# Downlink Massive Random Access With Lossy Source Coding

Ryan Song[*†] and Wei Yu[*]
[*] Electrical and Computer Engineering Department, University of Toronto, Canada
[†] School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Emails: ryan.song@epfl.ch, weiyu@ece.utoronto.ca

***Abstract*—This paper considers the coded downlink massive random access problem in which a base-station (BS) aims to communicate descriptions of the sources $(X_1, \cdots, X_k)$ to a randomly activated subset of $k$ users, among a large pool of $n$ potential users, via a common message in the downlink. Assuming that the downlink channel is noiseless, this paper investigates the lossy source coding setting where upon receiving the common message from the BS, each active user aims to recover a reconstruction $\hat{X}_i$ of their intended source $X_i$, such that the expected distortion between $(\hat{X}_1, \cdots, \hat{X}_k)$ and $(X_1, \cdots, X_k)$ is less than $D$. In this paper, we show that a previously proposed lossless coding strategy and its corresponding codebook construction for exchangeable sources, the urn codebook, can be extended to the lossy source coding setting using the Poisson functional representation. With this coding strategy, we show that for exchangeable sources $(X_1, \cdots, X_k)$, a common message length of $R(D)$ bits plus an overhead of $O(k)$ bits, independent of $n$, is achievable, where $R(D)$ is the rate-distortion function for compressing $(X_1, \cdots, X_k)$. If the sources are i.i.d., this overhead can be reduced to $O(\log(k))$ bits.**

## I. Introduction

This paper considers the problem setting of *coded downlink massive random access* [1], where a central base-station (BS) aims to communicate sources $(X_1, \cdots, X_k)$ to a randomly activated subset of $k$ users out of a large pool of $n$ potential users. Due to the sporadic activity of the users, only the activated $k \ll n$ users are listening to the BS. Although each active user has no knowledge of the identities of the other active users, the BS is assumed to know the identities of all the active users. Using a common message transmitted over a noiseless downlink channel, the BS aims to communicate to each active user a lossy reconstruction of the source intended for it. This problem arises naturally in the context of machine-type communications or Internet-of-Things, where users are often randomly activated and can request information from the BS in the downlink upon being detected by the BS [2]–[5].

As in [1], we assume in this paper that the user activity patterns are random and symmetric across the pool of all potential users, resulting in no subset of users being preferred over any other subset. Moreover, we assume that the sources intended for the active users are independent of the user activity pattern, and that for any fixed user activity pattern, all permutations of any source sequence are equally probable. This symmetry under permutations is known as *exchangeability*. Formally, sources $(X_1, \cdots, X_k)$ are exchangeable if their joint distribution $p(x_1, \cdots, x_k)$ is invariant under permutations, i.e.,

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

for all $(x_1, \cdots, x_k) \in \mathcal{X}^k$ and $\sigma \in \mathcal{S}_k$, where $\mathcal{S}_k$ is the set of all bijections $\sigma : [k] \to [k]$, where $[k] = \{1, \cdots, k\}$.

The lossless coding setting of this problem is studied in [1], where each active user is to recover their intended source $X_i$ without loss. It is shown that if the sources $\mathbf{X} = (X_1, \cdots, X_k)$ are exchangeable, then a common message length of $H(\mathbf{X})$ bits plus an overhead independent of $n$ is achievable. Specifically, for general exchangeable sources, an overhead of $O(k)$ bits is achievable. If the sources are i.i.d., then this overhead can be reduced to $O(\log(k))$ bits. This contrasts the naive scheme, which would have required an overhead of $k \log(n)$ bits to identify which users the sources are intended for.

This paper extends the results of [1] to the lossy source coding setting, where instead of requiring each active user to recover their intended source exactly, each active user is to recover a lossy reconstruction of their intended source such that a distortion criterion is satisfied. Let $\hat{X}_i$ denote the lossy reconstruction of source $X_i$. To quantify distortion, we use a distortion measure $\delta : \mathcal{X}^k \times \hat{\mathcal{X}}^k \to [0, \infty]$. We require the expected distortion to satisfy

$$\mathbb{E}[\delta((X_1, \cdots, X_k), (\hat{X}_1, \cdots, \hat{X}_k))] \leq D. \tag{2}$$

Similar to why the sources are assumed to be exchangeable, we also assume the distortion measure to be exchangeable, i.e., for all $\sigma \in \mathcal{S}_k$,

$$\delta(\mathbf{x}, \hat{\mathbf{x}}) = \delta(\mathbf{x}_\sigma, \hat{\mathbf{x}}_\sigma), \tag{3}$$

where $\mathbf{x}_\sigma = (x_{\sigma(1)}, \cdots, x_{\sigma(k)})$.

