MDPI

Article

The Strongly Asynchronous Massive Access Channel

Sara Shahi, Daniela Tuninetti and Natasha Devroye *

Electrical and Computer Engineering Department, University of Illinois at Chicago, Chicago, IL 60607, USA
* Correspondence: devroye@uic.edu

Abstract: This paper considers the Strongly Asynchronous, Slotted, Discrete Memoryless, Massive Access Channel (SAS-DM-MAC) in which the number of users, the number of messages, and the asynchronous window length grow exponentially with the coding blocklength with their respective exponents. A joint probability of error is enforced, ensuring that all the users' identities and messages are correctly identified and decoded. Achievability bounds are derived for the case that different users have similar channels, the case that users' channels can be chosen from a set which has polynomially many elements in the blocklength, and the case with no restriction on the users' channels. A general converse bound on the capacity region and a converse bound on the maximum growth rate of the number of users are derived. It is shown that reliable transmission with an exponential number of users with an exponential asynchronous exponent with joint error probability is possible at strictly positive rates.

Keywords: multiple access channels; massive access; asynchronous communication

1. Introduction

One of the main attributes of the Internet of the Things (IoT) paradigm is the wireless connectivity of large a number of distributed and uncoordinated devices. Each device is automated, and is allowed to transmit random bursts of data without network or human coordination. A contention-based/grant-based network access requires multiple rounds of back and forth transmissions between the device and the network to acquire synchronization, identify users, and grant them allocations. This signaling overhead and power consumption is not justified for power-limited devices with sporadic traffic patterns and motivates a grant-free network access implementation. An un-coordinated network access leads to a new traffic patterns and collision resolutions over the wireless network. A foundational study of such a system is needed.

We propose a novel communication and multiple-access model that captures these new network characteristics. The formal model definition is given in Section 2, but in order to demonstrate the differences between our proposed model and existing work in the literature, we briefly introduce our model parameters here for the Strongly Asynchronous Slotted Discrete Memoryless Massive Access Channel (SAS-DM-MAC). In a SAS-DM-MAC, the number of users $K_n := e^{n\nu}$ increases exponentially with blocklength n with *occupancy* exponent $\nu \geqslant 0$. Moreover, the users are strongly asynchronous, meaning, they transmit at a randomly chosen time slot within a window of length $A_n := e^{n\alpha}$ slots, each slot of length n, where $\alpha \geqslant 0$ is the *asynchronous* exponent. In addition, when active, each user chooses uniformly at random one message to transmit from a set of $M_n := e^{nR}$ messages, where $R \geqslant 0$ is the transmission *rate*. All transmissions are sent to an access point and the receiver is required to *jointly decode and identify all users*. The goal is to characterize the set of all achievable (R, α, ν) triplets.

1.1. Past Work

Strongly asynchronous communication was first introduced in [1] for synchronization of a single user, and later extended in [2] for synchronization with positive transmission rate.

In [3] the authors of [2] made a brief remark about a "multiple access collision channel" extension of their original single-user model. In this model, any collision of users (i.e., users who happen to transmit in the same block) is assumed to result in output symbols that appear distributed as noise. The error metric is taken to be the *per user* probability of error, which is required to be vanishing for all but a vanishing fraction of users. In this scenario, it is fairly easy to quantify the capacity region for the case that the number of users are less than the square root of the asynchronous window length (i.e., in our notation $\nu < \alpha/2$). However, finding the capacity of the "multiple access collision channel" for *joint* probability of error, as opposed to per user probability of error, is much more complicated and requires novel achievability schemes and novel analysis tools. This is the main subject and contribution of this paper.

Recently, motivated by the emerging machine-to-machine type communications and sensor networks, a large body of work has studied "many-user" versions of classical multiuser channels as pioneered in [4]. In [4] the number of users is allowed to grow at most linearly with the blocklength n. A full characterization of the capacity of the synchronous Gaussian (random) many access channel was given [4]. In [5], the author studied the synchronous massive random access channel where the total number of users increases linearly with the blocklength n. However, the users are restricted to use the same codebook and only a per user probability of error is enforced. In the model proposed here, the users are strongly asynchronous, the number of users grows exponentially with the blocklength, and we enforce a joint probability of error over all users.

In [6,7] a broadcast channel model was introduced where each receiver is only interested in decoding its intended messaged (and not all messages). For a per user/per pair of users probability of error criteria doubly exponential number of receivers was shown to be identifiable by using a randomized code. In this work we study a multi-access channel model with non-cooperative transmitters and a single receiver which needs to decode all messages from all users.

Training based synchronization schemes (i.e., pilot signals) was proven to be suboptimal for bursty communications in [2]. Rather, one can utilize the users' statistics at the receiver for synchronization or user identification purposes. The identification problem (defined in [8]) is a classical problem considered in hypothesis testing. In this problem, a finite number of distinct sources each generates a sequence of i.i.d. samples. The problem is to find the underlying distribution of each sample sequence, given the constraint that each sequence is generated by a distinct distribution.

All studies on identification problems assume a fixed number of sequences. In [9], authors study the Logarithmically Asymptotically Optimal (LAO) Testing of identification problem for a finite number of distributions. The authors in [10] also propose a test for a generalized Neyman-Pearson-like optimality criterion to match multiple yet finite number of sequences to their source distributions.

In contrast, we allow the number of users to increase exponentially with the blocklength. We assume that the users are strongly asynchronous and may transmit randomly anytime within a time window that is exponentially large in the blocklength. We require the receiver to recover both the transmitted messages and the users' identities under a joint probability of error criteria. By allowing the number of sequences to grow exponentially with the number of samples, the number of different possibilities (or hypotheses) would be doubly exponential in the blocklength and the analysis of the optimal decoder becomes much more challenging than classical (with constant number of distributions) identification problems. These differences in modeling the channel require a number of novel analytical tools.

1.2. Contribution

We consider (R, α, ν), the capacity region of the SAS-DM-MAC, where we require the *joint* probability of error to vanish. Our contributions are as follows:

- We propose a new achievability scheme that supports strictly positive values of (R, α, ν) for identical channels for the users.
- We define a new massive identification paradigm in which we allow the number of sequences in a classical identification problem to increase exponentially with the sequence blocklength (or sample size). We find asymptotically matching upper and lower bounds on the probability of identification error for this problem. We use this result in our SAS-DM-MAC model to recover the identity of the users.
- We propose a new achievability scheme for the case that the channels are chosen from a set of conditional distributions. The size of the set increases polynomially in the blocklength n. In this case, the channel statistics themselves can be used for user identification.
- We propose a new achievability scheme without imposing any restrictive assumptions on the users' channels. We show that strictly positive (R, α, ν) are possible.
- We propose a novel converse bound for the capacity of the SAS-DM-MAC.
- We show that for $\nu > \alpha$, not even reliable synchronization is possible.

These results were presented in parts in [11–13].

1.3. Paper Organization

In Section 2 we introduce the notation and problem formulation for SAS-DM-MAC and the identification problem. We present our main results in Section 3. Section 4 concludes the paper with a discussion and future work. Proofs are provided in the Appendix.

2. Notations and Problem Formulation

We first introduce the notation used in the SAS-DM-MAC and then formally define the problem.

2.1. Notation

Capital letters represent random variables that take on lower case letter values in calligraphic letter alphabets. The notation $a_n \doteq e^{nb}$ means $\lim_{n \to \infty} \frac{\log a_n}{n} = b$. We write $[M : N]$, where $M, N \in \mathbb{Z}, M \leqslant N$, to denote the set $\{M, M+1, \ldots, N\}$, and $[K] := [1 : K]$. y^n is used to denote $y^n := [y_1, \ldots, y_n]$. The n-fold cartesian product of a set S is denoted by S^n.

A transition probability / channel from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $Q(y|x)$ (or Q by short), $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$, and the output marginal distribution induced by $P \in \mathcal{P}_\mathcal{X}$ through the channel Q as

$$[PQ](y) := \sum_{x \in \mathcal{X}} P(x)Q(y|x), \forall y \in \mathcal{Y}, \tag{1}$$

where $\mathcal{P}_\mathcal{X}$ is the space of all distributions on $\mathcal{X}$. We define the shorthand notation

$$Q_x(y) := Q(y|x), \forall y \in \mathcal{Y}. \tag{2}$$

For a MAC channel $Q(y|x_1, \ldots, x_K), \forall (x_1, \ldots, x_K, y) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_K \times \mathcal{Y}$, we define the shorthand notation

$$Q_S(y|x_S) := Q(y|x_S, \star_{S^c}), \forall S \subseteq [K], \tag{3}$$

to indicate that users indexed by S transmit $x_i \in \mathcal{X}_i, \forall i \in S$, and users indexed by $S^c = [K] \backslash S$ transmit their respective idle symbol $\star_j \in \mathcal{X}_j \forall j \in S^c$. When $|S| = 1$ and $|S| = 0$, we respectively use

$$Q_i(y|x_i) := Q_{\{i\}}(y|x_i) = Q(y|\star_1, \ldots, \star_{i-1}, x_i, \star_{i+1}, \ldots, \star_K), \tag{4}$$

$$Q_\star(y) := Q_{\{\varnothing\}}(y|x_i) = Q(y|\star_1, \ldots \star_K).$$

The P-type set and the V-shell of the sequence x^n are defined, respectively, as

$$T(P) := \left\{ x^n : \frac{\mathcal{N}(a|x^n)}{n} = P(a), \forall a \in \mathcal{X} \right\}, \tag{5}$$

$$T_V(x^n) := \left\{ y^n : \frac{\mathcal{N}(a,b|x^n,y^n)}{\mathcal{N}(a|x^n)} = V(b|a), \forall (a,b) \in (\mathcal{X}, \mathcal{Y}) \right\}, \tag{6}$$

where $\mathcal{N}(a,b|x^n,y^n) = \sum_{i=1}^n \mathbb{1}_{\left\{ \substack{x_i=a \\ y_i=b} \right\}}$ is the number of joint occurrences of (a,b) in the pair of sequences (x^n, y^n) and where $\mathcal{N}(a|x^n)$ denotes the number of occurrences of $a \in \mathcal{X}$ in the sequence x^n. The empirical distribution of the sequence x^n is

$$\widehat{P}_{x^n}(a) := \frac{1}{n}\mathcal{N}(a|x^n) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{x_i=a\}}, \forall a \in \mathcal{X}. \tag{7}$$

If in using (7) the target sequence x^n is clear from the context, we drop the subscript x^n in $\widehat{P}_{x^n}(\cdot)$.

We also use the notation $X^n \overset{\text{i.i.d}}{\sim} P$ when all random variables in the random vector X^n are generated i.i.d according to distribution P.

We use $D(P_1 \| P_2)$ to denote the *Kullback Leibler divergence* between distribution P_1 and P_2, and

$$D(Q_1 \| Q_2|P) := \sum_{(x,y)\in(\mathcal{X}\times\mathcal{Y})} P(x)Q_1(y|x) \log \frac{Q_1(y|x)}{Q_2(y|x)} \tag{8}$$

for the conditional Kullback Leibler divergence. We let

$$I(P,Q) = D(Q \| [PQ] \mid P) \tag{9}$$

denote the *mutual information* between random variables X and Y jointly distributed according to $P_{X,Y}(x,y) = P(x)Q(y|x)$. As for binary functions, we use

$$d(p \| q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q},$$

$$p * q := p(1-q) + (1-p)q,$$

$$h(p) := -p \log(p) - (1-p) \log(1-p)$$

to represent binary divergence, convolution and entropy functions, respectively.

The *Chernoff distance* and the *Bhatcharrya distance* between two distributions $P_1, P_2 \in \mathcal{P}_\mathcal{X}$ are respectively defined as

$$C(P_1,P_2) := \sup_{0 \leqslant t \leqslant 1} -\log\left(\sum_{x\in\mathcal{X}} P_1(x)^t P_2(x)^{1-t} \right), \tag{10}$$

$$B(P_1,P_2) := -\log\left(\sum_{x\in\mathcal{X}} \sqrt{P_1(x)P_2(x)} \right). \tag{11}$$

We extend this definition and introduce the quantities

$$C(P_i, Q_i, P_j, Q_j) := \sup_{0 \leqslant t \leqslant 1} \mu_{i,j}(t), \tag{12a}$$

$$\mu_{i,j}(t) := -\log \sum_{x_i \in \mathcal{X}, x_j \in \mathcal{X}, y \in \mathcal{Y}} P_i(x_i) P_j(x_j) Q_i(y|x_i)^{1-t} Q_j(y|x_j)^t, \tag{12b}$$

$$C(., Q_\star, P_j, Q_j) := \sup_{0 \leqslant t \leqslant 1} -\log \left(\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_j(x) Q_\star(y)^{1-t} Q_j(y|x)^t \right). $$

We also use

$$E_r(R, P, Q) := \max_{0 \leqslant \rho \leqslant 1} [E_0(R, P, Q, \rho) - \rho R], \tag{13}$$

$$E_{sp}(R, P, Q) := \sup_{\rho > 0} [E_0(R, P, Q, \rho) - \rho R] \tag{14}$$

to denote the *random coding* and the *sphere packing* error exponent of channel $Q(y|x)$, with input distribution $P(x)$ and rate R as defined in [14] where

$$E_0(R, P, Q, \rho) := -\log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P(x) Q(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{15}$$

2.2. SAS-DM-MAC Problem Formulation

Let M be the number of messages (the same for each user), A be the number of blocks, and K be the number of users.

