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 MAC in [2], in the practically more relevant case of Rayleigh fading, when channel gains are unknown to the decoder. We propose a random coding achievability bound, which we analyze both non-asymptotically and asymptotically. As a candidate practical solution, we propose an explicit iterative coding scheme. The performance of such a solution is surprisingly close to the finite blocklength bounds. Our main findings are twofold. First, just like in the AWGN MAC, we see that jointly decoding a large number of users leads to a surprising phase transition effect, where, at spectral efficiencies below a critical threshold, a perfect multi-user interference cancellation is possible. Second, while the presence of Rayleigh fading significantly increases the minimal required energy-per-bit, the inherent randomization introduced by the channel makes it much easier to attain the optimal performance via iterative schemes. We hope that a principled definition of the random-access model, together with their information-theoretic analysis, will open the road towards unified benchmarking and performance comparison of various random-access solutions for the 5G/6G.

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 carried out at Skoltech and supported by the Russian Science Foundation (project no. 18-19-00673).

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

I. Introduction

Presently, wireless networks are starting to see a new type of communication traffic, in which hundreds of thousands of devices are serviced by a single base station, each communicating very small and infrequent data payloads. This scenario is known under the name of massive machine-type communications (mMTC). 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 it is done in today's systems). In this case, devices transmit independently from each other as well as without any centralized scheduling mechanism. In the literature, this problem is also known as *uncoordinated multiple access* or *random access*. The most popular techniques for uncoordinated multiple access include slotted ALOHA (SA) protocol [3] and its modifications with contention resolution methods [4]–[7]. The asynchronous case was considered in [8]–[10]. The main difference in our problem is in a huge number of devices which lead to a large collision probability. This requires new methods to decode colliding transmissions. In this work, we aim at understanding the fundamental tradeoffs of these dense random access systems and provide coding solutions that are close to achieving these fundamental limits. As energy efficiency is of critical importance for the mMTC scenario, we measure performance in terms of minimal energy-per-bit required to achieve the target per-user probability of error (PUPE). 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 equals to n complex-valued channel uses.

An information-theoretic formulation of this problem was done in [2] where the author considered an additive white Gaussian noise (AWGN) random access channel (RAC) model. In this formulation, the random access means the following: each of K_a active users encodes his k-bit message into an n-symbol codeword. The receiver observes the superposition of K_a codewords corrupted by the AWGN. There are many challenges in this model: finite blocklength (FBL) effects due to small payload size, a 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, 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* [11]–[13]. The unsourced random access is a type of uncoordinated random access with no need to identify the sender. 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 is 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 SA was considered in [14]. T-fold SA is a modification of the standard SA protocol [3], in which collisions of order up to T can be decoded in a slot. So, SA corresponds to $T = 1$. The idea of T-fold SA itself is not new as the idea of employing multi-packet receivers to resolve small order collisions has reappeared periodically [15] and more recently [6, Appendix A]. The gap between this low-complexity scheme in [14] and the FBL bound [2] was reduced in [11] by employing a successive interference cancellation scheme on top of an interleaved LDPC code. Achievability bounds for successive interference cancellation scheme (also known as irregular repetition SA [6]) were further improved in [16], where density evolution method [6] and a finite length random coding bound for the Gaussian MAC [2] were combined. In [17] the LDPC portion of [11] was improved by optimizing the protograph of LDPC code for Gaussian MAC using generalized PEXIT charts. Further improvements were obtained in [12] by developing a compressive sensing based algorithm. In [18] the idea of sparsifying collisions, inherent in T-fold SA, 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 [19], 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 [19] authors observed a perfect multi-user cancellation effect, shown to exist also for the fundamental limit in [20]. 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. [21].

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 [22]–[24] 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 in bits. However, already a moderate value of $k = 100$ bits precludes the straightforward usage of activity detection protocols.) In [25], scaling laws were derived for activity detection in a massive MIMO scenario. This and the ideas from [12] have been used to develop a low-complexity coding scheme in [13]. We also note here that our problem can be understood as a sparse support recovery in the compressed sensing literature [26]–[29]. Theoretical investigations in that literature predominantly consider iid Gaussian codebooks. In particular, in [27], 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 [2], [20], [30]–[32] where the number of active users scales linearly in blocklength.

The structure and main contributions of this paper are as follows. In Section II we formally define the problem of unsourced frame synchronized single antenna quasi-static Rayleigh fading RAC under per-user error. We assume that the channel realizations are not known to the receiver or the transmitters. A T-fold SA access method from [14] is reviewed in Section III. There are two ways we apply T-fold SA 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 [33]–[35] and a belief propagation decoder on a joint Tanner graph. In SectionVI 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.

II. System Model

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 onto the subspace *spanned* by S as P_S and its orthogonal complement as $P_S^{\perp}$. For $0 \leq p \leq 1$, let $h_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ and $h(p) = -p\ln(p) - (1-p)\ln(1-p)$, with $0\ln(0)$ defined to be 0. We denote by $\mathcal{N}(0,1)$ and $\mathcal{CN}(0,1)$ the standard normal and the standard circularly symmetric complex normal distributions, respectively. $\mathbb{P}$ and $\mathbb{E}$ denote probability measure and expectation operator respectively. For $n \in \mathbb{N}$, let $[n] = \{1, 2, ..., n\}$. Lastly, $\|\cdot\|$ represents the standard euclidean norm.

We follow the definition of a code from [2]. Fix an integer $K_a \geq 1$ – the number of active users. Let $\{P_{Y^n|X^n} = P_{Y^n|X_1^n, X_2^n, ..., X_{K_a}^n} : \times_{i=1}^{K_a} \mathcal{X}_i^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$ be a multiple access channel (MAC), which is also permutation invariant: for any permutation π on $[K_a]$, the distribution $P_{Y^n|X_1^n,...,X_{K_a}^n}(\cdot|x_1^n, ..., x_{K_a}^n)$ coincides with $P_{Y^n|X_1^n,...,X_{K_a}^n}(\cdot|x_{\pi(1)}^n, ..., x_{\pi(K_a)}^n)$. We also call this a random access channel (RAC).

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

$$P_e = \frac{1}{K_a}\sum_{j=1}^{K_a} \mathbb{P}\left[E_j\right] \leq \epsilon \tag{1}$$

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:

$$Y^n = \sum_{i=1}^{K_a} X_i^n \cdot \operatorname{diag}(H_i^n) + Z^n \tag{2}$$

where $X_i^n \in \mathbb{C}^n$ is the i-th transmitted codeword, $Z^n \sim \mathcal{CN}(0, I_n)$ is an additive white Gaussian noise (AWGN) and $\{H_i^n\}$ are the fading coefficients which are independent of $\{X_i^n\}$ and Z^n. We emphasize that our channel model is an approximation to a slow fading channel and assume the channel to be constant for

$n_1 < n$ channel uses (so-called quasi-static property or channel coherence time)[1]. Thus, in what follows we consider a frame of n channel uses divided into $L \in \mathbb{N}$ slots of length n_1, such that $n = Ln_1$. Finally, we require each codeword produced by the encoder f to satisfy a maximum power constraint:

$$\|f(w)\|^2 \leq nP\,, \qquad \forall w \in [M]\,. \tag{3}$$

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 Slotted Aloha

In this section, we discuss our main achievability bound based on T-fold SA protocol [14]. Let $T, n_1 \in \mathbb{N}$ such that $T < K_a$ and $n_1 < n$. The frame of length n complex-valued channel uses (either in time or in frequency domain) is partitioned into $L = n/n_1 \in \mathbb{N}$ subframes of length n_1 complex-valued channel uses. 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

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

and $H_i \overset{iid}{\sim} \mathcal{CN}(0,1)$, $i = 1, \dots, r$.

