Strongly Asynchronous Massive Access Communications

BY

Sara Shahi
B.S., Amir Kabir University of Technology, 2013

THESIS
Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Natasha Devroye, Chair and Co-Advisor
Daniela Tuninetti, Co-Advisor
Besma Smida
Mojtaba Soltanalian
Dongning Guo, Northwestern University

ACKNOWLEDGMENTS

First and foremost, I want to to express my sincerest gratitudes to my advisors, Professor Natasha Devroye and Professor Daniela Tuninetti for their guidance, motivation and endless support throughout the past five years. They guided me through my PhD path yet gave me the freedom and flexibility to explore my interests. I am especially happy that I got the chance to work and learn from them both; two knowlegable and smart individuals with a great sense of humor.

I would also like to thank my committee members Prof. Besma Smida, Prof. Mojtaba Soltanalian and Prof. Dongning Guo for their insightful comments and valuable time.

Thanks to my amazing friends and labmates, Alex, Tang and Narueporn, for their invaluable discussions. They made my graduate life at UIC a very memorable experience.

I am forever indebted to my parents, for their unconditional love, sacrifice and encouragement throughout my life and in pursuing my dreams thousands of miles away from them.

My heartfelt thanks goes to my husband, Mohammad. He has been my true companion and this would have never been possible without his love, patience and continuous support.

SS

Contribution of Authors

The contents of this thesis is the result of a joint effort between my PhD advisors Prof. Daniela Tuninetti, Prof. Natasha Devroye and I.

Chapter 1 presents the main motivation and contributions of our work with a detailed list of prior related work. In 1 we use, in parts, our previously published work [1–3]. In Chapter 2 we present parts of our previously published work [2] and study the strongly asynchronous capacity of a single bursty user. Chapter 3 also presents our result in massive identification problem which was previously published in [3]. In Chapter 4, which includes parts of our previous published result in [1], we introduce the Strongly Asynchronous Massive Access channel model and study its capacity region.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF FIGURES

SUMMARY

Current wireless networks are designed to accommodate a fixed and finite number of active users at any point in time with a centralized controlling mechanism. Recent years however have witnessed an exponential growth in the number of wireless services and users, especially machine-to-machine communications, also known as the Internet of Things (IoT). IoT drastically differs from human-initiated communications, not only because of the massive number of devices that will potentially need access to the network at the same time, but also because of the type of traffic they are expected to generate (i.e., very bursty, low payload, and requiring both mission-critical reliability and latency). Next generation of IoT wireless networks must therefore to be able to serve a massive number of users where

1. each user demonstrates a bursty traffic pattern in which it transmits short packets of data infrequently to an access point, hence the conventional synchronicity assumption in the network is distorted; and
2. arbitrarily large number of users among a massive number of users may be active at each time, which need to be reliably identified and decoded; and
3. devices have strict energy consumption limits due to their characteristics; and
4. devices have strict latency requirements.

These innate features and requirements of the devices in IoT requires a general reconsideration of the conventional information theoretic assumptions such as frame synchronization. This thesis is mostly focused on investigating challenges 1 and 2. More specifically we investigate the

(information theoretic) fundamental technology-independent performance of a novel channel model that captures the essence of IoT communications. This thesis mainly consists of two parts. The first part is dedicated to the study of a single bursty user which demonstrates a random-access traffic pattern–challenge 1. We propose a strongly asynchronous bursty user model based on the on-off uplink transmission pattern where the user may transmit sporadically within a large asynchronous window. In this model we do not assume the use of pilot signals and hence the receiver does not have a priori knowledge of the codeword transmission time. We therefore study the tradeoff between the size of the asynchronous window, the rate of transmission and the burstiness of the user.

The second part is mostly focused on the study of the effects of the number of bursty users in IoT–challenge 2. In this regard, we propose a strongly asynchronous massive access model in which we allow the number of users to grow exponentially with the codeword blocklength. In this problem, we do not assume the use of pilot signals for either user identification nor synchronization purposes. We study the tradeoff between the number of users, their rate and the length of the asynchronous window.

