Finite-Blocklength Results for the A-channel: Applications to Unsourced Random Access and Group Testing

Alejandro Lancho, Alexander Fengler and Yury Polyanskiy

***Abstract*—We present finite-blocklength achievability bounds for the *unsourced* A-channel. In this multiple-access channel, users noiselessly transmit codewords picked from a common codebook with entries generated from a q-ary alphabet. At each channel use, the receiver observes the set of different transmitted symbols but not their multiplicity. We show that the A-channel finds applications in unsourced random-access (URA) and group testing. Leveraging the insights provided by the finite-blocklength bounds and the connection between URA and non-adaptive group testing through the A-channel, we propose improved decoding methods for state-of-the-art A-channel codes and we showcase how A-channel codes provide a new class of structured group testing matrices. The developed bounds allow to evaluate the achievable error probabilities of group testing matrices based on random A-channel codes for arbitrary numbers of tests, items and defectives. We show that such a construction asymptotically achieves the optimal number of tests. In addition, every efficiently decodable A-channel code can be used to construct a group testing matrix with sub-linear recovery time.**

I. Introduction

We consider the problem where K users transmit symbols from a q-ary input alphabet $[q] = \{1, \dots, q\}$ over a noiseless channel. Specifically, let $c_{i,j} \in [q]$ be the transmitted symbol from user $j \in [K]$ at channel use i. The channel output Y_i at channel use i is given by

$$Y_i = \bigcup_{j=1}^{K} c_{i,j}. \tag{1}$$

In this channel, sometimes referred to as A-channel [1], [2], the receiver observes the set of transmitted symbols but not who transmitted them, and also not the multiplicity.[1] The A-channel was introduced by Chang and Wolf in [1] as the "T-user M-frequency channel without intensity information", and it is also known as the hyperchannel [3]. The mutual information of the A-channel under uniform inputs was obtained in [1]. Its limit when K and q tend to infinity but its ratio $\lambda = K/q$ is fixed was studied in [2]. Specifically, in [2], it was shown that in this limit the mutual information grows proportional to q. Also in [2], it was shown that uniform inputs are not optimal in general, although they become optimal in the limit $\lambda \to 0$ and when $\lambda = \ln 2$, where $\ln(\cdot)$ denotes the natural logarithm. Besides, when the input distributions of the users are constrained to be equal, uniform distributions become asymptotically optimal for all $\lambda \leq \ln 2$ [2]. The mutual information with uniform inputs in the sparse limit of $K, q \to \infty$ with fixed ratio $(\log K)/q$ was computed in [4] and it was shown that in this limit the mutual information grows proportional to $\log q$. Furthermore, in this regime the simplified *cover* decoder, which checks each codeword individually for consistency with the channel output, is optimal. For general K and q, the optimal input distribution as well as the capacity of the A-channel are still unknown.

In the case where all users transmit their messages from a common codebook, we will refer to (1) as the *unsourced* A-channel. Under this setup, the receiver can only recover a list of transmitted codewords up to permutation. The information theoretic question of multiple-access in the unsourced setting was first formulated in [5] for the AWGN multiple-access channel (MAC), where it was established that a relevant setup should consider the following aspects: i) the decoder only aims to return a list of messages without recovering users' identities; ii) the error event should be defined per user; iii) the error probability has to be averaged over the users; iv) each user sends a fixed amount of information bits within a finite frame length.

This formulation is well suited for short-packet random-access wireless communications since, in theory, it does not require coordination among users. As such, it captures the requirements of massive machine-type communications (mMTC), one of the new emerging communication scenarios in next generation wireless networks, where a huge amount of battery-limited devices is expected to connect sporadically to the network to send short information packets. Since its inception, this problem has been commonly referred in the literature as unsourced random access (URA). Several papers establishing fundamental limits for different relevant multiple-access channel models and setups appeared since then (see, e.g., [6]–[9]), and many transmission schemes trying to perform as close as possible to this fundamental limits has been proposed (e.g., [10]–[12]).

The authors are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge 02139, MA, USA (e-mails: {lancho,fengler,yp}@mit.edu). Alejandro Lancho has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 101024432. Alexander Fengler was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Grant 471512611. This work is also supported by the National Science Foundation under Grant No CCF-2131115.

[1]Note that, for the case where $K = 2$, the multiplicity can be inferred from the cardinality of Y, and thus, for $q = 2$, the A-channel is equivalent to the binary-adder channel (BAC).

The A-channel played an important role for codes design in URA. In [13], a coding scheme for AWGN URA termed coded compressed sensing (CCS) was introduced. It used a random inner code of size q concatenated with an outer q-ary A-channel code. The A-channel code constructed for this purpose was termed *tree code*. The flexibility of this code construction allowed it to be extended to different channel models. Several follow-up works on URA (e.g, [11], [14]–[17]) made use of an outer A-channel code. In [4] an asymptotically Bayesian optimal inner decoder for the AWGN channel was constructed and it was shown that the CCS construction can achieve the Shannon limit when K and q grow but its ratio $\lambda \to 0$. In practical applications the density λ is not zero.

In fact, the A-channel is of relevance to URA in a more general sense: *Every unsourced K-user code for B bits at blocklength n_0 can be extended to a code of length nn_0 for $nBR_A(n,K)$ bits by concatenating with an outer unsourced A-channel code of rate R_A with n A-channel uses.* The loss in rate of R_A does not appear in classical multiple-access where user identification is done based on the codebook. A system that can transmit 1 bit for each user with zero error can be used to transmit arbitrary many bits by simple repetition. For the unsourced channel this is not possible and an outer A-channel code is necessary to couple repeated transmissions.

Furthermore, the blocklength of the outer A-channel used for concatenated coding (e.g., [4], [13], [14]) is in the order of $10-40$. Therefore, the asymptotic results for the A-channel are not necessarily insightful for code design.

In this paper, we study the unsourced A-channel in the finite blocklength regime with arbitrary K and q. In particular, we present two novel non-asymptotic achievability bounds. Also, we provide a second-order asymptotic approximation whose relevance is validated by means of numerical examples in different scenarios of interest.

The A-channel finds interesting applications in noiseless non-adaptive group testing. The goal in group testing is to identify K defective items in a large population of N items by applying T binary tests. A group-testing design is a $T \times N$ binary matrix where each column specifies the test in which that item participates. A test is declared positive if at least one tested item is defective. Group testing was developed by Dorfman in 1943 [18] for syphilis testing. Dorfman discovered that it is possible to test more people with a limited number of tests by pooling blood samples together. The topic has seen a recent rise in popularity since the COVID-19 pandemic led to a shortage of available tests for which group testing provides an appropriate solution. Group testing finds further important applications in DNA screening, large scale manufacturing control, neighborhood discovery, random access, machine learning, anomaly detection in routing networks, etc. [19]–[23]. For a recent survey on group testing from an information theoretic view, see [24].

The connection to the A-channel is as follows: Each codebook for the unsourced q-ary A-channel with blocklength n and size M gives rise to a group-testing design for $N = M$ items with $T = nq$ tests. To convert the codebook to a group-testing design, each q-ary symbol c_i is converted to a binary vector of size q with a 1 at position c_i. The defective items take the role of the transmitting users and the set of defective items can be obtained by recovering the transmitted messages. This A-channel group-testing design has a fixed number n of tests per item. The pair (n,q) can be used to optimize the group-testing design.

It is known that a fixed number of tests leads to improved error probabilities compared to an independent and identically distributed (i.i.d.) Bernoulli test design, even if the average number of tests is the same [24]. A popular design, analyzed in [25], uses a fixed number of tests per item, which are chosen at random from all tests. Compared to that, an A-channel design offers more structure as each item participates in exactly one of each group of q tests. The Kautz-Singleton (KS) construction [26] is another popular group-testing design that naturally has a q-ary structure. In particular, it is based on a q-ary Reed-Solomon code of length n. The KS construction was recently shown to be optimal for probabilistic group testing in certain scaling regimes [27]. The random coding bound developed in this paper gives a concrete finite blocklength achievability result for a random, but highly structured, group-testing design.

Motivated by the insights of our results and the algorithms developed in group testing, we also propose an improved decoder for the tree code. Numerical simulations confirm that the improved decoder significantly increases the achievable rates of the tree code.