An (M, A, K, n, ϵ) code for the SAS-DM-MAC consists of:

- A message set $[M]$, for each user $i \in [K]$.
- An encoding function $f_i : [M] \to \mathcal{X}^n$, for each user $i \in [K]$. We define

$$x_i^n(m) := f_i(m). \tag{16}$$

- Each user $i \in [K]$ chooses a message $m_i \in [M]$ and a block index $t_i \in [A]$ (called 'active' block henceforth), both uniformly at random. It then transmits $[\star_i^{n(t_i-1)} x_i^n(m_i) \star_i^{n(A-t_i)}]$, where $\star_i \in \mathcal{X}$ is the designated idle symbol for user i.
- A destination decoding function

$$g : \mathcal{Y}^{nA} \to ([A] \times [M])^K \tag{17}$$

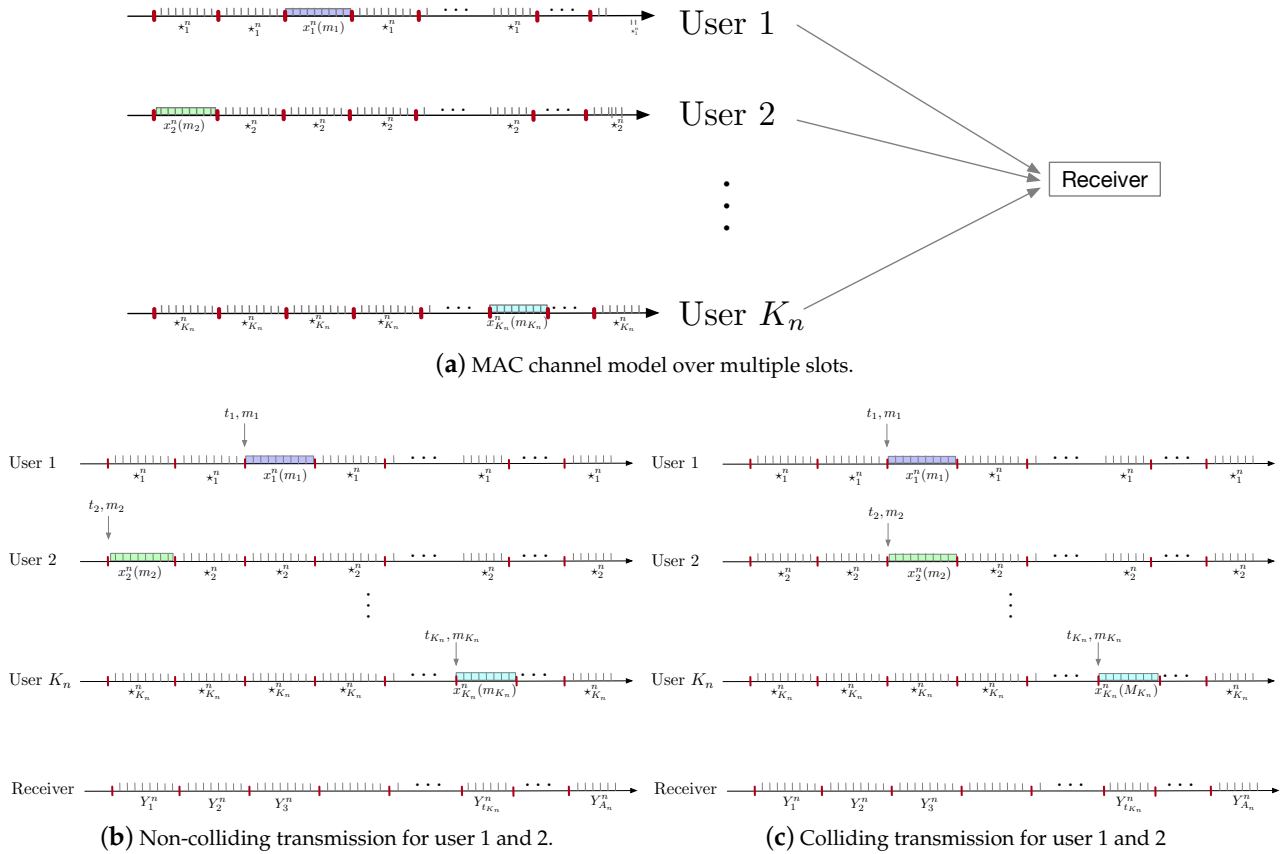such that its associated probability of error, $P_e^{(n)}$, satisfies $P_e^{(n)} \leqslant \epsilon$ where

$$P_e^{(n)} := \frac{1}{(AM)^K} \sum_{(t_1, m_1), \dots, (t_K, m_K)} \mathbb{P}\left[g(Y^{nA}) \neq ((t_1, m_1), \dots, (t_K, m_K)) | H_{((t_1, m_1), \dots, (t_K, m_K))} \right], \tag{18}$$

where the hypothesis that user $i \in [K]$ has chosen to transmit message $m_i \in [M]$ in block $t_i \in [A]$ is denoted by $H_{((t_1, m_1), \dots, (t_K, m_K))}$.

A tuple (R, α, ν) is said to be *achievable* if there exists a sequence of codes $(e^{nR}, e^{n\alpha}, e^{n\nu}, n, \epsilon_n)$ with $\lim_{n \to \infty} \epsilon_n = 0$. The capacity region of the SAS-DM-MAC at asynchronous exponent α, occupancy exponent ν, and rate R, is the closure of all possible achievable (R, α, ν) triplets.

Remark 1. *We should emphasize that we are focusing our attention on discrete memoryless channels (DMC) where a user not transmitting/being idle is equivalent to sending that user's idle symbol in the DMC model, just as, in a Gaussian channel, a user not sending corresponds to sending the symbol '0'. We do not study Continuous channels in this work.*

A depiction of a timeline for our channel model is provided in Figure 1a and for specific cases: one in which user 1 and 2 are not simultaneously transmitting (Figure 1b) and another in which user 1 and 2 are simultaneously transmitting (Figure 1c).

(**a**) MAC channel model over multiple slots.

(**b**) Non-colliding transmission for user 1 and 2.   (**c**) Colliding transmission for user 1 and 2

Figure 1. Transmitted sequence for each user and the received sequence in colliding and non-colliding transmission time for user 1 and 2.

2.3. Identification Problem Formulation

We first summarize the notation specifically used in this Section (and in Appendix B) in Table 1 and then introduce the problem formulation.

Table 1. Special notation for the identification problem.

S_A	A symmetric group which is the set of all possible permutations of A elements.
σ_i	The i-th element of the permutation $\sigma \in S_A$; $i \in [A]$.
K_k	$:= K_k(w_{\{1,2\}} \ldots w_{\{k,1\}})$; A complete graph with k nodes and edge weights $w_{\{i,j\}}$ between node i, j.
$c_r^{(V)}, c_r^{(E)}$	A cycle of length r represented by the set of its vertices and edges, respectively.
$c_r^{(V)}(i)$	i-th vertex of the cycle $c_r^{(V)}$.
$C_k^{(r)}$	Set of cycles of length r in the complete graph K_k.
$N_k^{(r)}$	Number of cycles of length r in a complete graph K_k.
$G(c)$	$:= \prod_{\forall i \neq j} w_{i,j}$; Gain of a cycle which we define as the product of the edge weights within the cycle c.
AM-GM inequality	Refers to the following inequality $n\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \leqslant \sum_{i=1}^{n} a_i$.

Assume $P := \{P_1, \ldots, P_{T_n}\} \subseteq \mathcal{P}_{\mathcal{X}}$ consist of T_n distinct distributions on $\mathcal{X}$ and let random variable Σ be uniformly distributed over S_{T_n} (see definition in Table 1). Let

$\sigma \in S_{T_n}$ and let $x^{nT_n} = [x_1^n, \ldots, x_{T_n}^n]$ be a sample vector of $X^{nT_n} = \left[X_1^n, \ldots, X_{T_n}^n \right]$, in which $X_i^n \overset{\text{i.i.d}}{\sim} P_{\sigma_i}^n, \forall i \in [T_n]$. Given x^{nT_n}, we would like to find the permutation σ. More specifically, we are interested in finding a permutation $\hat{\sigma} : \mathcal{X}^{nT_n} \rightarrow S_{T_n}$ in which $X_i^n \overset{\text{i.i.d}}{\sim} P_{\hat{\sigma}_i}, \forall i \in [T_n]$. Let $\hat{\Sigma} = \hat{\sigma}(X^{nT_n})$.

The average probability of error for the set of distributions P is defined by

$$P_e^{(n)} = \mathbb{P}[\hat{\Sigma} \neq \Sigma] = \frac{1}{(T_n)!} \sum_{\sigma \in S_{T_n}} \mathbb{P}[\hat{\Sigma} \neq \sigma], \tag{19}$$

where in (19) the measure $\mathbb{P}$ over $(X_1^n, ..., X_{T_n}^n)$ is $X_i^n \overset{\text{i.i.d}}{\sim} P_{\sigma_i}, \forall i \in [T_n]$.

We say that a set of distributions P is *identifiable* if $\lim_{n \to \infty} P_e^{(n)} \to 0$.

3. Main Results

We first introduce an achievable region for the case that different users have identical channels (in Theorem 1). We then derive the identification criteria (in Theorem 2) and use it in the study of a more general case of massive communications in which the users' channels belong to a set of conditional probability distributions of polynomial size in n (in Theorem 3). In this case, we use the output statistics to distinguish and identify the users. Afterwards, we remove all conditions on the users' channels and derive an achievability bound on the capacity of the SAS-DM-MAC in (Theorem 4). We also propose a converse bound on the capacity of general SAS-DM-MAC (in Theorem 5) and a converse bound on the number of users (in Theorem 6).

3.1. Users with Identical Channels

The following theorem is an achievable region for the SAS-DM-MAC for the case that different users have identical channels toward the base station when they are the sole active user.

Theorem 1. *For a SAS-DM-MAC with $Q_{\{i\}}(y|x) = W(y|x)$ (recall Definition (4)) for all users, the following (R, α, ν) region is achievable*

$$\bigcup_{\substack{P \in \mathcal{P}_{\mathcal{X}} \\ \lambda \in [0,1]}} \left\{ \begin{array}{ll} \nu & < \frac{\alpha}{2} \\ \nu & < \min\{D(W_\lambda \parallel W|P), E_r(R + \nu, P, W)\} \\ \alpha + R + \nu & < D(W_\lambda \parallel Q_\star|P) \\ R + \nu & < I(P, W) \end{array} \right\}, \tag{20}$$

where

$$W_\lambda(y|x) := \frac{W(y|x)^\lambda Q_\star(y)^{1-\lambda}}{\sum_{y' \in \mathcal{Y}} W(y'|x)^\lambda Q_\star(y')^{1-\lambda}}, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{21}$$

and where E_r was defined in (13).

In the scenario where users have identical channels toward the base station, one can combine the user identification and decoding stages, i.e., one can distinguish the user's identity from the decoded message. This appears as the combination $R + \nu$ in the bounds in (20).

One could also interpret the result in Theorem 1 as follows. There is one single "big codebook" with MK codewords of length n. User 1 uses codewords with indices 1 through M; User 2 uses codewords with indices $M + 1$ through $2M$, and so forth, concluding with User K using codewords with indices $(K - 1)M + 1$ through KM. In each active slot, the receiver decodes the "big codebook"; reliable decoding imposes $\log(MK)/n = R + \nu < I(P, W)$. However, since in our setting we need to decode K codewords (one per active slot, as

opposed to just one), in addition to $R + \nu \leqslant I(P, W)$ we also need the bound $\log(K)/n = \nu < E_r(\nu + R, P, W)$. Comparing the third and fourth bounds of (20) with the per-user probability of error achievability region in [3], we can see an additional ν penalty as a result of a joint probability of error criteria. However, we do not claim (via a matching converse bound) that the rate penalty (at least in the format that we prove) is conclusive.

We should also note that for $\nu < \frac{\alpha}{2}$ (first bound in (20)), with probability approaching one as the blocklength n goes to infinity, the users transmit in distinct blocks. Hence, regardless of the achievability scheme, the probability of error under a hypothesis where a collision has occurred is vanishing as the blocklength goes to infinity. As a result, in analyzing the joint probability of error in our achievability scheme, we need to focus on the hypothesis that users do not collide. In our achievability scheme, we use a two-stage decoder which first synchronizes the users (i.e., finds the location of active blocks) and then decodes the users' messages (which also identifies the users' identities). The complete proof of Theorem 1 is given in Appendix A.

Remark 2. *The existence of an error exponent bound on the occupancy exponent (i.e., ν) is intuitive (second argument in the second condition in (20)). Even if the decoder has side information about the location of noisy blocks (i.e., blocks with only idle symbols for all users), it still has to decode all the messages within active blocks. Due to independence of the output in different blocks and by the fact that there cannot be intra-block coding, the probability of decoding error over active blocks is asymptotically equal to the sum of the decoding error probabilities in each active block. This puts a restriction on ν based on the best decay rate for the decoding error in each block, i.e., the error exponent.*

3.2. Users with Different Choice of Channels

We now move on to a more general setting in which we remove the restriction that all users have identical channels towards the base station (when only one is active). Theorem 3 gives an achievable region when we allow the users' channels (when they are the sole active user toward the base station) to be chosen from a set of conditional distributions of polynomial size in the blocklength n.

In this scenario, one may use the users' statistics at the channel output to identify the set of users who have a similar channel. In this regard, before introducing Theorem 3, we introduce Theorem 2 in which we characterize the relation between the number of distributions and their pairwise distance for them to be identifiable. The proof of Theorem 2, given in Appendix B, is itself based on a novel graph theoretic technique to analyze the optimal Maximum Likelihood (ML) decoder, which is of independent interest.

Theorem 2. *A sequence of distributions $P = \{P_1, \ldots, P_{T_n}\}$ is identifiable iff*

$$\lim_{n \to \infty} \sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i, P_j)} = 0, \tag{22}$$

where $B(P_i, P_j)$ is the Bhatcharrya distance, which was defined in (11).

Remark 3. *From Theorem 2 one can see that when the number of distributions T_n is a constant or grows polynomially with n, the sequence of distributions are always identifiable and by the problem formulation in Section 2.3, it is implied that the probability of error in the identification problem decays to zero as the blocklength n goes to infinity.*

By using the criterion for identifiability of a massive number of distributions in Theorem 2, we move on to the SAS-DM-MAC problem. We adapt the concept of identification of distribution to identify the users' statistics at the output with an optimal ML test. We use the result of Theorem 2 in the derivation of Theorem 3.

Theorem 3. *For a SAS-DM-MAC where $Q_i(y|x) = W_{c(i)}(y|x)$ is the channel for user $i \in [K_n]$, for $\mathcal{W} := \left\{ \bigcup_{i \in [K_n]} W_{c_i} \right\}$ and $|\mathcal{W}| = poly(n)$ the region*

$$\liminf_{n \to \infty} \mathcal{A}_n, \tag{23}$$

is achievable where $\mathcal{A}_n$ is defined as

$$\mathcal{A}_n := \bigcup_{\substack{\epsilon > 0}} \bigcup_{\substack{(P_1, \dots P_{|\mathcal{W}|}) \in \mathcal{P}_{\mathcal{X}}^{|\mathcal{W}|} \\ (\lambda_1, \dots \lambda_{|\mathcal{W}|}) \in [0,1]^{|\mathcal{W}|}}} \bigcap_{j \in [|\mathcal{W}|]} \left\{ \begin{array}{cc} \nu_j + \epsilon & < \frac{\alpha}{2} \\ \nu_j + \epsilon & < \min \left\{ D([P_j W_j]_{\lambda_j} \| [P_j W_j]), E_r(R + \nu_j, P_j, W_j) \right\} \\ \alpha + \epsilon & < D([P_j W_j]_{\lambda_j} \| Q_\star) \\ 0 & < \inf_{\substack{i \in [|\mathcal{W}|] \\ i \neq j}} B([P_i W_i], [P_j W_j]) \\ R + \nu_j + \epsilon & < I(P_j, W_j) \end{array} \right\}, \tag{24}$$

and where E_r is defined in (13), $B(P_i, P_j)$ was defined in (11) and

$$\nu_j := \frac{1}{n} \log(\mathcal{N}_j),$$

$$\mathcal{N}_j := \sum_{i=1}^{K_n} \mathbb{1}_{\{Q_i = W_j\}} \quad \text{and such that} \quad \sum_{j=1}^{|\mathcal{W}|} \mathcal{N}_j = K_n, \tag{25}$$

$$[P_j W_j]_\lambda(y) := \frac{([P_j W_j](y))^\lambda (Q_\star(y))^{1-\lambda}}{\sum_{y' \in \mathcal{Y}} ([P_j W_j](y'))^\lambda (Q_\star(y'))^{1-\lambda}}. \tag{26}$$

The proof of Theorem 3 is given in Appendix F.