This thesis is organized into five chapters. Chapter 1 presents the motivation and contributions of our work with a detailed list of prior related work. In Chapter 2, we study the strongly asynchronous capacity of a single bursty user. Chapter 3 presents our result in massive identification problem, which we will use in the following chapter. In Chapter 4, we introduce the Strongly Asynchronous Massive Access channel model and study its capacity region. Finally,

SUMMARY (Continued)

Chapter 5 concludes the thesis and introduces some future research topics. Some of the proofs can be found in the Appendix.

CHAPTER 1

INTRODUCTION

Parts of this chapter has been previously published in [1–3].

1.0.1 Background

For the past few decades, wireless services were mostly utilized by Human-type Communications (HTC) subscribers with a centralized resource allocation policy. With the evolution of telecommunication technologies a new class of Machine-type communications (MTC) is emerging. Unlike the HTC, which need human intervention, the new Machine-type Devices (MTD) are fully automated and can be utilized over a wide class of devices such as sensors, actuators, tracking devices and meters.

The emerging new paradigms like Internet of Things (IoT) and Smart Cities also consist of many interconnected heterogenous objects which are equipped with MTDs. As a result of the wide range of applications for the MTDs, their population is steadily increasing. In 2016, there have been 6.4 Billion connected things worldwide and is forecasted to increase to 20.4 Billion connected things in 2020 [4].

The MTC have inherently different features and requirements in terms of number of devices, control signaling, subscriber traffic load and diverse delay, reliability and energy constraints from that of HTC [5], [6]. Most MTDs exhibit a sporadic on-off transmission pattern which is due to their intrinsic objective. For example, they may transmit to a base station only when some

incident has taken place such as in an advanced metering infrastructure and vending machines. More specifically, the traffic in a MTC network usually features small packets of infrequent data generated from a mass of MTDs which result in high uplink traffic volume [7]. Moreover, these networks rely on the massive deployment of MTDs to achieve their goals.

With the massive number of MTD deployments, most of which introduce a short length sporadic random-access type of traffic to the network, a novel modeling and investigation of the MTC networks is essential. The next generation of wireless network should therefore provide a solution for coexistence of a massive number of infrequently communicating devices in the same frequency band. This problem is respectively known as mMTC (massive machine-type communication) and LP-WANs (low-power wide-area networks) in the licensed spectrum (3GPP and 5G-PPP) and unlicensed spectrum communities.

As a first step in providing a solution, one must exactly define and formulate the problem. Several different approaches to the problem exist in the literature which include

- Compressed sensing: By exploiting the scarcity of the user activity, one can define the problem as a compressed sensing or sparse recovery problem whose goal is to detect the user identity and channel estimation. [8–11].
- Protocol design: By focusing on the fact that a large number of users transmit short packets of data, one may be challenged to propose reliable massive access protocols [12–15].
- Information theoretic analysis: By modeling and analyzing the communication system (from an information theoretic point of view) to capture the random traffic pattern and the massive deployment of users [16–19].

In this thesis, we investigate this problem from an information theoretic view point. The main contributions and structure of this thesis are as follows.

1.0.2 Objective and Contribution: Asynchronous capacity of a bursty user

Parts of this subsection was previously published in [1].

Motivated by the innate nature of the sensor networks, and their bursty on-off communication pattern, we introduce the model of a bursty user in a strongly asynchronous channel and we study its fundamental information theoretic limits.