The idea of T-fold SA is to resolve collisions of order up to T. In other words if the number of users r transmitting in a given slot is at most T, then, with good reliability the decoder can estimate all the messages. For $T = 1$ this corresponds to the usual "collision model" prevalent in the analysis of the SA. Thus, T-fold SA is a compromise between conventional SA and joint decoding of all active users, which transmit simultaneously with the use of a common codebook. We want to emphasize that T-fold SA is theoretically beneficial as it utilizes multi-user decoding and still has low complexity when the value of T is small enough.

[1] In the numerical experiments we vary n_1 in a small range as we may consider the channel to be constant for some set of n_1.

Let $B(N,\rho) = \{X \in \mathbb{C}^N : \|X\| \le \rho\}$ be an $\mathbb{C}^N$-ball of radius ρ. For a given common codebook $\mathcal{C} \subset B(n_1, \sqrt{n_1 LP})$ of size $|\mathcal{C}| = M$ we let $P_{e,\text{genie}}(\mathcal{C}, r)$ denote the following quantity:

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

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 SA access scheme then attains the overall PUPE for all of K_a users bounded by

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

To obtain this estimate, we first bound the probability that the i-th user's message is in collision:

$$\mathbb{P}\left[\exists j \ne i : W_j = W_i\right] \le \frac{K_a - 1}{M}.$$

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. We will use the genie bound for our random-coding constructions and upper bound $P_{e,\text{genie}}$ via (26) in appendix A.

Remark 1. *Note that the genie assumption prevents the above from being a true achievability bound. Consequently, our genie-based bound strictly speaking is only an optimistic estimate of the performance achievable within a T-fold SA 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

$$P_e(\mathcal{C}, r) = \frac{1}{r}\sum_{i=1}^{r} \mathbb{P}\left[W_i \notin \mathcal{L}\right] , \quad Q_e(\mathcal{C}, r) = \mathbb{P}\left[|\mathcal{L}| > r\right] . \tag{5}$$

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

$$\epsilon_T(\mathcal{C}) \triangleq 1 - \sum_{r=1}^{T}(1 - P_e(\mathcal{C}, r))\binom{K_a - 1}{r-1}\left(\frac{1}{L}\right)^{r-1}\left(1 - \frac{1}{L}\right)^{K_a - r} + \frac{K_a - 1}{M} + q , \tag{6}$$

where $q = L\sum_{r=0}^{K_a}\binom{K_a}{r}L^{-r}(1-\frac{1}{L})^{K_a-r}Q_e(\mathcal{C},r)$ is an upper bound on $\mathbb{P}\left[\cup_{j=1}^{L}F_j\right]$, 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 $r \geq T$.

IV. Converse bound

In this section we describe a simple converse bound based on results from [36] and the meta-converse from [37]. We omit the proof which is available in [38].

Theorem IV.1. *Let*

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

$$S_n = n\log(1+PG) + \sum_{i=1}^{n}\left(1 - \frac{|\sqrt{PG}Z_i - 1|^2}{1+PG}\right) \tag{8}$$

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

$$\log(M) \leq \log(K_a) + \log\frac{1}{\mathbb{P}\left[L_n \geq n\gamma_n\right]} \tag{9}$$

where γ_n is the solution of

$$\mathbb{P}\left[S_n \leq n\gamma_n\right] = \epsilon. \tag{10}$$

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 $[n_1, k]$ LDPC codebook and replace each 0 with $+\sqrt{P}$ and each 1 with $-\sqrt{P}$.

Recall that T is a design parameter of our algorithm, which means the maximal collision order we are going to resolve. Let us start with the case when the number of users transmitting in a slot $r = T$ (in what follows we will show how to generalize it for the case of unknown r) and show the bit-wise MAP decoding rule for the j-th bit of the i-th user below (in what follows we omit the superscript n_1)

$$\hat{X}_{i,j} = \arg\max_{X_{i,j}\in\pm\sqrt{P}} \mathbb{E}\left[\sum_{\sim X_{i,j}} p_{Y|X}\left(Y \mid \sum_{l=1}^{T} H_l X_l\right)\prod_{l=1}^{T}\mathbb{1}_{X_l\in\mathcal{C}}\right], \tag{11}$$

where the expectation is taken over $H_1, H_2, \ldots, H_T$. Following [35], 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, [39]), which is shown in Fig. 1 for the case $T = 2$.

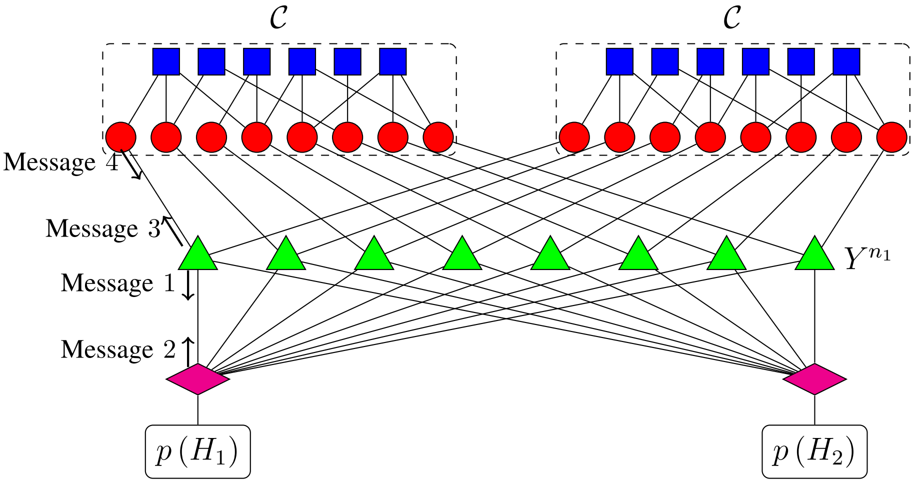

Fig. 1. Iterative joint decoding algorithm (alternating BP-decoder), factor graph for $T = 2$.

There are four types of nodes in the graph. Users' 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. In what follows, we refer to the iterations of our algorithm as outer iterations. By inner iterations, we mean iterations that are used for LDPC code decoding (see Algorithm 1). Within the outer iteration, the users' codewords are decoded sequentially. Let us consider the decoding of the first user. 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. Messages are 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 $y = Y_1$. By $x_i = X_{i,1} \in \{+\sqrt{P}, -\sqrt{P}\}$, $i = 1, \ldots, T$, 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

$$R_1^{(1)}(h_1) \propto \mathbb{E}\left[\sum_{x_1, x_2, \ldots x_T} p(y|\sum_{j=1}^{T} H_j x_j) \prod_{j=2}^{T} \Pr(x_j)\right], \tag{12}$$

where the expectations are taken over $H_2, \ldots, H_T$. Such updates are calculated at every functional node and denoted by $R_1^{(i)}$, $i = 1, \ldots, n_1$.

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

$$Q_1^{(1)}(h_1) = \prod_{i=1}^{n} R_1^{(i)}(h_1) p(h_1), \tag{13}$$

In a conventional message passing algorithm, the outgoing message is calculated based on messages which come through all the edges except the considered one. In our algorithm we use a randomized version of this step and calculate the product of a few randomly selected incoming messages (in our simulations we used 50 out $n_1 = 400$). Further numerical experiments show this approach to reduce the decoding complexity without any performance loss.