The main question this paper aims to answer is the following. Assuming that both the sources and the distortion measure are exchangeable, what is the minimum common message length needed to communicate lossy reconstructions of the sources to the intended active users? In this paper we show that a common message length of $R(D)$ bits plus a small overheard is achievable, where

$$R(D) = \min_{p(\hat{\mathbf{x}}|\mathbf{x}):\ \mathbb{E}[\delta(\mathbf{X}, \hat{\mathbf{X}})] \leq D} I(\mathbf{X}; \hat{\mathbf{X}}) \tag{4}$$

is the rate-distortion function for compressing $(X_1, \cdots, X_k)$. Analogous to the results of [1], if the sources are exchangeable, then an overhead of $O(k)$ bits is achievable. If the sources are i.i.d., then this overhead can be reduced to $O(\log(k))$ bits. In both cases, the overhead is independent of $n$, i.e., the proposed coding strategy is able to communicate the sources without explicitly transmitting the identities of the users for which the sources are intended.

In this paper, we propose a generalization of the coding scheme presented in [1] for the lossless setting to achieve the aforementioned common message lengths for the lossy case. The main idea behind the coding scheme of [1] is to use a codebook comprised of many length-$n$ codewords, each being different realizations of the sources. Each of the $n$ users is assigned a unique location in the codewords apriori, so that when the identities of the $k$ active users are revealed to the BS, the sources can be communicated to the active users by searching over the codebook for the *first* codeword that *matches* the sources intended for the $k$ active users. The BS can then transmit this index to the active users, who can then decode by referencing their assigned entry in the specified codeword.

The lossless coding scheme of [1] may lead one to a naive generalization, where the BS and users share a codebook consisting of many length-$n$ codewords, each being realizations of source reconstructions. Upon revealing the active users' identities, the BS finds the index of the first codeword that satisfies the distortion criterion and communicates this index to the active users. Although seemingly reasonable, this coding scheme fails to achieve a common message length on the order of $R(D)$. The main reason is that it enforces a maximum distortion instead of the average distortion criterion.

This paper shows that a more sophisticated index selection can remedy the above issue. Instead of selecting the first codeword that satisfies the distortion criterion, this paper proposes using a selection method inspired by the Poisson functional representation [6], [7]. The main contribution of this paper is to show that when the sources are exchangeable, this index selection method is an effective strategy for lossy source coding in the massive random access context.

The proposed coding strategy can be applied in a straightforward fashion to i.i.d. sources using an i.i.d. codebook. For non-i.i.d., but exchangeable sources, we propose to use the urn codebook construction from [1]. This paper analyzes a resource allocation problem as an example to show that the proposed coding scheme can achieve a common message length that can be significantly better than using an i.i.d. codebook.

## II. Problem Formulation

This paper considers a massive random access setting in which a random subset of $k$ users becomes active among a large pool of $n$ users. We assume that both $k$ and $n$ are fixed and known. The identities of the active users are known by the BS, but not by the other active users. Using a noiseless downlink channel, the BS aims to communicate lossy reconstructions $\hat{\mathbf{X}} = (\hat{X}_1, \cdots, \hat{X}_k)$ of sources $\mathbf{X} = (X_1, \cdots, X_k)$ to the $k$ active users via a downlink common message.

This paper aims to find the minimum length of common message required for the active users to produce reconstructions $\hat{\mathbf{X}}$ of the sources such that the expected distortion is less than $D$, i.e.,

$$\mathbb{E}[\delta(\mathbf{X}, \hat{\mathbf{X}})] \leq D, \tag{5}$$

where $\delta : \mathcal{X}^k \times \hat{\mathcal{X}}^k \to [0, \infty]$ is the distortion measure, $\mathcal{X}$ is the source alphabet, $\hat{\mathcal{X}}$ is the reconstruction alphabet, and the expectation is taken over $(\mathbf{X}, \hat{\mathbf{X}})$. This paper considers the lossy coding setting where both the sources and the distortion measure are exchangeable, as defined in (1) and (3).

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

$$\mathcal{A}^{(n,k)} = \{\mathbf{a} \in [n]^k \mid a_i \neq a_j, \forall i \neq j\}. \tag{6}$$

Here, $a_i \in [n]$ is the index of the $i$th active user. While $\mathbf{X}$ describes the source contents, the activity pattern $\mathbf{A}$ indicates the target user for each source, i.e., we want each user $a_i$ to reconstruct $x_i$ $\forall i \in [k]$. Together, they form a source-activity pair $(\mathbf{X}, \mathbf{A})$. We assume that $\mathbf{X}$ and $\mathbf{A}$ are independent. Notationally, we use $(\mathbf{x}, \mathbf{a})$ to represent a realization of $(\mathbf{X}, \mathbf{A})$.