II. Finite-Blocklength Framework

We consider the channel model introduced in (1), where K users transmit codewords from a common codebook with entries drawn from a q-ary input alphabet $[q] = \{1,\dots,q\}$ over n channel uses of a noiseless channel. To denote the n-length input-output relation, we shall also write

$$\boldsymbol{Y} = \bigcup_{j\in[K]} \mathbf{c}_j \tag{2}$$

where $\mathbf{c}_j \in [q]^n$ denotes the codeword transmitted by user j. We next define the notion of URA code for the A-channel.

Definition 1 (Code): Let $\binom{[a]}{b}$ denote the set of combinations of b-element subsets of $[a]$. Assume $q > K$, and let W_j, $j \in [K]$, denote the transmitted message by user j. An (M,n,ϵ)-code for the unsourced A-channel (2), where $\mathbf{c}_j \in [q]^n$, consists of an encoder-decoder pair,

- encoder: $f : [M] \mapsto [q]^n$;
- decoder: $g : \left\{\bigcup_{k=1}^{K} \binom{[q]}{k}\right\}^n \mapsto \binom{[M]}{K}$,

satisfying either the per-user probability of error (PUPE)

$$P_{\mathrm{e}}^{(\mathrm{p})} \triangleq \mathbb{P}[\{W_j \notin g(\boldsymbol{Y})\} \cup \{W_j = W_i, j \neq i\}] \leq \epsilon \tag{3}$$

or the joint probability of error (JPE)

$$P_{\mathrm{e}}^{(\mathrm{j})} \triangleq \mathbb{P}\big[\{\{W_j\}_{j=1}^{K} \neq g(\boldsymbol{Y})\} \cup \{W_j = W_i, j \neq i\}\big] \leq \epsilon. \tag{4}$$

We assume that $\{W_j\}_{j=1}^K$ are independent and uniformly distributed on $[M]$, and that $f(W_j) = \mathbf{c}_j \in [q]^n$. For each type of error probability, we say the code achieves a *rate* $R = \log_2 M/n$.

Hence, we have K users selecting randomly a codeword from a common codebook, and the decoder's task is to provide an estimate of the transmitted list of length K. In this paper, we assume K is known at the receiver.

A. Achievability Non-Asymptotic Bounds

In this section, we present our finite-blocklength achievability bounds for the unsourced A-channel. To do so, we consider a random-coding scheme where a codebook $\mathcal{C}$ contains M randomly generated codewords of length n distributed according to $P_{\boldsymbol{X}}(\mathbf{c}) = \prod_{i=1}^n P_X(c_i)$, where $P_X = \mathrm{Unif}[q]$. According to Definition 1, user j selects uniformly at random a message $W_j \in [M]$, and transmits the corresponding encoded codeword $f(W_j) = \mathbf{c}_j$. Due to symmetry, we assume without loss of generality that the first K codewords are transmitted. We shall consider two different decoders, which will lead to our two different achievability bounds:

Cover decoder: From the received sequence $\boldsymbol{Y}$, the decoder first discards all codewords from the codebook that are incompatible with the received sequence, i.e., those ones that are not covered by $\boldsymbol{Y}$. Then, the decoder outputs a list of K codewords chosen uniformly at random from the surviving codewords. Since the A-channel is noiseless, the list of surviving codewords always contains the transmitted list plus $N_{\mathrm{fa,c}} \in [0 : M-K]$ false alarms. Therefore, $P_{\mathrm{e}}^{(\mathrm{p})}$ can be upper-bounded by the PUPE achieved by this decoding rule, namely, $P_{\mathrm{e}}^{(\mathrm{p})} \le \mathbb{E}_{N_{\mathrm{fa,c}}}\left[\frac{N_{\mathrm{fa,c}}}{K+N_{\mathrm{fa,c}}}\right]$. Similarly, $P_{\mathrm{e}}^{(\mathrm{j})}$ can be upper-bounded by the probability of having at least one false alarm, i.e., $P_{\mathrm{e}}^{(\mathrm{j})} \le \mathbb{P}[N_{\mathrm{fa,c}} \ge 1]$.

Joint decoder: This decoder finds all combinations of K codewords from the codebook that can be selected to generate the output $\boldsymbol{Y}$. If there is more than one valid combination, the decoder chooses one, uniformly at random, and outputs the list of indices in that combination. Note, that this is exactly the maximum likelihood decoder. Since the A-channel is noiseless, the combination containing only the K transmitted codewords will always be valid. A wrong combination will differ from the correct one in $N_{\mathrm{fa,j}} = N_{\mathrm{md,j}}$ indices, i.e., same number of misdetections and false alarms. Hence, we can bound the error probability as $P_{\mathrm{e}}^{(\mathrm{p})} \le \mathbb{E}_{N_{\mathrm{fa,j}}}\left[\frac{N_{\mathrm{fa,j}}}{K}\right]$ and $P_{\mathrm{e}}^{(\mathrm{j})} \le \mathbb{P}[N_{\mathrm{fa,j}} \ge 1]$.

Remark 1: Recall that, in this paper, we assumed K to be known at the receiver. The cover decoder does not require this knowledge and works unaltered if K is unknown. The joint decoder can be adopted in two ways to deal with the missing information. One possibility is to extend the code design and use additional channel uses to estimate the number of users. Another way is to let the receiver find the smallest set of messages that recreate the channel output, as in the smallest satisfying set algorithm in group testing [24].

We are now ready to present our two achievability bounds.

Theorem 1 (Cover decoding): There exists an (M, n, ϵ)-code for the unsourced K-user A-channel with PUPE satisfying

$$\epsilon \le \sum_{\ell=1}^{K-1} \frac{\ell}{K+\ell} \mathbb{E}\left[\min\left\{1, \binom{M-K}{\ell} \prod_{k=1}^{K} \left(\frac{k}{q}\right)^{A_k \ell}\right\}\right] + \mathbb{E}\left[\min\left\{1, \binom{M-K}{K} \prod_{k=1}^{K} \left(\frac{k}{q}\right)^{A_k K}\right\}\right] + \frac{\binom{K}{2}}{M} \tag{5}$$

and there exists an (M, n, ϵ)-code with JPE satisfying

$$\epsilon \le \frac{\binom{K}{2}}{M} + \mathbb{E}\left[\min\left\{1, (M-K) \prod_{k=1}^{K} \left(\frac{k}{q}\right)^{A_k}\right\}\right]. \tag{6}$$

In both (5) and (6), A_k is the k-th element of $\boldsymbol{A} = [A_1, \ldots, A_K]^{\mathsf{T}}$, which is a multinomial-distributed random vector with n trials and K possible outcomes with probabilities $\{p_k\}_{k=1}^K$, which are given by

$$p_k = \frac{q! S(K,k)}{(q-k)! q^K} \tag{7}$$

where $S(K,k)$ denotes the Stirling number of the second kind [28, Sec. 26.8.6].