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.

$$L(x_1) = \log \frac{\mathbb{E}\left[\sum\limits_{x_1 = +\sqrt{P}, x_2, \ldots x_T} p(y|\sum\limits_{j=1}^{T} H_j x_j) \prod\limits_{j=2}^{T} \Pr(x_j)\right]}{\mathbb{E}\left[\sum\limits_{x_1 = -\sqrt{P}, x_2, \ldots x_T} p(y|\sum\limits_{j=1}^{T} H_j x_j) \prod\limits_{j=2}^{T} \Pr(x_j)\right]}, \tag{14}$$

where the expectations are taken over $H_1, H_2, \ldots, H_T$ 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, [35]) to decode an LDPC code.

Algorithm 1 Iterative decoding algorithm (alternating BP-decoder)

```
Input: $Y^{n_1}$, $T$    ▷ Received sequence and the value $T$
Output: Successfully decoded codewords list of size at most $T$.
initialize the LLR values of variable nodes for each user code with zero values assuming equal probability for $\sqrt{P}$ and $-\sqrt{P}$ values
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.
for $i_O = 1, \dots, I_O$ do    ▷ perform $I_O$ outer iterations
    for $u = 1, \dots, T$ do    ▷ decode users sequentially
        Propagate message type 1, eq. (12)    ▷ from functional nodes to fading nodes
        Propagate message type 2, eq. (13)    ▷ from fading nodes to functional nodes
        Sample fading coefficients for expectation estimation at (14) from the fading coefficients pdfs
        Propagate message type 3 using sampled fading coefficients, eq. (14)    ▷ from functional nodes to LDPC codes
        Propagate message type 4    ▷ run $I_I$ inner iterations of BP decoder for $u$-th user LDPC code.
    end for
    Update the list of successfully decoded codewords.
end for
return Subset of unique codewords among successfully decoded    ▷ If $r < T$, there can be duplicated codewords in the list.
```

The Algorithm 1 summarizes the above described procedure. We use the Gaussian mixtures (GM) to construct the practical implementation. Let us present the fading coefficient estimate pdf in the form of GM, i.e.

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

where ν is the number of components. This parameter controls the trade-off between the accuracy and complexity of the decoding algorithm. In our simulations, we used $\nu = 20$. The larger T is, the more GM components are required. Thus, one should choose the larger value of ν.

The message type 1 propagation (12) deals with the following subtraction (again consider the first functional node). Note that we consider real and imaginary components of H estimate pdfs as separate GMs.

$$H_1 x_1 = y - \left(\sum_{j=2}^{T} H_j x_j + z \right), \tag{15}$$

where j is the user index. Given H_j to be a GM, the $H_j x_j$ is also a GM with the number of components being doubled. This corresponds to two possible values of $x_j = \pm\sqrt{P}$ with the probabilities taken from LLR (note, that the same procedure also holds for extracting the H_j from $H_j x_j$). Thus, the RHS of (15) is a sum of random variables having GM pdfs. Hence, the resulting pdf is the convolution of GMs – also a GM. As soon as the result of convolution or a product of two GMs with ν_1 and ν_2 components is a GM with $\nu_1 \cdot \nu_2$ components, the fading coefficient estimates have the same form from iteration to iteration. Consider k-th component of the first GM and l-th component of the second GM. The convolution results in the GM with the components mean $\mu_{k,l} = \mu_k + \mu_l$ and the variance $\sigma^2_{k,l} = \sigma^2_k + \sigma^2_l$ indexed by the k, l pair. The GM product (message type 2) results in the $\mu_{k,l} = \frac{\mu_k}{\sigma^2_l} + \frac{\mu_k}{\sigma^2_l}$, $\quad \frac{1}{\sigma^2_{k,l}} = \frac{1}{\sigma^2_k} + \frac{1}{\sigma^2_l}$. Note that the procedures described above significantly increase the number of GM components. We use the components prune and merge procedures. The first procedure skips the components with the negligible weight ω_l, while the second one merges multiple components with small the distance (for example, the KL distance) smaller than some threshold into a single component. The next two steps in the user decoding procedure are sampling from GM and functional nodes decoding procedure (see eq. (14)).

B. 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 (users' LLRs and fading coefficients), one can expect this algorithm to converge to some local maximum of (11). 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 (so e.g. for $r = 3$ and $T = 4$ the decoder may return 4 codewords but only 3 of them are different). 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. We have limited the maximum number of decoding attempts to be equal to 10 in our numerical experiments.

The approach presented in this paper is similar to the approach from [40]. 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 $\epsilon = 0.1$[2] as a function of K_a for the channel (2). Fig. 2 shows plots of various schemes. The parameters used for evaluation are frame length $n = 30000$ and message size $k = 100$ bits. Next we describe how each of these curves was obtained.

For T-fold SA using FBL bound, we use the bound for p_t given in (26). 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 (6) is less than 0.1. Since directly optimizing the bound is not easy, we approximate PUPE for the fading channel as [41]

$$P_e(M, n_1, r, LP) \approx \mathbb{E}\left[\mathcal{Q}\left(\frac{n_1 C_{AWGN}(LP\sum_{i=1}^{r}|H_i|^2) - \log_2 M}{\sqrt{n_1 V_{AWGN}(LP\sum_{i=1}^{r}|H_i|^2)}}\right)\right] \tag{16}$$

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 (16) in (6). 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 (6) where $P_e(M, n_1, r, LP)$ is computed using brute-force

[2] We note that such value of ϵ is a regime of interest for LP-WANs such as LoRaWAN and Weightless.

Monte-Carlo simulation of (26) 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 $\max_{\substack{S_0 \subset [r] \\ |S_0|=t}} G(Y, S_0, c_{S_0}, t)$ (given in (27)) for each $t \leq r$. Then this smooth approximation is used to optimize over δ in (6).

For T-fold SA 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 PUPE $\leq \epsilon = 0.1$) for each value of K_a. Iterative decoder used the multiple component Gaussian mixture model. Note again, that in LDPC-based scheme we perform honest blind slot decoding (without assuming the knowledge of user count in a slot). 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 that we have $Q_e(\mathcal{C}, r) \approx 0$ (see (5)) 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.

It can be seen from Fig. 2 that the performance of T-fold SA for iterative decoding scheme is very close to that of T-fold SA 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 SA 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.

$$\epsilon \approx \mathbb{E}\left[\mathcal{Q}\left(\frac{nC_{AWGN}\left(\frac{P|H_1|^2}{1+P\sum_{i=2}^{T}|H_i|^2}\right) - k}{\sqrt{nV_{AWGN}\left(\frac{P|H_1|^2}{1+P\sum_{i=2}^{T}|H_i|^2}\right)}}\right)\right] \tag{17}$$

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 (17).

Also plotted for reference is the Shamai-Bettesh capacity bound from [42]. It is an asymptotic bound

($n \to \infty$) for the probability of error per-user in the case of symmetric rate and large K_a. The idea is the following. The joint decoder knows the realization of fading coefficients and users are ranked according to 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 the smallest 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 SA performance is off from the capacity bound of [42] by just 3 dB. Note that Shamai-Bettesh bound is only an achievable bound (i.e. not guaranteed to be tight) for the capacity under PUPE. It does not apply to our setting for two reasons: 1) it assumes different codebooks for different users, 2) it assumes asymptotically large blocklength. Note also that the asymptotics considered in Shamai-Bettesh is as follows: first n is taken to ∞ (under fixed K_a) and second the K_a is taken to infinity too. From our studies of the non-fading AWGN [20] we are convinced the correct asymptotic is to take K_a and n both tending to infinity at a given ratio – see Section VII below.

