , between our analysis of the projection decoder, the ML decoder in [19], the optimal decoder based on the true posteriors (see [19, Theorem 8] for instance, this assumes replica symmetry to hold) and finally a converse.

To formally state our results we modify the definition of  $(M, n, \epsilon)$  code for the  $K_a$  user channel  $P_{Y^n|X^n}$  given in (2) as follows.

**Definition 2.** An  $(M, n, \epsilon)$  random-access code for the  $K_a$  user MAC  $P_{Y^n|X^n}$  is a pair of (possibly randomized) maps  $f : [M] \rightarrow \mathcal{X}^n$  (the encoder) and  $g : \mathcal{Y}^n \rightarrow \binom{[M]}{K_a}$  such that if  $W_1, \dots, W_{K_a}$  are sampled uniformly without replacement from  $[M]$  and  $X_j = f(W_j)$  then the average (per-user) probability of error satisfies

$$P_e = \frac{1}{K_a} \sum_{j=1}^{K_a} \mathbb{P}[W_j \notin g(Y^n)] \leq \epsilon \quad (26)$$

where  $Y^n$  is the channel output.

Define  $(n, M, \epsilon, \mathcal{E}, K_a)$ -code as an  $(M, n, \epsilon)$  random access code (from definition 2) for the  $K_a$ -MAC with codebook  $\mathcal{C}$  such that  $\|c\|^2 \leq nP = \mathcal{E} \log M_1, \forall c \in \mathcal{C}$ . Then we can define the following fundamental limit

$$\mathcal{E}^*(M_1, \mu, \epsilon) = \limsup_{n \rightarrow \infty} \inf \{ \mathcal{E} : \exists (n, M = K_a M_1, \epsilon, \mathcal{E}, K_a = \mu n) \text{ - code} \}. \quad (27)$$

In appendix B we sandwich the fundamental limit between an achievability and a converse bound as follows:

$$\mathcal{E}_{conv} \leq \mathcal{E}^* \leq \mathcal{E}_{ach}. \quad (28)$$

For particular, quite cumbersome, expressions please refer to Appendix B.

These bounds are plotted in figures 4 and 5 for two different values of PUPE. The main achievability bound is from theorem B.1 and is based on the analysis of projection decoding described in appendix A. A different analysis of this decoder was performed in [19] and the result is plotted as well. We have also plotted predicted performance of the PUPE-optimal decoder for the iid codebook which is obtained via a non-rigorous (but highly likely to be correct) replica-method from statistical physics; see appendix B-B and [19]).

The converse bound plotted is based on Fano inequality and the single-user converse for AWGN channel from [37]. The details are in appendix B-C. A tighter converse (see theorem B.3) bound can be obtained if we assume that the codebook consists of iid entries of the form  $\frac{C}{K_a}$  where  $C$  is of zero mean and finite variance. This follows from [36, Theorem 37]. This bound, although only applicable to a special class of codes (iid codebooks), improves our converse bound by taking into account the penalty incurred due to absence of knowledge of the channel state information at the decoder (resulting in a need to spend some of the information on estimating the fading coefficients).

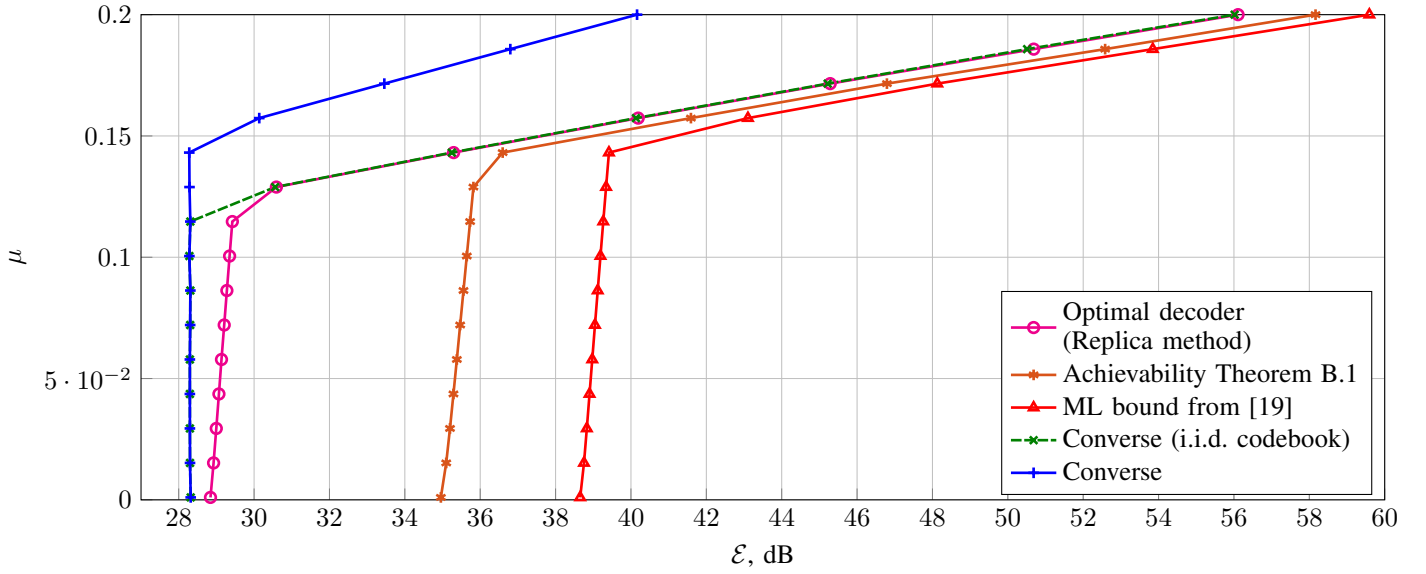


Fig. 4:  $\mu$  vs  $\mathcal{E}$  for  $\epsilon \leq 10^{-3}$ ,  $M_1 = 2^{100}$

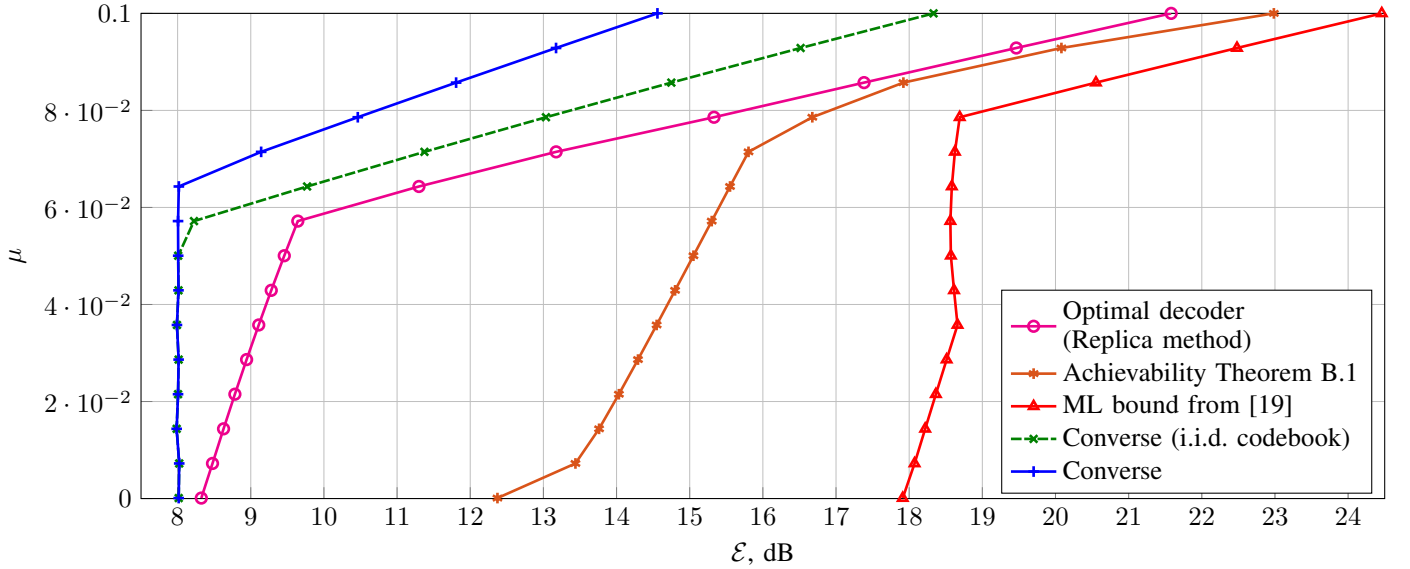


Fig. 5:  $\mu$  vs  $\mathcal{E}$  for  $\epsilon \leq 0.1$ ,  $M_1 = 2^{100}$

### VIII. CONCLUSION AND FUTURE WORK

In this work we considered random access for a quasi-static Rayleigh fading model. We developed low-complexity iterative decoding scheme using LDPC codes to decode up to  $T$ -users in a slot, and using  $T$ -fold ALOHA on top of it gave us a practical achievable scheme whose required  $E_b/N_0$  vs  $K_a$  trade-off is very close to that of a potential random coding bound. In terms of future work, one of the most important things is to figure out how to relax the assumption on the knowledge of the number of users in a slot in  $T$ -fold ALOHA to get a rigorous random coding achievability bound. Another important factor is frame-synchronization which we have assumed. Our rationale is that frame-synchronism can be achieved via regularly spaced beacons. However, to reduce complexity even further it would be interesting to develop a beacon-free (and, hence, frame-asynchronous) schemes. Finally, large gains in energy consumption can be attained via the use of MIMO, especially multiple receive antennas. Quantifying these gains is yet another interesting direction.

