Coded Downlink Massive Random Access

Ryan Song, Kareem M. Attiah, and Wei Yu
Electrical and Computer Engineering Department
University of Toronto, Toronto, Canada
E-mails: r.song@mail.utoronto.ca, kattiah@ece.utoronto.ca, weiyu@ece.utoronto.ca

***Abstract*—This paper considers a massive connectivity scenario in which a base-station (BS) aims to communicate k individual sources $(X_1, \cdots, X_k)$ to a random subset of k users among a large pool of n users via a common downlink message. The identities of the k active users are known at the BS, but each active user only knows whether it is active itself and does not know the identities of the other active users. The naive coding strategy of transmitting the source messages together with the indices of the users for which the messages are intended would require a rate of $H(X_1, \cdots, X_k) + k\log(n)$ bits. This paper shows that if the sources are jointly distributed according to an exchangeable distribution, better coding techniques can be used to eliminate the dependency of the overhead on $\log(n)$. Specifically, if the sources are independently and identically distributed (i.i.d.) or are i.i.d. mixture, then the overhead can be reduced to $O(\log(H(X_1, \cdots, X_k)))$ or at most $O(\log(k))$ bits. The overhead can be further reduced to $O(1)$ if the source distribution is uniform over its support. For a general exchangeable source not necessarily i.i.d. nor i.i.d. mixture, an overhead of $O(k + \log(k + H(X_1, \cdots, X_k)))$ bits is achievable; if the source distribution has finite support, the overhead can be further reduced to $O(\log(k))$. Moreover, for exchangeable distributions that are extendable, the rate can be further improved.**

I. Introduction

This paper considers the problem of providing downlink connectivity to a massive number of n users through a central base-station (BS). The user activities are assumed to be sporadic, so that at any given time, only $k < n$ users are actively listening to the BS. The BS knows the identities of the active users and wishes to communicate sources $(X_1, \cdots, X_k)$ to these active users via a common downlink message; but each user only knows whether it is active itself and does not know which other users are active. We ask the question: What is the minimum rate of the common message that allows each active user i to learn the source X_i intended for it?

The question is trivial if every active user knows the identities of the entire subset of active users. In this case, all n users can each be pre-assigned an index. The BS can then list the source messages for the k active users according to the order in which the active users appear in the index. Such a common message requires only $H(X_1, \cdots, X_k)$ bits, but relies on each user knowing the identities of all active users.

The question is much more interesting for the practical scenario in which every active user only knows whether it is active itself and does not know who else is active. In this case, it would appear that the BS needs to send not only $H(X_1, \cdots, X_k)$ bits for the source messages, but also a header of $\log(n)$ bits per user to describe which user each source message is intended for, thus resulting in a $k\log(n)$ overhead.

Such a $k\log(n)$ overhead can be significant if n is large (e.g., $n = 10^6$ means $\log(n) \approx 20$ bits), especially when the payload for each user is small in comparison. The main insight of this paper is that if the joint distribution of the sources $(X_1, \cdots, X_k)$ follows an exchangeable distribution, then the $\log(n)$ dependency in the overhead can be eliminated!

Exchangeable distributions arise naturally in the context of massive random access, because the user activity patterns are typically random and symmetric across the users. Consequently, no subset of users is preferred over any other subset. More formally, we assume that the sources $(X_1, \cdots, X_k)$ form a vector of exchangeable random variables, defined by the condition that if $p(x_1, \ldots, x_k)$ is the joint distribution of the sources and π is a permutation of $(1, \cdots, k)$, then

$$p(x_1, \cdots, x_k) = p(x_{\pi(1)}, \cdots, x_{\pi(k)}). \tag{1}$$

Classical examples of exchangeable distributions include independently and identically distributed (i.i.d.) random variables, mixture of i.i.d. random variables, and urn distributions (sampling with or without replacement from a population).

The main idea of this paper is that when $(X_1, \cdots, X_k)$ is exchangeable, it is possible to build a codebook comprised of many possible realizations of the sources $(x_1, \cdots, x_n)$ over all n users in a symmetric way, so that when the identities of the activity users and their associated sources messages are revealed, the BS only needs to search over the codebook and broadcasts the index of the first codeword that *matches* the actual sources intended for the k active users in their respective locations in the codeword. The main technical contributions of this paper are novel code constructions and analyses which show that the entropy of the first matching codeword index can be made to be close to $H(X_1, \cdots, X_k)$ bits, without the $O(\log(n))$ overhead. Since a common message rate of at least $H(X_1, \cdots, X_k)$ bits is clearly required, the proposed scheme is essentially optimal to within a small overhead term.