Proof: See Appendix A-B. ∎

Theorem 2 (Joint decoding): There exists an (M, n, ϵ)-code for the unsourced K-user A-channel with PUPE satisfying

$$\epsilon \le \frac{\binom{K}{2}}{M} + \sum_{\ell=1}^{K} \frac{\ell}{K} \mathbb{E}\left[\min\left\{1, \binom{K}{K-\ell}\binom{M-K}{\ell} \times \prod_{k=1}^{K} \left(\sum_{\eta=\underline{\eta}}^{\overline{\eta}} \bar{p}_\eta p(k,\ell,\eta)\right)^{A_k}\right\}\right] \tag{8}$$

and there exists an (M, n, ϵ)-code with JPE satisfying

$$\epsilon \le \frac{\binom{K}{2}}{M} + \sum_{\ell=1}^{K} \mathbb{E}\left[\min\left\{1, \binom{K}{K-\ell}\binom{M-K}{\ell} \times \prod_{k=1}^{K} \left(\sum_{\eta=\underline{\eta}}^{\overline{\eta}} \bar{p}_\eta p(k,\ell,\eta)\right)^{A_k}\right\}\right]. \tag{9}$$

In (8) and (9),

$$\bar{p}_\eta = \frac{k! S(K-\ell, \eta)}{(k-\eta)! k^{K-\ell} Z_\eta} \tag{10}$$

with Z_η being a normalizing constant ensuring that $\sum_{\eta=\underline{\eta}}^{\overline{\eta}} \bar{p}_\eta = 1$. Here $\underline{\eta} \triangleq \max\{0, k-\ell\}$ and $\overline{\eta} \triangleq \min\{k, K-\ell\}$. Finally

$$p(k,\ell,\eta) = \left(\frac{k}{q}\right)^{\ell} \pi(k,\ell,\eta) \tag{11}$$

where the first factor $(k/q)^\ell$ is the probability that the ℓ non-transmitted codewords hit one of the k output symbols, and $\pi(k,\ell,\eta)$ is the conditional probability that the ℓ non-transmitted codewords hit the remaining $k-\eta$ symbols given they all hit one of the k output symbols. Note that the

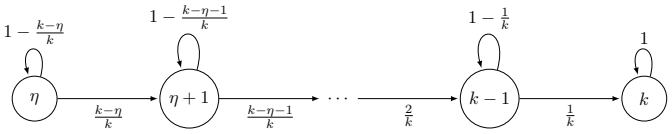


Fig. 1: Markov chain describing the state evolution yielding $\pi(k,\ell,\eta)$, which denotes the probability that ℓ non-transmitted codewords hit the remaining $k-\eta$ symbols of Y_i of cardinality k at channel use i, conditioned on the fact that the the ℓ non-transmitted codewords lie within the set of symbols in Y_i.

probability $\pi(k,\ell,\eta)$ resembles the classical coupon collector problem, which can be modelled by the Markov chain depicted in Fig. 1. Specifically, the problem is analogous to the coupon collector problem in the sense that $\pi(k,\ell,\eta)$ is the probability of collecting k out of k possible coupons in ℓ steps when starting with η coupons. The case $\eta=0$ can be computed in closed form as $\pi(k,\ell,0)=S(\ell,k)k!/k^\ell$. For $\eta>0$, $\pi(k,\ell,\eta)$ can be efficiently computed recursively. The specific formulas are given in Appendix B.

Proof: See Appendix A-C. ∎

B. Asymptotic Analysis

Let

$$\mu_\ell(K,q) \triangleq \mathbb{E}\left[\log \frac{P_{Y|\boldsymbol{X}_{[K]}}(Y|\boldsymbol{X}_{[K]})}{P_{Y|\boldsymbol{X}_{[K-\ell]}}(Y|\boldsymbol{X}_{[K-\ell]})}\right] = I(\boldsymbol{X}_{[K-\ell+1:K]};Y|\boldsymbol{X}_{[K-\ell]}), \tag{12}$$

$$\sigma^2_\ell(K,q) \triangleq \mathbb{V}\mathrm{ar}\left[\log \frac{P_{Y|\boldsymbol{X}_{[K]}}(Y|\boldsymbol{X}_{[K]})}{P_{Y|\boldsymbol{X}_{[K-\ell]}}(Y|\boldsymbol{X}_{[K-\ell]})}\right] \tag{13}$$

where $\boldsymbol{X}_S=(X_i)_{i\in S}$ for any $S\subset[K]$. We drop the explicit dependence on K,q for readability whenever it is clear from the context, so $\mu_\ell\equiv\mu_\ell(K,q)$. Since the channel is noiseless, $\mu_K=-\mathbb{E}[\log P_Y(Y)]$, i.e., the mutual information coincides with the output entropy, and $\sigma^2_\ell=-\mathbb{V}\mathrm{ar}[\log P_Y(Y)]$.[2] In the case $\ell=K$, each output sequence y with cardinality k has probability $P_{Y,|Y|}(y,k)=S(K,k)k!/q^K$. Since there are $\binom{q}{k}$ different outputs y for $|y|=k$,

$$\mu_K(K,q) = -\sum_{k=1}^{K}\sum_{y:|y|=k}\frac{S(K,k)k!}{q^K}\log\frac{S(K,k)k!}{q^K} \tag{14}$$

$$= -\sum_{k=1}^{K}\binom{q}{k}\frac{S(K,k)k!}{q^K}\log\frac{S(K,k)k!}{q^K} \tag{15}$$

$$= K\log q - \sum_{k=1}^{K} p_k \log(S(K,k)k!) \tag{16}$$

where in the last equality we used the definition of p_k in (7). The output entropy for the noiseless A-channel with uniform inputs (16) was already obtained in [1], [2]. By similar steps,

$$\sigma^2_K(K,q) = \sum_{k=1}^{K} p_k\left[\log\frac{S(K,k)k!}{q^K}\right]^2 - \mu_K(K,q)^2. \tag{17}$$

[2]When P_Y is the output distribution induced by a capacity achieving input distribution, μ_K is also the channel capacity.

Throughout the rest of this section, we will use

$$I_{K,q} \triangleq \frac{\mu_K(K,q)}{K\log q}, \tag{18}$$

$$V_{K,q} \triangleq \frac{\sigma^2_K(K,q)}{(K\log q)^2}. \tag{19}$$

Recall that in Theorem 2, A_k is the k-th entry of $\boldsymbol{A}=[A_1,\ldots,A_k]^\mathsf{T}$, which is multinomial distributed with parameters $\{p_k\}_{k=1}^K$ (with p_k given in (7)) and $\sum_{k=1}^K A_k=n$. Let $c_k \triangleq \log_2(\sum_{\underline{\eta}}^{\overline{\eta}}\bar{p}_\eta p(k,\ell,\eta))$. It follows that $\sum_{k=1}^K A_k c_k \overset{d}{=} \sum_{i=1}^n Z_i$, where $\overset{d}{=}$ denotes equality in distribution, and where $\{Z_i\}_{i=1}^n$ is a sequence of i.i.d. random variables taking values on c_k with probability p_k for $k\in[K]$. In the following, a generic realization of the random variable Z_i will be denoted simply by Z. Then, by applying the so-called normal approximation (Berry-Esseen theorem [29, Ch. XVI.5] and [30, Lemma 47]) to the expected value of (8), it follows that for some constant B independent of n (see, e.g., [30, Eqs. (255)-(267)]),

$$\epsilon \le \sum_{\ell=1}^{K}\frac{\ell}{K}Q\left(\frac{\frac{\mu_\ell}{\ell}-R\log_2 q-\frac{\log_2\left(\left(\frac{e}{\ell}\right)^\ell\binom{K}{K-\ell}\right)}{\ell n}}{\sigma_\ell/(\ell\sqrt{n})}\right) + \frac{B}{\sqrt{n}}+\frac{\binom{K}{2}}{M}. \tag{20}$$

It is shown in Appendix C that basic properties of the conditional mutual information and the symmetry of the X_i's imply

$$\frac{\mu_\ell}{\ell} \ge \frac{\mu_{\ell+1}}{\ell+1} \tag{21}$$

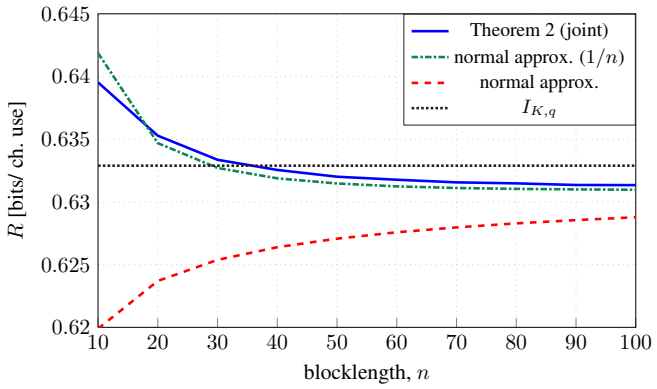
for every $\ell\in[K-1]$. Then, as n grows and the rate approaches $I_{K,q}$, the $\ell=K$ term in (8) becomes dominant while the $\ell<K$ terms still decay exponentially fast with n.