### REFERENCES

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

- [2] 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.
- [3] 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.
- [4] A. Fengler, G. Caire, P. Jung, and S. Haghghatshoar, "Massive MIMO Unsourced Random Access," *arXiv preprint arXiv:1901.00828*, 2019.
- [5] 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.
- [6] S. Ghez, S. Verdú, and S. C. Schwartz, "Stability properties of slotted Aloha with multipacket reception capability," *IEEE Trans. Autom. Control*, vol. 33, no. 7, pp. 640–649, 1988.
- [7] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 477–487, 2011.
- [8] A. Glebov, N. Matveev, K. Andreev, A. Frolov, and A. Turlikov, "Achievability Bounds for T-Fold Irregular Repetition Slotted ALOHA Scheme in the Gaussian MAC," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019.
- [9] 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.
- [10] A. Pradhan, V. Amalladinne, A. Vem, K. R. Narayanan, and J.-F. Chamberland, "A joint graph based coding scheme for the unsourced random access gaussian channel," *arXiv preprint arXiv:1906.05410*, 2019.
- [11] G. Caire and R. Muller, "The optimal received power distribution for ic-based iterative multiuser joint decoders," in *Proc. Allerton Conf. Comm. Control Comp.*, vol. 39, no. 2, 2001, pp. 1132–1141.
- [12] 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.
- [13] K. Stern, A. E. Kalør, B. Soret, and P. Popovski, "Massive random access with common alarm messages," *arXiv preprint arXiv:1901.06339*, 2019.
- [14] Z. Chen, F. Sahrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1890–1904, 2018.
- [15] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. De Carvalho, "Sparse Signal Processing for Grant-Free Massive IoT Connectivity," *arXiv preprint arXiv:1804.03475*, 2018.
- [16] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2933–2946, 2018.
- [17] S. Haghghatshoar, P. Jung, and G. Caire, "Improved scaling law for activity detection in massive MIMO systems," *arXiv preprint arXiv:1803.02288*, 2018.
- [18] 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.
- [19] G. Reeves and M. Gastpar, "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.
- [20] —, "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.
- [21] G. Reeves, "Sparse Signal Sampling using Noisy Linear Projections," Univ. Calif., Berkeley, Dept. of Electrical Engineering and Computer Science, Tech. Rep., 2008.
- [22] X. Chen, T.-Y. Chen, D. Guo *et al.*, "Capacity of gaussian many-access channels," *IEEE Trans. Information Theory*, vol. 63, no. 6, pp. 3516–3539, 2017.
- [23] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *arXiv preprint arXiv:1901.06732*, 2019.
- [24] —, "Quasi-static fading MAC with many users and finite payload," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019.
- [25] R. G. Gallager, *Low-Density Parity-Check Codes*. M.I.T. Press, 1963.
- [26] R. Tanner, "A recursive approach to low complexity codes," *IEEE Transactions on information theory*, vol. 27, no. 5, pp. 533–547, 1981.
- [27] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.
- [28] 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.
- [29] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits*. Princeton University, 2010.
- [30] V. Y. Tan and P. Moulin, "Fixed error asymptotics for erasure and list decoding," *arXiv preprint arXiv:1402.4881*, 2014.
- [31] 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.
- [32] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on information theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [33] D. E. Clark, K. Panta, and B. n. Vo, "The GM-PHD Filter Multiple Target Tracker," in *2006 9th International Conference*

on *Information Fusion*, July 2006, pp. 1–8.

- [34] M. Kobayashi, J. Boutros, and G. Caire, “Successive interference cancellation with SISO decoding and EM channel estimation,” *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1450–1460, Aug 2001.
- [35] 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.
- [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] 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.
- [38] W. Van Zwet, “A strong law for linear functions of order statistics,” *The Annals of Probability*, pp. 986–990, 1980.
- [39] L. Birgé, “An alternative point of view on Lepski’s method,” *Lecture Notes-Monograph Series*, pp. 113–133, 2001.

APPENDIX A  
FBL ACHIEVABILITY BOUNDS

In this section we state the random coding FBL achievability bounds for the model in (2). But first, we discuss the encoding and decoding which we use to derive achievability. For encoding, we use random coding with **Gaussian codebook**: for each message a  $\mathcal{CN}(0, P'I_n)$  vector is independently generated. That is  $X_i \stackrel{iid}{\sim} \mathcal{CN}(0, P'I_n)$  where  $P' \leq P$ . For a message  $W_j$  of user  $j$ , if  $\|X(W_j)\|^2 > nP$  then that user sends 0.

A. *Projection decoding*

Inspired from [28], we use a projection based decoder. The idea is the following. Suppose there was no additive noise. Then the received vector will lie in the subspace spanned by the sent codewords no matter what the fading coefficients are. Fix an output list size  $K_1$ . The decoder outputs a list of  $K_1$  codewords which form the subspace, such that projection of  $Y$  onto this subspace is maximum. Formally, 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  $\mathcal{C}$  denote the common codebook. Then, upon receiving  $Y$  from the channel, the decoder outputs  $g(Y)$  given by

$$\begin{aligned} g(Y) &= \{f^{-1}(c) : c \in \hat{C}\} \\ \hat{C} &= \arg \max_{C \subset \mathcal{C}: |C|=K_1} \|P_C Y\|^2 \end{aligned} \quad (29)$$

where  $f$  is the encoding function.

The projection decoding is also called nearest-subspace decoding, and has been used in the compressed sensing literature [18–21]. One might prefer to view it as a kind of maximum likelihood (ML) decoding as well (and is called as such), since it is equivalent to

$$\hat{C} = \arg \max_{C \subset \mathcal{C}: |C|=K_1} \max_{\{H_i: i \in C\}} P_{Y|X, H} \quad (30)$$

$$P_{Y|X, H}(y, \{x_i\}, \{h_i\}) = \frac{1}{\pi^n} e^{-\|y - \sum_i h_i x_i\|^2} \quad (31)$$

It can be shown that for the vanilla  $K_a$ -user quasi-static fading MAC (with different codebook and the usual joint probability of error) with no channel state information, projection decoding achieves  $\epsilon$ -capacity region  $C_\epsilon$  of the MAC [23].

B. *FBL Achievability bounds*

**Theorem A.1.** Fix  $P' < P$ . Let  $K_1 \leq K_2$ . Then there exists an  $(M, n, \epsilon)$  (with  $\epsilon \geq \frac{K_2 - K_1}{K_2}$ ) random access code for the  $K_2$ -MAC (2) satisfying power constraint  $P$  (see (3)) and

$$\epsilon \equiv P_e(M, n, K_2, P) \leq \frac{K_2 - K_1}{K_2} + \frac{1}{K_2} \sum_{t=1}^{K_1} K_{1,t} p_t + p_0 \quad (32)$$

with

$$p_0 = \frac{\binom{K_2}{2}}{M} + K_2 \mathbb{P} \left[ \frac{P'}{2} \sum_{i \in [2n]} W_i^2 > nP \right], \quad W_i \stackrel{iid}{\sim} \mathcal{N}(0, 1),$$

and

$$p_t \leq \inf_{\delta > 0} \left( \binom{K_2}{K_{1,t}} e^{-(n-K_1)\delta} + \mathbb{P} \left[ \bigcup_{\substack{S_0 \subset [K_2] \\ |S_0|=K_{1,t}}} \{G(Y, S_0, c_{S_0}, t) \geq V_{n,t}\} \right] \right) \quad (33)$$

where

$$G(Y, S_0, c_{S_0}, t) = \frac{\|Y\|^2 - \max_{\substack{S_2 \subset S_0 \\ |S_2|=t}} \|P_{c_{[S_2 \cup ([K_2] \setminus S_0)]}} Y\|^2}{\|Y\|^2 - \|P_{c_{[[K_2] \setminus S_0]}} Y\|^2} \quad (34)$$

$$K_{1,t} = K_2 - K_1 + t \quad (35)$$

$$V_{n,t} = e^{-\tilde{V}_{n,t}} \quad (36)$$

$$\tilde{V}_{n,t} = \delta + R_1 + s_t \quad (37)$$

$$s_t = \frac{\ln \binom{n'-1}{t-1}}{n - K_1} \quad (38)$$

$$R_1 = \frac{\ln \binom{M-K_2}{t}}{n - K_1} \quad (39)$$

$$n' = n - K_1 + t \quad (40)$$

and,  $\mathcal{C} = \{c_i : i \in [M]\}$  denotes the Gaussian codebook,  $\{c_i : i \in [K_2]\}$  are the transmitted codewords,  $c_S = \{c_i : i \in S\}$ ,  $Y$  is the received vector.