We assume the availability of infinite common randomness, with a fixed but arbitrary distribution, between the BS and active users, denoted as $\mathbf{M} \in \mathcal{M}$. Upon learning the identities of the active users and their respective sources, the BS uses encoder

$$f : \mathcal{X}^k \times \mathcal{A}^{(n,k)} \times \mathcal{M} \to \{0,1\}^* \tag{7}$$

to map the source-activity pair as well as the realization of the common randomness to a binary string. This binary string is then broadcast to all the active users, who then use

$$d_{a_i} : \{0,1\}^* \times \mathcal{M} \to \hat{\mathcal{X}} \tag{8}$$

to recover a reconstruction of their intended source. The encoder and the decoders $(f, d_1, \cdots, d_n)$ satisfy the distortion criterion if $\mathbb{E}[\delta(\mathbf{X}, \hat{\mathbf{X}})] \leq D$, where

$$\hat{\mathbf{X}} = (d_{a_1}(f(\mathbf{X}, \mathbf{A}, \mathbf{M}), \mathbf{M}), \cdots, d_{a_k}(f(\mathbf{X}, \mathbf{A}, \mathbf{M}), \mathbf{M})) \tag{9}$$

are the reconstructions of the sources by the active users.

To ensure unique decodability, we require that the set of all possible output binary strings of the encoder to be a *prefix-free code*. An optimal encoding scheme is defined as the encoder and decoders $(f^*, d_1^*, \cdots, d_n^*)$ that minimize $\mathbb{E}[\mathtt{len}(f(\mathbf{X}, \mathbf{A}, \mathbf{M}))]$ with $\mathtt{len}(\cdot)$ denoting the length of a string, while satisfying the distortion criterion. The optimal common message length is

$$R^* \triangleq \mathbb{E}\left[\mathtt{len}\left(f^*(\mathbf{X}, \mathbf{A}, \mathbf{M})\right)\right], \tag{10}$$

where $f^*$ is the optimal encoder.

## III. Encoder and Decoder Design

In this section, we develop a lossy coding scheme for downlink massive random access. We begin by introducing a one-shot lossy source coding scheme based on the Poisson functional representation [6], [7]. We then generalize this coding scheme towards downlink massive random access by integrating it with the codebook construction of [1].

### A. One-Shot Coding via Poisson Functional Representation

Consider the following two-stage strategy for the one-shot single-user lossy source coding problem. Upon observing source realization $x \in \mathcal{X}$ and common randomness realization $\mathbf{m} \in \mathcal{M}$, the encoder first maps $x$ and $\mathbf{m}$ to a positive index using

$$l : \mathcal{X} \times \mathcal{M} \to \mathbb{N}. \tag{11}$$

This positive index is then compressed into a binary string using an optimal prefix-free code $f(x, \mathbf{m})$. The binary string is then transmitted to the decoder, which first recovers the index $l(x, \mathbf{m})$ from the binary string and then, along with the common randomness, decodes to a source reconstruction. This two-stage decoding process is represented by the map

$$s : \{0,1\}^* \times \mathcal{M} \to \hat{\mathcal{X}}. \tag{12}$$

One approach towards one-shot lossy source coding is through the lens of the following distributed sampling problem. Given source $X \sim p(x)$, fix a conditional distribution $p(\hat{x}|x)$ such that $\mathbb{E}[\delta(X, \hat{X})] \leq D$. Upon observing source realization $x$ and common randomness, the encoder transmits a message to the decoder. Using this message along with the observed common randomness, the decoder generates a sample distributed according to $p(\hat{x}|X = x)$. Since the conditional distribution $p(\hat{x}|x)$ satisfies the distortion criterion, a coding scheme for this sampling problem is also a valid coding scheme for one-shot lossy source coding.

Although we assume the availability of infinite common randomness between the encoder and decoder, the distribution of the common randomness must be fixed beforehand. Taking inspiration from [6] and [7], we use points from a marked Poisson point process (PPP) as the common randomness. Let $U^{(1)} \leq U^{(2)} \leq \cdots$ be real-valued random variables such that the differences $U^{(t+1)} - U^{(t)}$ are i.i.d. $\mathrm{Exp}(1)$ for all $t \in \mathbb{N}$. The sequence $\{U^{(t)}\}_{t=1,2,\ldots}$ defined this way is known as a PPP with rate 1. Next, we mark each point $U^{(t)}$ with a sample $\hat{X}^{(t)}$ drawn i.i.d. from an arbitrary fixed distribution $q(\hat{x})$. We let this marked PPP be denoted as