Specifically, this paper shows that for i.i.d. and mixture i.i.d. sources, a rate of $H(X_1, \cdots, X_k) + \log(H(X_1, \cdots, X_k) + 1) + 1$ bits is achievable, thus the overhead is at most $O(\log(k))$. This rate can be further reduced to $O(1)$ if the source distribution is close to uniform over its support. Moreover, for any source distribution $p(x_1, \cdots, x_k)$, which is exchangeable but not necessarily i.i.d. or i.i.d. mixture, a rate of $H(X_1, \cdots, X_k) + k\log(e) + \log(H(X_1, \cdots, X_k) + k\log(e) + 1) + 1$ bits is achievable, thus the overhead for the general exchangeable sources is at most $O(k)$. If the alphabet size of the source is finite, the overhead can be

further improved to $O(\log(k))$. Finally, this paper also studies source distributions $p(x_1, \cdots, x_k)$ that are extendable to an exchangeable distribution over a vector of size d. In this case, if $d = O(k^{2+\epsilon})$, the overhead can be further improved to essentially $O(\log(k))$. Note that in all cases, the $O(\log(n))$ dependency is completely removed.

The intuitive reason why the coding strategy presented in this paper outperforms the naive scheme is because the naive scheme broadcasts too much redundant information. Specifically, each active user needs to recover only its own designated source message and should not care which other users are active, nor what their source messages are. The naive scheme broadcasts such information to everyone. In contrast, the coding strategy used in this paper takes advantage of the fact that each of the k active users is only interested in the source message pertaining to itself, so that each codeword can cover many different instances of the activity patterns and the associated source messages.

The coding strategy used in this paper has been used in previous work [1], [2], in which the problem of scheduling the active users into non-colliding transmission slots and the problem of assigning the active users into categories of fixed sizes are considered. Both of these previous works are examples of exchangeable source distributions, but the analyses of code rates in [1], [2] are specific to their respective problems; they do not generalize to all exchangeable distributions, and in particular do not apply to i.i.d. distributions. In contrast, this paper utilizes a new technique inspired by an analysis tool from the proof of the strong functional representation lemma [3] to provide a much more general result than that of [1], [2].

The codebook design and the subsequent achievability bounds of this paper can be applied to a number of interesting practical applications. For example, consider the problem in which the BS wishes to distribute a fixed amount of resource C amongst k users, i.e., $\sum_{i=1}^{k} X_i = C$. Assuming that each X_i is an non-negative integer, this vector of random variables forms an exchangeable source. As another example, suppose the BS wishes to assign the k users into d frequency bands with $d > k$ and with at most one user per slot. This vector of assignments is a sequence of random variables which are extendable to an exchangeable distribution over a vector of size d. The coding theorem of this paper provides a codebook construction for these cases with an overhead of at most $O(k + \log(k + H(X_1, \cdots, X_k)))$ bits. In the latter case, if $d = O(k^{2+\epsilon})$, then the overhead can be further improved to $O(\log(k))$.

Throughout this paper, we use lowercase letters to denote scalars, lowercase boldface letters to denote vectors, capital letters to denote random variables, boldface capital letters to denote random vectors, and calligraphic letters, i.e. $\mathcal{S}$, to denote sets and $|\mathcal{S}|$ to denote their cardinality. We let $\log(\cdot)$ denote the base 2 logarithm and $\ln(\cdot)$ denote the natural logarithm. All information measures are expressed in bits, including entropy $H(\cdot)$ and Kullback-Leibler divergence $D(\cdot\|\cdot)$. Lastly, we use short-hands $[n] = \{1, \ldots, n\}$ and $a^{\underline{b}} = a(a-1)\cdots(a-b+1)$.

II. Problem Formulation

Consider a massive random access setting in which a random subset of k users becomes active among a total number of n users. The identities of the active users are known to the BS but not among the users. We are interested in the setting in which the BS wishes to communicate a source to each of the k active users simultaneously using a common message. Upon receiving this common message, each active user should recover its respective source without error.

In this paper, we consider the class of source distributions that are exchangeable. Let $\mathbf{X} = (X_1, \cdots, X_k)$ be a sequence of exchangeable random variables taking values in discrete set $\mathcal{X}$ and let $p(\mathbf{x}) = p(x_1, \cdots, x_k)$ be their joint distribution.