Further, the right hand side of (33) can be upper bounded as

$$p_t \leq \inf_{\substack{\delta > 0 \\ \delta_1 > 0 \\ 0 < \delta_2 < 1}} \left[ \binom{K_2}{K_{1,t}} \left( e^{-(n-K_1)\delta} + e^{-n'f_n(\delta_1)} + e^{-n'\frac{\delta_2^2}{2}} \right) + \mathbb{P} \left[ \min_{1 \leq i \leq K_1-t+1} \frac{P' \sum_{j=i}^{i+t-1} |H_{(j)}|^2}{1 + P' \sum_{j=i+t}^{K_{1,t}-1+i} |H_{(j)}|^2} \leq \frac{(1 + \delta_1(1 - V_{n,t}))V_{n,t}^{-1} - 1}{1 - \delta_2} \right] \right] \quad (41)$$

where

$$f_n(\delta_1) = \delta_1 + 1 + \frac{2V_{n,t}}{1 - V_{n,t}}(1 + \delta_1) - \sqrt{1 + \frac{2V_{n,t}}{1 - V_{n,t}}(1 + \delta_1)} \sqrt{2\delta_1 + 1 + \frac{2V_{n,t}}{1 - V_{n,t}}(1 + \delta_1)} \quad (42)$$

and  $\{|H_{(j)}|^2 : j \in [K_2]\}$  denotes the order statistics of fading powers (in decreasing order).

*Proof:* See appendix C. ■

**Remark 5.** We note that (33) in the above theorem holds even in case of random coding with spherical codebook i.e., codewords distributed uniformly on the (complex) power shell with  $p_0 = \frac{\binom{K_2}{2}}{M}$ . But (41) requires that the codebook is (complex) Gaussian.

To compute (33) we use Monte-Carlo simulation described in section VI for small values of  $K_2$ . For moderated values of  $K_2$ , the computation of the probability of union of a combinatorially large number of events in (33) is prohibitive. However, there is a computationally tractable bound (which is worse than (33)) on  $p_t$  that we present in appendix C.

We make the following observation about  $K_1$ . When the number of active users  $K_2$  is large, it is hard to decode the message of the user with least fading power, since its expectation is  $\frac{1}{K_2}$ . Consequently, this user becomes a bottleneck. So, intuitively, it makes sense to drop the users with very bad channel gains and decode the rest, and the definition of per-user probability of error makes this possible. Indeed, this was proposed in [35] where the joint multiuser detector drops a fraction of users with smallest gains such that the rate tuple of the remaining users is inside the (random) capacity region. So for each  $K_2$ , we can find the optimum  $K_1$  which is the number of messages that are decoded in a frame.

## APPENDIX B ASYMPTOTICS OF RANDOM-ACCESS

In this section, we provide achievability and converse bounds on  $\mathcal{E}^*$ , defined in (27).

### A. Achievability

**Theorem B.1.** Consider the channel (2) with  $K_a = \mu n$  where  $\mu < 1$ . Fix  $M_1 > 1$  and target PUPE  $\epsilon$ . Let  $M = K_a M_1$  denote the size of the codebook and  $P_{tot} = K_a P$  be the total power. Let  $h(p) = -p \ln p - (1-p) \ln(1-p)$ ,  $p \in [0, 1]$ . Fix  $\nu \in (1 - \epsilon, 1]$ . Let  $\epsilon' = \epsilon - (1 - \nu)$ . Then if  $\mathcal{E} > \mathcal{E}_{ach} = \sup_{\frac{\epsilon'}{\nu} < \theta \leq 1} \sup_{\xi \in [0, \nu(1-\theta)]} \frac{P_{tot,\nu}(\theta, \xi)}{\mu \log M_1}$ , there exists a sequence of  $(n, M = K_a M_1, \epsilon_n, \mathcal{E}, K_a = \mu n)$  codes such that  $\limsup_{n \rightarrow \infty} \epsilon_n \leq \epsilon$ , where, for  $\frac{\epsilon'}{\nu} < \theta \leq 1$  and  $\xi \in [0, \nu(1-\theta)]$ ,

$$P_{tot,\nu}(\theta, \xi) = \frac{\hat{f}(\theta, \xi)}{1 - \hat{f}(\theta, \xi)\alpha(\xi + \nu\theta, \xi + 1 - \nu(1-\theta))} \quad (43)$$

$$\hat{f}(\theta, \xi) = \frac{f(\theta)}{\alpha(\xi, \xi + \nu\theta)} \quad (44)$$



$$f(\theta) = \frac{\frac{1+\delta_1^*(1-V_\theta)}{V_\theta} - 1}{1 - \delta_2^*} \quad (45)$$

$$V_\theta = e^{-\tilde{V}_\theta} \quad (46)$$

$$\tilde{V}_\theta = \delta^* + \mu \frac{(M_1 - 1)}{1 - \mu\nu} h\left(\frac{\theta\nu}{M_1 - 1}\right) + \frac{1 - \mu\nu(1 - \theta)}{1 - \mu\nu} h\left(\frac{\theta\mu\nu}{1 - \mu\nu(1 - \theta)}\right) \quad (47)$$

$$\delta^* = \frac{\mu h(1 - \nu(1 - \theta))}{1 - \mu\nu} \quad (48)$$

$$c_\theta = \frac{2V_\theta}{1 - V_\theta} \quad (49)$$

$$q_\theta = \frac{\mu h(1 - \nu(1 - \theta))}{1 - \mu\nu(1 - \theta)} \quad (50)$$

$$\delta_1^* = q_\theta(1 + c_\theta) + \sqrt{q_\theta^2(c_\theta^2 + 2c_\theta) + 2q_\theta(1 + c_\theta)} \quad (51)$$

$$\delta_2^* = \inf \left\{ x : 0 < x < 1, -\ln(1 - x) - x > \frac{\mu h(1 - \nu(1 - \theta))}{1 - \mu\nu(1 - \theta)} \right\} \quad (52)$$

$$\alpha(a, b) = a \ln(a) - b \ln(b) + b - a. \quad (53)$$

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

The proof of the above theorem follows from (41) (theorem A.1) and ideas very similar to [23, Theorem IV.1]. We omit the details.

### B. Optimal decoder

In this section we briefly describe the optimal decoder and its performance assuming replica symmetry. More details can be found in [19].

Let the codebook be  $\mathcal{C}$ . The optimal decoder for PUPE is the one which computes, for  $c \in \mathcal{C}$ , the posteriors  $P_{c|Y^n}$  which is the probability, conditional on received vector  $Y^n$ , that  $c$  is the list of transmitted codewords. Then, it outputs the list of codewords corresponding to top  $K_a$  posteriors. Further, the system model is slightly modified in that each message is transmitted with probability  $p = K_a/M = 1/M_1$ . In the limiting case, assuming replica symmetry, the posteriors converge to the posterior  $\mathbb{P}[X \neq 0|Y]$  of a scalar channel  $Y = X + \sigma Z$  where  $Z \sim \mathcal{CN}(0, 1)$ ,  $X$  is  $\mathcal{CN}(0, 1)$  with probability  $p$  and 0 with probability  $1 - p$  and is independent of  $Z$ . The value of  $\sigma$  is given by (see [19, Theorem 8], but modified here for complex case)

$$\sigma^2 = \arg \min_{\tau > 0} \left\{ \frac{1}{\mu M_1} \log \tau M_1 + \log(e) \frac{1}{\tau M_1 P_{tot}} + I(X; X + \sqrt{\tau} Z) \right\}. \quad (54)$$

The PUPE converges to  $\mathbb{P}[\mathbb{P}[X \neq 0|Y] < T | X \neq 0]$  where  $T$  satisfies  $\mathbb{P}[\mathbb{P}[X \neq 0|Y] > T] = p$ . Hence, we can find the minimum  $P_{tot}$  such that this PUPE of the scalar channel is at most  $\epsilon$ , and this gives another achievability bound (assuming replica symmetry) on  $\mathcal{E}^*$ .