In [20], the authors considered a user who transmits once and only once within a strong asynchronous window. The goal was to locate and decode the user transmission time and message. In our work, the user transmits exponentially many times in blocklength (or arbitrary number of times in our second model). Moreover our error metric is the global / joint probability of error (i.e., an error is declared if *any* of the user's transmissions is in error, and we have an exponential number of transmissions) and we require the exact recovery of the transmission time and codeword in *all* transmissions. The approach in [20] does not extend to the global probability of error criterion for exponential number of transmission in blocklength n (where the number of transmissions is equal to $K_n = e^{n\nu}, \nu > 0$). This is due to the fact that their achievability relies on the typicality decoder and the derived error bounds do not decay fast enough with blocklength n.

In [21], the authors considered the special case of the problem considered here where a user transmits one synchronization pattern of length n (hence the rate $R = 0$) only once (hence $K_n = 1$) in a window of length $A_n = e^{n\alpha}$ of n channel uses each. They showed that for any α

below the synchronization threshold, α_0, the user can detect the location of the synchronization pattern. In addition they showed that a synchronization pattern consisting of the repetition of a single symbol which induces an output distribution with the maximum divergence from the noise distribution, suffices. The typicality decoder introduced in [21] however, even in a slotted channel model, only retrieves one of the trade-off points that we obtain in this Thesis that corresponds to a sub-exponential number of transmissions. We propose new achievability and converse techniques to support an exponential number of transmissions ($K_n = e^{n\nu}, \nu > 0$). Interestingly, we show that the symbol used for synchronization may change for different values of α and ν.

The single user strongly asynchronous channel was also considered in [22], where it was shown that the exact transmission time recovery, as opposed to the error criterion in [20] (which allows a sub-exponential delay in n), does not change the capacity.

Recently, the *synchronous* Gaussian massive multiple access channel with random access has been modeled in [23] where the number of users is let to grow linearly in the code blocklength and a random subset of users may try to access the channel. In [23], the authors took advantage of the Gaussian channel structure to exactly derive matching upper and lower bounds on the capacity. Since then, other versions of "massive number of users" have been proposed in [17], [24].

In our model, we consider a slotted strongly asynchronous channel with $A_n = e^{n\alpha}$ blocks of n channel uses each. We also consider two scenarios to capture the essence of sporadic

communication. Our models, objective and contributions in this problem is summarized as follows.

- For the first bursty communication model we assume that the user transmits a randomly selected message among $M_n = e^{nR}$ different ones in exactly $K_n = e^{n\nu}$ randomly selected but distinct blocks in the window. The receiver must locate and decode, with vanishing error probability in n, *each and every* one of the transmitted messages. Our ultimate goal is to characterize the capacity region (R, α, ν). In this regard, we show:

  1. For synchronization and data transmission ($R \geq 0$), we propose novel bounding techniques to find upper (converse) and lower (achievability) bounds on the capacity region of (R, α, ν). Our bounds are tight for $R = 0$ (synchronization only) and $\nu = 0$ (sub-exponential number of transmissions).
  2. For synchronization only ($R = 0$), we propose a sequential synchronization scheme which achieves the optimal tradeoff between (α, ν). We show that using a repetition pattern for synchronization is optimal. Surprisingly, we show that the optimal synchronization pattern is not fixed and it may change depending on the considered value of asynchronization level α and burstiness level ν.
  3. For certain values of R, which are small enough, the achievability and converse bounds match.

- For the second bursty communication model we assume that the number of transmissions of the user is not fixed and the user may randomly with probability $p_n = e^{-n\beta}$ access

a block of n channel uses and transmit a randomly selected message among $M_n = e^{nR}$ different ones. In this case, we find:

4. The achievability and converse bounds on the capacity region (R, α, β) is derived. Our achievability result shows that the asynchronous window length $A_n = e^{n\alpha}$ can increases with the increase of β since the number of transmissions to be detected decreases. Moreover, for $\beta = \alpha$ and $R = 0$ the achievability and converse bound match.

1.0.3 Objective and contribution: Identifying a massive number of distributions

Information theory has close ties with statistical hypothesis testing and we can use the latter to solve information theoretic problems. In conventional data transmission schemes, an overhead signal is sent at the beginning of the data packets to act as user's identifier. However, the user's statistics at the receiver itself can be potentially used as the users identifier.