We let random variable $\mathbf{A} \in \mathcal{A}^{(n,k)}$ specify the identities of the k active users, where

$$\mathcal{A}^{(n,k)} = \{\mathbf{a} \in [n]^k \mid a_i \neq a_j, \forall i \neq j\}. \tag{2}$$

Here, $a_i \in [n]$ is the index of the ith active user. For example, if the BS wishes to transmit a source $\mathbf{x} \in \mathcal{X}^k$ to users $\mathbf{a} \in \mathcal{A}^{(n,k)}$, we require that every user a_i receives x_i for all $i \in [k]$. While the source $\mathbf{X}$ describes the contents of the messages, the activity pattern $\mathbf{A}$ indicates which users should receive which message. Together, they form a message-activity tuple $(\mathbf{X}, \mathbf{A})$. Throughout this paper we assume that n and k are fixed, $\mathbf{X}$ and $\mathbf{A}$ are independent, and that $\mathbf{A}$ is distributed uniformly over $\mathcal{A}^{(n,k)}$. Notationally, we use $(\mathbf{x}, \mathbf{a})$ to represent a realization of $(\mathbf{X}, \mathbf{A})$.

The problem of communicating the sources $\mathbf{X}$ to the active users in $\mathbf{A}$ can now be cast as a one-shot source coding problem consisting of a single encoder and multiple decoders. In this paper, we pursue the following two-stage strategy for the source coding problem. In the first stage, the BS uses encoder f to map a message-activity pair $(\mathbf{x}, \mathbf{a})$ to a natural number, i.e.,

$$f : \mathcal{X}^k \times \mathcal{A}^{(n,k)} \to \mathbb{N}. \tag{3}$$

In the second stage, entropy coding is used to compress $f(\mathbf{x}, \mathbf{a})$ into a variable-length prefix-free binary codeword, which is then broadcast to all active users.

On the decoding side, assuming error-free broadcasting of the message, each active user a_i first recovers $f(\mathbf{x}, \mathbf{a})$. It then uses its own decoder $d_{a_i} : \mathbb{N} \to \mathcal{X}$ to recover its intended message. The sources are received successfully if

$$d_{a_i}(f(\mathbf{x}, \mathbf{a})) = x_i, \quad \forall (\mathbf{x}, \mathbf{a}) \in \mathcal{X}^k \times \mathcal{A}^{(n,k)}, \ \forall i \in [k]. \tag{4}$$

The rate of the encoding and decoding scheme is defined as $R = H(f(\mathbf{X}, \mathbf{A}))$. This is a reasonable definition because entropy coding can be used to reach the rate R to within one bit. Lastly, we define the optimal encoding and decoding scheme as the scheme that minimizes $H(f(\mathbf{X}, \mathbf{A}))$, while satisfying the condition (4). We denote the optimal rate as

$$R^* \triangleq H\left(f^*(\mathbf{X}, \mathbf{A})\right). \tag{5}$$

The goal of this paper is to establish upper bounds on R^*.

III. Codebook Construction

In this section we propose a codebook-based encoding and decoding scheme inspired by previous work [1], [2]. This encoding and decoding scheme utilizes the following shared codebook between the BS and all the users

$$\mathbf{M} = (\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \cdots) \tag{6}$$

which consists of an infinite sequence of length-n vectors $\mathbf{m}^{(t)} \in \mathcal{X}^n$. We assign a unique entry location in the codewords to each of the n users. For a particular message-activity tuple $(\mathbf{x}, \mathbf{a})$, the BS encodes $(\mathbf{x}, \mathbf{a})$ by finding a codeword $\mathbf{m}^{(t)}$ such that every active user has the correct message in their designated entry in $\mathbf{m}^{(t)}$, i.e., $m_{a_i}^{(t)} = x_i$ for all $i \in [k]$. Since the codebook is of infinite size and $\mathcal{X}$ is a discrete alphabet, such a codeword always exists. The idea is to encode $(\mathbf{x}, \mathbf{a})$ as the index of the *first* such codeword in the codebook. Mathematically, the proposed encoder can be written as

$$f_{\mathbf{M}}(\mathbf{x}, \mathbf{a}) = \min_{t: m_{a_i}^{(t)} = x_i, \forall i \in [k]} t. \tag{7}$$

The decoders for each user $u \in [n]$ are defined as:

$$d_u(t) = m_u^{(t)}. \tag{8}$$

One can easily verify that this encoding and decoding scheme satisfies the condition (4).

It remains to discuss how to generate the codewords in $\mathbf{M}$. Similar to [2], we use random codebook construction in which the codewords are generated in an i.i.d. fashion. Moreover, the entries of each codeword are generated according to an i.i.d. mixture distribution. Fix a $q(x|\theta)$ where $\theta \sim r(\theta)$. Each length-n codeword $\mathbf{m}^{(t)} = [x_1, \cdots, x_n]^\top$ is generated according to

$$q(x_1, \cdots x_n) = \int_\theta r(\theta) \left[\Pi_{i=1}^n q(x_i|\theta)\right] d\theta. \tag{9}$$

Operationally, each codeword is generated in a two stage-process in which we first randomly select a θ according to distribution $r(\theta)$, then generate the codeword entries of the codeword in an i.i.d. fashion according to $q(x|\theta)$. Each subsequent codeword is then generated in the same way. Once the entire codebook is generated, it is shared amongst the BS and all the users.