In the proof of Theorem 3 we propose a three stage decoder which performs the task of synchronization, identification and decoding, sequentially. It is also worth noting that given that the number of users is exponential in the blocklength and the number of channels is only polynomial in the blocklength, there must exist at least a single channel that occupies exponentially many blocks. Hence, it is not surprising to see bounds on the occupancy level of each channel (i.e., ν_j) in (24).

Remark 4. *We should note that (as it is apparent in (24)) as long as we have the assumption that the number of user's channels increase with blocklength n, our optimization space over all users also increases with n and hence our achievability and converse bounds would be dependent on the exact definition of $n \to \infty$ (e.g, compared to (20) in Theorem 1). However, in order to compute these bounds, we only need to optimize with respect to $P_i, i \in [K_n]$ as opposed to P^n in the conventional n-letter capacity expressions.*

3.3. General Case

Now we investigate a SAS-DM-MAC with no restriction on the channels of the users toward the base station. The key ingredient in our analysis is a novel way to bound the probability of error reminiscent of Gallager's error exponent [14]. We show an achievability scheme that allows a positive lower bound on the R and on ν. This proves that reliable transmission with an exponential number of users with an exponential asynchronous exponent with joint error probability is possible. We use an ML decoder sequentially in each block to identify the active user and its message.

Theorem 4. *For a SAS-DM-MAC the following region is achievable*

$$
\liminf_{n\to\infty} \bigcup_{\epsilon>0} \bigcup_{(P_1,\dots P_{K_n})\in\mathcal{P}_{\mathcal{X}}^{K_n}} \bigcap_{i\in[K_n]}
\left\{
\begin{array}{ll}
\nu & < \frac{\alpha}{2} \\
\nu + R + \epsilon & < C(P_i,Q_i,P_i,Q_i), \\
2\nu + R + \epsilon & < \inf_{\substack{j\in[K_n]\\ j\neq i}} C(P_i,Q_i,P_j,Q_j), \\
\alpha + \nu + R + \epsilon & < C(\,.\,,Q_\star,P_i,Q_i),
\end{array}
\right\}
\tag{27}
$$

The proof of Theorem 4 is given in Appendix G.

Remark 5. *Note that one can show the following (see Appendix H):*

$$
C(P_i,Q_i,P_i,Q_i) = -\log\sum_{x,x',y} P_i(x)P_i(x')\sqrt{Q_i(y|x)Q_i(y|x')}, \tag{28a}
$$

$$
C(\,.\,,Q_\star,P_i,Q_i) \leqslant I(P_i,Q_i) + D([P_iQ_i]\,\|\,Q_\star), \tag{28b}
$$

where (28a) is the special case of $E_0(R,P,Q,\rho=1)$ in (15), and the right hand side of (28b) is the bound on $\alpha + R$ in the single user strong asynchronous channel [15]. From (28) and (27) we thus see that

$$
\nu < E_r(R,P_i,Q_i),
$$
$$
\alpha + R < I(P_i,Q_i) + D([P_iQ_i]\,\|\,Q_\star) - \nu,
$$

which shows that the second and fourth bounds in (27) are, respectively, less than the synchronous channel error exponent and the point-to-point strong asynchronous channel capacity.

3.4. Example

Consider the SAS-DM-MAC with input-output relationship $Y = \sum_{i\in[K_n]} X_i \oplus Z$ with $Z \sim \text{Bernoulli}(\delta)$ for some $\delta \in (0,1/2)$. In our notation

$$
Q(y|x) = \mathbb{P}[X_i \oplus Z = y | X_i = x] = \mathbb{P}[Z = x \oplus y]
$$
$$
= \begin{cases} 1-\delta & x\oplus y = 0 \ (i.e.,\ x=y) \\ \delta & x\oplus y = 1 \ (i.e.,\ x\neq y) \end{cases}.
$$

Assume that the input distribution used is $P = \text{Bernoulli}(p)$ for some $p \in (0,1/2)$. The achievability region in Theorem 1, is

$$
\bigcup_{\substack{p\in[0,\frac{1}{2}]\\ \lambda\in[0,1]}}
\left\{
\begin{array}{ll}
\nu & < \alpha/2 \\
\nu & < \min\left\{p\cdot d(\epsilon_\lambda\,\|\,\delta), \min_{q\in[0,1]}\left\{d(\delta\,\|\,q) + \max\{0, d(q\,\|\,\tfrac{1}{2}) - R - \nu\}\right\}\right\} \\
\alpha + R + \nu & < p\cdot d(\epsilon_\lambda\,\|\,1-\delta) \\
R + \nu & < h(p*\delta) - h(\delta)
\end{array}
\right\},
\tag{29}
$$

where $\epsilon_\lambda := \frac{\delta^\lambda(1-\delta)^{(1-\lambda)}}{\delta^\lambda(1-\delta)^{(1-\lambda)} + (1-\delta)^\lambda\delta^{(1-\lambda)}}$.

Moreover, by assuming $P_i = \text{Bernoulli}(p_i)$ for all $i \in [K_n]$, we can show that the optimal t in $C(P_i,Q_i,P_j,Q_j) = \sup_t \mu_{i,j}(t)$ is $t = 1/2$ and hence the achievability region in Theorem 4 is

$$
\liminf_{n\to\infty} \bigcup_{\epsilon>0} \bigcup_{(P_1,\dots P_{K_n})\in\mathcal{P}_{\mathcal{X}}^{K_n}} \bigcap_{i\in[K_n]}
\left\{
\begin{array}{ll}
\nu & < \frac{\alpha}{2} \\
\nu + R + \epsilon & < B(P_i,Q) = g(p_i*p_i,\delta) \\
2\nu + R + \epsilon & < \inf_{i\neq j} C(P_i,Q,P_j,Q) = \inf_{i\neq j} g(p_i*p_j,\delta) \\
\alpha + \nu + R + \epsilon & < C(\,.\,,Q_\star,P_i,Q) = g(p_i,\delta)
\end{array}
\right\},
$$

where

$$g(a,b) := -\log\left(1 - a + 2a\sqrt{b(1-b)}\right).$$

Finally, by symmetry, we can see that the optimal $p_i = \frac{1}{2}$, $\forall i \in [K_n]$ and hence $g(\frac{1}{2},\delta) = -\log\left(1/2 + \sqrt{\delta(1-\delta)}\right) > 0$. So on the BSC($\delta$) strictly positive rates and ν are achievable. In this regard, the region in Theorem 4 reduces to

$$\alpha + \nu + R < -\log\left(1/2 + \sqrt{\delta(1-\delta)}\right). \tag{30}$$

The achievable region in (29) for (α, ν, R) is shown in Figure 2a. In addition, the achievable region for (α, ν, R) with the achievable scheme in Example (30) is also plotted in Figure 2b for comparison. Figure 2 shows that the achievable scheme in Theorem 1 indeed results in a larger achievable region than the one in Theorem 4.

(**a**) Achievable region in (29).   (**b**) Achievable region in (30).
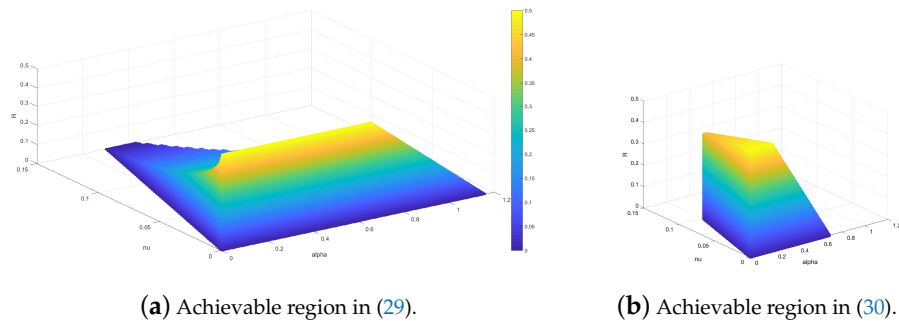
Figure 2. Comparison of the achievable region in Theorems 1 and 4, for the Binary Symmetric Channel with cross over probability $\delta = 0.11$.

The fact that the achievability region for Theorem 1 is larger than the achievability region of Theorem 4 for identical channels is not surprising. In Theorem 1 we separated the synchronization and decoding steps, whereas in Theorem 4 synchronization and codeword decoding was done the same time but sequentially for each block. The sequential block decoding step results in a smaller achievability region in Theorem 4.

3.5. Converse on the Capacity Region of the SAS-DM-MAC

Thus far, we have provided achievable regions for the SAS-DM-MAC when different users have identical channels; when their channels belong to a set of channels with size that grows polynomially in the blocklength; and without any restriction on the users' channels. Theorem 5 next provides a converse to the capacity region of the general SAS-DM-MAC.

Theorem 5. *For the SAS-DM-MAC, such that $\nu < \alpha/2$, the following region is impermissible*

$$\liminf_{n\to\infty} \bigcup_{\epsilon>0} \bigcap_{\substack{(P_1,\dots P_{K_n})\in\mathcal{P}_\mathcal{X}^{K_n} \\ (\lambda_1,\dots\lambda_{K_n})\in[0,1]^{K_n}}} \left\{ \begin{array}{c} \left\{ \nu > \frac{1}{K_n}\sum_{i=1}^{K_n} D(Q_{i\lambda_i} \| Q_i|P_i) + \epsilon, \right. \\ \left. \alpha > \frac{1}{K_n}\sum_{i=1}^{K_n} D(Q_{i\lambda_i} \| Q_\star|P_i) - (1-\bar{r}_n)(\nu+R) + \epsilon \right\} \\ \cup \\ \left\{ \nu > \frac{1}{K_n}\sum_{i=1}^{K_n} E_{sp}(R, P_i, Q_i) + \epsilon \right\} \\ \cup \\ \{R > I(P_i, Q_i) + \epsilon\} \end{array} \right\}, \tag{31}$$

where $\bar{r}_n$ is the infimum probability of error, over all estimators T, in distinguishing different hypothesis $Q_{i\lambda_i}(y^n|x_i^n(m)), i \in [K_n], m \in [M_n]$, i.e.,

$$\bar{r}_n := \inf_T \frac{1}{K_n M_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i\lambda_i}(T \neq i, m|x_i^n(m)). \tag{32}$$

Proof. The complete proof is given in Appendix I. □

The first bound in (31) corresponds to synchronization, while the second and third bounds correspond to decoding. Also, as it was noted before in Remark 2, one can again see that the value of ν is restricted by the decoding error exponent.

3.6. Converse on the Number of Users in a SAS-DM-MAC

In previous sections, we restricted ourselves to the regime where $\nu < \frac{\alpha}{2}$. However, an interesting question is how large a ν can be. Theorem 6 provides a converse bound on the value of ν such that for $\nu > \alpha$, not even reliable synchronization is possible.

Theorem 6. *For a SAS-DM-MAC with $\nu > \alpha$, reliable synchronization is not possible, i.e., even with $M = 1$, one has $P_e^{(n)} > 0$.*

The complete proof can be found in Appendix J.

4. Discussion and Conclusions

In this paper we studied a Strongly Asynchronous and Slotted Massive Access Channel (SAS-DM-MAC) where $K_n := e^{n\nu}$ different users transmit a randomly selected message among $M_n := e^{nR}$ ones within a strong asynchronous window of length $A_n := e^{n\alpha}$ blocks of n channel uses each. We found inner and outer bounds on the (R, α, ν) tuples. Our analysis is based on a joint probability of error in which we required all users' messages and identities to be jointly correctly decoded at the base station. Our results are focused on the regime $\nu < \frac{\alpha}{2}$, where the probability of user collision vanishes. We proved in Theorem 6 that for the regime $\nu > \alpha$, not even synchronization is possible. We now discuss some of the difficulties in analyzing the region $\frac{\alpha}{2} \leqslant \nu \leqslant \alpha$.

For the region $\nu < \frac{\alpha}{2}$, with probability $\frac{\binom{A_n}{K_n}}{(A_n)^{K_n}} \to 1$ as blocklength $n \to \infty$, the users transmit in distinct blocks. Hence, in analyzing the probability of error of our achievable schemes, we only need to bound the error under the hypothesis that users do not collide. For $\frac{\alpha}{2} \leqslant \nu \leqslant \alpha$, we need to consider the events where collisions occur. In particular, based on Lemma 1 (proved in the Appendix L), for $\frac{\alpha}{2} \leqslant \nu \leqslant \alpha$, the probability of every arrangement of users is itself vanishing in the blocklength.

Lemma 1. *For the region $\frac{\alpha}{2} \leqslant \nu \leqslant \alpha$ the non-colliding arrangement of users has the highest probability among all possible arrangements, yet, the probability of this event is also vanishing as blocklength n goes to infinity.*

As a consequence of Lemma 1, one needs to propose an achievable scheme that accounts for several arrangements (the number of which is non-trivial to the authors) and collision of users and drives the probability of error in these arrangements to zero. It is also worth noting that the number of possible hypotheses is doubly exponential in the blocklength. Finally, it is worth emphasizing that the authors in [16] can get to $\nu \leqslant \alpha$ since they require the recovery of the messages of a *large fraction* of users, and require the per-user probability of error to be vanishing (rather than the overall or joint probability of error, which is a much stronger condition). To prove whether or not strictly positive (R, α, ν) are possible in the region $\frac{\alpha}{2} \leqslant \nu \leqslant \alpha$, with vanishing joint probability of error, is a challenging open problem.

Appendix A. Proof of Theorem 1

We first introduce the codebook generation and then introduce and analyze the decoder.