Remark 2: Usually, the capacity region of the multiple access channel is the union of K-dimensional pentagon constrained by the different conditional mutual information terms μ_ℓ. In the unsourced case, where all input distributions are constrained to be equal, it is apparent from (21) that μ_K is the most constraining limit and therefore it dominates the $n\to\infty$ limit. Formula (20) shows that the conditional mutual information terms still influence the random coding error probabilities in the finite blocklength regime. Nonetheless, their contribution vanishes exponentially with the blocklength.

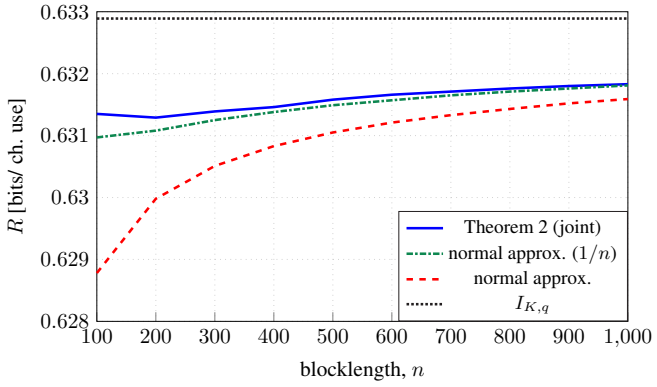
By collecting the $\ell<K$ terms and $\binom{K}{2}/M$ in (20) in a $o(1/n)$ term, after some standard manipulations, (20) can be expressed in terms of the rate as

$$R = I_{K,q} - \sqrt{\frac{V_{K,q}}{n}}Q^{-1}\left(\epsilon-\frac{B}{\sqrt{n}}+o\left(\frac{1}{n}\right)\right)+\frac{\log_2(K/e)}{n\log_2(q)}. \tag{22}$$

The constant B is determined by the Berry-Esseen theorem [29, Ch. XVI.5] and [30, Lemma 47]. For sufficiently large n, it follows that $Q^{-1}(\epsilon-B/\sqrt{n}+o(n^{-1}))=Q^{-1}(\epsilon)+\overline{B}/\sqrt{n}+\mathcal{O}(1/n)$, for some $\overline{B}$ independent of n. Numerical



(a) $10 \leq n \leq 100$.



(b) $100 \leq n \leq 1000$.

Fig. 2: Rate versus blocklength n for $q = 16$, $K = 5$ and $\epsilon = 0.05$.

experiments suggest that for the A-channel, the value of B that can be obtained by applying the Berry-Esseen theorem [29, Ch. XVI.5] and [30, Lemma 47] is not tight. In other words, $\sum_{i=1}^{n} Z_i$ converges much faster to the Gaussian distribution than the speed suggested by B. In Fig. 2, we show that the approximation

$$R = I_{K,q} - \sqrt{\frac{V_{K,q}}{n}} Q^{-1}(\epsilon) + \frac{\log_2(K/e)}{n \log_2(q)} \tag{23}$$

can indeed provide accurate estimates of the bound provided in Theorem 2 for small values of n. This approximation is tight as long as the true value of B is sufficiently small so that the resulting $\overline{B}$ is much smaller than $\log_2(K/e)/\log_2(q)$. When this is true, ignoring the term $\overline{B}/\sqrt{n}$ does not compromise the accuracy of the approximation for small n.

This is shown in Fig. 2, where we compare the non-asymptotic random coding bound with joint decoding given in Theorem 2, and the normal approximation (23) with and without the $1/n$-term. We further plot the maximum coding rate achievabile with uniform inputs $I_{K,q}$. We can observe that the $1/n$-term of the normal approximation is necessary to capture the behaviour of the non-asymptotic bound in the small blocklength regime, where rates are higher than $I_{K,q}$ (Fig. 2a). As n grows large the dispersion term becomes dominant and the achievability curve starts to show the typical $1/\sqrt{n}$ convergence to the asymptotic limit from below (Fig. 2b).

III. A-Channel Code: Tree Code

A. Code Construction

A B-bit message is divided into blocks of size $\{b_i\}_{i=1}^{n}$ such that $\sum_{i=1}^{n} b_i = B$ and such that $b_1 = J$ and $b_i < J$ for all $i \in [2 : n]$. Each subblock $i \in [2 : n]$ is augmented to size J by appending $\pi_i = J - b_i$ parity bits, obtained using pseudo-random linear combinations of the information bits of the previous blocks $i' < i$. Note that there is a one-to-one association between the set of all sequences of coded blocks and the paths of a tree of depth n. The pseudo-random parity-check equations generating the parity bits are identical for all users, i.e., each user makes use exactly of the same outer *tree code*. This makes the code compatible with the unsourced paradigm. Each user then transmits the n coded symbols over the 2^J-ary A-channel.

Let Y_i, $i \in [n]$, be the channel outputs of the A-channel. Since the sections contain parity bits with parity profile $\{0, \pi_2, \ldots, \pi_n\}$, not all message sequences in $Y_1 \times Y_2 \times \cdots \times Y_n$ are possible. The role of the outer decoder is to identify all possible message sequences, i.e., those corresponding to paths in the tree of the outer tree code [13]. The output list $\mathcal{L}$ is initialized as an empty list. Starting from $i = 1$ and proceeding in order, the decoder converts all the integer indices in Y_i back to their binary representation, separates data and parity bits, computes the parity checks for all the combinations with messages from the list $\mathcal{L}$, and extends only the paths in the tree which fulfill the parity checks. A precise analysis of the error probability in various asymptotic regimes as well as an algorithm to optimize the parity profile for a target complexity and error probability are provided in [13].

The analysis in [13] and [4] showed that the tree code performs well in the regime of vanishing sparsity, i.e., $K/q \to 0$, which is the regime where both joint and cover decoding bounds (see Theorems 1 and 2) perform similarly. However, for moderate sparsity, our numerical evaluation of Theorems 1 and 2 reveals that the joint and cover decoding bounds exhibit a considerable gap (See Fig. 3). Since the original tree decoder outputs all codewords that satisfy the parity checks, the tree code described above cannot outperform the cover decoding bound. In the next section, we propose enhanced decoding strategies for the original tree code based on ideas from group testing and insights from the analysis of the joint decoder.

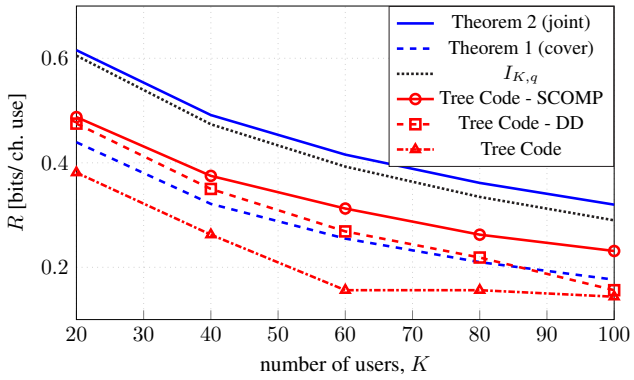
B. Enhanced Decoding

The proof of Theorem 2 shows that joint decoding can improve upon cover decoding by considering combinations of codewords instead of just individual codewords. In this section, we use this concept to develop two improved decoding algorithms for the tree code. These methods strictly improve the performance of the tree code since they consists of a post-processing step of the output list when the output list is greater than K. In earlier works such as [14] and [4], codewords

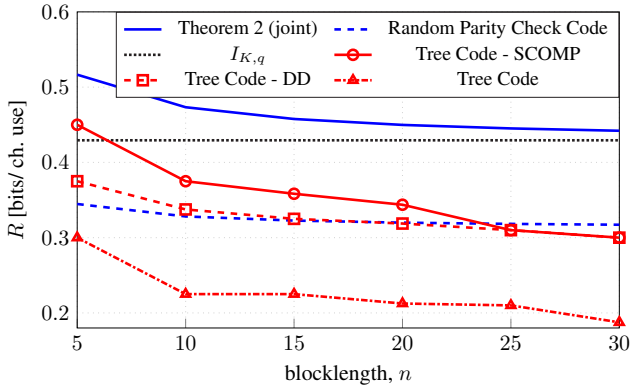
were discarded at random to reduce the output list to the required size. This necessarily results in a large number of errors when the output list is significantly larger than K. Since the output list contains only false alarms and no misdetections, the decoding performance can be improved by filtering the output list to remove false alarms. Let the size of the cover decoder output list be $K+\Delta$. A valid strategy is to check all $\binom{K+\Delta}{K}$ combinations of K codewords from the list. Of course this leads to a complexity that grows exponentially in Δ. When K is not known, one can search for the combination with the least codewords that produces the channel output. In the group-testing literature, this approach is called the smallest-satisfying set (SSS) method [24]. Note that finding the SSS is in general NP hard, as it can be shown to be equivalent to the set cover problem [24, Remark 2.1]. In the following we describe two methods, developed for group testing, that approximate the combinatorial search in a greedy manner. In particular, we will consider the so-called definitive defectives (DD) and sequential combinatorial pursuit (SCOMP) algorithms [31].[3] They both work by filtering the original output list. Specifically, SCOMP is a strict improvement over DD, in the sense that it consists of applying DD followed by an additional processing step. Therefore, the algorithm can be chosen based on complexity and/or rate requirements, since each processing step increases the decoding complexity, but also increases the performance.