The "Optimal decoder (replica method)" curve is computed by non-rigorously estimating performance of the optimal decoder applied to a random Gaussian code under the asymptotics $K_a,\ n \to \infty$, $K_a/n \to \mu$ and $M_n = 2^{100}K_a$. This estimation is based the replica method from statistical physics. See section VII and appendix B-B for more details.

The converse from (9) and (10) is also plotted. This is in essence a single user[3] based converse bound. The converse presented here illustrates the fact the E_b/N_0 requirements are necessarily higher compared to the AWGN channel in [2].

[3] We can also derive a Fano type converse, but for the range of parameters we work with, it is worse than the presented one.

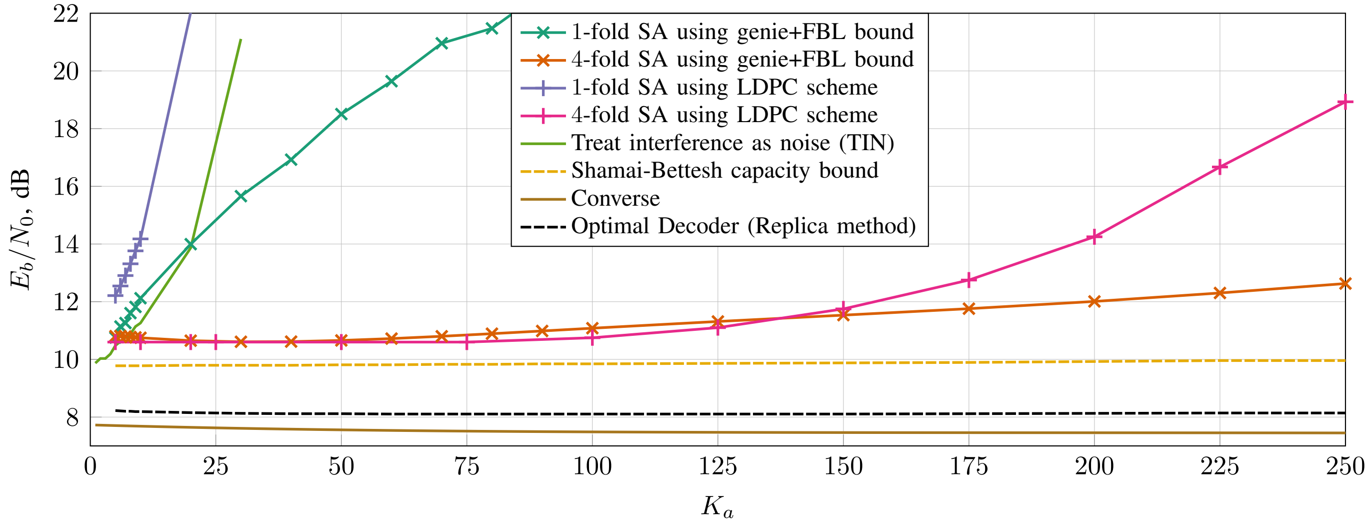

Fig. 2. K_a vs E_b/N_0 for $\epsilon \leq 0.1$, $n = 30000$, $k = 100$ bits. Dashed lines correspond to asymptotic approximation obtained by taking $n \to \infty$ and are shown only for reference.

VII. Asymptotics of Random-Access

In [2] the authors evaluated a random coding bound for AWGN RAC with $n = 30000$ and $K_a = 1, ..., 300$. 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 [20], 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 [2], [20]. 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 $n \to \infty$ with K_a, the number of active users, scaling linearly with blocklength (similar to the *many-access* regime [2], [30], [31]) i.e., $K_a = \mu n$. At the same time, the size of the common codebook is also scaling linearly: $M = M_1 K_a$. 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 $\{W_1, ..., W_{K_a}\}$ 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 $\mathrm{Unif}[M]$ 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

$$E_b/N_0 = \frac{nP}{\log M_1}. \tag{18}$$

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

$$\mathcal{E} = \frac{P_{tot}}{\mu \log M_1}. \tag{19}$$

We note that E_b/N_0 is defined this way for the reason that $\log \binom{M}{K_a} \approx K_a \log M_1$ 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 μ as $n \to \infty$. This setup is equivalent to the support recovery in compressed sensing considered in [27], [43]. Here, we provide a comparison of the fundamental trade-off of energy-per-bit with user density, for given PUPE and ρ, between our analysis of the projection decoder, the ML decoder in [27], the optimal decoder based on the true posteriors (see [27, 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, ϵ) code for the K_a user channel $P_{Y^n|X^n}$ given in (2) as follows.

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

$$P_e = \frac{1}{K_a} \sum_{j=1}^{K_a} \mathbb{P}\left[W_j \notin g(Y^n)\right] \le \epsilon \tag{20}$$

where Y^n is the channel output.

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

$$\mathcal{E}^*(M_1, \mu, \epsilon) = \limsup_{n \to \infty} \inf \left\{ \mathcal{E} : \exists \left(n, M = K_a M_1, \epsilon, \mathcal{E}, K_a = \mu n\right) - \text{code} \right\}. \tag{21}$$

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

$$\mathcal{E}_{conv} \leq \mathcal{E}^* \leq \mathcal{E}_{ach}\,. \tag{22}$$

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

These bounds are plotted in Fig. 3 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 [27] 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 [27]). The idea is that in the limit considered in this section, given a codebook, the posterior probability of a particular set of messages being transmitted undergo decoupling and converge to the posterior $\mathbb{P}\left[X \neq 0|Z\right]$ of the scalar channel $Y = X + \sigma Z$ where $Z \sim \mathcal{CN}(0,1)$ and X is $\mathcal{CN}(0,1)$ with probability $1/M_1$ and 0 with probability $1 - 1/M_1$ and $X \perp Z$. Here, the value of σ is given by [27]

$$\sigma^2 = \arg\min_{\tau > 0} \left\{ \frac{1}{\mu M_1} \log \tau M_1 + \log(e) \frac{1}{\tau M_1 P_{tot}} + I(X; X + \sqrt{\tau} Z) \right\}. \tag{23}$$

The above result in [27] was calculated for the compressed sensing (sparse support recovery) problem, and was based on the calculations in [44] using the replica method. However our setup can be directly translated to the sparse support recovery setup as can be seen in [20] for instance. Using (23), the PUPE can be computed as $\epsilon = \mathbb{P}\left[\mathbb{P}\left[X \neq 0|Y\right] < T|X \neq 0\right]$ where T satisfies $\mathbb{P}\left[\mathbb{P}\left[X \neq 0|Y\right] > T\right] = 1/M_1$.

The "optimal decoder" curves in Fig. 3 correspond to what we believe to be a fundamentally best achievable performance of a random Gaussian codebook. Unfortunately, at present it is impossible to rigorously establish this fact. Specifically, while the replica method predictions for the MMSE have been established rigorously (see [45] and [46]), a similar result about PUPE is not possible to extract from those works.