Appendix A.1. Codebook Generation

Let $K_n = e^{n\nu}$ be the number of users, $A_n = e^{n\alpha}$ be the number of blocks, and $M_n = e^{nR}$ be the number of messages. Each user $i \in [K_n]$ generates a constant composition codebook with composition P by drawing each message's codeword uniformly and independently at random from the P-type set $T(P)$ (recall definition in (5)). The codeword of user $i \in [K_n]$ for message $m \in [M_n]$ is denoted as $x_i^n(m)$.

Appendix A.2. The Decoder and Analysis of the Probability of Error

A two-stage decoder is used, to first synchronize and then decode (which also identifies the users' identities) the users' messages. The probability of error of this two-stage decoder is thus given by

$$\mathbb{P}[\text{ Error }] \leqslant \mathbb{P}[\text{ Error } | \text{ No-collision}] + \mathbb{P}[\text{ collision }]$$

$$\leqslant \mathbb{P}[\text{ Synchronization error } | \text{ No-collision }] \tag{A1a}$$

$$+ \mathbb{P}[\text{ Decoding error } | \text{ No synchronization error, No-collision }] \tag{A1b}$$

$$+ \mathbb{P}[\text{ collision }]. \tag{A1c}$$

Synchronization Step. We perform a sequential likelihood test as follows. Fix a threshold

$$T \in [-D(Q_\star \parallel W|P), D(W \parallel Q_\star|P)]. \tag{A2}$$

For each block $j \in [A_n]$ if there exists any message $m \in [M_n]$ for any user $i \in [K_n]$ such that

$$L(x_i^n(m), y_j^n) := \frac{1}{n} \log \frac{W(y_j^n|x_i^n(m))}{Q_\star(y_j^n)} \geqslant T, \tag{A3}$$

then declare that block j is an 'active' block, and an 'idle' block otherwise. Let

$$H^{(1)} := H_{((1,1),(2,1),\dots,(K_n,1))}, \tag{A4}$$

be the hypothesis that user i (for each $i \in [K_n]$) is active in block i and sends message $m_i = 1$. As explained earlier (in the paragraph after Theorem 1), the probability of collision in (A1c) is vanishing and we can focus on calculating (A1a) and (A1b). For (A1a), by symmetry of the different no-collision hypotheses, the average probability of synchronization error (over different such hypotheses) is upper bounded by

$$\mathbb{P}\Big[\text{Synchronization error}|H^{(1)}\Big]$$

$$\leqslant \sum_{j=1}^{K_n} \mathbb{P}\left[\bigcap_{i=1}^{K_n} \bigcap_{m=1}^{M_n} L(x_i^n(m), Y_j^n) < T|H^{(1)}\right] + \sum_{j=K_n+1}^{A_n} \mathbb{P}\left[\bigcup_{i=1}^{K_n} \bigcup_{m=1}^{M_n} L(x_i^n(m), Y_j^n) \geqslant T|H^{(1)}\right]$$

$$\leqslant \sum_{j=1}^{K_n} \mathbb{P}\Big[L(x_j^n(1), Y_j^n) < T|H^{(1)}\Big] + \sum_{j=K_n+1}^{A_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \mathbb{P}\Big[L(x_i^n(m), Y_j^n) \geqslant T|H^{(1)}\Big]$$

$$\leqslant e^{n\nu} e^{-nD(W_\lambda \| W|P)} + e^{n(\alpha+\nu+R)} e^{-nD(W_\lambda \| Q_\star|P)}, \tag{A5}$$

where (A5) can be derived as in Chapter 11 [17]. The upper bound on the probability of error for the synchronization error in (A5) vanishes as *n* goes to infinity if the second and third bound in (20) hold.

Decoding Step. In this stage we evaluate the decoding error under the no-collision, no synchronization error hypothesis using a maximum likelihood decoder. Moreover, since in each block, we have to distinguish among $e^{n(R+\nu)}$ messages, we can upper bound the probability of decoding error in (A1b) as

$$\mathbb{P}[\text{Decoding error}| \text{ No synchronization error, No-collision}] \leqslant \sum_{i=1}^{K_n} \mathbb{P}[\text{Decoding error in block } i]$$

$$\leqslant e^{n\nu} e^{-nE_r(R+\nu,P,W)} \tag{A6}$$

where (A6) is by [14] and E_r is defined in (13). This retrieves the second bound in (20). Finally, as a synchronous channel and since in each block we have to distinguish among $e^{n(R+\nu)}$ messages we retrieve the fourth bound in (20).

Appendix B. Proof of Theorem 2

We now introduce a graph theoretic Lemma, the proof of which is given in Appendix C.

Lemma A1. *In a complete graph $K_k(w_{1,2}, \dots, w_{k,1})$, for the set of cycles of length r, $C_k^{(r)}$, we have*

$$\frac{1}{N_k^{(r)}} \left(\sum_{c \in C_k^{(r)}} G(c) \right) \leqslant \left(\frac{w_{1,2}^2 + \dots + w_{k,1}^2}{\binom{k}{2}} \right)^{\frac{r}{2}}$$

where $C_k^{(r)}$ and $N_k^{(r)}$ are the set and number of all cycles of length r in a complete graph K_k, respectively.

Next, we use Lemma A1 to derive upper and lower bounds on the probability of error.

Appendix B.1. Upper Bound on the Probability of Identification Error

We use the optimal ML decoder, which minimizes the average probability of error, given by

$$\hat{\sigma}(x_1^n, \dots, x_{T_n}^n) := \arg\max_{\sigma \in S_{T_n}} \sum_{i=1}^{T_n} \log\big(P_{\sigma_i}(x_i^n)\big), \tag{A7}$$

where $P_{\sigma_i}(x_i^n) = \prod_{t=1}^n P_{\sigma_i}(x_{i,t})$. Let $\hat{\Sigma} = \hat{\sigma}(X_1^n, \dots, X_{T_n}^n)$ be the estimate of the permutation of the distributions of $X_i^n, i \in [T_n]$. The average probability of error associated with the ML decoder can also be written as

$$P_e^{(n)} = \mathbb{P}\left[\hat{\Sigma} \neq [T_n]\right]$$

$$= \mathbb{P}\left[\bigcup_{\hat{\sigma} \neq [T_n]} \hat{\Sigma} = \hat{\sigma}\right]$$

$$= \mathbb{P}\left[\bigcup_{r=2}^{T_n} \bigcup_{\substack{\hat{\sigma}: \\ \left\{\sum_{i=1}^{T_n} \mathbb{1}_{\{\hat{\sigma}_i \neq i\}} = r\right\}}} \hat{\Sigma} = \hat{\sigma}\right] \tag{A8a}$$

$$= \mathbb{P}\left[\bigcup_{r=2}^{T_n} \bigcup_{\substack{\hat{\sigma}: \\ \left\{\sum_{i=1}^{T_n} \mathbb{1}_{\{\hat{\sigma}_i \neq i\}} = r\right\}}} \sum_{i=1}^{T_n} \log \frac{P_{\hat{\sigma}_i}(X_i^n)}{P_i(X_i^n)} \geq 0\right], \tag{A8b}$$

where $\mathbb{P}$ over $\left(X_1^n, ..., X_{T_n}^n\right)$ is used when $X_i^n \overset{\text{i.i.d}}{\sim} P_{\sigma_i}, \forall i \in [T_n]$ and where (A8a) is due to the requirement that each sequence is distributed according to a distinct distribution and hence the number of incorrect distributions ranges from $[2 : T_n]$. In order to avoid considering the same set of error events multiple times in the union over $\left\{\hat{\sigma} : \sum_{i=1}^{T_n} \mathbb{1}_{\{\hat{\sigma}_i \neq i\}} = r\right\}$, we incorporate a graph theoretic interpretation of $\left\{\sum_{i=1}^{T_n} \mathbb{1}_{\{\hat{\Sigma}_i \neq i\}} = r\right\}$ in (A8b) which is used to denote the fact that we have identified r distributions incorrectly. It is apparent from Figure A1a that the event that r distributions are incorrectly identified either forms a single cycle of length r ($r = 4$ in Figure A1a) or multiple cycles with sum length equal to r (2 cycles of length 2 in Figure A1b). Hence, it is straightforward to show that we can replace the union over $\left\{\hat{\sigma} : \sum_{i=1}^{T_n} \mathbb{1}_{\{\hat{\sigma}_i \neq i\}} = r\right\}$ in (A8b) with the union over cycles of length r in a complete graph K_{T_n}, i.e., $\left\{c \in C_{T_n}^{(r)}\right\}$.
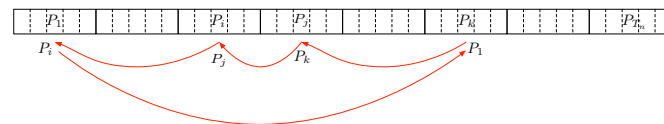
(**a**) 4 incorrectly identified distributions with a single cycle of length 4.
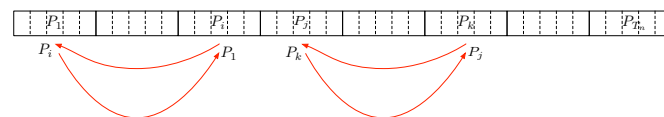
(**b**) 4 incorrectly identified distributions with 2 cycles of length 2 each.

Figure A1. *r* incorrectly identified distribution will form cycles with sum length *r*.

Hence, we can re-write the probability of error in (A8b) as

$$P_e^{(n)} = \mathbb{P}\left[\bigcup_{r=2}^{T_n} \bigcup_{c \in C_{T_n}^{(r)}} \sum_{i \in c(V)} \log \frac{P_{c(V)(i)}(X_i^n)}{P_i(X_i^n)} \geq 0\right]$$

$$\leq \sum_{r=2}^{T_n} \sum_{c \in C_{T_n}^{(r)}} e^{-n \sum_{i \in c(V)} B\left(P_{c(V)(i)}, P_i\right)}, \tag{A9}$$

where the inequality in (A9) follows from Appendix D. Next, we define $w_{i,j} := e^{-nB(P_i,P_j)}$ to be the edge weight between vertices (i,j) in the complete graph $K_{T_n}(w_{(1,2)}, \ldots w_{(K_n,1)})$ shown in Figure A2.
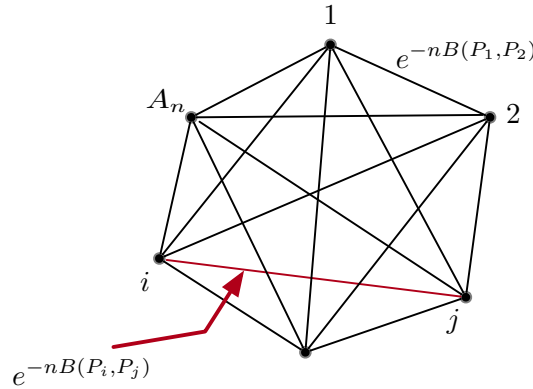
Figure A2. Complete graph K_{T_n} with edge weight $e^{-nB(P_i,P_j)}$ for every pair of vertices $i \neq j \in [K_n]$.

With this definition of the complete graph K_{T_n}, it is easy to see that $G(c) = e^{-n\sum_{i\in c(V)} B(P_{c(V)(i)},P_i)}$ is the gain of cycle c in the complete graph K_{T_n}. Based on this, we re-write the probability of error in (A9) as

$$P_e^{(n)} \leqslant \sum_{r=2}^{T_n} \sum_{c\in C_{A_n}^{(r)}} G(c) \tag{A10a}$$

$$\leqslant \sum_{r=2}^{T_n} \frac{N_{T_n}^{(r)}}{\left(\binom{T_n}{2}\right)^{\frac{r}{2}}} \left(w_{1,2}^2 + \ldots + w_{T_n,1}^2\right)^{r/2} \tag{A10b}$$

$$\leqslant \sum_{r=2}^{T_n} 4^r \left(\sum_{1\leqslant i<j\leqslant T_n} e^{-2nB(P_i,P_j)}\right)^{r/2} \tag{A10c}$$

$$\leqslant \frac{16\left(\sum_{1\leqslant i<j\leqslant T_n} e^{-2nB(P_i,P_j)}\right)}{1 - 4\sqrt{\sum_{1\leqslant i<j\leqslant T_n} e^{-2nB(P_i,P_j)}}}, \tag{A10d}$$

where r enumerates the number of incorrectly identified distributions and where (A10b) is by Lemma A1 and (A10c) is by Fact 1 (see Appendix C) and

$$\frac{N_{T_n}^{(r)}}{(n_{T_n})^{r/2}} = \frac{\binom{T_n}{r}(r-1)!/2}{\left(\binom{T_n}{2}\right)^{r/2}} \leqslant 4^r.$$

Finally, (A10d) will go to zero as n goes to infinity if

$$\lim_{n\to\infty} \sum_{1\leqslant i<j\leqslant T_n} e^{-2nB(P_i,P_j)} = 0.$$

Remark A1. *As a result of Lemma A1, it can be seen from (A10c) that the sum of probabilities that $r \geqslant 3$ distributions are incorrectly identified is dominated by the probability that only $r = 2$ distributions are incorrectly identified. This shows that the most probable error event is indeed an error event with two wrong distributions.*

Appendix B.2. Lower Bound on the Probability of Identifiability Error

For our converse, we use the optimal ML decoder, and as a lower bound to the probability of error in (A8b), we only consider the set of error events with exactly two incorrect distributions, i.e., the set of events with $r = 2$. In this case we have

$$
P_e^{(n)} \geqslant \mathbb{P}\left[\bigcup_{1 \leqslant i < j \leqslant T_n} \log \frac{P_i(X_j^n)}{P_j(X_j^n)} + \log \frac{P_j(X_i^n)}{P_i(X_i^n)} \geqslant 0 \right]
$$

$$
\geqslant \frac{\left(\sum_{1 \leqslant i < j \leqslant T_n} \mathbb{P}[\xi_{i,j}] \right)^2}{\sum_{\substack{(i,j),(j,k) \\ (i,j) \neq (l,k) \\ i \neq j, l \neq k}} \mathbb{P}[\xi_{i,j}, \xi_{k,l}]}, \tag{A11}
$$

where $\mathbb{P}$ is taken over $\left(X_1^n, ..., X_{T_n}^n \right)$ and is used when $X_i^n \overset{\text{i.i.d}}{\sim} P_{\sigma_i}, \forall i \in [T_n]$ and where (A11) is by [18] and finally

$$
\xi_{i,j} := \left\{ \log \frac{P_i}{P_j}(X_j^n) + \log \frac{P_j}{P_i}(X_i^n) \geqslant 0 \right\}. \tag{A12}
$$

We prove in Appendix E that a lower bound on $P_e^{(n)}$ is given by

$$
P_e^{(n)} \geqslant \frac{\left(\sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)} \right)^2}{\sum_{1 \leqslant i < j < k \leqslant T_n} e^{-nB(P_i,P_j)-nB(P_i,P_k)-nB(P_k,P_j)} + \left(\sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)} \right)^2} \tag{A13a}
$$

$$
\geqslant \frac{\left(\sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)} \right)^2}{8 \left(\sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)} \right)^{\frac{3}{2}} + \left(\sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)} \right)^2} \tag{A13b}
$$

$$
= \frac{\sqrt{\sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)}}}{8 + \sqrt{\sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)}}}, \tag{A13c}
$$

where (A13b) is by Lemma A1. As can be seen from (A13c), if $\lim_{n \to \infty} \sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)} \neq 0$, the probability of error is bounded away from zero. As a result

$$
\lim_{n \to \infty} \sum_{1 \leqslant i < j \leqslant T_n} e^{-2nB(P_i,P_j)} = 0,
$$

must hold, which also matches our upper bound on the probability of error in (A10d).