DD : As a first step we re-encode all the messages in the output list of the tree decoder, which we denote by $\mathbf{m}_1, ..., \mathbf{m}_{|\mathcal{L}|}$. For $i \in [n]$, the DD algorithm isolates all indices i for which $m_{j,i}$ is unique among $\{m_{1,i}, ..., m_{|\mathcal{L}|,i}\}$. The messages with indices isolated this way have for sure been transmitted since they were the only ones in the list that can explain the observed channel output. Let $\mathcal{L}_{\text{DD}}$ be the list of isolated messages and let $\mathcal{L}_R$ denote the remaining messages that were not isolated. If $|\mathcal{L}_{\text{DD}}| < K$ we choose random messages from $\mathcal{L}_R$ to fill the output list up to size K.

SCOMP : The SCOMP algorithm proceeds by scanning the list of remaining entries $\mathcal{L}_R$ after DD processing for appropriate candidates using the following greedy heuristic: i) The symbols in the channel output that have been covered by the DD list are removed. The remaining symbols are called *unexplained*. ii) The index $j_{\max}$ is searched for which $\mathbf{m}_j$ covers the *most* unexplained symbols. This index is added to the output list $\mathcal{L}_{\text{SCOMP}} = \mathcal{L}_{\text{DD}} \cup j_{\max}$. iii) The symbols covered by $\mathbf{m}_{j_{\max}}$ are removed from the list of unexplained symbols. The algorithm repeats this process until no unexplained symbols are left. If $|\mathcal{L}_{\text{SCOMP}}| < K$ we again add messages at random. This algorithm will always terminate in at most K steps, since the transmitted messages are always contained in the original output list.



(a) $J = 8$, $n = 20$, $\epsilon = 0.05$



(b) $J = 8$, $K = 50$, $\epsilon = 0.05$

Fig. 3: Rate versus number of active users K (Fig. 3a) and versus blocklength n (Fig. 3b) including the tree code with DD and SCOMP post-processing.

C. Numerical Results

In Fig. 3, we compare the performance of the original tree code described in Section III-A with the enhanced versions described in Section III-B. As performance benchmarks, we use the finite-blocklength bounds derived in Theorems 1 (cover decoding) and 2 (joint decoding), and the maximum coding rate achievable asymptotically by uniform inputs $I_{K,q}$ (18). We use $q = 2^J$ with $J = 8$. Let the rate $R = B/(B+P)$, where B denotes the number of information bits, and P the number of parity check bits. We fix the error constraint $\epsilon \leq 0.05$, and select the largest rate R (smallest value of P) such that the error constraint is satisfied. The parity profile is set by choosing $\pi_n = J$ and dividing the remaining parity check bits evenly between sections $2, ..., n$. If the remaining parity check bits cannot be divided evenly, the later sections are prioritized. We remark that the resulting parity profile provides a good balance between decoding complexity and error probability.

We can observe that there is a considerable gap between joint and cover decoding. Furthermore, we can observe that

[3] An alternative approach is based on linear programming [32]. It is very similar to SCOMP in terms of achievable rates and complexity, so we exclude it from the comparison in this paper. A more detailed comparison is left for future work.

$I_{K,q}$ is exceeded for small blocklengths as are the achievable rates of all tree code variants. We can also observe that the suggested group-testing-motivated post-processing strategies (Tree code - DD, Tree code - SCOMP) allow to increase the achievable rates of the tree code significantly. Remarkably, both DD and SCOMP post-processing strategies allow to outperform the cover decoding bound.

IV. A-Channel Designs in Group Testing

Recall from Section I that an unsourced q-ary A-channel code of blocklength n and size M can be thought of as a group-testing matrix for $N = M$ items with $T = nq$ tests. Here, the number of active users is analog to the number of defective items. The tests are divided into n groups of size q so that each item participates in exactly n tests, i.e., in one test per group. Even though A-channel-based group-testing constructions are less flexible (they require the number of tests T to be a multiple of q), they also provide more structure, which allows for efficient recovery and an easier analysis.

The finite-blocklength achievability bounds given in Theorems 1 and 2 allow to compute concrete achievable test numbers for a fixed q and a fixed error probability ϵ. In particular, q can be seen as an optimization parameter that can be chosen to minimize the number of required tests. The analogy between unsourced A-channel codes and group testing motivates the following results.

Corollary 1: There exist group-testing matrices, constructed from unsourced A-channel codes, such that d defective items out of N items can be recovered with $T = nq$ tests and the error probability given by Theorems 1 and 2 without the penalty term $\binom{K}{2}/M$ (since random collisions among items are not possible).

The following theorem shows that it is possible to achieve the optimal number of tests $T = \mathcal{O}(d \log N)$.

Theorem 3: There exists a sequence of group-testing matrices, constructed from unsourced A-channel codes, such that d defective items out of N items can be recovered with an error probability that vanishes in the limit $d, N, T \to \infty$ if

$$T = d \log N. \tag{24}$$

Proof: Let $K, q \to \infty$ with $\lambda = K/q$ fixed. The mutual information for the K-user A-channel with uniform inputs in this limit is given by [2]:

$$\lim_{K,q\to\infty} \frac{I(\boldsymbol{X}_{[K]}; Y)}{q} = h(1 - e^{-\lambda}) \tag{25}$$

where $h(\cdot)$ is the binary entropy function. By the channel coding theorem [33, Ch. 7.7], there exist codes with sum-rates $R_{\text{sum}} = K \log M/(nq)$ for which the error probability vanishes as long as $R_{\text{sum}} < h(1 - e^{-\lambda})$. The right hand side is maximised for $\lambda = \ln 2$. Assuming that a code achieving this performance is used, we obtain (24) by replacing $n = K \log M/q$ in $T = nq$, and using that in the standard group-testing notation $M = N$. ∎

If the optimal sparsity $\lambda = \ln 2$ cannot be attained, the number of required tests becomes $T = h(1 - e^{-\lambda})^{-1} d \log N$. This result lies in the realm of probabilistic group testing [24] as for finite values of N and T, there is always a non-zero chance of failure, albeit it can be made arbitrary small by increasing N and T. Note also that the relative scaling of N and d is not specified in Theorem 3. It is implicitly assumed though, through the order of limits (first $N, T \to \infty$ then $d, q \to \infty$), that N is much larger than d. Instead of taking the second limit, we can generalize (24) to hold for all d by setting

$$T = d \log N \min_q \frac{q}{\mu_d(d, q)} \tag{26}$$

where $\mu_d(d, q)$ is given in (16).

It is known that group-testing designs with a constant number w of tests per item perform better than unstructured random designs, even when the average number of test per item is the same [25]. A commonly analyzed setup consists in fixing w and choosing the tests randomly from the set of all w-weight vectors. An A-channel design also has a fixed number of test per item $w = T/q$ but has even more structure, which provides some advantages. In particular, the q-ary structure of the A-channel allows to represent the group-testing matrix in an efficient way using only $n \log q$ bits to specify the test in which each item participates. If a structured code is used, such as the tree code, the group-testing matrix does not need to be stored explicitly as each column can be constructed in $\mathcal{O}(\log N)$ time. Furthermore, the reconstruction of the defective set can be done in $\mathcal{O}(d^2 \log N)$ time. As such, the tree code falls into the category of sub-linear group-testing designs [24]. They are especially useful in problems where the recovery time is the limiting factor, rather than the acquisition of tests. This is the case, for example, in big data and computer science applications. Theorem 3 shows the existence of A-channel codes achieving the optimal test scaling, but it requires q to scale proportional to d. The analysis of the tree code in such a scaling regime is an interesting open problem, which is left for future work.