If we use this codebook construction, then the distribution of any k distinct entries of a single codeword is

$$q(x_1, \cdots x_k) = \int_\theta r(\theta) \left[\Pi_{i=1}^k q(x_i|\theta)\right] d\theta. \tag{10}$$

This means that the output of k distinct entries of the codewords can be designed according to any i.i.d. mixture.

This paper focuses attention to the encoding and decoding functions and the codebooks generated in this way. The main result of this paper is that for any exchangeable source distribution $p(x_1, \cdots, x_k)$, we can choose an appropriate i.i.d. mixture distribution $q(x|\theta)r(\theta)$ such that there exists a codebook, $\mathbf{M}$ generated in an i.i.d. fashion according to (9), whose encoder output entropy $H(f_{\mathbf{M}}(\mathbf{X}, \mathbf{A}))$ is upper bounded by $H(X_1, \cdots, X_k)$ plus a small overhead term.

IV. Entropy of the Encoder Output

In this section, we provide upper bounds on the optimal rate R^* of the downlink message for massive random access, defined in (5), by analyzing the entropy of the output of the encoder for a source $\mathbf{X} = (X_1, \cdots, X_k)$ distributed according to an exchangeable source distribution $p(x_1, \cdots, x_k)$ using an i.i.d. codebook generated according to a i.i.d. mixture distribution $q(x|\theta)r(\theta)$.

Theorem 1: Consider a massive access scenario with a total of n users and a random subset of k active users. Let sources $\mathbf{X} = (X_1, \cdots, X_k)$ take values in a discrete set $\mathcal{X}^k$ and be distributed according to an exchangeable distribution $p(\mathbf{x})$. Then for any i.i.d. mixture distribution $q(\mathbf{x})$, we have that the minimum achievable rate R^* is bounded above as

$$R^* \leq H(\mathbf{X}) + D(p\|q) + \log(H(\mathbf{X}) + D(p\|q) + 1) + 1 \tag{11}$$

$$R^* \leq H(\mathbf{X}) + D(p\|q) + \log\left(\log\left(\frac{q_{\max}}{q_{\min}}\right) + 1\right) + 3, \tag{12}$$

where

$$q_{\max} = \max_{\mathbf{x} \in \mathcal{X}^k : p(\mathbf{x}) > 0} q(\mathbf{x}); \quad q_{\min} = \min_{\mathbf{x} \in \mathcal{X}^k : p(\mathbf{x}) > 0} q(\mathbf{x}). \tag{13}$$

Note that the $D(p\|q)$ term corresponds to the cost of using a codebook constructed according to $q(\mathbf{x})$ to compress a source with distribution $p(\mathbf{x})$, just as in classical source coding.

Corollary 1: Under the setting of Theorem 1, if the source distribution is i.i.d. $p(x)$, then the minimum entropy of the encoder output is bounded above by

$$R^* \leq kH(p) + \log(kH(p) + 1) + 1 \tag{14}$$

$$R^* \leq kH(p) + \log\left(k \log\left(\frac{p_{\max}}{p_{\min}}\right) + 1\right) + 3, \tag{15}$$

where

$$p_{\max} = \max_{x \in \mathcal{X} : p(x) > 0} p(x); \quad p_{\min} = \min_{x \in \mathcal{X} : p(x) > 0} p(x). \tag{16}$$

This result follows directly from Theorem 1 by using a codebook constructed according to $q(x) = p(x)$. This is a natural choice for i.i.d. sources, because it maximizes the probability of matching codewords with the source. The result shows that for i.i.d. sources, the proposed code construction and encoding/decoding scheme can achieve the source entropy to within at most an $O(\log(k))$ overhead. The overhead can be further reduced if the distribution is close to uniform. In particular, for an i.i.d. uniform distribution with $p_{\max} = p_{\min}$, the entropy can be achieved to within $O(1)$ bits. Similar results can be obtained for mixture of i.i.d. source distributions, where it can be shown that a rate of $H(X_1, \cdots, X_k)$ plus an overhead at most $O(\log(k))$ bits is achievable.

It is worth noting that in the proposed codebook construction, the distribution of the codeword entries is required to be an i.i.d. mixture. Thus, for general exchangeable source distributions that are not i.i.d. mixtures, one can no longer simply set the codeword distribution to be the source distribution. We develop additional tools in Section V to construct an appropriate codebook and to upper bound the rate for general exchangeable sources.