$$\mathbf{M} = \{(\hat{X}^{(t)}, U^{(t)})\}_{t=1,2,\cdots}. \tag{13}$$

Using the observed common randomness $\mathbf{m}$, the encoder can communicate a lossy reconstruction $\hat{x}$ by specifying an index $l(x, \mathbf{m})$. Upon receiving the index and observing $\mathbf{m}$, the decoder recovers a reconstruction by referencing $\hat{x}^{(l(x,\mathbf{m}))}$ in $\mathbf{m}$. The index selection $l(x, \mathbf{m})$ must simultaneously satisfy two objectives. The first is that the distortion criterion must be satisfied. Supposing that $\mathbf{M}$ is random, then the output $l(x, \mathbf{M})$ is also random. As the decoded reconstruction is described by the random variable $\hat{X}^{(l(x,\mathbf{M}))}$, one way to satisfy the distortion criterion is to require that

$$\hat{X}^{(l(x,\mathbf{M}))} \sim p(\hat{x}|X = x). \tag{14}$$

The second objective is for the entropy $H(l(X, \mathbf{M}))$ to be small. This ensures that the expected message length $R$ is small because

$$R < H(l(X, \mathbf{M})) + 1, \tag{15}$$

assuming that an optimal prefix-free code is used to compress the index $l(x, \mathbf{m})$,

Leveraging the marking and displacement properties of PPPs (see [8]), we propose the following index selection inspired by [6]:

$$l(x, \mathbf{m}) = \arg\min_t \left\{ u^{(t)} \cdot \frac{q\left(\hat{x}^{(t)}\right)}{p\left(\hat{x}^{(t)}|X = x\right)} \right\}. \tag{16}$$

This index selection satisfies both of the aforementioned objectives. The first being that $\hat{X}^{(l(x,\mathbf{M}))} \sim p(\hat{x}|X = x)$ and the second being that the entropy $H(l(X, \mathbf{M}))$ is approximately $I(X; \hat{X}) + D_{\mathrm{KL}}(p(\hat{x}) \| q(\hat{x}))$ bits plus a log factor. This is captured formally in Lemma 1, which states a result similar to that of [6, Theorem 2], but with the added detail of accounting for the setting where the distribution of the marking process $q(\hat{x})$ is mismatched from the distribution $p(\hat{x})$.

This mismatch results in a penalty of $D_{\mathrm{KL}}(p(\hat{x}) \| q(\hat{x}))$ bits. As seen in the next section, when coding for exchangeable sources in massive random access, it is not always possible to set the distribution of the marking process $q(\hat{x})$ to be equal to the optimal distribution of the reconstructions. Hence, it is important to quantify the penalty for having such a mismatch.

*Lemma 1:* Let source $X$ have distribution $p(x)$ and fix a conditional distribution $p(\hat{x}|x)$. Let $\mathbf{M} = \{(\hat{X}^{(t)}, U^{(t)})\}_{t=1,2,\cdots}$, where $\hat{X}^{(t)}$ are i.i.d. according to $q(\hat{x})$ and $U^{(t)}$ are points from a PPP with rate 1. Let $l(x, \mathbf{m})$ be defined as in (16). Then, $\hat{X}^{(l(x,\mathbf{M}))} \sim p(\hat{x}|X = x)$ and

$$\begin{aligned} H(l(X,\mathbf{M})) \leq\ & I(X; \hat{X}) + D_{\mathrm{KL}}(p(\hat{x}) \| q(\hat{x})) \\ & + \log(I(X; \hat{X}) + D_{\mathrm{KL}}(p(\hat{x}) \| q(\hat{x})) + 1) + 4. \end{aligned} \tag{17}$$

*Proof:* We present only a proof outline, as the proof is largely similar to that of [6, Theorem 1]. The distribution of $\hat{X}^{(l(x,\mathbf{M}))}$ can be shown to be $p(\hat{x}|X = x)$ using the definition of $l(x, \mathbf{m})$ and the displacement property of PPPs (see [8]). To upper bound $H(l(X, \mathbf{M}))$, we use an intermediate result from [6, Appendix A], which implies that

$$\mathbb{E}[\log(l(x, \mathbf{M}))] \leq D_{\mathrm{KL}}(p(\hat{x}|X = x) \| q) + \frac{\log(e)}{e} + 1. \tag{18}$$

By expanding the divergence term, taking an expectation with respect to $X$, and applying the maximum entropy argument of [6, Proposition 4], we get the desired upper bound on $H(l(X, \mathbf{M}))$. ■

### B. Encoder for Downlink Massive Random Access

We now utilize the coding scheme from the previous section for downlink massive random access in a two-stage process. In the first stage, the BS uses a function $g$ to map the source-activity pair to a positive index

$$g : \mathcal{X}^k \times \mathcal{A}^{(n,k)} \times \mathcal{M} \to \mathbb{N}. \tag{19}$$

In the second stage, the BS compresses the output of $g(\mathbf{x}, \mathbf{a}, \mathbf{m})$ into a variable-length binary string $f(\mathbf{x}, \mathbf{a}, \mathbf{m})$ using an optimal prefix-free code. On the decoding side, each active user $a_i$ first recovers $g(\mathbf{x}, \mathbf{a}, \mathbf{m})$ based on the received binary string, then recovers its respective source reconstructions.