Appendix C. Proof of Lemma A1

We first consider the case that r is an even number and then prove

$$
r(n)^{\frac{r}{2}-1}(G(c_1) + \ldots G(c_N)) \leqslant \frac{Nr}{n} \left(a_1^2 + \ldots + a_n^2 \right)^{\frac{r}{2}}, \tag{A14}
$$

where as shorthand notation we defined and use $N := N_k^{(r)}$ and $n := \binom{k}{2}$ and $\{w_{1,2} \ldots w_{k,1}\} = \{a_1, \ldots, a_n\}$.

Our goal is to expand the right hand side (RHS) of (A14) such that all elements have coefficient 1. Then, we parse these elements into N groups (details provided later) such that using the AM-GM inequality (defined in Table 1) on each group, we obtain one of the

N terms on the LHS of (A14). Before stating the rigorous proof and to shed some light onto the proof strategy, we provide an example for the graph with $k = 4$ vertices shown in Figure A3. In this example, we consider the Lemma for $r = 4$ cycles (for which $N = 3$).
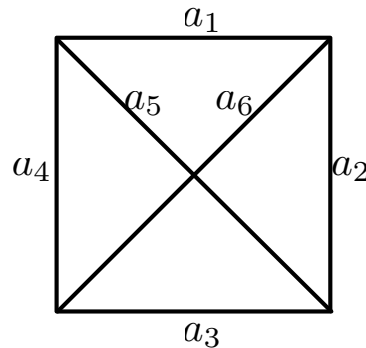

Figure A3. A complete graph with 4 vertices.

We may expand the RHS in (A14) as

$$2\left(a_1^2 + \ldots + a_6^2\right)^2 = \Theta_1 + \Theta_2 + \Theta_3,$$

$$\Theta_1 = \{a_1^4 + a_2^4 + a_3^4 + a_4^4 + a_1^2 a_3^2 + a_1^2 a_3^2 + a_2^2 a_4^2 + a_2^2 a_4^2$$
$$+ a_1^2 a_2^2 + a_1^2 a_2^2 + a_1^2 a_2^2 + a_1^2 a_2^2 + a_1^2 a_4^2 + a_1^2 a_4^2 + a_1^2 a_4^2 + a_1^2 a_4^2$$
$$+ a_2^2 a_3^2 + a_2^2 a_3^2 + a_2^2 a_3^2 + a_2^2 a_3^2 + a_3^2 a_4^2 + a_3^2 a_4^2 + a_3^2 a_4^2 + a_3^2 a_4^2\}$$

$$\Theta_2 = \{a_1^4 + a_6^4 + a_3^4 + a_5^4 + a_5^2 a_6^2 + a_5^2 a_6^2 + a_1^2 a_3^2 + a_1^2 a_3^2$$
$$+ a_1^2 a_6^2 + a_1^2 a_6^2 + a_1^2 a_6^2 + a_1^2 a_6^2 + a_1^2 a_5^2 + a_1^2 a_5^2 + a_1^2 a_5^2 + a_1^2 a_5^2$$
$$+ a_3^2 a_6^2 + a_3^2 a_6^2 + a_3^2 a_6^2 + a_3^2 a_6^2 + a_3^2 a_5^2 + a_3^2 a_5^2 + a_3^2 a_5^2 + a_3^2 a_5^2\}$$

$$\Theta_3 = \{a_4^4 + a_5^4 + a_2^4 + a_6^4 + a_5^2 a_6^2 + a_5^2 a_6^2 + a_2^2 a_4^2 + a_2^2 a_4^2$$
$$+ a_4^2 a_5^2 + a_4^2 a_5^2 + a_4^2 a_5^2 + a_4^2 a_5^2 + a_4^2 a_6^2 + a_4^2 a_6^2 + a_4^2 a_6^2 + a_4^2 a_6^2$$
$$+ a_2^2 a_5^2 + a_2^2 a_5^2 + a_2^2 a_5^2 + a_2^2 a_5^2 + a_2^2 a_6^2 + a_2^2 a_6^2 + a_2^2 a_6^2 + a_2^2 a_6^2\}.$$

If we use the AM-GM inequality on Θ_1, Θ_2 and Θ_3, we can obtain the lower bound equal to $24(a_1 a_2 a_3 a_4)$, $24(a_1 a_6 a_3 a_5)$ and $24(a_4 a_5 a_2 a_6)$, respectively where $r n^{\frac{r}{2}-1} = 24$ and hence (A14) holds in this example.

We proceed to prove Lemma A1 for arbitrary k and (even) $r \geqslant 2$. We propose the following scheme to group the elements on the RHS of (A14) and then prove that this grouping indeed leads to the claimed inequality.

Grouping Scheme

For each cycle $c_i = \{a_{i_1} \ldots, a_{i_r}\}$, we need a group of elements, Θ_i, from the RHS of (A14). In this regard, we consider all possible subsets of the edges of cycle c_i with 1 to $\frac{r}{2}$ elements (e.g., $\left\{ \{a_{i_1}\}, \ldots \{a_{i_1}, a_{i_2}\}, \ldots \{a_{i_1} \ldots, a_{i_{r/2}}\}, \ldots \right\}$). For each one of these subsets, we find the respective elements from the RHS of (A14) that is the multiplication of the elements in that subset. For example, for the subset $\{a_{i_1}, a_{i_2}, a_{i_3}\}$, we consider the elements like $a_{i_1}^{n_{i_1}} a_{i_2}^{n_{i_2}} a_{i_3}^{n_{i_3}}$ for all possible $n_{i_1}, n_{i_2}, n_{i_3} > 0$ from the RHS of (A14). However, note that we do not assign all such elements to cycle c_i only. If there are ℓ cycles of length r that all contain $\{a_{i_1}, a_{i_2}, a_{i_3}\}$, we should assign $\frac{1}{\ell}$ of the elements like $a_{i_1}^{n_{i_1}} a_{i_2}^{n_{i_2}} a_{i_3}^{n_{i_3}}$, $n_{i_1}, n_{i_2}, n_{i_3} > 0$ to cycle c_i (note the symmetry of the strategy over cycles).

We state some facts, which can be easily.

Fact 1. In a complete graph K_k, there are $N = N_k^{(r)} = \binom{k}{r} \frac{(r-1)!}{2}$ cycles of length r.

Fact 2. By expanding the RHS of (A14) such that all elements have coefficient 1, we end up with $\left(\frac{N_k^{(r)}r}{n}\right)n^{\frac{r}{2}}$ elements.

Fact 3. Expanding the RHS of (A14) such that all elements have coefficient 1, and finding their product yields

$$(a_1 \times \ldots \times a_n)^{\left(\frac{Nr}{n}\right)rn^{\frac{r}{2}-1}}.$$

Fact 4. In the above grouping scheme each element on the RHS of (A14) is summed in exactly one group. Hence, by symmetry and Fact 2, each group is the sum of $rn^{\frac{r}{2}-1}$ elements.

Now, consider any two cycles $c_i = \{a_{i_1}, \ldots, a_{i_r}\}, c_j = \{a_{j_1}, \ldots, a_{j_r}\}$. Assume that using the above grouping scheme, we obtain the group of elements Θ_i, Θ_j (where by fact 3 each one is the sum of $rn^{\frac{r}{2}-1}$ elements). If we apply the AM-GM inequality on each of the two groups, we obtain

$$\Theta_i \geqslant rn^{\frac{r}{2}-1}\left(a_{i_1}^{n_{i_1}} \times \ldots \times a_{i_r}^{n_{1_r}}\right)^{\left(\frac{1}{rn^{\frac{r}{2}-1}}\right)},$$

$$\Theta_j \geqslant rn^{\frac{r}{2}-1}\left(a_{j_1}^{n_{j_1}} \times \ldots \times a_{j_r}^{n_{j_r}}\right)^{\left(\frac{1}{rn^{\frac{r}{2}-1}}\right)},$$

where $\prod_{t=1}^r a_{i_t}^{n_{i_t}}$ is the product of the elements in Θ_i. By symmetry of the grouping scheme for different cycles, it is obvious that $\forall t \in [r], n_{i_t} = n_{j_t}$. Hence $n_{i_t} = n_{j_t} = p_t, \forall i, j \in [N]$. i.e., we have

$$\Theta_i \geqslant rn^{\frac{r}{2}-1}\left(a_{i_1}^{p_1} \times \ldots \times a_{i_r}^{p_r}\right)^{\left(\frac{1}{rn^{\frac{r}{2}-1}}\right)}. \tag{A15}$$

By symmetry of the grouping scheme over the elements of each cycle, we also get that $n_{i_k} = n_{i_l} = q_i, \forall k, l \in [r]$, i.e.,

$$\Theta_i \geqslant rn^{\frac{r}{2}-1}\left(a_{i_1}^{q_i} \times \ldots \times a_{i_r}^{q_i}\right)^{\left(\frac{1}{rn^{\frac{r}{2}-1}}\right)}. \tag{A16}$$

It can be seen from (A15) and (A16) that all elements of all groups have the same power $n_{i_t} = p, \forall i \in [N], t \in [r]$. i.e.,

$$\Theta_i \geqslant rn^{\frac{r}{2}-1}\left(a_{i_1}^{p} \times \ldots \times a_{i_r}^{p}\right)^{\left(\frac{1}{rn^{\frac{r}{2}-1}}\right)}.$$

Since each element on the RHS of (A14) is assigned to one and only one group and since $\prod_{t=1}^r a_{i_t}^{n_{i_t}} = \prod_{t=1}^r a_{i_t}^{p}$ is the product of the elements of each group Θ_i, the product of all elements in $\Theta_1 + \ldots + \Theta_N$ (which is equal to product of the elements in the expanded version of the RHS of (A14)) is $\prod_{i=1}^N \prod_{t=1}^r a_{i_t}^{p}$.

In addition, since each a_i appears in exactly $\frac{Nr}{n}$ of the cycles, by Fact 3 and a double counting argument, we have

$$p \times \frac{Nr}{n} = \left(\frac{Nr}{n}\right)rn^{\frac{r}{2}-1},$$

and hence $p = rn^{\frac{r}{2}-1}$. Hence, the lower bound of the AM-GM inequality on the $\Theta_1 + \ldots + \Theta_N$, will result in

$$rn^{\frac{r}{2}-1}G(c_1) + \ldots + rn^{\frac{r}{2}-1}G(c_{N_k^{(r)}}),$$

and the Lemma is proved for even r.

For odd values of r, the problem that may arise by using the grouping strategy in its current form, is when $r < \frac{k}{2}$. In this case, some of the terms on the RHS of (A14) may contain

multiplication of a_i's that are not present in any of the $G(c_i)$'s. To overcome this, take both sides to the power of $2m$ for the smallest m such that $rm > \frac{k}{2}$. Then the RHS of (A14) is at most the multiplication of rm different a_i's and on the LHS of (A14), there are $2m$ cycles of length r multiplied together. By our choice of $2m$, now, all possible combinations of a_i's on the RHS are present in at least one cycle multiplication in the LHS. Hence, it is straightforward to use the same strategy as even values of r to prove the theorem for the odd values of r.

Appendix D. Proof of (A9)

The inequality in (A9) is by

$$\mathbb{P}\left[\sum_{i \in c^{(V)}} \log \frac{P_{c^{(V)}(i)}\left(X_i^n\right)}{P_i\left(X_i^n\right)} \geqslant 0\right] \tag{A17a}$$

$$\leqslant \exp\left\{n \inf_t \log \mathbb{E}\left[\prod_{i \in c^{(V)}} \left(\frac{P_{c^{(V)}(i)}\left(X_i^n\right)}{P_i\left(X_i^n\right)}\right)^t\right]\right\}$$

$$\leqslant \exp\left\{n \sum_{i \in c^{(V)}} \log \mathbb{E}\left[\left(\frac{P_{c^{(V)}(i)}\left(X_i^n\right)}{P_i\left(X_i^n\right)}\right)^{1/2}\right]\right\} \tag{A17b}$$

$$= \exp\left\{-n \sum_{i \in c^{(V)}} B\left(P_{c^{(V)}(i)}, P_i\right)\right\},$$

where (A17a) is by the Chernoff inequality. The fact that we used $t = 1/2$ in (A17b) instead of finding the exact optimizing t, comes from the fact that $t = 1/2$ is the optimal choice for $r = 2$ and as we will explain in Remark A1. The rest of the error events are dominated by the set of only 2 incorrectly identified distributions. This can be seen as follows for $X_1^n \overset{\text{i.i.d}}{\sim} P_1, X_2^n \overset{\text{i.i.d}}{\sim} P_2$

$$\mathbb{P}\left[\log \frac{P_1(X_2^n)}{P_2(X_2^n)} + \log \frac{P_2(X_1^n)}{P_1(X_1^n)} \geqslant 0\right]$$

$$= \sum_{\substack{\hat{P}_1, \hat{P}_2: \\ \sum_{x \in \mathcal{X}} \hat{P}_1(x) \log \frac{P_2(x)}{P_1(x)} + \\ \hat{P}_2(x) \log \frac{P_1(x)}{P_2(x)} \geqslant 0}} \exp\left\{-nD\left(\hat{P}_1 \parallel P_1\right) - nD\left(\hat{P}_2 \parallel P_2\right)\right\}$$

$$\doteq e^{-nD\left(\tilde{P} \parallel P_1\right) - nD\left(\tilde{P} \parallel P_2\right)} = e^{-2nB(P_1, P_2)}, \tag{A18}$$

where $\tilde{P}$ in the first equality in (A18), by using the Lagrangian method, can be shown to be equal to $\tilde{P}(x) = \frac{\sqrt{P_1(x)P_2(x)}}{\sum_{x'} \sqrt{P_1(x')P_2(x')}}$ and subsequently the second inequality in (A18) is proved.