### C. Converse

We present a converse for  $\mathcal{E}^*$  based on Fano inequality and using the results from [30, 37]

**Theorem B.2.** *Let  $M = K_a M_1$  be the codebook size. Given  $\epsilon \leq 1 - \frac{K_a}{M}$  and  $\mu$  such that  $M_1 > 2$  then  $\mathcal{E}^*(M_1, \mu, \epsilon) > \mathcal{E}_{conv}$  where  $\mathcal{E}_{conv} = \max\{\mathcal{E}_{conv,1}, \mathcal{E}_{conv,2}\}$  satisfies the following two bounds*

1)

$$\mathcal{E}_{conv,1} = \inf \frac{P_{tot}}{\mu \log M_1} \quad (55)$$

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

$$\mu \theta \log M_1 - \epsilon \mu \log(M_1 - 1) - \mu h_2(\epsilon) \leq \log(1 + \alpha(1 - \theta, 1)P_{tot}), \forall \theta \in [0, 1] \quad (56)$$

and  $\alpha$  is defined in (53).

2)

$$\mathcal{E}_{conv,2} = \inf \frac{P_{tot}}{\mu \log M_1} \quad (57)$$

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

$$\epsilon \geq 1 - \mathbb{E} \left[ Q \left( Q^{-1} \left( \frac{1}{M_1} \right) - \sqrt{\frac{2P_{tot}}{\mu}} |H|^2 \right) \right] \quad (58)$$

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

*Proof:* The proof of (55), (56) is based of Fano inequality and genie argument. Let  $W = (W_1, \dots, W_{K_a})$  where  $W_i \stackrel{iid}{\sim} \text{Unif}[M]$  denote the transmitted messages of  $K_a$  users. Let  $\hat{W}$  be the decoded list of messages. Then  $\epsilon = P_e = \frac{1}{K_a} \sum_{j \in [K_a]} \mathbb{P} [W_j \notin \hat{W}]$ .

Suppose a genie  $G$  reveals to the decoder a set  $S_1 \subset [K_a]$  of transmitted messages  $W_{S_1} = \{W_i : i \in S_1\}$  along with corresponding fading coefficients  $H_{S_1} = \{H_i : i \in S_1\}$ . A converse bound in this case is a converse for the actual problem (when there is no Genie). Hence the equivalent channel at the receiver becomes

$$Y_G = \sum_{i \in S_2} H_i X_i + Z \quad (59)$$

where  $S_2 = [K_a] \setminus S_1$  with the decoder outputting a list  $L_G = L(Y_G, W_{S_1}, H_{S_1})$  of messages of size at most  $K_a$  and PUPE

$$P_e^G = \frac{1}{K_a} \sum_{j \in [K_a]} \mathbb{P} [W_j \notin L_G]$$

First note that the optimal decoder (for PUPE) outputs a list of size exactly  $K_a$  since otherwise PUPE can be strictly reduced by extending the list to size  $K_a$  by adding random messages. Further, it must contain  $W_{S_1}$  because if there is  $j \in S_1$  such that  $W_j \notin L_G$  then replacing one non-transmitted message in  $L_G$  by  $W_j$  strictly decreases PUPE. Let  $E_i = 1[W_i \notin L_G]$  and  $\epsilon_i^G = \mathbb{E}[E_i]$ . Note that  $\epsilon_i^G = 0$  for  $i \in S_1$ . Now standard Fano type arguments give, for  $i \in S_2$ ,

$$I(W_i; L_G) \geq \log M - h_2(\epsilon_i^G) - \epsilon_i^G \log(M - K_a) - (1 - \epsilon_i^G) \log K_a. \quad (60)$$

Since

$$I(W_{S_2}; L_G) \geq \sum_{i \in S_2} I(W_i; L_G),$$

we have

$$I(W_{S_2}; L_G) \geq |S_2| \log M - \sum_{i \in S_2} h_2(\epsilon_i^G) - \log \left( \frac{M}{K_a} - 1 \right) \sum_{i \in S_2} \epsilon_i^G - |S_2| \log K_a. \quad (61)$$

Further,

$$I(W_{S_2}; L_G) \leq n \mathbb{E} \left[ \log \left( 1 + \frac{P_{tot}}{K_a} \sum_{i \in S_2} |H_i|^2 \right) \right].$$

Let

$$P_e(S_2) = \frac{1}{|S_2|} \sum_{i \in S_2} \epsilon_i^G.$$

Then

$$\frac{|S_2|}{K_a} P_e(S_2) = P_e^G \leq P_e.$$

Hence we have

$$\frac{n}{K_a} \mathbb{E} \left[ \log \left( 1 + \frac{P_{tot}}{K_a} \sum_{i \in S_2} |H_i|^2 \right) \right] \geq \frac{|S_2|}{K_a} \log M - \frac{1}{K_a} \sum_{i \in S_2} h_2(\epsilon_i^G) - \log \left( \frac{M}{K_a} - 1 \right) \frac{|S_2|}{K_a} P_e(S_2) - \frac{|S_2|}{K_a} \log K_a \quad (62)$$

$$\geq \frac{|S_2|}{K_a} \log \frac{M}{K_a} - h_2(P_e^G) - \log \left( \frac{M}{K_a} - 1 \right) P_e^G \quad (63)$$

where the second inequality follows from Jensen's inequality and the fact that  $M_1 = \frac{M}{K_a} > 2$ . Since  $P_e^G \leq P_e \leq 1 - \frac{K_a}{M}$ ,  $h_2(P_e^G) + \log \left( \frac{M}{K_a} - 1 \right) P_e^G \leq h_2(P_e) + \log \left( \frac{M}{K_a} - 1 \right) P_e$ . The above equation holds for all  $S_2$ . Taking limit, with  $|S_2| = \theta K_a$  and using results on strong laws of order statistics [38] (see proof of [23, Theorem IV.6]) gives the first part of the theorem.

For the second part, we have the following converse for a single user AWGN MAC  $Y = X + Z$ ,  $X, Y \in \mathbb{R}^\infty$ ,  $Z_i \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ . Define an  $(E, M, \epsilon)$  code for this channel: codewords  $(c_1, \dots, c_M)$  with  $\|c_i\|^2 \leq E$  and a decoder such that probability of error is smaller than  $\epsilon$ . Then from [37] we have that any  $(E, M, \epsilon)$  code satisfies

$$\frac{1}{M} \geq Q\left(\sqrt{2E} + Q^{-1}(1 - \epsilon)\right). \quad (64)$$

Now, if the decoder were to output a list of size at most  $K_a$  in the above and the error is defined as the probability that the transmitted message is not in the output list, then from the proof of (64) in [37] and ideas of meta-converse for list decoding [30], it can be easily verified that the above equation modifies to

$$\frac{K_a}{M} \geq Q\left(\sqrt{2E} + Q^{-1}(1 - \epsilon)\right). \quad (65)$$

Hence using the ideas in proof of theorem IV.1 to reduce the problem to single user case with list decoding and translating (65) to quasi-static case as in the proof of [23, Theorem IV.6], the result in (58) follows. ■

Tighter converse bounds can be obtained if further assumptions are made on the codebook. For example, if we assume that each codebook consists of iid entries of the form  $\frac{C}{K_a}$  where  $C$  is sampled from a distribution with zero mean and finite variance, then we have the following converse bound from [36, Theorem 3] (see [36, Remark 3] as well).