For the common randomness, we use a PPP marked with length-$n$ vectors of possible reconstructions. Let $\{U^{(t)}\}_{t=1,2,\cdots}$ be points from a PPP with rate 1. For each point $U^{(t)}$, we mark it with a $n$-vector $\mathbf{C}^{(t)} \in \hat{\mathcal{X}}^n$, where each $\mathbf{C}^{(t)}$ is distributed i.i.d. according to an i.i.d. mixture distribution

$$q(\hat{x}_1, \cdots \hat{x}_n) = \int_\theta w(\theta) \left( \prod_{i=1}^{n} q(\hat{x}_i|\theta) \right) d\theta. \tag{20}$$

Putting these together, the common randomness is

$$\mathbf{M} = \{(\mathbf{C}^{(t)}, U^{(t)})\}_{t=1,2,\cdots} \tag{21}$$

Note that $\mathbf{C}^{(t)}$ is made to take values in $\hat{\mathcal{X}}^n$ so that a unique entry location can be assigned to each of the $n$ users. Since $\mathbf{C}^{(t)}$ is distributed according to an i.i.d. mixture, every $k$ distinct entries of $\mathbf{C}^{(t)}$ is distributed according to $q(\hat{x}_1, \cdots \hat{x}_k)$.

To encode a realization of $(\mathbf{x}, \mathbf{a})$ and $\mathbf{m}$, the BS computes

$$g(\mathbf{x}, \mathbf{a}, \mathbf{m}) = \arg\min_t \left\{ u^{(t)} \cdot \frac{q\left(c^{(t)}_{a_1}, \cdots, c^{(t)}_{a_k}\right)}{p^*\left(c^{(t)}_{a_1}, \cdots, c^{(t)}_{a_k} | \mathbf{X} = \mathbf{x}\right)} \right\}, \tag{22}$$

where $p^*(\hat{\mathbf{x}}|\mathbf{x})$ is a conditional distribution that satisfies the distortion criterion. The index $g(\mathbf{x}, \mathbf{a}, \mathbf{m})$ is then compressed using an optimal variable-length prefix-free code and broadcast to the active users. Since an optimal prefix-free code is used, the common message length is bounded from above as

$$R < H(g(\mathbf{X}, \mathbf{A}, \mathbf{M})) + 1. \tag{23}$$

Each active user first recovers the index $t = g(\mathbf{x}, \mathbf{a}, \mathbf{m})$, then decodes by referencing their entry of $\mathbf{c}^{(t)}$ in the common randomness $\mathbf{m}$, i.e., each active user $a_i$ decodes by computing

$$d_{a_i}(t, \mathbf{m}) = c^{(t)}_{a_i}. \tag{24}$$

## IV. Achievable Common Message Length

In this section, we discuss the common message length achievable by the coding scheme introduced in Section III-B. Applying Lemma 1 to the coding scheme, we have the following result.

*Theorem 1:* Consider a massive access scenario with a total of $n$ users and a random subset of $k$ active users. Let sources $\mathbf{X} = (X_1, \cdots, X_k) \in \mathcal{X}^k$ be exchangeable with distribution $p(\mathbf{x})$ and $\delta : \mathcal{X}^k \times \hat{\mathcal{X}}^k \to [0, \infty]$ be an exchangeable distortion measure, where $\hat{\mathcal{X}}$ is the reconstruction alphabet. The minimum common message length is bounded above by

$$\begin{aligned} R^* < \min_{q \in \mathcal{Q}} \Big( & R(D) + D_{\mathrm{KL}}(p^*(\hat{\mathbf{x}}) \| q(\hat{\mathbf{x}})) \\ & + \log(R(D) + D_{\mathrm{KL}}(p^*(\hat{\mathbf{x}}) \| q(\hat{\mathbf{x}})) + 1) + 5 \Big), \end{aligned} \tag{25}$$

where $\mathcal{Q}$ is the family of all i.i.d. mixture distributions on $\hat{\mathcal{X}}^k$ and $p^*$ is the marginal distribution that attains $R(D)$, i.e.,

$$p^*(\hat{\mathbf{x}}|\mathbf{x}) = \underset{p(\hat{\mathbf{x}}|\mathbf{x}):\ \mathbb{E}[\delta(\mathbf{X}, \hat{\mathbf{X}})] \le D}{\arg\min} I(\mathbf{X}; \hat{\mathbf{X}}) \tag{26}$$

and

$$p^*(\hat{\mathbf{x}}) = \sum_{\mathbf{x}} p^*(\hat{\mathbf{x}}|\mathbf{x}) p(\mathbf{x}). \tag{27}$$