The converse bound plotted is based on Fano inequality and the single-user converse for AWGN channel from [47]. 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 is C is of zero mean and

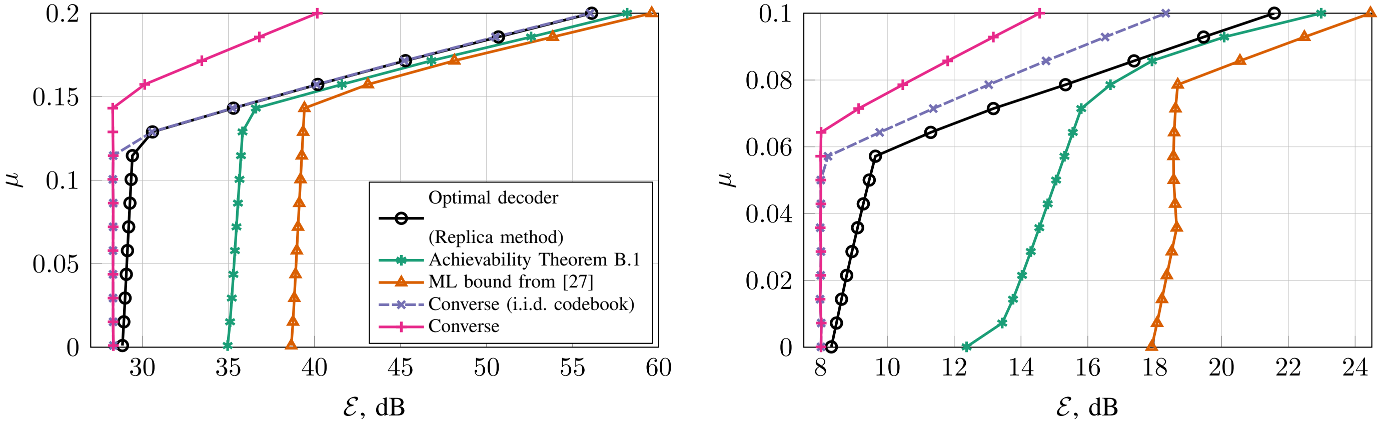

Fig. 3. μ vs $\mathcal{E}$ for $\epsilon \leq 10^{-3}$ (left) and $\epsilon \leq 0.1$ (right), $M_1 = 2^{100}$

finite variance. This follows from [43, 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).

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 SA 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 SA to get a rigorous random coding achievability bound. To be implemented in hardware the complexity of our algorithm should be significantly reduced. In the further research we are going to investigate a decoder based on successive interference cancellation which can be considered as a simplified version of our joint decoder in which the operation in the functional node is replaced with hard subtraction. We also mention that the subtraction in our setup is not straightforward as we need to know a fading coefficient with high accuracy. We also plan to consider another classes of codes e.g. polar codes which are better for the short length regime. 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] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading MAC," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019.

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

[3] L. G. Roberts, "ALOHA packet system with and without slots and capture," *ACM SIGCOMM Computer Communication Review*, vol. 5, no. 2, pp. 28–42, 1975.

[4] B. S.Tsybakov, "Survey of USSR Contributions to Random Multiple-Access Communications," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 143–165, 1985.

[5] E. Casini, R. De Gaudenzi, and O. Del Rio Herrero, "Contention Resolution Diversity Slotted ALOHA (CRDSA): An Enhanced Random Access Schemefor Satellite Access Packet Networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1408–1419, April 2007.

[6] 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.

[7] E. Paolini, G. Liva, and M. Chiani, "Coded Slotted ALOHA: A Graph-Based Method for Uncoordinated Multiple Access," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6815–6832, Dec 2015.

[8] R. De Gaudenzi, O. del Río Herrero, G. Acar, and E. Garrido Barrabés, "Asynchronous Contention Resolution Diversity ALOHA: Making CRDSA Truly Asynchronous," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6193–6206, Nov 2014.

[9] F. Clazzer, C. Kissling, and M. Marchese, "Enhancing Contention Resolution ALOHA Using Combining Techniques," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2576–2587, June 2018.

[10] E. Sandgren, A. Graell i Amat, and F. Brännström, "On Frame Asynchronous Coded Slotted ALOHA: Asymptotic, Finite Length, and Delay Analysis," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 691–704, Feb 2017.

[11] 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.

[12] 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.

[13] A. Fengler, G. Caire, P. Jung, and S. Haghighatshoar, "Massive MIMO Unsourced Random Access," *arXiv preprint arXiv:1901.00828*, 2019.

[14] 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.

[15] S. Ghez, S. Verdu, and S. C. Schwartz, "Stability properties of slotted Aloha with multipacket reception capability," *IEEE Transactions on Automatic Control*, vol. 33, no. 7, pp. 640–649, 1988.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] K. Stern, A. E. Kalør, B. Soret, and P. Popovski, "Massive Random Access with Common Alarm Messages," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1–5.

[22] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1890–1904, 2018.

[23] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, Sep. 2018.

[24] 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.

[25] S. Haghighatshoar, P. Jung, and G. Caire, "Improved Scaling Law for Activity Detection in Massive MIMO Systems," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 381–385.

[26] 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.

[27] 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.

[28] ——, "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.

[29] G. Reeves, "Sparse Signal Sampling using Noisy Linear Projections," Univ. Calif., Berkeley, Dept. of Electrical Engineering and Computer Science, Tech. Rep., 2008.

[30] 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.

[31] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *arXiv preprint arXiv:1901.06732*, 2019.

[32] ——, "Quasi-static fading MAC with many users and finite payload," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019.

[33] R. G. Gallager, *Low-Density Parity-Check Codes*. M.I.T. Press, 1963.

[34] R. Tanner, "A recursive approach to low complexity codes," *IEEE Transactions on information theory*, vol. 27, no. 5, pp. 533–547, 1981.

[35] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[36] 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.

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

[38] S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient coded random access for the wireless uplink," *arXiv preprint arXiv:1907.09448*, 2019.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

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

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

[46] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 665–669.

[47] 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.

[48] V. Y. Tan and P. Moulin, "Fixed error asymptotics for erasure and list decoding," *arXiv preprint arXiv:1402.4881*, 2014.

[49] 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 $X_i \overset{iid}{\sim} \mathcal{CN}(0, P' I_n)$ 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 [36], 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

$$g(Y) = \{f^{-1}(c) : c \in \hat{C}\}$$

$$\hat{C} = \arg \max_{C \subset \mathcal{C} : |C| = K_1} \|P_C Y\|^2 \tag{24}$$

where f is the encoding function.

The projection decoding is also called nearest-subspace decoding, and has been used in the compressed sensing literature [26]–[29]. 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

$$\hat{C} = \arg \max_{C \subset \mathcal{C} : |C| = K_1} \max_{\{H_i : i \in C\}} P_{Y|X,H}, \quad P_{Y|X,H}(y, \{x_i\}, \{h_i\}) = \frac{1}{\pi^n} e^{-\|y - \sum_i h_i x_i\|^2}.$$

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 ϵ–capacity region C_ϵ of the MAC [31].