**Theorem B.3.** *Let  $\mu = K_a/n < 1$  be the user density and  $M = K_a M_1$  be the codebook size such that  $M_1 > 2$ , and let the common codebook be generated such that each code symbol iid of the form  $\frac{C}{K_a}$  where  $C$  is of zero mean and variance  $P_{tot}$ . Then in order for the codebook to achieve PUBE  $\epsilon$  with high probability, the energy-per-bit  $\mathcal{E}$  should satisfy*

$$\mathcal{E} \geq \inf \frac{P_{tot}}{\mu \log M_1} \quad (66)$$

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

$$h_2\left(\frac{1}{M_1}\right) - \frac{1}{M_1} h_2(\epsilon) - \left(1 - \frac{1}{M_1}\right) h_2\left(\frac{\epsilon}{M_1 - 1}\right) \leq \left(\mathcal{V}\left(\frac{1}{\mu M_1}, P_{tot}\right) - \frac{1}{M_1} \mathcal{V}\left(\frac{1}{\mu}, P_{tot}\right)\right) \log e \quad (67)$$

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

$$\mathcal{V}(r, \gamma) = r \ln(1 + \gamma - \mathcal{F}(r, \gamma)) + \ln(1 + r\gamma - \mathcal{F}(r, \gamma)) - \frac{\mathcal{F}(r, \gamma)}{\gamma} \quad (68)$$

$$\mathcal{F}(r, \gamma) = \frac{1}{4} \left( \sqrt{\gamma(\sqrt{r} + 1)^2 + 1} - \sqrt{\gamma(\sqrt{r} - 1)^2 + 1} \right)^2 \quad (69)$$

## APPENDIX C PROOF OF THEOREM A.1

In this section, we present the proof of theorem A.1. We remark that (72) and (79) prove (33).

*Proof:* Let the common (complex) Gaussian codebook  $\mathcal{C}$  of size  $M$  and power  $P' < P$  be generated, that is, for each  $j \in [M]$ , generate  $c_j \stackrel{iid}{\sim} \mathcal{CN}(0, P' I_n)$ . Let  $W_j$  denote the random (in  $[M]$ ) the message of user  $j$ . The transmitted channel input is given by  $X_j = c_{W_j} 1\left\{\|c_{W_j}\|^2 \leq nP\right\}$ . Let  $K_1 \leq K_a$  be the number of messages in the received signal that are decoded. The decoder searches of all  $K_1$  sized subsets of  $[M]$ . The decoder output  $g_D(Y) \in \mathcal{C}$  is given by

$$\begin{aligned} \hat{C} &= \arg \max_{\substack{\mathcal{C} \subseteq \mathcal{C} \\ |C|=K_1}} \|P_{\{c:c \in C\}} Y\|^2 \\ g_D(Y) &= \left\{ f^{-1}(c) : c \in \hat{C} \right\} \end{aligned} \quad (70)$$

where  $f$  is the encoding function. The probability of error is given by

$$P_e = \frac{1}{K_2} \sum_{j=1}^{K_2} \mathbb{P}[W_j \notin g_D(Y), \text{ or } \exists i \neq j, W_j = W_i]. \quad (71)$$

Note that  $W_1, \dots, W_{K_2}$  are sampled independently with replacement from  $[M]$ . We perform a change of measure by sampling  $W_1, \dots, W_{K_2}$  from  $[M]$  *without* replacement, and also change the measure of transmitted message from

$$X_j = c_{W_j} 1 \left\{ \|c_{W_j}\|^2 \leq nP \right\}$$

to  $X_j = c_{W_j}^j$ . 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 be bounded by

$$p_0 = \frac{\binom{K_2}{2}}{M} + K_2 \mathbb{P} \left[ \frac{\chi_2(2n)}{2n} > \frac{P}{P'} \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where  $\chi_2(d)$  is the distribution of sum of squares of  $d$  iid standard normal random variables (the chi-square distribution). This follows from the same reasoning used in the main theorem in [1]. Henceforth we only consider the new measure. Now,  $P_e$  can be bounded as

$$P_e \leq \mathbb{E} \left[ \frac{1}{K_2} \sum_{j=1}^{K_2} 1[W_j \notin g(Y)] \right] + p_0 \leq \frac{K_2 - K_1}{K_2} + \frac{1}{K_2} \sum_{t=1}^{K_1} p_{1,t} K_{1,t} + p_0 \quad (72)$$

where  $K_{1,t}$  is given by (35) and  $p_{1,t} = \mathbb{P} \left[ \sum_{j=1}^{K_2} 1[W_j \notin g(Y)] = K_{1,t} \right]$ .

Let  $F_t = \left\{ \sum_{j=1}^{K_2} 1[W_j \notin g(Y)] = K_{1,t} \right\}$ . W.l.o.g, we will assume that the transmitted message list is  $S = [K_2]$  and hence the corresponding codewords are  $\{c_1, c_2, \dots, c_{K_2}\}$ . Let  $c_{[S_0]} \equiv \{c_i : i \in S_0\}$  and  $H_{[S_0]} \equiv \{H_i : i \in S_0\}$ , where  $S_0 \subset [K_2]$ . Further, let  $c_{[S_1][S_2]} = c_{[S_1 \cup S_2]}$ . Conditioning on  $c_{[K_2]}, H_{[K_2]}$  and  $Z$ , we have (73)

$$\begin{aligned} \mathbb{P} [F_t | c_{[K_2]}, H_{[K_2]}, Z] &\leq \mathbb{P} [\exists S_0 \subset [K_2] : |S| = K_{1,t}, \exists S_1 \subset [M] \setminus [K_2] : |S_1| = t : \\ &\quad \left\| P_{c_{[S_1][K_2] \setminus S_0]} Y \right\|^2 > \max_{\substack{S_2 \subset S_0 \\ |S_2|=t}} \left\| P_{c_{[S_2][K_2] \setminus S_0]} Y \right\|^2 \mid c_{[K_2]}, H_{[K_2]}, Z \right] \\ &\leq \mathbb{P} \left[ \bigcup_{\substack{S_0 \subset [K_2] \\ |S_0|=K_{1,t}}} \bigcup_{\substack{S_1 \subset [M] \setminus [K_2] \\ |S_1|=t}} F(S_0, S_2^*, S_1, t) \mid c_{[K_2]}, H_{[K_2]}, Z \right], \end{aligned} \quad (73)$$

where

$$F(S_0, S_2^*, S_1, t) = \left\{ \left\| P_{c_{[S_1][K_2] \setminus S_0]} Y \right\|^2 > \left\| P_{c_{[S_2^*][K_2] \setminus S_0]} Y \right\|^2 \right\},$$

and  $S_2^* \subset S_0$  is a possibly random (depending only on  $H_{[K_2]}$ ) subset of size  $t$ , to be chosen later. Next we will bound  $\mathbb{P} [F(S_0, S_2^*, S_1, t) | c_{[K_2]}, H_{[K_2]}, Z]$ .

For the sake of brevity, let  $A_0 = c_{[S_2^*][K_2] \setminus S_0}$ ,  $A_1 = c_{[K_2] \setminus S_0}$  and  $B_1 = c_{[S_1]}$ . We have the following claim which follows from [23, Claim 1].

**Claim 1** ([23]). *For any  $S_1 \subset [M] \setminus [K_2]$  with  $|S_1| = t$ , conditioned on  $c_{[K_2]}, H_{[K_2]}$  and  $Z$ , the law of  $\left\| P_{c_{[S_1][K_2] \setminus S_0]} Y \right\|^2$  is same as the law of  $\|P_{A_1} Y\|^2 + \|(I - P_{A_1})Y\|^2 \text{Beta}(t, n - K_1)$  where  $\text{Beta}(a, b)$  is a beta distributed random variable with parameters  $a$  and  $b$ .*

Therefore we have,

$$\mathbb{P} [F(S_0, S_2^*, S_1, t) | c_{[K_2]}, H_{[K_2]}, Z] = \mathbb{P} [\text{Beta}(n - K_1, t) < G_{S_0} | c_{[K_2]}, H_{[K_2]}, Z] = F_\beta(G_{S_0}; n - K_1, t) \quad (74)$$

where

$$G_{S_0} = \frac{\|Y\|^2 - \|P_{A_0} Y\|^2}{\|Y\|^2 - \|P_{A_1} Y\|^2}. \quad (75)$$

Since  $t \geq 1$ , we have  $F_\beta(G_{S_0}; n - K_1, t) \leq \binom{n'-1}{t-1} G_{S_0}^{n-K_1}$ , where  $n'$  is given by (40).