Appendix E. Proof of (A13a)

We upper bound the denominator of (A11) by

$$\mathbb{P}[\xi_{i,j}, \xi_{i,k}] = \mathbb{P}\left[\log \frac{P_i(X_j^n)}{P_j(X_j^n)} + \log \frac{P_j(X_i^n)}{P_i(X_i^n)} \geqslant 0 \cap \log \frac{P_i(X_k^n)}{P_k(X_k^n)} + \log \frac{P_k(X_i^n)}{P_i(X_i^n)} \geqslant 0\right]$$

$$\leqslant \mathbb{P}\left[\log \frac{P_i(X_j^n)}{P_j(X_j^n)} + \log \frac{P_j(X_i^n)}{P_i(X_i^n)} + \log \frac{P_i(X_k^n)}{P_k(X_k^n)} + \log \frac{P_k(X_i^n)}{P_i(X_i^n)} \geqslant 0\right]$$

$$\leqslant \exp\left\{n \inf_t \log\left(\mathbb{E}\left[\left(\frac{P_i(X_j^n)}{P_j(X_j^n)} \cdot \frac{P_j(X_i^n)}{P_i(X_i^n)} \cdot \frac{P_i(X_k^n)}{P_k(X_k^n)} \cdot \frac{P_k(X_i^n)}{P_i(X_i^n)}\right)^t\right]\right)\right\}$$

$$\leqslant \exp\left\{n \log \mathbb{E}\left[\left(\frac{P_i(X_j^n)}{P_j(X_j^n)} \cdot \frac{P_j(X_i^n)}{P_i(X_i^n)} \cdot \frac{P_i(X_k^n)}{P_k(X_k^n)} \cdot \frac{P_k(X_i^n)}{P_i(X_i^n)}\right)^{\frac{1}{2}}\right]\right\}$$

$$= \exp\{-nB(P_i, P_j) - nB(P_j, P_k) - nB(P_i, P_k)\}. \tag{A19}$$

An upper bound for $\mathbb{P}[\xi_{i,j}, \xi_{k,l}]$ can be derived similarly.

Appendix F. Proof of Theorem 3

We propose a three-stage achievability scheme where we should again note that with similar arguments as the ones in Theorem 1, by imposing the first bound in (24), we can mainly focus on the no-collision hypothesis. The three achievability stages perform the tasks of synchronization, identification, and decoding, respectively and the joint probability of error can be decomposed as

$$\mathbb{P}[\text{Error}] \leqslant \mathbb{P}[\text{ Synchronization error | No-collision }]$$
$$+ \mathbb{P}[\text{ Identification error | No synchronization error, No-collision }]$$
$$+ \mathbb{P}[\text{ Decoding error | No synchronization, No identification error, No-collision}]$$
$$+ \mathbb{P}[\text{collision}].$$

Appendix F.1. Codebook Generation

Let $K_n = e^{n\nu}$ be the number of users, $A_n = e^{n\alpha}$ be the number of blocks, $M_n = e^{nR}$ be the number of messages, and $|\mathcal{W}| = \text{poly}(n)$ be the number of channels. Each user $i \in [K_n]$ generates a random i.i.d codebook according to distribution $P_{c(i)}$ where the index $c(i) \in [|\mathcal{W}|]$ is chosen based on the channel $Q_i = W_{c(i)}$. For each user $i \in [K_n]$, the codeword for each message $m \in [M_n]$ is denoted as $x_i^n(m)$.

Appendix F.2. Probability of Error Analysis

A three-stage decoder is used. We now introduce the three stages and bound the probability of error for each stage.

Synchronization Step. We perform a sequential likelihood ratio test for synchronization as follows. Recall $Q_i(\cdot|\cdot) = W_{c(i)}(\cdot|\cdot)$ for all user $i \in [K_n]$. Fix thresholds

$$T_{c(i)} \in \left[-D\left(Q_\star \parallel [P_{c(i)} W_{c(i)}]\right), D\left([P_{c(i)} W_{c(i)}] \parallel Q_\star\right)\right], \ i \in [K_n]. \tag{A20}$$

For each block $j \in [A_n]$ if there exists any user $i \in [K_n]$ such that

$$L_{c(i)}(y_j^n) := \frac{1}{n} \log \frac{[P_{c(i)} W_{c(i)}](y_j^n)}{Q_\star(y_j^n)} \geqslant T_{c(i)}, \tag{A21}$$

then declare that block j is an 'active' block. Else, declare that block j is an 'idle' block. Note that were able to calculate the probabilities of error corresponding to (A3) by leveraging the constant composition construction of codewords in Theorem 1. In here, we can leverage the i.i.d. construction of the codewords and calculate the probability of error corresponding to (A21).

We now find an upper bound on the average probability of synchronization error for this scheme over different hypotheses. With the same argument as the one after (A4) we have

$$\mathbb{P}\Big[\text{Synchronization error}|H^{(1)}\Big]$$

$$\leqslant \mathbb{P}\left[\bigcup_{i=1}^{K_n} L_{c(i)}(Y_i^n) < T_{c(i)}|H^{(1)}\right] + \mathbb{P}\left[\bigcup_{s=K_n+1}^{A_n} \bigcup_{i=1}^{K_n} L_{c(i)}(Y_s^n) \geqslant T_{c(i)}|H^{(1)}\right]$$

$$\leqslant \sum_{i=1}^{K_n} \mathbb{P}\Big[L_{c(i)}(Y_i^n) < T_{c(i)}|H^{(1)}\Big] + e^{n\alpha}\mathbb{P}\left[\bigcup_{i=1}^{K_n} L_{c(i)}(Y^n) \geqslant T_{c(i)}|H^{(1)}\right]$$

$$= \sum_{i=1}^{K_n} \mathbb{P}\Big[L_{c(i)}(Y_i^n) < T_{c(i)}|H^{(1)}\Big] + e^{n\alpha}\mathbb{P}\left[\bigcup_{j=1}^{|\mathcal{W}|} L_j(Y^n) \geqslant T_j|H^{(1)}\right] \tag{A22a}$$

$$\leqslant \sum_{i=1}^{K_n} e^{-nD\Big([P_{c(i)}W_{c(i)}]_{\lambda_i}\|[P_{c(i)}W_{c(i)}]\Big)} + e^{n\alpha}\sum_{j=1}^{|\mathcal{W}|} e^{-nD\Big([P_jW_j]_{\lambda_j}\|Q_\star\Big)} \tag{A22b}$$

$$= \sum_{j=1}^{|\mathcal{W}|} \mathcal{N}_j e^{-nD\Big([P_jW_j]_{\lambda_j}\|[P_jW_j]\Big)} + e^{n\alpha}e^{-nD\Big([P_jW_j]_{\lambda_j}\|Q_\star\Big)}, \tag{A22c}$$

where (A22a) is by taking $j := c(i)$ and (A22b) is by the Chernoff inequality in which $[P_jW_j]_\lambda$ was defined in (26) and where (A22c) is again by taking $j := c(i)$ and re-counting the events.

The probability of error in this stage will decay to zero as blocklength n goes to infinity if for all $j \in [|\mathcal{W}|]$ and for a $\epsilon > 0$

$$\nu_j + \epsilon < D\Big([P_jW_j]_{\lambda_j} \parallel [P_jW_j]\Big), \tag{A23a}$$

$$\alpha + \epsilon < D\Big([P_jW_j]_{\lambda_j} \parallel Q_\star\Big). \tag{A23b}$$

This corresponds to the second and third bounds in (24).

Identification Step. Having found the location of the 'active' blocks, we move on to the second stage of the achievability scheme to identify which user is active in which block. We note that, by the random codebook generation and the memoryless property of the channel, the output of the block occupied by user $i \in [K_n]$ is i.i.d distributed according to the marginal distribution

$$[P_{c(i)}Q_{c(i)}](y) := \sum_{x\in\mathcal{X}} P_{c(i)}(x)Q_{c(i)}(y|x).$$

We leverage this property and customize the result in Theorem 2 to identify the different distributions of the different users. Note that at this point, we only distinguish the users with different channels from one another. Users with the same channel are distinguished from each other in the decoding stage as will be explained next. In Theorem 2, it was assumed that all the distributions are distinct, while in here, our distributions are not necessarily distinct. The only modification that is needed in order to use the result of Theorem 2 is as follows. We need to consider a graph in which the edge between every two similar distributions has edge weights equal to zero (as opposed to $e^{B(P,P)} = e^0 = 1$). By doing so, we can easily conclude that the probability of identification error in our problem using an ML decoder is upper bounded by

$$\mathbb{P}[\text{ Identification error } | \text{ No synchronization error, No-collision }]$$

$$\leqslant \sum_{1 \leqslant i < j \leqslant |\mathcal{W}|} e^{-2nB\big([P_i W_i],[P_j W_j]\big)} \tag{A24}$$

$$\leqslant \sum_{1 \leqslant i < j \leqslant |\mathcal{W}|} e^{-2n \inf_{i,j} B\big([P_i W_i],[P_j W_j]\big)} \tag{A25}$$

which goes to zero for $|\mathcal{W}| = \mathrm{poly}(n)$ if $0 < \inf_{i,j} B\big([P_i W_i],[P_j W_j]\big)$. This corresponds to the fourth bound in (24).

Decoding Step. After finding the permutation of users in the active blocks, we can go ahead with the third stage of the achievable scheme to find the transmitted messages of the users using a maximum likelihood decoder. In this stage, we can group each set of users (with group size $\mathcal{N}_j$, $j \in [|\mathcal{W}|]$ defined in (25)) who have similar channel W_j. In each block of each group we have to distinguish among $e^{R+\nu_j}$ messages and with similar steps as the proof of Theorem 1, we have

$$\mathbb{P}[\text{ Decoding error } | \text{ No synchronization, No identification error, No-collision}]$$

$$\leqslant \sum_{j=1}^{|\mathcal{W}|} \mathbb{P}\Big[\substack{\text{Decoding error for} \\ \text{users with channel } W_j} \Big| \text{ No synchronization and identification error, No-collision}\Big]$$

$$\leqslant \sum_{j=1}^{|\mathcal{W}|} \mathcal{N}_j e^{-nE_r\big(R+\nu_j, P_j, W_j\big)} = \sum_{j=1}^{|\mathcal{W}|} e^{n\big(\nu_j - E_r\big(R+\nu_j, P_j, W_j\big)\big)}, \tag{A26}$$

which retrieves the second bound in (24). Finally, as for a synchronous channel and since we have to distinguish among $e^{R+\nu_j}$ messages in each block for some $\epsilon > 0$, we must have

$$R + \nu_j + \epsilon < I(P_j, W_j), \forall j \in [|\mathcal{W}|], \tag{A27}$$

which corresponds to the last bound in (24).

Finally to conclude the proof, we should note that the bounds derived in (A23), (A26) and (A27) are all achievable bounds for each user for large enough block length n (say for $n \geqslant n_0$). Since we need the intersection of the achievable bounds for all users while the number of users itself increases with n, we need to express the asymptotic bound with the definition of n going to infinity. By expressing the achievable set for all users as $\mathcal{A}_n$ defined in (24) by the equivalence of $\liminf_{n\to\infty} \mathcal{A}_n \equiv \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \mathcal{A}_n$ and by noting that for some n_0 every region $\mathcal{A}_{n \geqslant n_0}$ is achievable by our achievability scheme previously introduced here, we see that (23) is indeed achievable.

Remark A2. *Note that the existence of $\epsilon > 0$ in our bounds are due to the fact that we have a summation that depends on n and our exponent also decays exponentially with n while the n multiplier depends on the summation arguments. By including a constant $\epsilon > 0$ that multiplies n, we guarantee that the summation as a whole decays to zero.*

Remark A3. *The achievability proof of Theorem 1 relies on constant composition codes whereas the achievability proof of Theorem 3 relies on i.i.d. codebooks. The reason for these restrictions is that in Theorem 3 we also need to distinguish different users and in order to use the result of [13], we focused our attention on i.i.d. codebooks.*

Appendix G. Proof of Theorem 4

For the proof of Theorem 4, we use an ML decoder sequentially to perform the synchronization, identification and decoding all together, for each block.

Proof. *Codebook generation:* Each user $i \in [K_n]$ generates an i.i.d. random codebook according to the distribution P_i.

Probability of error analysis: For each block $s \in [A_n]$, the decoder outputs

$$i^* = \arg \max_{i \in [0:K_n], m \in [M_n]} Q_i(y_s^n | x_i^n(m)),$$

where $x_0^n = \varnothing$.

We now find an upper bound on the probability of error given the hypothesis $H^{(1)}$ in (A4) for this decoder as follows

$$P_e^{(n)} \leqslant \sum_{i \in [K_n]} \sum_{m \in [2:M_n]} \mathbb{P}\left[\log \frac{Q_i(Y_i^n \mid x_i^n(m))}{Q_i(Y_i^n \mid x_i^n(1))} > 0 | H^{(1)} \right] \tag{A28a}$$

$$+ \sum_{i \in [K_n]} \sum_{\substack{j \in [0:K_n] \\ j \neq i}} \sum_{m \in [M_n]} \mathbb{P}\left[\log \frac{Q_j(Y_i^n | x_j^n(m))}{Q_i(Y_i^n | x_i^n(1))} > 0 | H^{(1)} \right] \tag{A28b}$$

$$+ \sum_{s \in [K_n+1:A_n]} \sum_{j \in [K_n]} \sum_{m \in [M_n]} \mathbb{P}\left[\log \frac{Q_j(Y_s^n | x_j^n(m))}{Q_\star(Y_s^n)} > 0 | H^{(1)} \right] \tag{A28c}$$

$$\leqslant \sum_{i \in [K_n]} e^{nR} e^{-n \sup_t - \log \mathbb{E}\left[\left(\frac{Q_i(Y_i | \overline{X}_i)}{Q_i(Y_i | X_i)} \right)^t \right]}$$

$$+ \sum_{i \in [K_n]} \sum_{\substack{j \in [0:K_n] \\ j \neq i}} e^{nR} e^{-n \sup_t - \log \mathbb{E}\left[\left(\frac{Q_j(Y_i | X_j)}{Q_i(Y_i | X_i)} \right)^t \right]}$$

$$+ e^{n\alpha} \sum_{j \in [K_n]} e^{nR} e^{-n \sup_t - \log \mathbb{E}\left[\left(\frac{Q_j(Y_s | X_j)}{Q_\star(Y_s)} \right)^t \right]},$$

where $P_{X,\overline{X}}(x, x') = P_X(x)P_X(x')$. The term given in (A28a) is due to decoding error, (A28b) is due to identification error and finally (A28c) is due to synchronization error. The last inequality is by the Chernoff bound. In order for each term in the probability of error upper bound to vanish as n grows to infinity and by the explanation in the proof of Theorem 3 for the inclusion of $n \to \infty$ in our bounds, we find the conditions stated in the theorem. $\square$

Appendix H. Proof of (28)

To see (28a), we note that due to the symmetry in

$$C(P, Q, P, Q) = \sup_t - \log \sum_{x,x',y} P(x)P(x')Q(y|x)^{1-t}Q(y|x')^t,$$

the supremum is achieved at the midpoint $t = \frac{1}{2}$. hence $C(P, Q, P, Q) = \mu(\frac{1}{2})$. Moreover, we find an upper bound on $C(\,.\,, Q_\star, P_i, Q_i)$ by noting that $\mu_{0,i}(t)$ defined in (12b) is concave in t with $\mu_{0,i}(1) = 0$ and

$$\frac{\partial \mu_{0,i}(t)}{\partial t} \Big|_{t=1} = -I(P_i, Q_i) - D([P_iQ_i] \parallel Q_\star) \leqslant 0.$$

Hence $\mu_{0,i}(t)$ is always less than $(I(P_i, Q_i) + D([P_iQ_i] \parallel Q_\star))(1-t)$ and for $0 \leqslant t \leqslant 1$ it is always less than $I(P_i, Q_i) + D([P_iQ_i] \parallel Q_\star)$.