From Theorem 1, a common message length of $R(D) + D_{\mathrm{KL}}(p^*(\hat{\mathbf{x}}) \| q(\hat{\mathbf{x}}))$ bits plus a logarithmic term is achievable, where $q(\hat{\mathbf{x}})$ is an arbitrary i.i.d. mixture distribution on $\hat{\mathcal{X}}^k$. The divergence term depends not only on the choice of $q(\hat{\mathbf{x}})$, but also on the distribution $p^*(\hat{\mathbf{x}})$ which attains $R(D)$. We further analyze some properties of $p^*(\hat{\mathbf{x}})$ under different assumptions on the sources and distortion measure.

*Lemma 2:* Let $(X_1, \cdots, X_k)$ be exchangeable sources taking values from alphabet $\mathcal{X}$ and $\delta : \mathcal{X}^k \times \hat{\mathcal{X}}^k \to [0, \infty]$ be an exchangeable distortion measure, where $\hat{\mathcal{X}}$ is the reconstruction alphabet. There exists a conditional distribution $p^*(\hat{\mathbf{x}}|\mathbf{x})$ which attains the rate-distortion function $R(D)$ and has a marginal distribution $p^*(\hat{\mathbf{x}})$ which is exchangeable.

Further, if the sources are i.i.d. and the distortion measure $\delta(\cdot, \cdot)$ is not only exchangeable but also tensorizable, i.e.,

$$\delta(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{i=1}^{k} \tilde{\delta}(x_i, \hat{x}_i) \tag{28}$$

for some $\tilde{\delta} : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty]$, then there exists a conditional distribution $p^*(\hat{\mathbf{x}}|\mathbf{x})$ which attains the rate-distortion function $R(D)$ and has marginal distribution $p^*(\hat{\mathbf{x}})$ which is i.i.d..

*Proof:* Let $\tilde{p}(\hat{\mathbf{x}}|\mathbf{x})$ be a conditional distribution that attains $R(D)$. Let

$$p^*(\hat{\mathbf{x}}|\mathbf{x}) = \sum_{\sigma \in \mathcal{S}_k} \frac{1}{k!} \tilde{p}(\hat{\mathbf{x}}_\sigma | \mathbf{x}_\sigma), \tag{29}$$

where $\mathcal{S}_k$ is the set of all bijections $\sigma : [k] \to [k]$. We claim that $p^*(\hat{\mathbf{x}}|\mathbf{x})$ attains $R(D)$. We begin by showing that $p^*(\hat{\mathbf{x}}|\mathbf{x})$ satisfies the distortion criterion. Let $(\mathbf{X}, \hat{\mathbf{X}}) \sim p(\mathbf{x}) p^*(\hat{\mathbf{x}}|\mathbf{x})$. By the exchangeability of the sources and the distortion measure,

$$\begin{aligned} \mathbb{E}[\delta(\mathbf{X}, \hat{\mathbf{X}})] &= \int \delta(\mathbf{x}, \hat{\mathbf{x}}) p(\mathbf{x}) p^*(\hat{\mathbf{x}}|\mathbf{x}) & (30) \\ &= \sum_{\sigma \in \mathcal{S}_k} \frac{1}{k!} \int \delta(\mathbf{x}_\sigma, \hat{\mathbf{x}}_\sigma) p(\mathbf{x}_\sigma) \tilde{p}(\hat{\mathbf{x}}_\sigma | \mathbf{x}_\sigma) & (31) \\ &\le D, & (32) \end{aligned}$$

where the integrals are taken over $\mathcal{X}^k \times \hat{\mathcal{X}}^k$. Since all terms are non-negative, the order of integration and summation can be swapped.