A. Proof of (11) in Theorem 1

Instead of analyzing the output entropy of a particular codebook, we analyze the output entropy over an ensemble of codebooks. Define $T = f_{\mathcal{M}}(\mathbf{x}, \mathbf{a})$ to be the encoder output of first randomly choosing a codebook $\mathbf{M} \in \mathcal{M}$ generated according to $q(\mathbf{x})$, then finding the first match according to (7). To upper bound R^*, we first upper bound $H(T)$.

Since the source distribution $p(\mathbf{x})$ is exchangeable and the codewords are generated i.i.d. according to distribution $q(\mathbf{x})$, it follows that when conditioned on a source realization $\mathbf{x}$, the probability that the first match occurs at $T = t$ is a geometric distribution with parameter $q(\mathbf{x})$. Therefore, the overall distribution of T is a mixture of geometric distributions:

$$\mathsf{Pr}(T = t) = \sum_{\mathbf{x} \in \mathcal{X}^k} p(\mathbf{x})(1 - q(\mathbf{x}))^{t-1} q(\mathbf{x}). \tag{17}$$

A direct entropy calculation for a mixture of geometric distributions is nontrivial. To circumvent this, we upper bound $H(T)$ using the fact that

$$H(T) \leq \mathbb{E}[\log(T)] + \log(\mathbb{E}[\log(T)] + 1) + 1. \tag{18}$$

A proof of (18) can be found in [3], where a maximum entropy argument is used. The problem is now reduced to bounding $\mathbb{E}[\log(T)]$. To this end, we use the fact that

$$\log(t) < \frac{1}{\ln(2)} \left(\sum_{\ell=1}^{t} \frac{1}{\ell} - \frac{1}{2} \right) \leq \frac{1}{\ln(2)} \sum_{\ell=1}^{t-1} \frac{1}{\ell}, \text{ for all } t > 1 \tag{19}$$

to compute the following bound for $\mathbb{E}[\log(T)]$:

$$\begin{aligned} \mathbb{E}[\log(T)] &= \sum_{t=1}^{\infty} \left(\sum_{\mathbf{x} \in \mathcal{X}^k} p(\mathbf{x})(1 - q(\mathbf{x}))^{t-1} q(\mathbf{x}) \right) \log(t) \\ &= \sum_{\mathbf{x} \in \mathcal{X}^k, p(\mathbf{x}) > 0} p(\mathbf{x}) q(\mathbf{x}) \sum_{t=2}^{\infty} (1 - q(\mathbf{x}))^{t-1} \log(t) \\ &\leq \sum_{\mathbf{x} \in \mathcal{X}^k, p(\mathbf{x}) > 0} p(\mathbf{x}) q(\mathbf{x}) \frac{1}{\ln(2)} \sum_{t=2}^{\infty} \sum_{\ell=1}^{t-1} (1 - q(\mathbf{x}))^{t-1} \frac{1}{\ell}. \end{aligned} \tag{20}$$

The next step is to compute the inner summations. Notice that for $0 < \alpha < 1$ we have that

$$\begin{aligned} \sum_{t=2}^{\infty} \sum_{\ell=1}^{t-1} \alpha^{t-1} \frac{1}{\ell} &= \sum_{\ell=1}^{\infty} \frac{1}{\ell} \sum_{t=\ell+1}^{\infty} \alpha^{t-1} = \frac{1}{1-\alpha} \sum_{\ell=1}^{\infty} \frac{\alpha^{\ell}}{\ell} \\ &= \frac{1}{1-\alpha} \int_0^{\alpha} \frac{1}{1-z} dz = \frac{-\ln(2) \log(1-\alpha)}{1-\alpha}, \end{aligned} \tag{21}$$

where the third equality follows from the fact that

$$\frac{d}{d\alpha} \sum_{\ell=1}^{\infty} \frac{\alpha^{\ell}}{\ell} = \sum_{\ell=1}^{\infty} \alpha^{\ell} = \frac{1}{1-\alpha}, \text{ for } 0 < \alpha < 1. \tag{22}$$

Using (21), we set $\alpha = 1 - q(\mathbf{x})$ in (20) to simplify it to

$$\begin{aligned} \mathbb{E}[\log(T)] &\leq \sum_{\mathbf{x} \in \mathcal{X}^k, p(\mathbf{x}) > 0} p(\mathbf{x}) q(\mathbf{x}) \frac{-\log(q(\mathbf{x}))}{q(\mathbf{x})} \\ &= H(\mathbf{X}) + D(p \| q). \end{aligned} \tag{23}$$

Combining with (18) yields

$$H(T) \leq H(\mathbf{X}) + D(p \| q) + \log(H(\mathbf{X}) + D(p \| q) + 1) + 1.$$