One of the main areas in hypothesis testing is identification and ranking problems. In the classical identification 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. With this constraint the number of hypothesis is exponential in the number of distributions. If one neglects the fact that the sequences are generated by distinct distributions, the problem boils down to multiple

M-ary hypothesis testing problems. This approach is suboptimal as it fails to exploit some of the (possibly useful) constraints.

Comprehensive studies on identification and ranking problems can be found in [25, 26]. In [27–30], the authors study the Logarithmically Asymptotically Optimal (LAO) Testing of identification problem for a finite number of distributions. In particular, the identification of only several different objects has been studied in detail and one can find the reliability matrix, which consist of the error exponents of all error types. Their optimality criterion is to find the largest error exponent for a set of error types for given values of the other error types' error exponents. The same problem with a different optimality criterion was also studied in [31], where multiple, finite, sequences were matched to the source distributions. More specifically, they proposed a test for a generalized Neyman-Pearson-like optimality criterion to minimize the rejection probability given that all other error probabilities decay exponentially with a pre-specified slope. The identification problem is also closely related to anomaly detection [32–35].

In here, we assume A sequences of length n are generated i.i.d according to A distinct distributions; i.e., random vectors $X_i^n \overset{\text{i.i.d}}{\sim} P_{\sigma_i}, i \in [A]$, for some unknown permutation σ of the distributions. The goal is to reliably identify the permutation σ with vanishing error probability as n goes to infinity, from an observation of $[X_1^n, \ldots, X_A^n]$. This problem has close ties with de-anonymization of anonymized data [31, 36]. A different motivation is the identification of users using only channel output sequences, without the use of pilot / explicit identification signals [1]. In both scenarios, the problem's difficulty increases with the number of users. In addition, in modeling the systems with a massive number of users (such as the Internet of

Things), it may be reasonable to assume that the number of users grow with the transmission blocklength [1], [16], and that the user's identities must be distinguished from the received data. As a result, it is useful to understand exactly how the number of distributions affects the system performance, in particular for the case that the cardinality of the distributions grows with the blocklength. Notice that in this scenario, the number of hypotheses, would be doubly exponential in blocklength and the analysis of the optimal decoder becomes much harder than the classical (with constant number of distributions) identification problems.

We consider the identification problem for the case that the number of distributions grow with the observation blocklength n as motivated by the massive user identification problem in the Internet of Things paradigm. The key novel element in this work consist of analyzing and reducing the complexity of the optimal maximum likelihood decoder, with doubly exponential number of hypothesis, using a graph theoretic result. In particular, we show

1. Find a novel approach to analyze the probability of error in the ML decoder. In particular, we are able to transform the probability of identification error into a graph theoretic problem. By doing so, we are able to obtain matching upper and lower bounds on the probability of error. This result specifies the relation between the growth rate of the number of distributions and the pairwise distance of the distributions for reliable identification.

2. We show that the probability that more than two distributions are incorrectly identified is dominated by the probability of the event that only two distributions are incorrectly identified.

3. We also derive a novel graph theoretic result on the sum of the graph cycles. More specifically we show that the arithmetic mean of the cycles gains (where we define the cycle gain as the product of the edge weights within the cycle) in a graph can be upper bounded by a function of the sum of the squares of the edge weights.

1.0.4 Objective and contribution: Massive asynchronous communication

In the literature, different levels of asynchronism have been studied. The level of asynchronism in a system is defined by the length of a window A_n, with respect to the codeword blocklength n, within which the transmission can initiate. *Mildly asynchronous* MAC was first introduced in [37], where $A_n = o(n)$ and the capacity region was proved to remain the same as that of the classical synchronous MAC. A *totally asynchronous* MAC was also defined in [38] where $A_n = n$ and users continuously send their messages after transmission initiation. In this setting, the authors proved that the time sharing is no longer feasible; as a result, the capacity region lacks the convex hull operation seen in the capacity region of the synchronous MAC.