Let us denote  $\bigcup_{\substack{S_0 \subset [K_2] \\ |S|=K_{1,t}}}$  as  $\bigcup_{S_0, K_1}$ ; similarly for  $\sum$  and  $\bigcap$  for the ease of notation. Using the above claim, we get,

$$\mathbb{P} [F_t | c_{[K_2]}, H_{[K_2]}, Z] \leq \sum_{S_0, K_1} \binom{M - K_2}{t} \binom{n' - 1}{t - 1} G_{S_0}^{n - K_1}. \quad (76)$$

Therefore  $p_{1,t}$  can be bounded as

$$\begin{aligned} p_{1,t} &= \mathbb{P} [F_t] \\ &\leq \mathbb{E} \left[ \min \left\{ 1, \sum_{S_0, K_1} \binom{M - K_2}{t} \binom{n' - 1}{t - 1} G_{S_0}^{n - K_1} \right\} \right] \\ &= \mathbb{E} \left[ \min \left\{ 1, \sum_{S_0, K_1} e^{(n - K_1)(s_t + R_1)} G_{S_0}^{n - K_1} \right\} \right] \end{aligned} \quad (77)$$

where  $s_t$  and  $R_1$  are given by (38) and (39) respectively.

For  $\delta > 0$ , define  $V_{n,t}$  as in (36). Let  $E_1$  be the event

$$\begin{aligned} E_1 &= \bigcap_{S_0, K_1} \{-\ln G_{S_0} - s_t - R_1 > \delta\} \\ &= \bigcap_{S_0, K_1} \{G_{S_0} < V_{n,t}\}. \end{aligned} \quad (78)$$

Let  $p_{2,t} = \mathbb{P} \left[ \bigcup_{S_0, K_1} \{G_{S_0} > V_{n,t}\} \right]$ . Then

$$\begin{aligned} p_{1,t} &\leq \mathbb{E} \left[ \min \left\{ 1, \sum_{S_0, K_1} e^{(n - K_1)(s_t + R_1)} G_{S_0}^{n - K_1} \right\} (1[E_1] + 1[E_1^c]) \right] \\ &\leq \mathbb{E} \left[ \sum_{S_0, K_1} e^{-(n - K_1)\delta} \right] + p_{2,t} \\ &= \binom{K_2}{K_{1,t}} e^{-(n - K_1)\delta} + p_{2,t}. \end{aligned} \quad (79)$$

**Note:** This proves (33).

Let us bound  $p_{2,t}$ . Let  $\hat{Z} = Z + \sum_{i \in S_0 \setminus S_2^*} H_i c_i$ . From [23, Claim 2] we have

**Claim 2** ([23]).  $p_{2,t}$  is bounded as

$$\begin{aligned} p_{2,t} &= \mathbb{P} \left[ \bigcup_{S_0, K_1} \{G_{S_0} > V_{n,t}\} \right] \\ &\leq \mathbb{P} \left[ \bigcup_{S_0, K_1} \left\{ \left\| (1 - V_{n,t}) P_{A_1}^\perp \hat{Z} - V_{n,t} P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i \right\|^2 \geq V_{n,t} \left\| P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i \right\|^2 \right\} \right]. \end{aligned} \quad (80)$$

Let  $\chi'_2(\lambda, d)$  denote the non-central chi-squared distributed random variable with non-centrality  $\lambda$  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 from [23, Claim 3].

**Claim 3** ([23]). Conditional on  $H_{[K_2]}$  and  $A_0$ ,

$$\left\| P_{A_1}^\perp \left( \hat{Z} - \frac{V_{n,t}}{1 - V_{n,t}} \sum_{i \in S_2^*} H_i c_i \right) \right\|^2 \sim \left( 1 + P' \sum_{i \in S_0 \setminus S_2^*} |H_i|^2 \right) \frac{1}{2} \chi'_2(2F, 2n') \quad (81)$$

where

$$F = \frac{\left\| \frac{V_{n,t}}{1-V_{n,t}} P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i \right\|^2}{\left(1 + P' \sum_{i \in S_0 \setminus S_2^*} |H_i|^2\right)} \quad (82)$$

$$(83)$$

Hence its conditional expectation is

$$\mu = n' + F. \quad (84)$$

Now let

$$T = \frac{1}{2} \chi_2'(2F, 2n') - \mu \quad (85)$$

$$U = \frac{V_{n,t}}{(1 - V_{n,t})} \frac{\left\| P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i \right\|^2}{\left(1 + P' \sum_{i \in S_0 \setminus S_2^*} |H_i|^2\right)} - n' \quad (86)$$

$$U^1 = \frac{1}{1 - V_{n,t}} (V_{n,t} W_{S_0} - 1) \quad (87)$$

where  $W_{S_0} = \left(1 + \frac{\left\| P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i \right\|^2}{n' \left(1 + P' \sum_{i \in S_0 \setminus S_2^*} |H_i|^2\right)}\right)$ . Notice that  $U = n' U^1$  and  $F = \frac{V_{n,t}}{1 - V_{n,t}} n' (1 + U^1)$ .

Then we have (88).

$$\begin{aligned} \text{RHS of (80)} &= \mathbb{P} \left[ \bigcup_{S_0, K_1} \left\{ \left\| P_{A_1}^\perp \hat{Z} - \frac{V_{n,t}}{(1 - V_{n,t})} P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i \right\|^2 - \mu \geq U \right\} \right] \\ &= \mathbb{P} \left[ \bigcup_{S_0, K_1} \{T \geq U\} \right]. \end{aligned} \quad (88)$$

Now, let  $\delta_1 > 0$ , and  $E_2 = \cap_{S_0, K_1} \{U^1 > \delta_1\}$ . Taking expectations over  $E_1$  and its complement, we have

$$\begin{aligned} \mathbb{P} \left[ \bigcup_{S_0, K_1} \{T \geq U\} \right] &\leq \sum_{S_0, K_1} \mathbb{P} [T > U, U^1 > \delta_1] + \mathbb{P} [E_2^c] \\ &= \sum_{S_0, K_1} \mathbb{E} [\mathbb{P} [T > U | H_{[K_2]}, A_0] 1[U^1 > \delta_1]] + \mathbb{P} [E_2^c] \end{aligned} \quad (89)$$

which follows from the fact that  $\{U^1 > \delta_1\} \in \sigma(H_{[K_2]}, A_0)$ . To bound this term, we use the following concentration result from [39, Lemma 8.1].

**Lemma C.1** ([39]). *Let  $\chi = \chi_2'(\lambda, d)$  be a non-central chi-squared distributed variable with  $d$  degrees of freedom and non-centrality parameter  $\lambda$ . Then  $\forall x > 0$*

$$\begin{aligned} \mathbb{P} \left[ \chi - (d + \lambda) \geq 2\sqrt{(d + 2\lambda)x} + 2x \right] &\leq e^{-x} \\ \mathbb{P} \left[ \chi - (d + \lambda) \leq -2\sqrt{(d + 2\lambda)x} \right] &\leq e^{-x} \end{aligned} \quad (90)$$

Hence, for  $x > 0$ , we have

$$\mathbb{P} [\chi - (d + \lambda) \geq x] \leq e^{-\frac{1}{2}(x + d + 2\lambda - \sqrt{d + 2\lambda} \sqrt{2x + d + 2\lambda})}. \quad (91)$$

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

$$\mathbb{P} [\chi \leq x] \leq e^{-\frac{1}{4} \frac{(d + \lambda - x)^2}{d + 2\lambda}}. \quad (92)$$

Observe that, in (91), the exponent is always negative for  $x > 0$  and finite  $\lambda$  due to AM-GM inequality. When  $\lambda = 0$ , we can get a better bound for the lower tail in (92) by using [19, Lemma 25].

**Lemma C.2** ([19]). Let  $\chi = \chi_2(d)$  be a chi-squared distributed variable with  $d$  degrees of freedom. Then  $\forall x > 1$

$$\mathbb{P} \left[ \chi \leq \frac{d}{x} \right] \leq e^{-\frac{d}{2}(\ln x + \frac{1}{x} - 1)} \quad (93)$$

Therefore, from (80), (88), (89) and (91), we have

$$p_{2,t} \leq \sum_{S_0, K_1} \mathbb{E} \left[ e^{-n' f_n(U^1)} 1[U^1 > \delta_1] \right] + \mathbb{P} \left[ \bigcup_{S_0, K_1} \{U^1 \leq \delta_1\} \right] \quad (94)$$

where  $f_n$  is given by (42).

Next, from [23, Claim 4] we have that for  $0 < V_{n,t} < 1$  and  $x > 0$ ,  $f_n(x)$  is a monotonically increasing function of  $x$ . From this, we obtain

$$p_{2,t} \leq \sum_{S_0, K_1} e^{-n' f_n(\delta_1)} + p_{3,t} \quad (95)$$

where  $p_{3,t} = \mathbb{P}[E_2^c]$ .