B. FBL Achievability bounds

Theorem A.1. *Fix $P' < P$. Let $K_1 \leq K_2$. Then there exists an (M, n, ϵ) (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*

$$\epsilon \equiv P_e(M, n, K_2, P) \leq \frac{K_2 - K_1}{K_2} + \frac{1}{K_2} \sum_{t=1}^{K_1} K_{1,t} p_t + p_0 \tag{25}$$

with

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

and

$$p_t \leq \inf_{\delta>0}\left(\binom{K_2}{K_{1,t}} e^{-(n-K_1)\delta} + \mathbb{P}\left[\bigcup_{\substack{S_0\subset[K_2]\\|S_0|=K_{1,t}}}\{G(Y,S_0,c_{S_0},t)\geq V_{n,t}\}\right]\right) \tag{26}$$

where

$$G(Y,S_0,c_{S_0},t) = \frac{\|Y\|^2 - \max_{\substack{S_2\subset S_0\\|S_2|=t}}\left\|P_{c_{[S_2\cup([K_2]\setminus S_0)]}}Y\right\|^2}{\|Y\|^2 - \left\|P_{c_{[[K_2]\setminus S_0]}}Y\right\|^2} \tag{27}$$

$$K_{1,t} = K_2 - K_1 + t \tag{28}$$

$$V_{n,t} = e^{-\tilde{V}_{n,t}} \tag{29}$$

$$\tilde{V}_{n,t} = \delta + R_1 + s_t \tag{30}$$

$$s_t = \frac{\ln\binom{n'-1}{t-1}}{n-K_1} \tag{31}$$

$$R_1 = \frac{\ln\binom{M-K_2}{t}}{n-K_1} \tag{32}$$

$$n' = n - K_1 + t \tag{33}$$

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 (26) *can be upper bounded as*

$$p_t \leq \inf_{\substack{\delta>0\\ \delta_1>0\\ 0<\delta_2<1}}\left[\binom{K_2}{K_{1,t}}\left(e^{-(n-K_1)\delta} + e^{-n'f_n(\delta_1)} + e^{-n'\frac{\delta_2^2}{2}}\right) + \mathbb{P}\left[\min_{1\leq i\leq K_1-t+1}\frac{P'\sum_{j=i}^{i+t-1}|H_{(j)}|^2}{1+P'\sum_{j=i+t}^{K_{1,t}-1+i}|H_{(j)}|^2} \leq \frac{(1+\delta_1(1-V_{n,t}))V_{n,t}^{-1}-1}{1-\delta_2}\right]\right] \tag{34}$$

where

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

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

Proof: See appendix C. ■

Remark 2. *We note that* (26) *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* (34) *requires that the codebook is (complex) Gaussian.*

To compute (26) 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 (26) is prohibitive. However, there is a computationally tractable bound (which is worse than (26)) 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 [42] 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 (21).

A. Achievability

Theorem B.1. *Consider the channel* (2) *with* $K_a = \mu n$ *where* $\mu < 1$*. Fix* $M_1 > 1$ *and target PUPE* ϵ*. Let* $M = K_a M_1$ *denote the size of the codebook and* $P_{tot} = K_a P$ *be the total power. 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* $\limsup_{n\to\infty} \epsilon_n \le \epsilon$*, where, for* $\frac{\epsilon'}{\nu} < \theta \le 1$ *and* $\xi \in [0, \nu(1-\theta)]$*,*

$$\begin{aligned}
P_{tot,\nu}(\theta,\xi) &= \frac{\hat{f}(\theta,\xi)}{1-\hat{f}(\theta,\xi)\alpha\left(\xi+\nu\theta, \xi+1-\nu(1-\theta)\right)} && (35)\\
\hat{f}(\theta,\xi) &= \frac{f(\theta)}{\alpha(\xi, \xi+\nu\theta)} && (36)\\
f(\theta) &= \frac{\frac{1+\delta_1^*(1-V_\theta)}{V_\theta}-1}{1-\delta_2^*} && (37)\\
V_\theta &= e^{-\tilde{V}_\theta} && (38)\\
\tilde{V}_\theta &= \delta^* + \mu\frac{(M_1-1)}{1-\mu\nu}h\left(\frac{\theta\nu}{M_1-1}\right) + \frac{1-\mu\nu(1-\theta)}{1-\mu\nu}h\left(\frac{\theta\mu\nu}{1-\mu\nu(1-\theta)}\right) && (39)\\
\delta^* &= \frac{\mu h(1-\nu(1-\theta))}{1-\mu\nu} && (40)\\
c_\theta &= \frac{2V_\theta}{1-V_\theta} && (41)\\
q_\theta &= \frac{\mu h(1-\nu(1-\theta))}{1-\mu\nu(1-\theta)} && (42)\\
\delta_1^* &= q_\theta(1+c_\theta) + \sqrt{q_\theta^2(c_\theta^2+2c_\theta)+2q_\theta(1+c_\theta)} && (43)\\
\delta_2^* &= \inf\left\{x : 0<x<1, -\ln(1-x)-x > \frac{\mu h(1-\nu(1-\theta))}{1-\mu\nu(1-\theta)}\right\} && (44)\\
\alpha(a,b) &= a\ln(a) - b\ln(b) + b - a. && (45)
\end{aligned}$$

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

The proof of the above theorem follows from (34) (theorem A.1) and ideas very similar to [31, 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 [27]. 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 in 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 $\mathbb{P}\left[X \neq 0|Y\right]$ 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 σ is given by (see [27, Theorem 8], but modified here for complex case)

$$\sigma^2 = \arg\min_{\tau > 0} \left\{ \frac{1}{\mu M_1} \log \tau M_1 + \log(e) \frac{1}{\tau M_1 P_{tot}} + I(X; X + \sqrt{\tau} Z) \right\}. \tag{46}$$

The PUPE converges to $\mathbb{P}\left[\mathbb{P}\left[X \neq 0|Y\right] < T | X \neq 0\right]$ where T satisfies $\mathbb{P}\left[\mathbb{P}\left[X \neq 0|Y\right] > T\right] = p$. Hence, we can find the minimum P_{tot} such that this PUPE of the scalar channel is at most ϵ, 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 [47], [48]

Theorem B.2. *Let $M = K_a M_1$ be the codebook size. Given $\epsilon \leq 1 - \frac{K_a}{M}$ and μ such that $M_1 > 2$ then $\mathcal{E}^*(M_1, \mu, \epsilon) > \mathcal{E}_{conv}$ where $\mathcal{E}_{conv} = \inf \frac{P_{tot}}{\mu \log M_1} \max\{\mathcal{E}_{conv,1}, \mathcal{E}_{conv,2}\}$ where infimum is taken over all $P_{tot} > 0$ that satisfies the following two bounds*

$$\mu\theta \log M_1 - \epsilon\mu \log\left(M_1 - 1\right) - \mu h_2(\epsilon) \leq \log\left(1 + \alpha(1-\theta, 1) P_{tot}\right), \forall \theta \in [0,1] \tag{47}$$

$$\epsilon \geq 1 - \mathbb{E}\left[Q\left(Q^{-1}\left(\frac{1}{M_1}\right) - \sqrt{\frac{2P_{tot}}{\mu}|H|^2}\right)\right] \tag{48}$$

where Q is the complementary CDF function of the standard normal distribution and α is defined in (45).

Proof: The proof of (47) is based of Fano inequality and genie argument. ■

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 [43, Theorem 3] (see [43, 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 PUPE ϵ with high probability, the energy-per-bit $\mathcal{E}$ should satisfy*

$$\mathcal{E} \geq \inf \frac{P_{tot}}{\mu \log M_1} \tag{49}$$

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

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

where $\mathcal{V}$ is given by [43]

$$\mathcal{V}(r,\gamma) = r\ln\left(1+\gamma-\mathcal{F}(r,\gamma)\right) + \ln\left(1+r\gamma-\mathcal{F}(r,\gamma)\right) - \frac{\mathcal{F}(r,\gamma)}{\gamma} \tag{50}$$

$$\mathcal{F}(r,\gamma) = \frac{1}{4}\left(\sqrt{\gamma\left(\sqrt{r}+1\right)^2+1} - \sqrt{\gamma\left(\sqrt{r}-1\right)^2+1}\right)^2 \tag{51}$$

APPENDIX C

PROOF OF THEOREM A.1

In this section, we present the proof of theorem A.1. We remark that (52) and (59) prove (26). Note that $W_1, ..., W_{K_2}$ are sampled independently with replacement from $[M]$. We perform a change of measure by sampling $W_1, ..., W_{K_2}$ from $[M]$ *without* replacement, and also change the measure of transmitted message from $X_j = c_{W_j} 1\left\{\left\|c_{W_j}\right\|^2 \leq nP\right\}$ to $X_j = c_{W_j}$. Since P_e is the expectation of a non-negative random variable bounded by 1, this measure change adds a total variation distance which can bounded by