More recently, *strongly asynchronous* communication was studied in [21, 39–42] with $A_n = e^{n\alpha}$ for some $\alpha \geq 0$ where users only transmit once within each window. In [21] it was shown that reliable communication is indeed possible for $0 < \alpha < \alpha_0$; α_0 being the synchronization threshold. In [42] the suboptimality of preamble based synchronization schemes was shown. The capacity of a strong-asynchronous point-to-point (SA-P2P) channel is [40]

$$C_{\text{SA-P2P}} = \max_{P_X : D([P_X Q] \| Q_\star) > \alpha} I(P_X, Q), \tag{1.1}$$

where the maximum is is defined to be zero for $\alpha > \alpha_0 = \max_{x \in \mathcal{X}} \mathsf{D}(\mathsf{Q}_x \| \mathsf{Q}_\star)$ (please refer to 4.0.1 for special notation convention).

The capacity in (Equation 1.1) may be interpreted as follows: while in the synchronous point-to-point channel the maximization is over all input distributions P_X, the maximization is now restricted to those input distributions that induce output distributions $\mathsf{P}_Y = [\mathsf{P}_X \mathsf{Q}]$ that are sufficiently different from the 'idle output distribution' $\mathsf{Q}_\star$.

After the introduction of the strong asynchronous point to point channel, several different studies has been dedicated to this channel model. In [42] the strong asynchronous point to point channel capacity was evaluated under the requirement of correct decoding only (and not necessarily synchronization); in [22] the author showed that imposing exact transmission time recovery does not change the capacity in (Equation 1.1). The authors in [43] extended the achievability of proof of [42] on discrete channels (which was based on method of types) to continuous channels. In [44, 45] the authors studied the capacity of the same channel model with delay and sampling constraint. Improved second order statistics was later studied in [46]. In [47–49] also, the authors studied the same model with the emphasize on "channel detection" and provided random coding error exponents for different error types.

In addition, while [41] mainly focussed on point-to-point communication per unit cost, the authors briefly discussed in [41, Remark 3] the capacity of the *strong-asynchronous collision MAC* with exponentially many users and with a *per-user* probability of error (though their parametrization is different from ours). In this model, simultaneous transmission of two or more users results in a 'collision' that produces as output distribution the same as if all users where

idle, regardless of the number of colliding users. The probability of error at the MAC receiver is evaluated for each user individually, as opposed to the classical (stronger) requirement that all messages are jointly reliably decoded. In the proposed achievability scheme for $K_n = e^{n\nu}$ number of users with $\nu < \alpha/2$, all users employ the same codebook and thus users are not distinguishable unless an identifier is sent along with the message.

A related line of work, but not dealing with asynchronism, is the so-called *many-user MAC*. In [50] the authors considered a synchronous MAC with random user activity where the number of users increases linearly with the blocklength. This many-user model, while different from ours, faces some challenges, as we do in here, which arise from the fact that the number of users increases with the blocklength. In [50] one of these challenges is that the number of possible error events is exponential (in the blocklength), which prevents them from using a simple union bound for bounding the probability of error. Here we encounter the same problem as the number of possible error events scales faster than exponentially in blocklength n. In [17] the authors studied the massive random access channel where the total number of users increases linearly with blocklength n. They however restricted the users to use the same codebook and required only recovering the transmitted messages (as opposed to recovering the messages and transmitters identity) and employed the per user probability of error in their model.