We now argue that $I(\mathbf{X}; \hat{\mathbf{X}}) = R(D)$. Let $(\mathbf{X}, \tilde{\mathbf{X}}_\sigma) \sim p(\mathbf{x}) \tilde{p}(\hat{\mathbf{x}}_\sigma | \mathbf{x}_\sigma)$. Since $p(\mathbf{x})$ is exchangeable, $p(\mathbf{x}) \tilde{p}(\hat{\mathbf{x}}_\sigma | \mathbf{x}_\sigma) = p(\mathbf{x}_\sigma) \tilde{p}(\hat{\mathbf{x}}_\sigma | \mathbf{x}_\sigma)$, which implies that $I(\mathbf{X}; \tilde{\mathbf{X}}_\sigma) = I(\mathbf{X}; \tilde{\mathbf{X}})$, as permuting indices does not affect mutual information. By the convexity of mutual information,

$$I(\mathbf{X}; \hat{\mathbf{X}}) \le \sum_{\sigma \in \mathcal{S}_k} \frac{1}{k!} I(\mathbf{X}; \tilde{\mathbf{X}}_\sigma) = R(D). \tag{33}$$

Lastly, we verify that $p^*(\hat{\mathbf{x}})$ is exchangeable. Notice that

$$p^*(\hat{\mathbf{x}}) = \int_{\mathcal{X}^k} p(\mathbf{x}) \sum_{\sigma \in \mathcal{S}_k} \frac{1}{k!} \tilde{p}(\hat{\mathbf{x}}_\sigma | \mathbf{x}_\sigma) \tag{34}$$

$$= \int_{\mathcal{X}^k} \sum_{\sigma \in \mathcal{S}_k} \frac{1}{k!} p(\mathbf{x}_\sigma) \tilde{p}(\hat{\mathbf{x}}_\sigma | \mathbf{x}_\sigma). \tag{35}$$

Since the summation is over all permutations, $p^*(\hat{\mathbf{x}}) = p^*(\hat{\mathbf{x}}_\sigma)$ for all $\sigma \in \mathcal{S}_k$. Therefore, $p^*(\hat{\mathbf{x}})$ is exchangeable.

Next, consider the setting where the sources are i.i.d. and the distortion measure is exchangeable and tensorizable. Let $\tilde{R}(D)$ be the single source rate-distortion function

$$\tilde{R}(D) = \min_{p(\hat{x}|x):\ \mathbb{E}[\tilde{\delta}(X,\hat{X})] \leq D} I(X; \hat{X}). \tag{36}$$

Since the sources are i.i.d., we have that

$$I(\mathbf{X}; \hat{\mathbf{X}}) \geq \sum_{i=1}^{k} h(X_i) - \sum_{i=1}^{k} h(X_i | \hat{X}_i) \geq k \tilde{R}\left(\frac{D}{k}\right). \tag{37}$$

Let $p^*(\hat{x}|x)$ be the conditional distribution that attains $\tilde{R}\left(\frac{D}{k}\right)$. Then $R(D)$ is attained by $p^*(\hat{\mathbf{x}}|\mathbf{x}) = \prod_{i=1}^k p^*(\hat{x}_i | x_i)$. ■

Following immediately from Theorem 1 and Lemma 2, we have the following result for i.i.d. sources.

*Corollary 1:* Under the setting of Theorem 1, if the sources are i.i.d. and the distortion measure is exchangeable and tensorizable, then the minimum common message length is bounded from above as

$$R^* < R(D) + \log(R(D) + 1) + 5. \tag{38}$$

Beyond the i.i.d. case, we know from Lemma 2 that if the sources and the distortion measure are both exchangeable, then there exists a $p^*(\hat{\mathbf{x}})$ that is exchangeable. But being exchangeable is not sufficient to allow us to set $q = p^*$, since the codebook construction requires $q(\hat{\mathbf{x}})$ to be an i.i.d. mixture. To deal with this issue, we adopt the urn codebook of [1], which constructs an i.i.d. mixture distribution through sampling with replacement from realizations of $p^*(\hat{\mathbf{x}})$. From [1, Theorem 2], we know that if the urn codebook distribution is used, then the divergence term is bounded from above as

$$D_{\text{KL}}(p^*(\hat{\mathbf{x}}) \| q_{\text{urn}}(\hat{\mathbf{x}})) \leq \min\{k \log(e), |\hat{X}| \log(k+1)\}. \tag{39}$$

This leads to the following result for general exchangeable sources and distortion measures.

*Corollary 2:* Under the setting of Theorem 1, if both the sources and the distortion measure are exchangeable, then the minimum common message length is bounded from above by

$$R^* < R(D) + \eta + \log(R(D) + \eta + 1) + 5, \tag{40}$$

where $\eta = \min\{k \log(e), |\hat{X}| \log(k+1)\}$.

Corollaries 1 and 2 are the main results of this paper. Together they show that i.i.d. sources can be communicated to the $k$ active users in a downlink massive random access setting using $R(D)$ bits plus an overhead of $O(\log(k))$ bits, and general exchangeable sources incur an overhead of at most $O(k)$ bits — in both cases, the overhead is independent of $n$.

## V. Examples

In this section, we give a non-trivial example of communicating exchangeable, but non-i.i.d. sources in a massive random access setting.