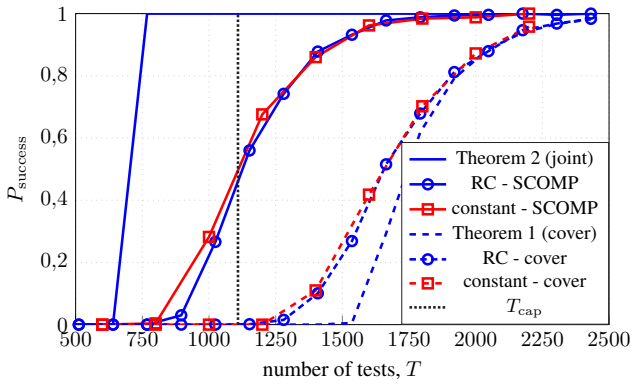


Fig. 4: P_{success} versus number of tests T in a group-testing setup with $d = 100$ and $N = 2000$. In the A-channel designs, $q = 2^7$.

A. Numerical Results

Fig. 4 shows the performance of Theorems 1 and 2 (without the $\binom{K}{2}/M$ term) in the group-testing setup in terms of

probability of success $P_{\text{success}} = 1 - P_e^{(\text{j})}$, where a success is declared if the set of defective items is perfectly recovered.[4] We compare our achievability bounds with empirical error rates achieved by a random A-channel code under cover (RC - cover) and SCOMP (RC - SCOMP) decoding. The black dotted line shows $T_{\text{cap}} \triangleq d \log Nq/\mu_d(d,q)$ for $q = 2^7$, which provides an asymptotic achievability bound since for $T > T_{\text{cap}}$, by the channel coding theorem, there exist A-channel codes achieving $P_{\text{success}} \to 1$ in the limit $N, T \to \infty$ with $(\log N)/T$ fixed. We also compare our bounds with a constant design (constant - cover; constant - SCOMP in Fig. 4) with exactly $w = (\ln 2)T/d$ test per item given in [24].

We assume that d out of N items are defective and set $d = 100$ and $N = 2000$. We choose $q = 2^7$, which was found empirically to give the best results. As we can observe, the A-channel design, which has $w = T/q$, exhibits almost the same performance as the constant weight designs, when using both the cover and the SCOMP decoders.

V. Conclusions

We present finite-blocklength achievability bounds for the *unsourced* A-channel, and we propose easy-to-evaluate refined asymptotic approximations, which are accurate from block-lengths as small as $n = 10$. Motivated by the analytical solution of the finite-blocklength bounds and the connection between URA and group testing through the unsourced A-channel, we introduce improved decoding algorithms of the so-called tree codes used as part of coding schemes for URA. We show that the proposed decoding algorithms allow to improve the rates achieved by off-the-shelf tree codes significantly at the cost of a moderate increase in decoding complexity. Finally, we adapt our finite-blocklength bounds so that they can be compared against well-known group-testing bounds and schemes. We show that A-channel constructions can perform close to constant tests-per-item constructions, albeit with a much more structured test matrix, which can enable its use in applications such as big data and computer science that usually demand stringent recovery times. For example, A-channel tree-codes test-matrices can be constructed in $\mathcal{O}(\log n)$ time, and the defective set can be reconstructed in $\mathcal{O}(d^2 \log n)$ time.

Appendix A
Proofs of Achievability bounds

A. Preliminaries

In both error definitions (3) and (4), we assumed that any collision among the transmitted codewords automatically results in error. It follows that

$$\mathbb{P}[\cup_{j\neq i}\{W_j = W_i\}] \leq \frac{\binom{K}{2}}{M}. \tag{27}$$

We shall replace the measure under which (3) and (4) are computed by the one under which $\{W_j\}_{j=1}^K$ are uniformly sampled without replacement from $[M]$, at the expense of adding a penalty term equal to $\binom{K}{2}/M$ to the upper bounds on the error probability.

Due to symmetry, we assume without loss of generality that the first K codewords are transmitted. For any set $S \in [M]$, let $\mathbf{c}(S) \triangleq \bigcup_{j\in S} \mathbf{c}_j$. Similarly, for any set $S \in [M]$, we shall use $c_i(S)$ to denote $\bigcup_{j\in S} c_{i,j}$, where $c_{i,j}$ indicates the input of $\mathbf{c}_j$ at channel use i. We shall omit the subindeces i and j when immaterial. Furthermore, we let S_ℓ denote a generic subset of $K - \ell$ elements in $[K]$, and S'_ℓ denote a generic subset of ℓ elements in $[M]\backslash[K]$.

Finally, the following definition will turn out useful throughout the proofs. Let $\mathcal{A}_k \triangleq \{i \in [n] : |Y_i| = k\}$ for $k \in [K]$. In words, $\mathcal{A}_k$ is the set of channel uses where the channel output $\boldsymbol{Y}$ has cardinality k. Note that $\sum_{k=1}^K |\mathcal{A}_k| = n$. Hence, A_k is the k-th element of $\boldsymbol{A} = [A_1, \ldots, A_K]^{\mathsf{T}}$, which is a multinomial-distributed random vector with parameters n and $\{p_k\}_{k=1}^K$, where n denotes the number of trials, K the number of possible outcomes in each trial, and p_k the probability that the cardinality of the output is k at channel use i, which is given by

$$p_k = \mathbb{P}[|Y_i| = k] = \frac{q! S(K,k)}{q^K (q-k)!}. \tag{28}$$

B. Proof of Theorem 1

It follows that

$$P_{\text{e}}^{(\text{p})} \leq \mathbb{E}\left[\frac{N_{\text{fa,c}}}{K + N_{\text{fa,c}}}\right] + \frac{\binom{K}{2}}{M} \tag{29}$$

$$= \sum_{\ell=1}^{M-K} \frac{\ell}{K+\ell}\mathbb{P}[N_{\text{fa,c}} = \ell] + \frac{\binom{K}{2}}{M} \tag{30}$$

$$\leq \sum_{\ell=1}^{K-1} \frac{\ell}{K+\ell}\mathbb{P}[N_{\text{fa,c}} \geq \ell] + \mathbb{P}[N_{\text{fa,c}} \geq K] + \frac{\binom{K}{2}}{M}. \tag{31}$$

Hence, to complete the proof of Theorem 1, we next show that

$$\mathbb{P}[N_{\text{fa,c}} \geq \ell] \leq \mathbb{E}\left[\min\left\{1, \binom{M-K}{\ell}\prod_{k=1}^K \left(\frac{k}{q}\right)^{\ell A_k}\right\}\right]. \tag{32}$$

Since the messages are independent and uniform on $[M]$ (see Def. 1), it follows that

$$\mathbb{P}[N_{\text{fa,c}} \geq \ell] = \mathbb{P}\left[\bigcup_{S'_\ell} \mathbf{c}(S'_\ell) \in \boldsymbol{Y}\right] \tag{33}$$

$$= \mathbb{P}\left[\bigcup_{S'_\ell} \bigcap_{k\in[K]} \bigcap_{i\in\mathcal{A}_k} \{c_i(S'_\ell) \in Y_i\}\right]. \tag{34}$$

[4] This corresponds to probabilistic group testing. For the PUPE bounds in Theorems 1 and 2, this would correspond to partial recovery [24][Ch. 5.1].