Note that

$$p_{3,t} = \mathbb{P}[E_2^c] = \mathbb{P} \left[ \bigcup_{S_0, K_1} \{V_{n,t} W_{S_0} - 1 \leq \delta_1(1 - V_{n,t})\} \right]. \quad (96)$$

Conditional on  $H_{[K_2]}$ ,  $\left\| P_{A_1}^\perp \sum_{i \in S_2^*} H_i c_i \right\|^2 \sim \frac{1}{2} P' \sum_{i \in S_2^*} |H_i|^2 \chi_2^{S_2^*}(2n')$ , 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  $E_4 = \bigcap_{S_0, K_1} \left\{ \frac{\chi_2^{S_2^*}(2n')}{2n'} > 1 - \delta_2 \right\}$ . Using (93), we can bound  $p_{3,t}$  as

$$p_{3,t} \leq \sum_t \binom{K_2}{K_{1,t}} e^{-n'(-\ln(1-\delta_2)-\delta_2)} + p_{4,t} \quad (97)$$

where

$$p_{4,t} = \mathbb{P}[E_4^c] = \mathbb{P} \left[ \bigcup_{S_0, K_1} \left\{ V_{n,t} \left( 1 + \frac{P' \sum_{i \in S_2^*} |H_i|^2 (1 - \delta_2)}{(1 + P' \sum_{i \in S_0 \setminus S_2^*} |H_i|^2)} \right) \leq 1 + \delta_1(1 - V_{n,t}) \right\} \right]. \quad (98)$$

We make an important observation here. The union bound over  $S_0$  is the minimum over  $S_0$ , and it can be seen that optimum  $S_0$  i.e., the minimizer should be contiguous amongst the indices arranged according the decreasing order of fading powers. Then the best upper bound is got by choosing  $S_2^*$  to be correspond to the top  $t$  fading powers in  $S_0$ . Hence, we get

$$p_4 = \mathbb{P} \left[ \min_{1 \leq i \leq K_1 - t + 1} \frac{P' \sum_{j=i}^{i+t-1} |H_{(j)}|^2}{1 + P' \sum_{j=i+t}^{K_{1,t}-1+i} |H_{(j)}|^2} \leq \frac{(1 + \delta_1(1 - V_{n,t})) V_{n,t}^{-1} - 1}{1 - \delta_2} \right] \quad (99)$$

Finally, combining (72), (79), (95), (97) and (99), and optimizing over  $\delta$ ,  $\delta_1$  and  $\delta_2$ , we are done. ■

## APPENDIX D RESULTS FOR BLIND SLOT DECODING

Here we present the numerical results for blind slot decoding. Let us fix the following parameters:

- [400, 100] LDPC code for 4-user case, obtained by PEXIT method in [9];
- 25 outer iterations, 50 inner (LDPC) iterations;
- $T = 4$ , which means that we can decode at most 4 users in a slot;

We present the curves for 2, 3 and 4 users, recall, that  $T = 4$  for all the cases. We compare these curves with the following “ideal” curves

- fading channel coefficients are unknown, number of users is known (i.e.  $T$  is selected to be equal to the actual number of users);
- fading channel coefficients are known, number of users is known (full CSI).

Frame error rate performance for listed above scenarios are presented on Fig. 6, Fig. 7 and Fig. 8 for  $K = 2, 3, 4$  respectively. We see, that the performance curves for our coding scheme coincide with “ideal” curves and achievability bound and very close (the loss is less, than 2 dB) to the converse bound. So we conclude, that LDPC-based scheme is good for resolving collisions of small order.

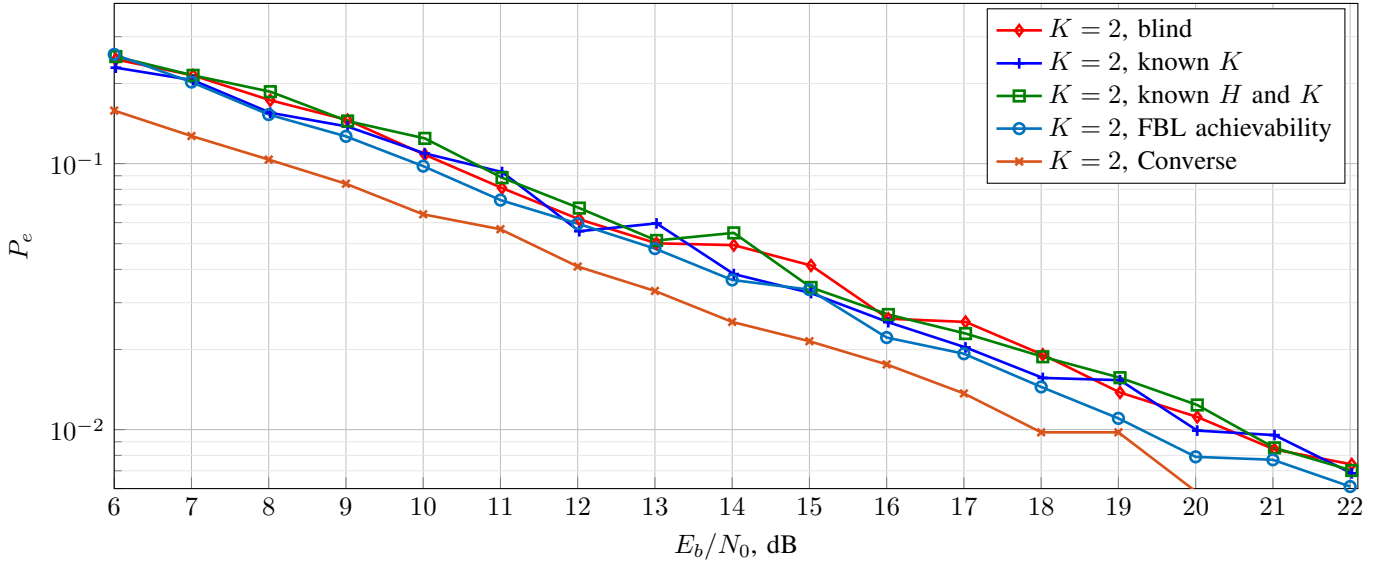


Fig. 6: Simulation results for  $K = 2$  users

## APPENDIX E SINGLE-COMPONENT GM PERFORMANCE

TABLE I: GM parameters for single and multiple component GM model

Parameter	Multi-component GM	Single-component GM
Gaussian mixture merge distance ( $d_{min}$ )	1	—
Gaussian mixture maximum component count ( $\nu$ )	500	1
GM sample count to evaluate (16)	20	20
Maximum cumulative weight to drop at prune (The components with the least weights are dropped before prune)	$10^{-3}$	—

Let us consider how the GM configuration affects the overall decoding performance. The algorithm complexity highly depends on the maximum number of components  $\nu$  allowed in the GM. Merge and prune procedures keep the maximum component count under some threshold. To address this issue, we have evaluated the frame error rate performance for  $K = 4$  users with the decoder having different settings. In the first setup, we have utilized multiple-component GM with merge and prune procedures (as before). The second setup assumes single-component GM with the merge procedure being disabled. Let's again consider the same [400, 100] LDPC code as in the previous Appendix D.