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

where $\chi_2(d)$ is the distribution of sum of squares of d iid standard normal random variables (the chi-square distribution). This follows from the same reasoning used in the main theorem in [2]. Henceforth we only consider the new measure. Now, P_e can be bounded as

$$P_e \leq \mathbb{E}\left[\frac{1}{K_2}\sum_{j=1}^{K_2} 1[W_j \notin g(Y)]\right] + p_0 \leq \frac{K_2 - K_1}{K_2} + \frac{1}{K_2}\sum_{t=1}^{K_1} p_{1,t}K_{1,t} + p_0 \tag{52}$$

where $K_{1,t}$ is given by (28) 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 $\{c_1, c_2, ..., c_{K_2}\}$. 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 (53)

$$\begin{aligned}\mathbb{P}\left[F_t|c_{[K_2]}, H_{[K_2]}, Z\right] &\leq \mathbb{P}\left[\exists S_0 \subset [K_2] : |S| = K_{1,t}, \exists S_1 \subset [M] \setminus [K_2] : |S_1| = t : \right.\\ &\qquad \left. \left\|P_{c_{[S_1][[K_2]\setminus S_0]}} Y\right\|^2 > \max_{\substack{S_2 \subset S_0\\ |S_2| = t}} \left\|P_{c_{[S_2][[K_2]\setminus S_0]}} Y\right\|^2 \right| c_{[K_2]}, H_{[K_2]}, Z \Bigg]\\ &\leq \mathbb{P}\left[\bigcup_{\substack{S_0 \subset [K_2]\\ |S_0| = K_{1,t}}} \bigcup_{\substack{S_1 \subset [M]\setminus[K_2]\\ |S_1| = t}} F(S_0, S_2^*, S_1, t) |\, c_{[K_2]}, H_{[K_2]}, Z\right], \qquad (53)\end{aligned}$$

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 $\mathbb{P}\left[F(S_0, S_2^*, S_1, t)|c_{[K_2]}, H_{[K_2]}, Z\right]$.

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 [31, Claim 1].

Claim 1 ([31]). *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 $\|P_{A_1} Y\|^2 + \|(I - P_{A_1})Y\|^2\,\mathrm{Beta}(t, n - K_1)$ where $\mathrm{Beta}(a, b)$ is a beta distributed random variable with parameters a and b.*

Therefore we have,

$$\mathbb{P}\left[F(S_0, S_2^*, S_1, t)|c_{[K_2]}, H_{[K_2]}, Z\right] = \mathbb{P}\left[Beta(n - K_1, t) < G_{S_0}|c_{[K_2]}, H_{[K_2]}, Z\right] = F_\beta\left(G_{S_0}; n - K_1, t\right) \qquad (54)$$

where

$$G_{S_0} = \frac{\|Y\|^2 - \|P_{A_0} Y\|^2}{\|Y\|^2 - \|P_{A_1} Y\|^2}. \qquad (55)$$

Since $t \geq 1$, we have $F_\beta\left(G_{S_0}; n - K_1, t\right) \leq \binom{n'-1}{t-1} G_{S_0}^{n-K_1}$, where n' is given by (33).

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,

$$\mathbb{P}\left[F_t|c_{[K_2]}, H_{[K_2]}, Z\right] \leq \sum_{S_0, K_1} \binom{M - K_2}{t} \binom{n' - 1}{t - 1} G_{S_0}^{n-K_1}. \qquad (56)$$

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

$$p_{1,t} = \mathbb{P}\left[F_t\right] \leq \mathbb{E}\left[\min\left\{1, \sum_{S_0, K_1} \binom{M - K_2}{t} \binom{n' - 1}{t - 1} G_{S_0}^{n-K_1}\right\}\right]$$

$$= \mathbb{E}\left[\min\left\{1, \sum_{S_0,K_1} e^{(n-K_1)(s_t+R_1)} G_{S_0}^{n-K_1}\right\}\right] \tag{57}$$

where s_t and R_1 are given by (31) and (32) respectively.

For $\delta > 0$, define $V_{n,t}$ as in (29). Let E_1 be the event

$$E_1 = \bigcap_{S_0,K_1} \{-\ln G_{S_0} - s_t - R_1 > \delta\} = \bigcap_{S_0,K_1} \{G_{S_0} < V_{n,t}\}. \tag{58}$$

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

$$\begin{aligned} p_{1,t} &\le \mathbb{E}\left[\min\left\{1, \sum_{S_0,K_1} e^{(n-K_1)(s_t+R_1)} G_{S_0}^{n-K_1}\right\}(1[E_1] + 1[E_1^c])\right] \\ &\le \mathbb{E}\left[\sum_{S_0,K_1} e^{-(n-K_1)\delta}\right] + p_{2,t} = \binom{K_2}{K_{1,t}} e^{-(n-K_1)\delta} + p_{2,t}. \end{aligned} \tag{59}$$

Note: This proves (26). Let us bound $p_{2,t}$. Let $\hat{Z} = Z + \sum_{i\in S_0\setminus S_2^*} H_i c_i$. From [31, Claim 2] we have

Claim 2 ([31]). *$p_{2,t}$ is bounded as*

$$\begin{aligned} p_{2,t} &= \mathbb{P}\left[\bigcup_{S_0,K_1} \{G_{S_0} > V_{n,t}\}\right] \\ &\le \mathbb{P}\left[\bigcup_{S_0,K_1} \left\{\left\|(1-V_{n,t})P_{A_1}^{\perp}\hat{Z} - V_{n,t}P_{A_1}^{\perp}\sum_{i\in S_2^*} H_i c_i\right\|^2 \ge V_{n,t}\left\|P_{A_1}^{\perp}\sum_{i\in S_2^*} H_i c_i\right\|^2\right\}\right]. \end{aligned} \tag{60}$$

Let $\chi_2'(\lambda, d)$ denote the non-central chi-squared distributed random variable with non-centrality λ and degrees of freedom d. That is, if $W_i \sim \mathcal{N}(\mu_i, 1), i \in [d]$ and $\lambda = \sum_{i\in[d]} \mu_i^2$, then $\chi_2'(\lambda, d)$ has the same distribution as that of $\sum_{i\in[d]} W_i^2$. We have the following claim from [31, Claim 3].

Claim 3 ([31]). *Conditional on $H_{[K_2]}$ and A_0,*

$$\left\|P_{A_1}^{\perp}\left(\hat{Z} - \frac{V_{n,t}}{1-V_{n,t}}\sum_{i\in S_2^*} H_i c_i\right)\right\|^2 \sim \left(1 + P'\sum_{i\in S_0\setminus S_2^*} |H_i|^2\right)\frac{1}{2}\chi_2'(2F, 2n') \tag{61}$$

where

$$F = \frac{\left\|\frac{V_{n,t}}{1-V_{n,t}} P_{A_1}^{\perp}\sum_{i\in S_2^*} H_i c_i\right\|^2}{\left(1 + P'\sum_{i\in S_0\setminus S_2^*} |H_i|^2\right)} \tag{62}$$

(63)

Hence its conditional expectation is $\mu = n' + F$.