Next, we show the existence of a single good codebook. By Jensen's inequality, we have that

$$\mathbb{E}[H(f_{\mathbf{M}}(\mathbf{X}, \mathbf{A}))] \leq H(f_{\mathcal{M}}(\mathbf{X}, \mathbf{A})) = H(T), \tag{24}$$

where the expectation is taken over $\mathbf{M} \in \mathcal{M}$. It then follows that there exists at least one codebook $\mathbf{M}^*$ such that

$$H(f_{\mathbf{M}^*}(\mathbf{X}, \mathbf{A})) \leq \mathbb{E}[H(f_{\mathbf{M}}(\mathbf{X}, \mathbf{A}))]. \tag{25}$$

Therefore

$$R^* \leq H(f_{\mathbf{M}^*}(\mathbf{X}, \mathbf{A})) \leq H(T). \tag{26}$$

B. Proof of (12) in Theorem 1

Let $T = f_{\mathcal{M}}(\mathbf{X}, \mathbf{A})$ be defined as before. By (17), we know that the distribution of T is a mixture of geometric distribution with parameters $q(\mathbf{x})$. We bound the entropy of T by defining a new variable L and using the fact that

$$H(T) \leq H(T, L) = H(L) + H(T|L). \tag{27}$$

The idea is to use an L that quantizes the parameters of the possible geometric distributions induced by $\mathbf{X}$. To this end, we define L based on the following partition $\mathcal{S}_1, \mathcal{S}_2 \cdots \subseteq \mathcal{X}^k$, where

$$\mathcal{S}_{\ell} = \left\{ \mathbf{x} \in \mathcal{X}^k \;\middle|\; \frac{1}{2^{\ell}} \leq q(\mathbf{x}) < \frac{1}{2^{\ell-1}}, p(\mathbf{x}) > 0 \right\}, \tag{28}$$

and let L be the index of the set $\mathcal{S}_{\ell}$ that contains $\mathbf{X}$. It follows that for any $\mathbf{x} \in \mathcal{X}^k$ and $p(\mathbf{x}) > 0$, the index of the set that $\mathbf{x}$ belongs to is $\ell = \lceil \log \left(\frac{1}{q(\mathbf{x})} \right) \rceil$. This means that L takes on at most $\lceil \log \left(\frac{1}{q_{\min}} \right) \rceil - \lceil \log \left(\frac{1}{q_{\max}} \right) \rceil$ values, and therefore has an entropy bounded above as

$$\begin{aligned} H(L) &\leq \log \left(\left\lceil \log \left(\frac{1}{q_{\min}} \right) \right\rceil - \left\lceil \log \left(\frac{1}{q_{\max}} \right) \right\rceil \right) \\ &\leq \log \left(\log \left(\frac{q_{\max}}{q_{\min}} \right) + 1 \right). \end{aligned} \tag{29}$$

It now remains to bound $H(T|L)$ in (27). Recall that the codebook is constructed according to $q(\mathbf{x})$. So, given $L = \ell$, sources $\mathbf{x}$ in $\mathcal{S}_{\ell}$ would induce a distribution of T close to a geometric distribution with parameter $q(\mathbf{x})$, where $q(\mathbf{x})$ is within the interval for each ℓ as defined in (28).

Now, the entropy of a geometric distribution with parameter $q(\mathbf{x})$ is essentially $-\log(q(\mathbf{x}))$. Thus, after averaging over ℓ, which can be written equivalently as averaging over $\mathbf{x}$, $H(T|L)$ essentially becomes $-\sum_{\mathbf{x}} p(\mathbf{x}) \log q(\mathbf{x})$. This gives rise to the $H(\mathbf{X}) + D(p \| q)$ term. This analysis can be made precise by carefully bounding all the approximation errors. We omit the details here and only state the final result:

$$H(T|L) \leq -\sum_{\mathbf{x}} p(\mathbf{x}) \left(\log \left(q(\mathbf{x}) \right) + \log(e) \right) + 1, \tag{30}$$

which together with (29) gives the desired result (12):

$$R^* \leq H(\mathbf{X}) + D(p \| q) + \log \left(\log \left(\frac{q_{\max}}{q_{\min}} \right) + 1 \right) + 3. \tag{31}$$

V. Coding for Exchangeable Sources

We now aim to generalize the result of Corollary 1 from i.i.d. sources to exchangeable sources with distribution $p(\mathbf{x})$. The challenge lies in constructing length-n codewords whose arbitrary subsequences of length k all "look like" $\mathbf{x}$. Recall from Theorem 1 that the achievable rate for communicating a general exchangeable source $\mathbf{X} = (X_1, \cdots, X_k)$ distributed as $p(\mathbf{x})$ is $H(\mathbf{X})$ plus an overhead term that depends on $D(p||q)$, where q is an i.i.d. mixture distribution used for codebook construction. The natural question is then: how should we design an i.i.d. mixture $q(\mathbf{x})$ such that $D(p||q)$ is small?