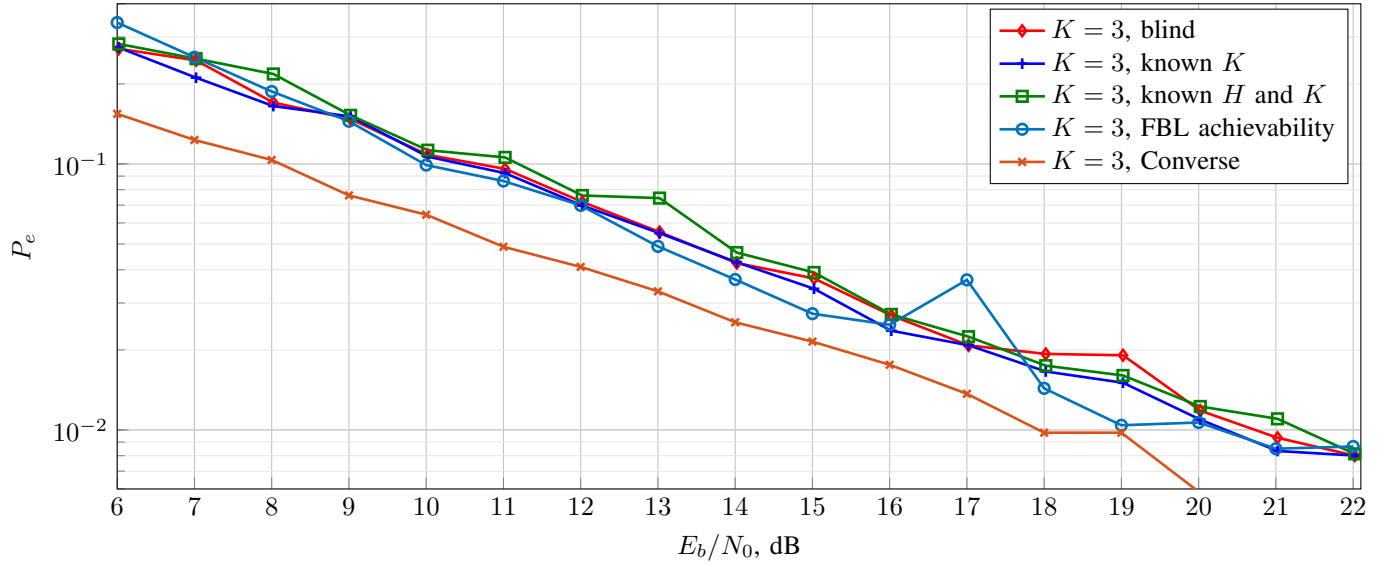


Fig. 7: Simulation results for  $K = 3$  users

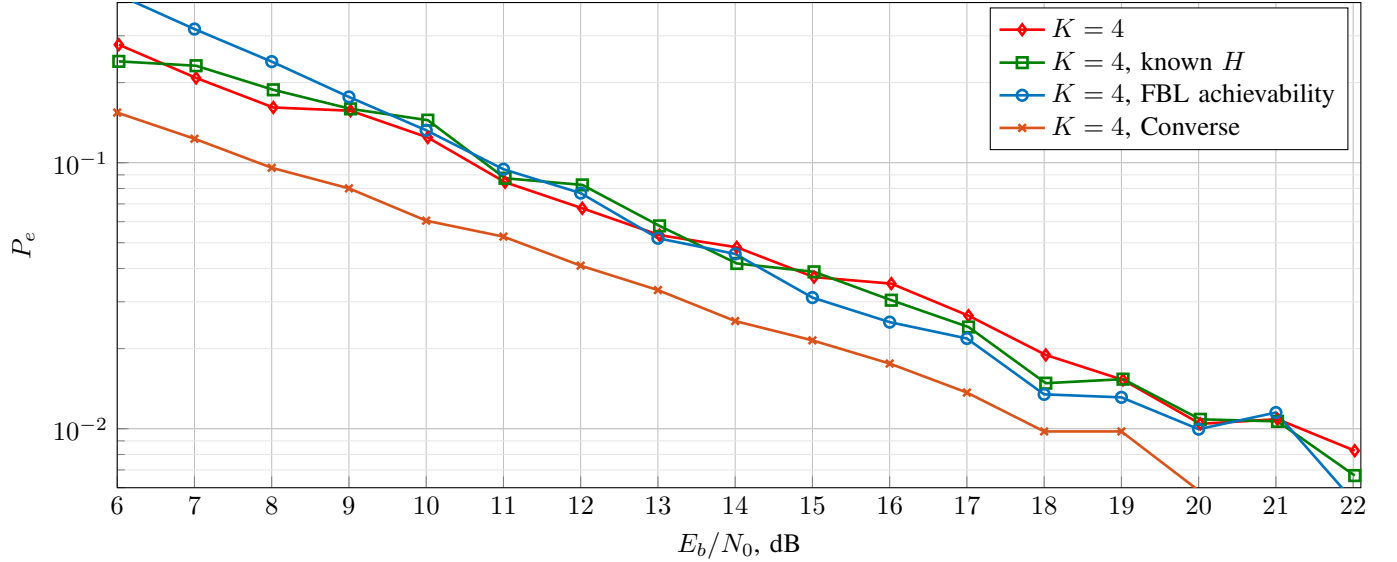


Fig. 8: Simulation results for  $K = 4$  users

The second setup should be explained in more details. Recall to the four message types described in section V. Each GM at every message passing step consists of a single component with the highest probability and the sampling (required to evaluate (16)) is performed from a single Gaussian distribution. As soon as the only GM component retains after every message type passing, there is no need to perform the merge procedure. It is worth to note that the merge procedure can only increase the covariance of the components to retain in the merge list. The most important change in the decoding algorithm (see section V) with single-component GM is equation (14). The most probable symbol is considered in this case (because the second alternative for BPSK constellation point will be immediately dropped by prune procedure under the limit  $\nu = 1$ ). Detailed difference in the GM parameters is shown in Table I.

The frame error rate performance is shown on Figure 9. Let us explain all the curves in the figure.

- Red curve corresponds to our most complex decoder from Appendix D, which utilizes Gaussian mixtures with a large number of components ( $\nu = 500$ , see Table I). The decoder performs merge and prune operations to guarantee that the number of components is less or equal than  $\nu$ . In this case each message is a pair of vectors  $(\bar{\mu}, \bar{\sigma})$  – means and variances, each vector is of length  $\nu$ . Real and imaginary components of fading coefficients estimates were represented

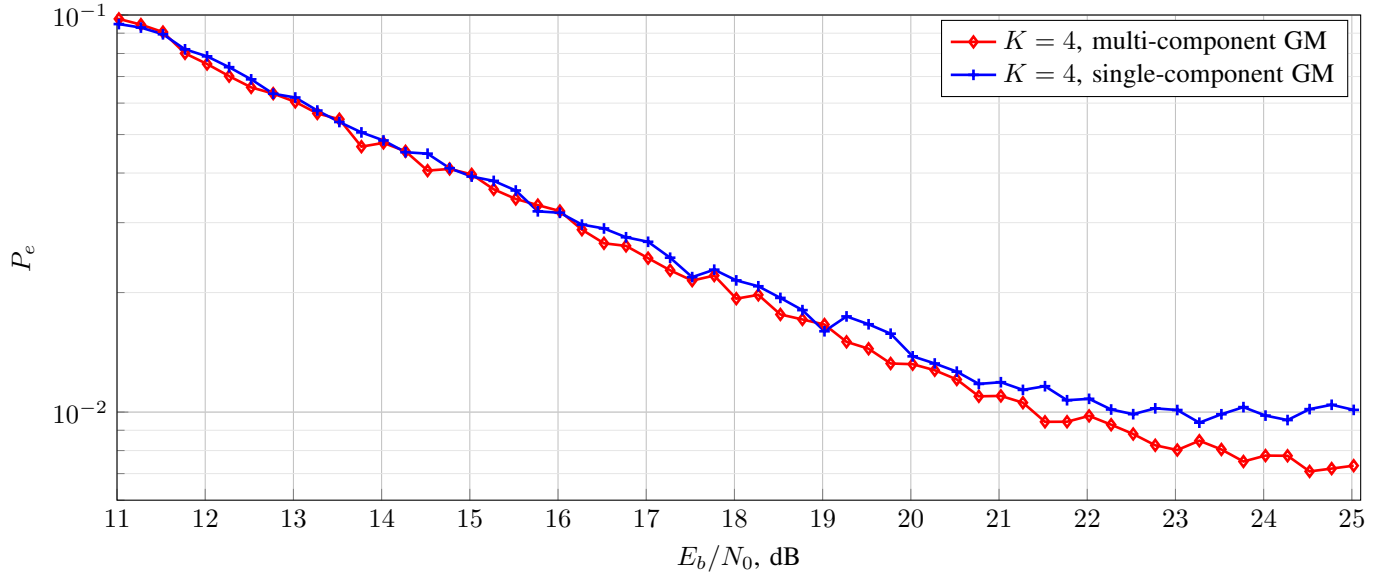


Fig. 9: Simulation results for  $K = 4$  users. Single component GM and multiple-component GM model (including component merge and prune model) frame error rate performance

by different mean and covariance vectors.

- Blue curve corresponds to the case when  $\nu = 1$ . We still perform prune operation but do not perform merge operation. At each step, the most probable component is chosen. So, in this case, the message is a pair of scalar values  $(\mu, \sigma)$  (again, real and imaginary parts are considered separately). The decoder has a surprisingly good performance.

One can see that the GM configuration affects the performance only at higher  $E_b/N_0$ . We see that the simpler the decoder the higher the error floor. For the blue and red curves we decided to perform simulation in  $E_b/N_0$  range  $[20, 25]$  dB to verify if error floor of the blue curve is higher.

An important moment is that all the decoders do several independent decoding attempts as described above. As described in section V-D, multiple attempts are needed to guarantee that decoder will not fall in to local maximum of (13). Otherwise, the performance is bad. This can be the explanation of the fact that single-component GM works fine.