We next use that $\{\mathcal{A}_k\}_{k=1}^K$ are disjoint sets together with the law of total probability to write

$$\begin{aligned}
&\mathbb{P}[N_{\text{fa,c}} \geq \ell] \\
&\quad = \mathbb{E}_{\boldsymbol{A}}\left[\mathbb{P}\left[\bigcup_{S'_\ell}\bigcap_{k\in[K]}\bigcap_{i\in\mathcal{A}_k}\{c_i(S'_\ell)\in Y_i\} \;\middle|\; |\mathcal{A}_k| = A_k\right]\right] && (35)\\
&\quad \leq \mathbb{E}_{\boldsymbol{A}}\left[\min\left\{1, \binom{M-K}{\ell}\prod_{k=1}^K (\mathbb{P}[c(S'_\ell)\in Y])^{A_k}\right\}\right] && (36)\\
&\quad = \mathbb{E}_{\boldsymbol{A}}\left[\min\left\{1, \binom{M-K}{\ell}\prod_{k=1}^K \left(\frac{k}{q}\right)^{\ell A_k}\right\}\right] && (37)
\end{aligned}$$

where the first inequality follows from the union bound, because the messages are independent and uniform on $[M]$ (see Def. 1), and because the probability that $c_i(S'_\ell) \in Y_i$ is independent of i, which also justifies why we omitted the subscript i. Finally, (37) follows since

$$\mathbb{P}[c(S'_\ell)\in Y] = \mathbb{P}\left[\bigcap_{j\in S'_\ell}\{c_j \in Y\}\right] = (\mathbb{P}[\bar{c}\in Y])^\ell \qquad (38)$$

for some generic non-transmitted symbol $\bar{c}$, and because $\mathbb{P}[\bar{c}\in Y] = k/q$.

C. Proof of Theorem 2

It follows that

$$\mathbb{P}[N_{\text{fa,j}} = \ell] \leq \mathbb{P}\left[\bigcup_{S_\ell}\bigcup_{S'_\ell}\{\mathbf{c}(S_\ell)\cup\mathbf{c}(S'_\ell) = \boldsymbol{Y}\}\right] + \frac{\binom{K}{2}}{M} \qquad (39)$$

and

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{S_\ell}\bigcup_{S'_\ell}\{\mathbf{c}(S_\ell)\cup\mathbf{c}(S'_\ell) = \boldsymbol{Y}\}\right] \\
&\quad = \mathbb{E}_{\boldsymbol{A}}\Bigg[\mathbb{P}\Bigg[\bigcup_{S_\ell}\bigcup_{S'_\ell}\bigcap_{k=1}^K\bigcap_{i\in\mathcal{A}_k} \\
&\qquad\qquad \{c_i(S_\ell)\cup c_i(S'_\ell) = \boldsymbol{Y}\} \;\Bigg|\; |\mathcal{A}_k| = A_k\Bigg]\Bigg]. \qquad (40)
\end{aligned}$$

Since the messages are independent and uniform on $[M]$ (see Def. 1), by applying the union bound on the right-hand side of (40), we have

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{S_\ell}\bigcup_{S'_\ell}\{\mathbf{c}(S_\ell)\cup\mathbf{c}(S'_\ell) = \boldsymbol{Y}\}\right] \\
&\quad \leq \mathbb{E}\Bigg[\min\Bigg\{1, \binom{K}{K-\ell}\binom{M-K}{\ell} \\
&\qquad \times\mathbb{P}\left[\bigcap_{k=1}^K\bigcap_{i\in\mathcal{A}_k}\{c_i(S_\ell)\cup c_i(S'_\ell) = Y_i\} \;\middle|\; |\mathcal{A}_k| = A_k\right]\Bigg\}\Bigg] && (41)\\
&\quad = \mathbb{E}\Bigg[\min\Bigg\{1, \binom{K}{K-\ell}\binom{M-K}{\ell} \\
&\qquad\qquad \times\prod_{k=1}^K (\mathbb{P}[\{c(S_\ell)\cup c(S'_\ell) = Y\}])^{A_k}\Bigg\}\Bigg] && (42)
\end{aligned}$$

where the last equality follows because $\{\mathcal{A}_k\}_{k=1}^K$ are disjoint sets together with the law of total probability, and because the considered input distribution is a product distribution. We conclude the proof by showing that

$$\mathbb{P}[\{c(S_\ell)\cup c(S'_\ell) = Y\}] = \sum_{\eta=\underline{\eta}}^{\overline{\eta}} \bar{p}_\eta p(k,\ell,\eta). \qquad (43)$$

Recall that, in the statement of Theorem 2, we defined $\eta = |c(S_\ell)|$, i.e., the cardinality of the subset of transmitted codewords $c(S_\ell)$ at a given channel use. Furthermore, we defined $\underline{\eta} = \max\{0, k-\ell\}$ and $\overline{\eta} = \min\{k, K-\ell\}$, where $\ell \in [K]$ denotes the number of elements from the subset of non-transmitted codewords $c(S'_\ell)$. Thus, $K-\ell$ corresponds to the number of elements from the subset of transmitted codewords $c(S_\ell)$. In words, $\underline{\eta}$ represents the minimum number of symbols in channel uses of cardinality k that that need to be covered by the subset of $K-\ell$ transmitted symbols to create a valid output together with the symbols of the subset of ℓ non-transmitted codewords. Similarly, $\overline{\eta}$ represents the maximum number of symbols that the subset of $K-\ell$ transmitted codewords could hit in channel uses of cardinality k, when we consider ℓ non-transmitted codewords. The probability term in (42) can be expressed as

$$\begin{aligned}
&\mathbb{P}[\{c(S_\ell)\cup c(S'_\ell)\} = Y] \\
&\quad = \sum_{\eta=\underline{\eta}}^{\overline{\eta}} \mathbb{P}\big[\{c(S_\ell)\cup c(S'_\ell)\} = Y \;\big|\; |c(S_\ell)| = \eta\big] \\
&\qquad\qquad \times\mathbb{P}[|c(S_\ell)| = \eta] && (44)\\
&\quad = \sum_{\eta=\underline{\eta}}^{\overline{\eta}} \bar{p}_\eta p(k,\ell,\eta). && (45)
\end{aligned}$$

where $p(k,\ell,\eta) = \mathbb{P}\big[\{c(S_\ell)\cup c(S'_\ell) = Y\} \;\big|\; |c(S_\ell)| = \eta\big]$, and

$$\bar{p}_\eta = \mathbb{P}[|c(S_\ell)| = \eta] = \frac{k!S(K-\ell,\eta)}{(k-\eta)!k^{K-\ell}Z_\eta} \qquad (46)$$

with Z_η being a normalizing constant used to make sure that $\sum_{\eta=\underline{\eta}}^{\overline{\eta}} \bar{p}_\eta = 1$. Note that $\bar{p}_\eta$ is similar to p_k in (28), except that in $\bar{p}_\eta$ not all the values of $\eta \in [K-\ell]$ are possible, since we are considering channel uses of cardinality k, and the number of symbols hit by the subset of transmitted codewords needs to be sufficiently large so that the subset of ℓ non-transmitted codewords can hit the remaining symbols. Also, η cannot be larger than the cardinality k. This implies that without Z_η, $\sum_{\eta=\underline{\eta}}^{\overline{\eta}} \bar{p}_\eta$ could be different from one.

Given the definition of $p(k,\ell,\eta)$ in (11), it remains to show how to compute $\pi(k,\ell,\eta)$, which can be computed in closed form only when $\eta = 0$ (see Theorem 2). In Appendix B, we present a possible way to compute $\pi(k,\ell,\eta)$ for $\eta > 0$.