Definition 1 (Urn Codebook): Given an exchangeable source $\mathbf{X}$ with distribution $p(\mathbf{x})$, an urn codebook $\mathbf{M}_{\text{URN}} = \left(\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \cdots, \right)$ is a codebook consisting of codewords $\mathbf{m}^{(j)}$ generated in the following fashion:

1) Sample a realization of $\mathbf{x} = (x_1, \cdots, x_k)$ using $p(\mathbf{x})$.
2) Generate each entry m of $\mathbf{m}^{(j)}$ in an i.i.d. fashion according to $\hat{p}_{\mathbf{x}}$, where $\hat{p}_{\mathbf{x}}$ is the empirical distribution of $\mathbf{x}$:

$$\hat{p}_{\mathbf{x}}(m) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{1}_{\{m\}} x_i. \tag{32}$$

It is easy to see that the urn codebook is constructed from a mixture of i.i.d. distributions. In particular, the mixture weights are given by the probability that the sampled realization is of a particular type. Further, we argue that this i.i.d mixture is close to the source distribution. To see this, note that this codebook generation process can be alternatively viewed as repeatedly sampling with replacement from an urn containing the elements of $\mathbf{x}$. On the other hand, if the k entries of a codeword are sampled without replacement from the entries of $\mathbf{x}$, then the k entries will look as if they are distributed according to $p(\mathbf{x})$. Thus, the difference between the source distribution and the i.i.d. mixture is precisely the difference between sampling with and without replacement.

Lemma 1: Let $p(\mathbf{x})$ be an exchangeable distribution. Let $q(\mathbf{x})$ be the distribution generated by choosing k distinct entries from the codewords in the urn codebook as described in Definition 1, then $D(p||q) \leq \min\{k \log(e), |\mathcal{X}| \log(k+1)\}$.

Proof: From the urn codebook construction, any k distinct entries in a codeword can be represented by a tuple $\mathbf{Y} = (X_{W_1}, \cdots, X_{W_k})$, where $\mathbf{W} = (W_1, \cdots, W_k)$ is i.i.d. over $[k]$. Now, if the sampling pattern were collision-free, it would have generated sequences that are distributed according to the original $p(\mathbf{x})$. Define the set of all collision-free $\mathbf{w}$ as: $\mathcal{F} = \{\mathbf{w} \in [k]^k \mid w_i \neq w_j \text{ for all } i \neq j\}$. We have

$$\begin{aligned} q(\mathbf{x}) &= \sum_{\mathbf{w} \in [k]^k} \mathsf{Pr}(\mathbf{W} = \mathbf{w}) \mathsf{Pr}(\mathbf{Y} = \mathbf{x} | \mathbf{W} = \mathbf{w}) \\ &\geq \sum_{\mathbf{w} \in \mathcal{F}} \mathsf{Pr}(\mathbf{W} = \mathbf{w}) \mathsf{Pr}(\mathbf{Y} = \mathbf{x} | \mathbf{W} = \mathbf{w}) \\ &= \sum_{\mathbf{w} \in \mathcal{F}} \frac{1}{k^k} p(\mathbf{x}) = \frac{k!}{k^k} p(\mathbf{x}). \end{aligned} \tag{33}$$

where in the third line we used the fact that $\mathsf{Pr}(\mathbf{W} = \mathbf{w}) = \frac{1}{k^k}$ since all sampling patterns are equally likely.

Noting that $k! > k^k e^{-k}$, we have $\frac{k!}{k^k} < k \log(e)$. This implies that $D(p||q) = \sum p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \leq k \log(e)$.

Next, we show that $D(p||q) \leq |\mathcal{X}| \log(k+1)$. This bound is a consequence of the method of types and its relation to exchangeability. For vector $\mathbf{x} \in \mathcal{X}^k$, let $\mathcal{T}_{\mathbf{x}}^{(k)} = \{\mathbf{s} \in \mathcal{X}^k \mid \hat{p}_{\mathbf{s}} = \hat{p}_{\mathbf{x}}\}$ denote its type class. Due to the exchangeability of $p(\mathbf{x})$, if $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathcal{T}_{\mathbf{x}}^{(k)}$ then $p(\mathbf{x}^{(1)}) = p(\mathbf{x}^{(2)})$. The idea is to lower bound $q(\mathbf{x})$ by restricting attention to the sequences within $\mathcal{T}_{\mathbf{x}}^{(k)} = \{\mathbf{s} \in \mathcal{X}^k \mid \hat{p}_{\mathbf{s}} = \hat{p}_{\mathbf{x}}\}$:

$$\begin{aligned} q(\mathbf{x}) &= \sum_{\mathbf{s} \in \mathcal{X}^k} p(\mathbf{s}) \left[\Pi_{i=1}^{k} \hat{p}_{\mathbf{s}}(x_i)\right] \\ &= \sum_{\mathbf{s} \in \mathcal{X}^k} p(\mathbf{s}) 2^{-k(H(\hat{p}_{\mathbf{x}}) + D(\hat{p}_{\mathbf{x}} \| \hat{p}_{\mathbf{s}}))} \geq \sum_{\mathbf{s} \in \mathcal{T}_{\mathbf{x}}^{(k)}} p(\mathbf{x}) 2^{-kH(\hat{p}_{\mathbf{x}})} \\ &= |\mathcal{T}_{\mathbf{x}}^{(k)}| p(\mathbf{x}) 2^{-kH(\hat{p}_{\mathbf{x}})} \geq \frac{1}{(k+1)^{|\mathcal{X}|}} p(\mathbf{x}), \end{aligned} \tag{34}$$

where the second line follows from theorem 11.1.2 in [4] and the last line follows from the fact that $|\mathcal{T}_{\mathbf{x}}| \geq \frac{1}{(k+1)^{|\mathcal{X}|}} 2^{kH(\hat{p}_{\mathbf{x}})}$ [4]. This implies that $D(p||q) \leq |\mathcal{X}| \log(k+1)$. ∎

Corollary 2: Consider a massive access scenario with a total of n users and a random subset of k active users. Let sources $\mathbf{X} = (X_1, \cdots, X_k)$ take values in a discrete set $\mathcal{X}^k$ and be distributed according to an exchangeable distribution $p(\mathbf{x})$. Then the minimum achievable rate R^* is bounded above as

$$R^* \leq H(\mathbf{X}) + k \log(e) + \log(H(\mathbf{X}) + k \log(e) + 1) + 1 \tag{35}$$

$$\begin{aligned} R^* \leq H(\mathbf{X}) &+ |\mathcal{X}| \log(k+1) \\ &+ \log(H(\mathbf{X}) + |\mathcal{X}| \log(k+1) + 1) + 1. \end{aligned} \tag{36}$$

Corollary 2 follows immediately from Lemma 1 and Theorem 1. The implication here is that for any exchangeable source, the overhead beyond $H(\mathbf{X})$ is at most $\min\{k \log e, |\mathcal{X}| \log(k+1)\} + \log(k + H(\mathbf{X}))$.

VI. Concluding Remarks

This paper shows that exchangeable sources can be efficiently communicated to a random subset of users in massive random access with at most constant overhead per user. This is achieved through a clever code construction that exploits the relationship between exchangeable distributions and mixture of i.i.d. distributions.

In fact, if we consider the case in which the exchangeable source $(X_1, \cdots, X_k)$ is extendable to a longer exchangeable sequence $(X_1, \cdots, X_k, \cdots, X_d)$, code rates with even lower overheads are achievable. This can be done by modifying the codebook construction in Definition 1 to sample from realizations of $(x_1, \cdots, x_d)$ instead of $(x_1, \cdots, x_k)$. The same arguments used to prove Lemma 1 can be used to show that if the codeword is generated through the sampling of this extended sequence, then $D(p||q) \leq \log\left(\frac{d^k}{d^{\underline{k}}}\right) \leq -\log\left(1 - \frac{k(k-1)}{2d}\right)$. This means that if $d = O(k^{2+\epsilon})$, then $D(p||q)$ essentially vanishes, which removes the $k \log(e)$ term in Corollary 2. It is worth mentioning that this result is similar to results found in the literature pertaining to exchangeable sequences and finite de Finetti theorems [5]–[7].

REFERENCES

[1] J. Kang and W. Yu, "Minimum feedback for collision-free scheduling in massive random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 12, pp. 8094–8108, Dec. 2021.

[2] R. Song, K. M. Attiah, and W. Yu, "Coded categorization in massive random access," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 2868–2873.

[3] C. T. Li and A. E. Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

[5] P. Diaconis and D. Freedman, "Finite Exchangeable Sequences," *The Annals of Probability*, vol. 8, no. 4, pp. 745 – 764, 1980. [Online]. Available: https://doi.org/10.1214/aop/1176994663

[6] A. J. Stam, "Distance between sampling with and without replacement," *Stat. Neerl.*, vol. 32, no. 2, pp. 81–91, Jun. 1978.

[7] P. Harremoës and F. Matúš, "Bounds on the information divergence for hypergeometric distributions," 2020. [Online]. Available: https://arxiv.org/abs/2002.03002