Now let

$$T = \frac{1}{2}\chi_2'(2F, 2n') - \mu \tag{64}$$

$$U = \frac{V_{n,t}}{(1-V_{n,t})} \frac{\left\| P_{A_1}^{\perp} \sum_{i \in S_2^*} H_i c_i \right\|^2}{\left(1 + P' \sum_{i \in S_0 \setminus S_2^*} |H_i|^2\right)} - n' \tag{65}$$

$$U^1 = \frac{1}{1 - V_{n,t}} \left(V_{n,t} W_{S_0} - 1\right) \tag{66}$$

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 (67).

$$\text{RHS of (60)} = \mathbb{P}\left[\bigcup_{S_0, K_1} \left\{ \left\| P_{A_1}^{\perp} \hat{Z} - \frac{V_{n,t}}{(1 - V_{n,t})} P_{A_1}^{\perp} \sum_{i \in S_2^*} H_i c_i \right\|^2 - \mu \geq U \right\}\right] = \mathbb{P}\left[\bigcup_{S_0, K_1} \{T \geq U\}\right] \tag{67}$$

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

$$\begin{aligned} \mathbb{P}\left[\bigcup_{S_0, K_1} \{T \geq U\}\right] &\leq \sum_{S_0, K_1} \mathbb{P}\left[T > U, U^1 > \delta_1\right] + \mathbb{P}\left[E_2^c\right] \\ &= \sum_{S_0, K_1} \mathbb{E}\left[\mathbb{P}\left[T > U \middle| H_{[K_2]}, A_0\right] 1[U^1 > \delta_1]\right] + \mathbb{P}\left[E_2^c\right] \end{aligned} \tag{68}$$

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 [49, Lemma 8.1].

Lemma C.1 ([49])**.** *Let $\chi = \chi_2'(\lambda, d)$ be a non-central chi-squared distributed variable with d degrees of freedom and non-centrality parameter λ. Then $\forall x > 0$*

$$\begin{aligned} &\mathbb{P}\left[\chi - (d + \lambda) \geq 2\sqrt{(d + 2\lambda)x} + 2x\right] \leq e^{-x} \\ &\mathbb{P}\left[\chi - (d + \lambda) \leq -2\sqrt{(d + 2\lambda)x}\right] \leq e^{-x} \end{aligned} \tag{69}$$

Hence, for $x > 0$, we have

$$\mathbb{P}\left[\chi - (d + \lambda) \geq x\right] \leq e^{-\frac{1}{2}\left(x + d + 2\lambda - \sqrt{d + 2\lambda}\sqrt{2x + d + 2\lambda}\right)}. \tag{70}$$

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

$$\mathbb{P}\left[\chi \leq x\right] \leq e^{-\frac{1}{4}\frac{(d + \lambda - x)^2}{d + 2\lambda}}. \tag{71}$$

Observe that, in (70), the exponent is always negative for $x > 0$ and finite λ due to AM-GM inequality. When $\lambda = 0$, we can get a better bound for the lower tail in (71) by using [27, Lemma 25].

Lemma C.2 ([27]). *Let $\chi = \chi_2(d)$ be a chi-squared distributed variable with d degrees of freedom. Then* $\forall x > 1$

$$\mathbb{P}\left[\chi \leq \frac{d}{x}\right] \leq e^{-\frac{d}{2}\left(\ln x + \frac{1}{x} - 1\right)} \tag{72}$$

Therefore, from (60), (67), (68) and (70), we have

$$p_{2,t} \leq \sum_{S_0, K_1} \mathbb{E}\left[e^{-n' f_n(U^1)} 1[U^1 > \delta_1]\right] + \mathbb{P}\left[\bigcup_{S_0, K_1} \left\{U^1 \leq \delta_1\right\}\right] \tag{73}$$

where f_n is given by (A.1).

Next, from [31, 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

$$p_{2,t} \leq \sum_{S_0, K_1} e^{-n' f_n(\delta_1)} + p_{3,t} \tag{74}$$

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

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

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 (72), we can bound $p_{3,t}$ as

$$p_{3,t} \leq \sum_t \binom{K_2}{K_{1,t}} e^{-n'(-\ln(1-\delta_2) - \delta_2)} + p_{4,t} \tag{75}$$

where

$$p_{4,t} = \mathbb{P}\left[E_4^c\right] = \mathbb{P}\left[\bigcup_{S_0, K_1} \left\{V_{n,t}\left(1 + \frac{P' \sum_{i \in S_2^*} |H_i|^2 (1 - \delta_2)}{\left(1 + P' \sum_{i \in S_0 \setminus S_2^*} |H_i|^2\right)}\right) \leq 1 + \delta_1(1 - V_{n,t})\right\}\right]. \tag{76}$$

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

$$p_4 = \mathbb{P}\left[\min_{1 \leq i \leq K_1 - t + 1} \frac{P' \sum_{j=i}^{i+t-1} |H_{(j)}|^2}{1 + P' \sum_{j=i+t}^{K_{1,t}-1+i} |H_{(j)}|^2} \leq \frac{(1 + \delta_1(1 - V_{n,t})) V_{n,t}^{-1} - 1}{1 - \delta_2}\right] \tag{77}$$

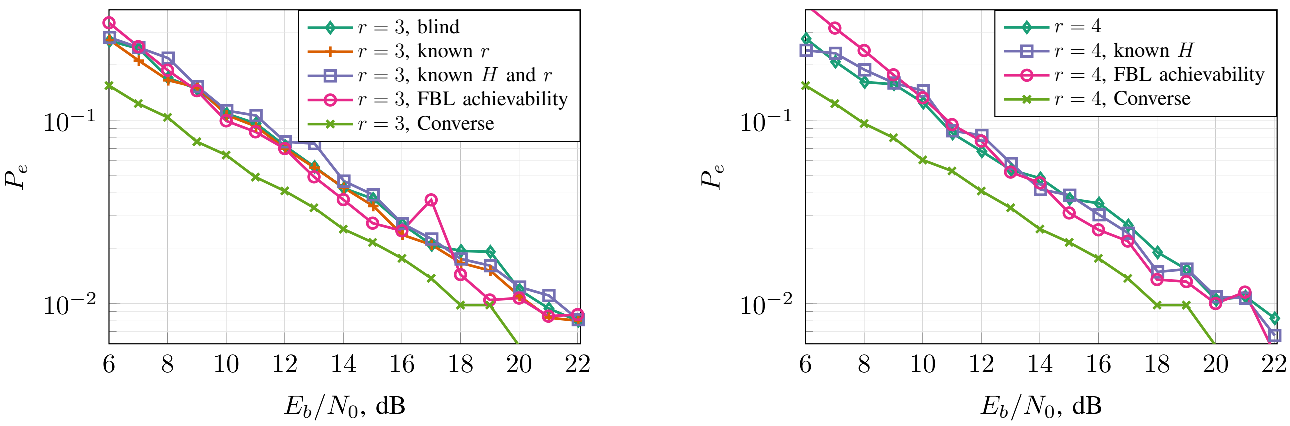

Fig. 4. Simulation results for $r = 3$ (left) and $r = 4$ (right) users, $T = 4$.

Finally, combining (52), (59), (74), (75) and (77) , and optimizing over δ, δ_1 and δ_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 [17]; 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 3 and 4 simultaneously active users in a slot, recall, that $T = 4$ for all the cases. We compare these curves with the following “ideal” curves: (a) fading channel coefficients are unknown, number of users is known (i.e. T is selected to be equal to the actual number of users); (b) fading channel coefficients are known, number of users is known (full CSI). Frame error rate performance for listed above scenarios are presented in Fig. 4 for $r = 3, 4$. 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 the LDPC-based scheme is good for resolving collisions of small order.