Appendix B
Computation of $\pi(k,\ell,\eta)$

Recall that $\pi(k,\ell,\eta)$ denotes the conditional probability, given that $c(S'_\ell) \in Y$, that the subset of non-transmitted symbols cover the remaining $k-\eta$ symbols. This problem resembles the classical coupon collector problem where $\pi(k,\ell,\eta)$ is exactly the probability to draw k out of k coupons in ℓ steps when one starts with η coupons and each coupon appears with probability $1/k$. The evolution of coupons can be modeled by the Markov Chain depicted in Fig. 1. The inter-arrival times in this chain are independent geometrically distributed random variables. The probability generating function of the final arrival time can be expressed as

$$G_{k,\ell,\eta}(z) = z^{k-\eta}\frac{(k-\eta)!}{k^{k-\eta}}\prod_{i=\eta}^{k}\frac{1}{1-\frac{i}{k}z}. \tag{47}$$

Finally, $\pi(k,\ell,\eta)$ can be obtained as the sum of the first ℓ coefficients of the polynomial representation of $G_{k,\ell,\eta}(z)$. These terms can be calculated recursively to avoid numerical issues. Since $G_{k,\ell,k}(z) = 1$ we have $\pi(k,\ell,k) = 1$. Then

$$G_{k,\ell,\eta-1}(z) = z\frac{k-\eta}{k}\left(1-\frac{\eta}{k}z\right)^{-1}G_{k,\ell,\eta}(z) \tag{48}$$

$$= z\frac{k-\eta}{k}\sum_{i=0}^{\infty}\left(\frac{\eta}{k}z\right)^{i}G_{k,\ell,\eta}(z). \tag{49}$$

Therefore, the polynomial representation of $G_{k,\ell,\eta-1}(z)$ can be computed from $G_{k,\ell,\eta}(z)$ by convolution with the polynomial $\sum_{i=0}^{\infty}\left(\frac{\eta}{k}z\right)^i$. Note that only the first ℓ coefficients of $G_{k,\ell,\eta}(z)$ are relevant, so it suffices to compute the convolution with $\sum_{i=0}^{\ell}\left(\frac{\eta}{k}z\right)^i$.

Appendix C
Proof of (21)

By symmetry we write $\mu_\ell = I(\boldsymbol{X}_{[\ell]}; Y|\boldsymbol{X}_{[\ell+1:K]})$. We next show that

$$\frac{\mu_\ell}{\ell} \geq \frac{\mu_K}{K} \tag{50}$$

for every $\ell \leq K$, which implies (21).

First, note that, since all X_i are iid, it holds that $I(\boldsymbol{X}_S; Y|\boldsymbol{X}_{S'}) \leq I(\boldsymbol{X}_S; Y|\boldsymbol{X}_{S''})$ for $S, S', S'' \subset [K]$ whenever $S \cap S' = S \cap S'' = \emptyset$ and $S' \subset S''$ which follows from $I(X_2;Y) \leq I(X_2;Y|X_1)$ for independent X_1, X_2. In other words, for independent random variables, conditioning increases mutual information. The latter follows from the convexity of $I(X_2;Y)$ in $p(y|x_2) = \sum_{x_1} p(y|x_1,x_2)p(x_1)$. Second, again, due to the iid property, the elements of $\boldsymbol{X}$ can be arbitrary permuted. With these two properties and repeated use of the chain rule for mutual information we can show that $\frac{\mu_\ell}{\ell} \geq \frac{\mu_{\ell+1}}{\ell+1}$:

$$\begin{aligned}\ell\mu_{\ell+1} &= \ell I(\boldsymbol{X}_{[\ell]}; Y|\boldsymbol{X}_{[\ell+1:K]}) + \ell I(X_{\ell+1}; Y|\boldsymbol{X}_{[\ell+2:K]}) \\ &= \ell\mu_\ell + \ell I(X_1; Y|\boldsymbol{X}_{[\ell+2:K]}).\end{aligned} \tag{51}$$

By the chain rule μ_ℓ can be expressed as

$$\mu_\ell = \sum_{i=1}^{\ell} I(X_i; Y|\boldsymbol{X}_{[i+1:K]}). \tag{52}$$

It is apparent that the righ-hand side of (51) can be upper bound by μ_ℓ by conditioning on additional X_i's, which shows that $\ell\mu_{\ell+1} \leq (\ell+1)\mu_\ell$.

References

[1] S.-C. Chang and J. Wolf, "On the T-user M-frequency noiseless multiple-access channel with and without intensity information," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 41–48, Jan. 1981.

[2] L. Bassalygo and M. Pinsker, *Calculation of the Asymptotically Optimal Capacity of a T-User M-Frequency Noiseless Multiple-Access Channel*. Boston, MA: Springer US, 2000, pp. 177–180. [Online]. Available: https://doi.org/10.1007/978-1-4757-6048-4_16

[3] L. A. Bassalygo and V. V. Rykov, "Multiple-access hyperchannel," *Problems of Information Transmission*, vol. 49, no. 4, pp. 299–307, Oct. 2013.

[4] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6894–6915, May 2021.

[5] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2523–2527.

[6] I. Zadik, Y. Polyanskiy, and C. Thrampoulidis, "Improved bounds on Gaussian MAC and sparse regression via Gaussian inequalities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 430–434.

[7] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 5853–5884, Jun. 2021.

[8] K.-H. Ngo, A. Lancho, G. Durisi, and A. Graell i Amat, "Unsourced multiple access with random user activity," Feb. 2022. [Online]. Available: https://arxiv.org/abs/2202.06365

[9] J. Ravi and T. Koch, "Scaling laws for Gaussian random many-access channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2429–2459, Apr. 2022.

[10] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "LDPC codes with soft interference cancellation for uncoordinated unsourced multiple access," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021.

[11] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "Unsourced random access with coded compressed sensing: Integrating AMP and belief propagation," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2384–2409, Apr. 2022.

[12] D. Truhachev, M. Bashir, A. Karami, and E. Nassaji, "Low-complexity coding and spreading for unsourced random access," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 774–778, Mar. 2021.

[13] V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6509–6533, Jul. 2020.

[14] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021.

[15] K. Andreev, P. Rybin, and A. Frolov, "Reed-Solomon coded compressed sensing for the unsourced random access," in *in Proc. IEEE Int. Symp. Wirel. Comm. Syst. (ISWCS)*, Sep. 2021.

[16] Z. Liang, J. Zheng, and J. Ni, "Index modulation–aided mixed massive random access," *Frontiers in Communications and Networks*, vol. 2, 2021.

[17] J. Che, Z. Zhang, Z. Yang, X. Chen, C. Zhong, and D. W. K. Ng, "Unsourced random massive access with beam-space tree decoding," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1146–1161, Apr. 2022.

[18] R. Dorfman, "The detection of defective members of large populations," *Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436–440, Dec. 1943.

[19] D.-Z. Du and F. K. Hwang, *Pooling Designs and Nonadaptive Group Testing: Important Tools for DNA Sequencing*, ser. Series on Applied Mathematics. WORLD SCIENTIFIC, Jun. 2006, vol. 18.

[20] V. K. Amalladinne, K. R. Narayanan, J.-F. Chamberland, and D. Guo, "Asynchronous neighbor discovery using coupled compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2019, pp. 4569–4573.

[21] T. Berger, N. Mehravari, D. Towsley, and J. Wolf, "Random multiple-access communication and group testing," *IEEE Trans. Commun.*, vol. 32, no. 7, pp. 769–779, Jul. 1984.

[22] D. M. Malioutov and K. R. Varshney, "Exact rule learning via Boolean compressed sensing," in *Proc. Int. Conf. Machine Learning*, ser. ICML'13, vol. 28. Atlanta, GA, USA: JMLR.org, Jun. 2013, pp. III–765–III–773.

[23] Y. Xuan, I. Shin, M. T. Thai, and T. Znati, "Detecting application denial-of-service attacks: A group-testing-based approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 8, pp. 1203–1216, Aug. 2010.

[24] M. Aldridge, O. Johnson, and J. Scarlett, "Group Testing: An Information Theory Perspective," *Foundations and Trends® in Communications and Information Theory*, vol. 15, no. 3-4, pp. 196–392, Dec. 2019.

[25] O. Johnson, M. Aldridge, and J. Scarlett, "Performance of group testing algorithms with near-constant tests-per-item," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 707–723, Feb. 2019.

[26] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Trans. Inf. Theory*, vol. 10, no. 4, pp. 363–377, Oct. 1964.

[27] H. A. Inan, P. Kairouz, M. Wootters, and A. Özgür, "On the optimality of the Kautz-Singleton construction in probabilistic group testing," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5592–5603, Mar. 2019.

[28] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *The NIST Handbook of Mathematical Functions*. Cambridge Univ. Press, 2010.

[29] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York, NY, USA: Wiley, 1971, vol. II.

[30] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[31] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: Bounds and simulations," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, Jun. 2014.

[32] D. Malioutov and M. Malyutov, "Boolean compressed sensing: LP relaxation for group testing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2012, pp. 3305–3308.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, U.S.A.: Wiley, 2006.