Appendix I. Proof of Theorem 5

We first define the following special shorthand notations that we will use throughout this proof

$$F_n := M_n K_n, \qquad\qquad\qquad Q_{i,m}^n(y^n) := Q_i(y^n | x_i^n(m)), \tag{A29a}$$

$$\overline{P_{Y^n}}(y^n) := \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m}^n(y^n), \qquad Q_{i,m_{\lambda_i}}^n(y^n) := \frac{\left(Q_{i,m}^n(y^n) \right)^{\lambda_i} (Q_\star^n(y^n))^{1-\lambda_i}}{\sum_{y^n} \left(Q_{i,m}^n(y^n) \right)^{\lambda_i} (Q_\star^n(y^n))^{1-\lambda_i}}, \tag{A29b}$$

$$\overline{P_{Y^n}^{(\lambda)}}(y^n) := \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m_{\lambda_i}}^n(y^n), \quad \left(\overline{P_{Y^n}} \right)_t(y^n) := \frac{\left(\overline{P_{Y^n}}(y^n) \right)^t (Q_\star^n(y^n))^{1-t}}{\sum_{y^n} \left(\overline{P_{Y^n}}(y^n) \right)^t (Q_\star^n(y^n))^{1-t}}, \tag{A29c}$$

$$Q_\star^n(y^n) := \prod_{i=1}^{n} Q_\star(y_i). \tag{A29d}$$

We use the optimal ML decoder to find the location of the 'active' blocks. In this stage, we are not concerned about the identity or the message of the users. In this regard, the decoder output is determined via

$$\arg \max_{\substack{(l_1, \dots, l_{K_n}) \\ l_i \neq l_j, \forall i \neq j \\ l_i \in [A_n], i \in [K_n]}} \sum_{i=1}^{K_n} \log \frac{\overline{P_{Y^n}}(Y_{l_i}^n)}{Q_\star^n(Y_{l_i}^n)}. \tag{A30}$$

Given the hypothesis that the users are active in blocks $[K_n]$, denoted by $H^{(1)}$ in (A4), the corresponding error events for the ML decoder are given by

$$\left\{ \text{error} | H^{(1)} \right\} = \bigcup_{\substack{(l_1, \dots, l_{K_n}) \\ \neq (1, \dots, K_n)}} \left\{ \sum_{i=1}^{K_n} \log \frac{\overline{P_{Y^n}}(Y_{l_i}^n)}{Q_\star^n(Y_{l_i}^n)} > \sum_{i=1}^{K_n} \log \frac{\overline{P_{Y^n}}(Y_i^n)}{Q_\star^n(Y_i^n)} \right\}$$

$$\supseteq \bigcup_{\substack{i \in [K_n] \\ j \in [K_n+1:A_n]}} \left\{ \log \frac{\overline{P_{Y^n}}(Y_j^n)}{Q_\star^n(Y_j^n)} \geqslant \log \frac{\overline{P_{Y^n}}(Y_i^n)}{Q_\star^n(Y_i^n)} \right\}$$

$$\supseteq \left\{ \bigcup_{j \in [K_n+1:A_n]} \log \frac{\overline{P_{Y^n}}(Y_j^n)}{Q_\star^n(Y_j^n)} \geqslant T \right\} \bigcap \left\{ \bigcup_{i \in [K_n]} T \geqslant \log \frac{\overline{P_{Y^n}}(Y_i^n)}{Q_\star^n(Y_i^n)} \right\}, \tag{A31}$$

for any $T \in \mathbb{R}$. We take T to be

$$T := \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \left(D\left(Q_{i,m_{\lambda_i}}^n \| Q_\star^n \right) - D\left(Q_{i,m_{\lambda_i}}^n \| \overline{P_{Y^n}} \right) \right), \tag{A32}$$

for different $\lambda_i \in [0,1], i \in [K_n]$.

We also find the following lower bounds

$$Q_\star^n \left[\log \frac{\overline{P_{Y^n}}}{Q_\star^n}(Y^n) \geqslant T \right] \geqslant e^{-\frac{n}{K_n} \left(\sum_{i=1}^{K_n} D(Q_{i\lambda_i} \| Q_\star | P_i) - (R+\nu)(1-\bar{r}_n) + \frac{h(\bar{r}_n)}{n} \right)}, \tag{A33a}$$

$$\overline{P_{Y^n}} \left[\log \frac{\overline{P_{Y^n}}}{Q_\star^n}(Y^n) \leqslant T \right] \geqslant e^{-\frac{n}{K_n} \sum_{i=1}^{K_n} D(Q_{i\lambda_i} \| Q_i | P_i)}, \tag{A33b}$$

which are proved in Appendix K.

By using the inequalities in (A33a) and (A33b), we find a lower bound on the probability of (A31) as follows

$$\mathbb{P}\left[\bigcup_{j\in[K_n+1A_n]}\log\frac{\overline{P_{Y^n}}(Y_j^n)}{Q_\star^n(Y_j^n)}\geqslant T\cap\bigcup_{i\in[K_n]}T\geqslant\log\frac{\overline{P_{Y^n}}(Y_i^n)}{Q_\star^n(Y_i^n)}\Big|H^{(1)}\right]\tag{A34a}$$

$$=\mathbb{P}\left[\bigcup_{j\in[K_n+1:A_n]}\log\frac{\overline{P_{Y^n}}(Y_j^n)}{Q_\star^n(Y_j^n)}\geqslant T|H^{(1)}\right]\mathbb{P}\left[\bigcup_{i\in[K_n]}T\geqslant\log\frac{\overline{P_{Y^n}}(Y_i^n)}{Q_\star^n(Y_i^n)}|H^{(1)}\right]\tag{A34b}$$

$$=:\mathbb{P}[Z_1\geqslant1]\mathbb{P}[Z_2\geqslant1]\tag{A34c}$$

$$\geqslant\left(1-\frac{\mathrm{Var}[Z_1]}{\mathbb{E}^2[Z_1]}\right)\left(1-\frac{\mathrm{Var}[Z_2]}{\mathbb{E}^2[Z_2]}\right)\tag{A34d}$$

$$\geqslant\left(1-\frac{1}{\sum_{j=K_n+1}^{A_n}\mathbb{P}[\xi_j=1]}\right)\left(1-\frac{1}{\sum_{i=1}^{K_n}\mathbb{P}[\zeta_i=1]}\right)\tag{A34e}$$

$$\geqslant\left(1-e^{-n\alpha+\frac{n}{K_n}\left(\sum_{i=1}^{K_n}D(Q_{i\lambda_i}\|Q_\star|P_i)-(R+\nu)(1-\bar{r}_n)+\frac{h(\bar{r}_n)}{n}\right)}\right)\left(1-e^{-n\nu+\frac{n}{K_n}\sum_{i=1}^{K_n}D(Q_{i\lambda_i}\|Q_i|P_i)}\right),\tag{A34f}$$

where (A34b) follows by independence of Y_i^n and Y_j^n whenever $i\neq j,\forall i,j\in[A_n]$ and the inequality in (A34d) is by Chebyshev's inequality, where we have defined

$$Z_1:=\sum_{j=K_n+1}^{A_n}\xi_j,\qquad\xi_j:=Ber\left(Q_\star^n\left[\log\frac{\overline{P_{Y^n}}}{Q_\star^n}(Y_j^n)\geqslant T\right]\right),\tag{A35a}$$

$$Z_2:=\sum_{i=1}^{K_n}\zeta_i,\qquad\zeta_i:=Ber\left(\overline{P_{Y^n}}\left[\log\frac{\overline{P_{Y^n}}}{Q_\star^n}(Y_i^n)\leqslant T\right]\right),\tag{A35b}$$

where $\{\xi_j,\zeta_i\}$ are mutually independent. We can see from (A34f) that if for some $\epsilon>0$

$$\nu>\frac{1}{K_n}\sum_{i=1}^{K_n}D(Q_{i\lambda_i}\|Q_i|P_i)+\epsilon,\tag{A36a}$$

$$\alpha>\frac{1}{K_n}\sum_{i=1}^{K_n}D(Q_{i\lambda_i}\|Q_\star|P_i)-(1-\bar{r}_n)(\nu+R)+\epsilon,\tag{A36b}$$

then the probability of error is strictly bounded away from zero and hence the region specified by (A36) (first bound in (31)) is impermissible.

We now focus only on decoding error probability, with the assumption that we have identified the active blocks. This is hence a lower bound on the overall probability of error. We have

$$\mathbb{P}[\text{Decoding error }|\text{No synchronization error}]\tag{A37a}$$

$$=\mathbb{P}\left[\bigcup_{i=1}^{K_n}\text{Decoding error in block }i\right]\tag{A37b}$$

$$\geqslant1-\frac{1}{\sum_{i=1}^{K_n}\mathbb{P}[D_i^c]}\tag{A37c}$$

$$\geqslant1-\frac{1}{\sum_{i=1}^{K_n}e^{-nE_{sp}(R,P_i,Q_i)}},\tag{A37d}$$

$$\geqslant1-\frac{1}{K_ne^{-n\frac{1}{K_n}\sum_{i=1}^{K_n}E_{sp}(R,P_i,Q_i)}},\tag{A37e}$$

$$=1-e^{-n\left(\nu-\frac{1}{K_n}\sum_{i=1}^{K_n}E_{sp}(R,P_i,Q_i)\right)}\tag{A37f}$$

where

$$D_i^c := \text{ Decoding error in block } i,$$

and where (A37c) can be derived using the same technique used in (A34). Finally, (A37d) is by lower bounding the probability of decoding error as in [14] where we defined E_{sp} in (14). (A37e) is by Jensen's inequality and finally as it can be seen in (A37f), if $\nu > \left(\frac{1}{K_n} \sum_{i=1}^{K_n} E_{sp}(R, P_i, Q_i) \right) + \epsilon$ for an $\epsilon > 0$ the lower bound to the probability of decoding error is bounded away from zero for large values of n and this retrieves the second bound in (31). In addition, the usual converse bound on the rate of a synchronous channel also applies to any asynchronous channel and hence the region where $R > I(P, Q) + \epsilon$ is also impermissible. This concludes the proof.

Appendix J. Proof of Theorem 6

User $i \in [K_n]$ has a codebook with $M_n = e^{nR}$ codewords of length n. Define for $i \in [K_n]$ an 'extended codebook' consisting of $A_n M_n$ codewords of length $n A_n$ constructed such that $\forall m \in [M_n]$ and $\forall t_i \in [A_n]$

$$\widetilde{X}_i^{nA_n}(m_i, t_i) := \left[\, \star_i^{n(t_i-1)} \; f_i(m_i) \; \star_i^{n(A_n-t_i)} \, \right],$$

as depicted in Figure A4.
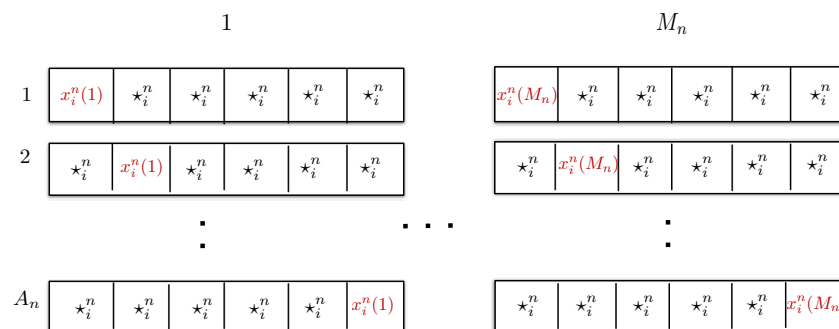
Figure A4. Extended codebook.

By using Fano's inequality, i.e., $H(X_1^{nA_n}, \ldots, X_{K_n}^{nA_n}|Y^{nA_n}) \leqslant n\epsilon_n : \epsilon_n \to 0$ as $n \to \infty$, for any codebook of length $n A_n$ we have

$$
\begin{aligned}
H(X_1^{nA_n}, \ldots, X_{K_n}^{nA_n}) &= H(m_1, t_1, \ldots, m_{K_n}, t_{K_n}) \\
&= nK_n(\alpha + R) \\
&= H(X_1^{nA_n}, \ldots, X_{K_n}^{nA_n}|Y^{nA_n}) + I(X_1^{nA_n}, \ldots, X_{K_n}^{nA_n}; Y^{nA_n}) \\
&\leqslant n\epsilon_n + ne^{n\alpha} |\mathcal{Y}| \quad \Longleftrightarrow \\
\nu + \frac{\log\left(1 + \frac{1}{\alpha K_n} \sum_{i \in [K_n]} R_i\right)}{n} &\leqslant \alpha + \frac{\log\left(1 + \frac{\epsilon_n}{e^{n\alpha}|\mathcal{Y}|}\right)}{n},
\end{aligned}
$$

where $\frac{\log\left(1 + \frac{1}{\alpha K_n} \sum_{i \in [K_n]} R_i\right)}{n} \geqslant 0$ and $\frac{\log\left(1 + \frac{\epsilon_n}{e^{n\alpha}|\mathcal{Y}|}\right)}{n} \geqslant 0$ vanish as n goes to infinity.