Suppose that $\mathbf{X} = (X_1, \cdots, X_k)$ represents an allocation of $r$ units of resource among the $k$ active users, i.e., $\sum_{i=1}^k X_i = r$ and $X_i \geq 0$. Let $\mathbf{X}$ be uniformly distributed on the simplex

$$\mathcal{T}^{(k-1)} = \left\{ \mathbf{x} \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = r, x_i \geq 0 \right\} \tag{41}$$

and let the distortion measure $\delta(\cdot, \cdot)$ be the $\ell_1$-distance

$$\delta(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{i=1}^{k} |x_i - \hat{x}_i|. \tag{42}$$

From Corollary 2, the urn codebook can be used to achieve a common message length of

$$R_{\text{urn}} \approx R(D) + k \log(e). \tag{43}$$

If instead, we restrict ourselves to using an i.i.d. codebook, the achievable common message length would be

$$R_{\text{i.i.d.}} \approx k \tilde{R}\left(\frac{D}{k}\right), \tag{44}$$

where

$$\tilde{R}\left(\frac{D}{k}\right) = \min_{p(\hat{x}_1|x_1):\ \mathbb{E}[|X_1 - \hat{X}_1|] \leq \frac{D}{k}} I(X_1; \hat{X}_1). \tag{45}$$

We now show that $R_{\text{urn}}$ can be less than $R_{\text{i.i.d.}}$. Consider the case of $k = 2$. We can upper bound $R(D)$ using the following coding scheme. The encoder uses $\tilde{R}\left(\frac{D}{2}\right)$ bits which allows the decoder to recover $\hat{X}_1$ such that $\mathbb{E}[|X_1 - \hat{X}_1|] \leq \frac{D}{2}$. Instead of encoding $X_2$, the decoder first recovers $\hat{X}_1$ and then computes $\hat{X}_2 = r - \hat{X}_1$. The distortion criterion is satisfied since

$$\mathbb{E}[|X_2 - \hat{X}_2|] = \mathbb{E}[|(r - X_1) - (r - \hat{X}_1)|] \leq \frac{D}{2}. \tag{46}$$

Therefore, $R(D) \leq \tilde{R}\left(\frac{D}{2}\right)$ and $R_{\text{urn}} \approx \tilde{R}\left(\frac{D}{2}\right) + 2\log(e)$. Comparing this to $R_{\text{i.i.d.}} \approx 2\tilde{R}\left(\frac{D}{2}\right)$, we can see that

$$R_{\text{i.i.d.}} - R_{\text{urn}} \approx \tilde{R}\left(\frac{D}{2}\right) - 2\log(e). \tag{47}$$

Since $\tilde{R}\left(\frac{D}{2}\right) \to \infty$ as $D \to 0$, we see that in the low distortion regime the urn codebook significantly outperforms an i.i.d. codebook.

## VI. Conclusion

This paper develops a coding scheme for lossy coding in the downlink massive random access which is inspired by both the Poisson functional representation and prior work on lossless coding for downlink massive random access. Using this coding scheme, we show that lossy reconstructions of sources can be communicated to a random subset of $k$ users out of a large pool of $n$ users using $R(D)$ bits, plus an overhead independent of $n$. For general exchangeable sources, the overhead is shown to be $O(k)$ bits. If the sources are i.i.d., then the overhead can be reduced to $O(\log(k))$ bits.

## References


[1] R. Song, K. M. Attiah, and W. Yu, "Coded downlink massive random access and a finite de Finetti theorem," *IEEE Trans. Inf. Theory*, 2025, to be published. [Online]. Available: https://arxiv.org/abs/2405.08301

[2] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[3] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Phase transition analysis for covariance-based massive random access with massive MIMO," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 1696–1715, 2022.

[4] A. E. Kalør, G. Durisi, S. Coleri, S. Parkvall, W. Yu, A. Mueller, and P. Popovski, "Wireless 6G connectivity for massive number of devices and critical services," *Proc. IEEE*, pp. 1–23, 2024.

[5] D. Gündüz, F. Chiariotti, K. Huang, A. E. Kalør, S. Kobus, and P. Popovski, "Timely and massive communication in 6G: Pragmatics, learning, and inference," *IEEE BITS Inf. Theory Mag.*, vol. 3, no. 1, pp. 27–40, 2023.

[6] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

[7] C. Maddison, "A Poisson process model for Monte Carlo," in *Perturbations, Optimization, and Statistics*. MIT Press, 2017, pp. 193–231.

[8] J. F. C. Kingman, *Poisson Processes*, ser. Oxford Studies in Probability. Oxford, England: Clarendon Press, Dec. 1992.