Appendix K. Proof of (A33a) and (A33b)

Before deriving lower bounds on (A33a) and (A33b), we note that by the Type-counting Lemma [19], at the expense of a small decrease in the rate (which vanishes in the limit for large blocklength) we may restrict our attention to constant composition codewords. Henceforth, we assume that the composition of the codewords for user $i \in [K_n]$ is given by P_i. Moreover, to make this paper self-contained, we restate the following Lemmas that we use in the rest of the proof.

Lemma A2 (Compensation Identity). *For arbitrary $\pi_i : \sum_{i=1}^{K} \pi_i = 1$ and arbitrary probability distribution functions $P_i \in \mathcal{P}_\mathcal{X}, i \in [K]$, we define $\bar{P}(x) = \sum_{i=1}^{K} \pi_i P_i(x)$. Then for any probability distribution function R we have*

$$D(\bar{P} \parallel R) + \sum_{i=1}^{K} \pi_i D(P_i \parallel \bar{P}) = \sum_{i=1}^{K} \pi_i D(P_i \| R). \tag{A38}$$

Lemma A3 (Fano). *Let F be an arbitrary set of size N. For $\bar{P} = \frac{\sum_{\theta \in F} P_\theta}{N}$ we have*

$$\frac{1}{N} \sum_{\theta \in F} D(P_\theta \parallel \bar{P}) \geqslant (1 - \bar{r}) \log(N(1 - \bar{r})) + \bar{r} \log\left(\frac{N\bar{r}}{N - 1}\right), \tag{A39}$$

where

$$\bar{r} := \inf_T \frac{1}{N} \sum_{\theta \in F} P_\theta \{T \neq \theta\} \tag{A40}$$

in which the infimum is taken over all possible estimators T.

We now continue with the proof of (A33a). Using the Chernoff bound we can write

$$Q_\star^n \left[\log \frac{\overline{P_{Y^n}}}{Q_\star^n}(Y^n) \geqslant \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \parallel Q_\star^n\right) - D\left(Q_{i,m_{\lambda_i}}^n \parallel \overline{P_{Y^n}}\right) \right] \overset{\text{Chernoff}}{\doteq} e^{-\sup_t A(t)}. \tag{A41}$$

The Chernoff bound exponent, $\sup_t A(t)$, is expressed and simplified as follows

$$
\begin{aligned}
A(t) &:= \frac{t}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \parallel Q_\star^n\right) - D\left(Q_{i,m_{\lambda_i}}^n \parallel \overline{P_{Y^n}}\right) - \log \mathbb{E}_{Q_\star^n}\left[\left(\frac{\overline{P_{Y^n}}}{Q_\star^n}(Y^n)\right)^t \right] \\
&= \frac{t}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m_{\lambda_i}}^n(y^n) \log \frac{\overline{P_{Y^n}}(y^n)}{Q_\star^n(y^n)} - \log \sum_{y^n} \left(\overline{P_{Y^n}}(y^n)\right)^t (Q_\star^n(y^n))^{1-t} \\
&= \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m_{\lambda_i}}^n(y^n) \log \frac{\frac{\left(\overline{P_{Y^n}}(y^n)\right)^t (Q_\star^n(y^n))^{1-t}}{\sum_{y^n} \left(\overline{P_{Y^n}}(y^n)\right)^t (Q_\star^n(y^n))^{1-t}}}{Q_\star^n(y^n)} \\
&= \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m_{\lambda_i}}^n(y^n) \log \frac{\left(\overline{P_{Y^n}}\right)_t(y^n)}{Q_\star^n(y^n)} \\
&= \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \parallel Q_\star^n\right) - D\left(Q_{i,m_{\lambda_i}}^n \parallel \left(\overline{P_{Y^n}}\right)_t\right) \\
&= \frac{1}{K_n} \sum_{i=1}^{K_n} n D\left(Q_{i\lambda_i} \parallel Q_\star | P_i\right) - \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \parallel \left(\overline{P_{Y^n}}\right)_t\right), \tag{A42}
\end{aligned}
$$

where (A42) is the result of the constant composition structure of the codewords. As a result,

$$\sup_t A(t) = \frac{1}{K_n} \sum_{i=1}^{K_n} n D\left(Q_{i\lambda_i} \parallel Q_\star | P_i\right) - \inf_t \left\{ \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \parallel \left(\overline{P_{Y^n}}\right)_t\right) \right\}. \tag{A43}$$

Moreover,

$$\inf_t \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \,\|\, (\overline{P_{Y^n}})_t\right)$$

$$= \inf_t \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m_{\lambda_i}}^n(y^n) \log \frac{Q_{i,m_{\lambda_i}}^n(y^n)}{(\overline{P_{Y^n}})_t(y^n)}$$

$$= \inf_t \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m_{\lambda_i}}^n(y^n) \log \frac{Q_{i,m_{\lambda_i}}^n(y^n)}{(\overline{P_{Y^n}})_t(y^n)} \cdot \frac{\overline{P_{Y^n}^{(\lambda)}}(y^n)}{\overline{P_{Y^n}^{(\lambda)}}(y^n)}$$

$$= \inf_t \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \,\|\, \overline{P_{Y^n}^{(\lambda)}}\right) + D\left(\overline{P_{Y^n}^{(\lambda)}} \,\|\, (\overline{P_{Y^n}})_t\right)$$

$$\geqslant \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \,\|\, \overline{P_{Y^n}^{(\lambda)}}\right).$$

Note that $\overline{P_{Y^n}^{(\lambda)}}$ is the average of $Q_{i,m_{\lambda_i}}^n$ over m, i's ($m \in [M_n], i \in [K_n]$) and hence based on Lemma A3, we have

$$\frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m}^n \,\|\, \overline{P_{Y^n}^{(\lambda)}}\right) \geqslant (1 - \bar{r}_n) \log(F_n(1 - \bar{r}_n)) + \bar{r}_n \log\left(\frac{F_n \bar{r}_n}{F_n - 1}\right)$$

$$\geqslant (1 - \bar{r}_n) \log F_n - h(\bar{r}_n).$$

As a result

$$\sup_t A(t) \leqslant \frac{1}{K_n} \sum_{i=1}^{K_n} n D(Q_{i\lambda_i} \,\|\, Q_\star | P_i) - n(R + \nu)(1 - \bar{r}_n) + h(\bar{r}_n).$$

Now we continue with the proof of (A33b). Again, using the Chernoff bound we have

$$\overline{P_{Y^n}}\left[\log \frac{P_{Y^n}}{Q_\star^n}(Y^n) \leqslant \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \,\|\, Q_\star^n\right) - D\left(Q_{i,m_{\lambda_i}}^n \,\|\, \overline{P_{Y^n}}\right)\right]$$

$$= \overline{P_{Y^n}}\left[\log \frac{Q_\star^n}{P_{Y^n}}(Y^n) \geqslant \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \,\|\, \overline{P_{Y^n}}\right) - D\left(Q_{i,m_{\lambda_i}}^n \,\|\, Q_\star^n\right)\right] \doteq e^{-\sup_t B(t)},$$

where

$$\sup_t B(t) := \sup_t \frac{t}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m_{\lambda_i}}^n(y^n) \log \frac{Q_\star^n(y^n) \overline{P_{Y^n}}(y^n)}{\overline{P_{Y^n}}(y^n)} - \log \sum_{y^n} (Q_\star^n(y^n))^t \left(\overline{P_{Y^n}}(y^n)\right)^{1-t}$$

$$= \sup_t \sum_{y^n} \overline{P_{Y^n}^{(\lambda)}}(y^n) \log \frac{(\overline{P_{Y^n}})_{1-t}(y^n)}{\overline{P_{Y^n}}(y^n)}$$

$$= \sup_t D\left(\overline{P_{Y^n}^{(\lambda)}} \,\|\, \overline{P_{Y^n}}\right) - D\left(\overline{P_{Y^n}^{(\lambda)}} \,\|\, (\overline{P_{Y^n}})_{1-t}\right)$$

$$\leqslant D\left(\overline{P_{Y^n}^{(\lambda)}} \,\|\, \overline{P_{Y^n}}\right) = \sum_{y^n} \left(\frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m_{\lambda_i}}^n(y^n)\right) \log\left(\frac{\frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m_{\lambda_i}}^n(y^n)}{\frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m}^n(y^n)}\right)$$

$$\leqslant \sum_{y^n} \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m_{\lambda_i}}^n(y^n) \log \frac{Q_{i,m_{\lambda_i}}^n(y^n)}{Q_{i,m}^n(y^n)} = \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m_{\lambda_i}}^n \,\|\, Q_{i,m}^n\right) \tag{A44}$$

$$= \frac{1}{K_n} \sum_{i=1}^{K_n} n D(Q_{i\lambda_i} \,\|\, Q | P_i),$$

and where the inequality in (A44) is by the Log-Sum inequality.

Appendix L. Proof of Lemma 1

We will prove the Lemma by contradiction. Define

$$t_i \triangleq \text{Number of users in block } i.$$

Assume that the arrangement with highest probability (let us call it $\mathcal{A}$) has at least two blocks, say blocks $1, 2$, for which $t_1 - t_2 > 1$. This assumption means that the arrangement with the highest probability is *not* the non-overlapping arrangement.

The probability of this arrangement, $\mathbb{P}(\mathcal{A})$, is proportional to

$$\mathbb{P}(\mathcal{A}) \propto \binom{K_n}{t_1}\binom{K_n - t_1}{t_2} = \frac{K_n!}{t_1, (K_n - t_1)!}\frac{(K_n - t_1)!}{t_2!\,(K_n - t_1 - t_2)!}$$
$$= \frac{K_n!}{t_1!\,t_2!\,(K_n - t_1 - t_2)!}.$$

We now consider a new arrangement, $\mathcal{A}_{\text{new}}$, in which $t_{1,\text{new}} = t_1 - 1$ and $t_{2,\text{new}} = t_2 + 1$ and all other blocks remain unchanged. This new arrangement is also feasible since we have not changed the number of users. The probability of this new arrangement is proportional to

$$\mathbb{P}(\mathcal{A}_{\text{new}}) \propto \binom{K_n}{t_1 - 1}\binom{K_n - t_1 - 1}{t_2 + 1}$$
$$= \frac{K_n!}{(t_1 - 1)!\,(K_n - t_1 + 1)!}\frac{(K_n - t_1 + 1)!}{(t_2 + 1)!\,(K_n - t_1 - t_2)!}$$
$$= \frac{K_n!}{(t_1 - 1)!\,(t_2 + 1)!\,(K_n - t_1 - t_2)!}.$$

Comparing $\mathbb{P}(\mathcal{A})$ and $\mathbb{P}(\mathcal{A}_{\text{new}})$ we see that $\mathbb{P}(\mathcal{A}) < \mathbb{P}(\mathcal{A}_{\text{new}})$ which is a contradiction to our primary assumption that $\mathcal{A}$ has the highest probability among all arrangements. Hence there do not exist two blocks which differ more than one in the number of active users within them in the arrangement with the highest probability.

References

1. Chandar, V.; Tchamkerten, A.; Wornell, G. Optimal Sequential Frame Synchronization. *IEEE Trans. Inf. Theory* **2008**, *54*, 3725–3728. [CrossRef]
2. Tchamkerten, A.; Chandar, V.; Wornell, G.W. Asynchronous Communication: Capacity Bounds and Suboptimality of Training. *IEEE Trans. Inf. Theory* **2013**, *59*, 1227–1255. [CrossRef]
3. Tchamkerten, A.; Chandar, V.; Caire, G. Energy and Sampling Constrained Asynchronous Communication. *IEEE Trans. Inf. Theory* **2014**, *60*, 7686–7697. [CrossRef]
4. Chen, X.; Chen, T.Y.; Guo, D. Capacity of Gaussian many-access channels. *IEEE Trans. Inf. Theory* **2017**, *63*, 3516–3539. [CrossRef]
5. Polyanskiy, Y. A perspective on massive random-access. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 2523–2527. [CrossRef]
6. Ahlswede, R.; Dueck, G. Identification via channels. *IEEE Trans. Inf. Theory* **1989**, *35*, 15–29. [CrossRef]
7. Han, T.S.; Verdu, S. Identification Plus Transmission. In Proceedings of the 1991 IEEE International Symposium on Information Theory, Budapest, Hungary, 24–28 June 1991; p. 267. [CrossRef]
8. Kiefer, J.; Sobel, M. *Sequential Identification and Ranking Procedures, with Special Reference to Koopman-Darmois Populations*; University of Chicago Press: Chicago, IL, USA, 1968.
9. Ahlswede, R.; Haroutunian, E. On logarithmically asymptotically optimal testing of hypotheses and identification. In *General Theory of Information Transfer and Combinatorics*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 553–571.
10. Unnikrishnan, J. Asymptotically optimal matching of multiple sequences to source distributions and training sequences. *IEEE Trans. Inf. Theory* **2015**, *61*, 452–468. [CrossRef]
11. Shahi, S.; Tuninetti, D.; Devroye, N. On the capacity of strong asynchronous multiple access channels with a large number of users. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 1486–1490. [CrossRef]

12. Shahi, S.; Tuninetti, D.; Devroye, N. On the capacity of the slotted strongly asynchronous channel with a bursty user. In Proceedings of the 2017 IEEE Information Theory Workshop (ITW), Kaohsiung, Taiwan, 6–10 November 2017; pp. 91–95. [CrossRef]

13. Shahi, S.; Tuninetti, D.; Devroye, N. On Identifying a Massive Number of Distributions. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018.

14. Gallager, R.G. *Information Theory and Reliable Communication*; Springer: Berlin/Heidelberg, Germany, 1968; Volume 588.

15. Polyanskiy, Y. Asynchronous Communication: Exact Synchronization, Universality, and Dispersion. *IEEE Trans. Inf. Theory* **2013**, *59*, 1256–1270. [CrossRef]

16. Chandar, V.; Tchamkerten, A. A Note on Bursty MAC. 2015. Available online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiL8qDGgZ78AhXgmVYBHcCSAtoQFnoECA4QAQ&url=https%3A%2F%2Fperso.telecom-paristech.fr%2Ftchamker%2Fburstymac.pdf&usg=AOvVaw0ferJC5-OOrr-KzrA7XG1V (accessed on 29 August 2022).

17. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.

18. Chung, K.L.; Erdos, P. On the application of the Borel-Cantelli lemma. *Trans. Am. Math. Soc.* **1952**, *72*, 179–186. [CrossRef]

19. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.