We propose a strongly asynchronous massive access model which consist of a strongly asynchronous window of $A_n = e^{n\alpha}$ blocks of length n and $K_n = e^{n\nu}$ different users. We allow the number of users to also increase exponentially in n to capture the *massive* number of users. Each user selects a message among $M_n = e^{nR}$ possible ones and a transmission block among A_n

possible blocks, both uniformly at random. Our goal is to find the trade off between $(\mathsf{R}, \alpha, \nu)$. In characterizing the capacity of this model, we require its *global* probability of error to be vanishing. We show

1. Model Strongly Asynchronous Slotted Massive Access Channel (SAS-MAC) with global probability of error constraint. More specifically, in our modeling we require the decoder to correctly identify the users identity and decode their messages.
2. We show that for a sub exponential number of users K_n with $\log \mathsf{K}_n = o(n)$, each user can achieve its point-to-point strong-asynchronous capacity.
3. We show that when $\nu \geq \alpha$, users can not even be synchronized when transmitting a single codeword.
4. We propose a sequential decoder for the case that users can potentially have different channels. We show that strictly positive $(\mathsf{R}, \alpha, \nu)$ is achievable.
5. We propose a non-sequential achievability scheme for the case that users have identical channels. This proves strictly positive values of $(\mathsf{R}, \alpha, \nu)$ is achievable. We show, by means of an example, that this new non-sequential scheme will result in larger achievability region than the one in item 4.
6. We propose a new non-sequential achievability scheme for the case that the channels of different users is chosen from a set of conditional distributions of polynomial size in blocklength n. In this case, the channel statistics itself, can be used for user identification.
7. We find a new converse bound on capacity of general SAS-MAC.

.

1.1 Notation

Throughout the following chapters, we will adopt the following notation convention.

- Capital lettes represent random variables that take on lower case letter values in calligraphic letter alphabets.
- The space of all distributions on $\mathcal{X}$ is denoted by $\mathcal{P}_{\mathcal{X}}$.
- A stochastic kernel / transition probability from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $Q(y|x), \forall (x, y) \in \mathcal{X} \times \mathcal{Y}$.
- The output marginal distribution induced by $P \in \mathcal{P}_{\mathcal{X}}$ through the channel Q is denoted as $[PQ](y) := \sum_{x \in \mathcal{X}} P(x)Q(y|x), \forall y \in \mathcal{Y}$.
- The empirical / joint empirical distribution of a sequence x^n/ (x^n, y^n) are respectively defined as

$$\hat{P}_{x^n}(a) := \frac{1}{n}\mathcal{N}(a|x^n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i=a\}}, \forall a \in \mathcal{X}, \tag{1.2}$$

$$\widehat{P}_{x^n,y^n}(a,b) := \frac{1}{n}\mathcal{N}(a,b|x^n,y^n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{\substack{x_i=a \\ y_i=b}\right\}}, \forall a,b \in \mathcal{X} \times \mathcal{Y}, \tag{1.3}$$

where $\mathcal{N}(a|x^n)$ and $\mathcal{N}(a,b|x^n,y^n)$ denotes the number of occurrences of letter $a \in \mathcal{X}$ in sequence x^n and the number of joint occurrences of (a,b) in the pair of sequences (x^n,y^n). When the sequence x^n is clear from the context, we may drop the subscript x^n from $\hat{P}_{x^n}(a)$.

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

$$T(P) := \{x^n : \mathcal{N}(a|x^n) = nP(a), \forall a \in \mathcal{X}\}, \tag{1.4}$$

$$T_V(x^n) := \left\{y^n : \frac{\mathcal{N}(a,b|x^n,y^n)}{\mathcal{N}(a|x^n)} = V(b|a), \forall (a,b) \in (\mathcal{X},\mathcal{Y})\right\}. \tag{1.5}$$

- We say that (x^n, y^n) are jointly strongly ϵ-typical according to $P_{X,Y}$, and write $(x^n, y^n) \in T_\epsilon^n(P_{X,Y})$, if

$$|\widehat{P}_{x^n y^n}(a,b) - P_{X,Y}(a,b)| \leq \epsilon P_{X,Y}(a,b), \forall (a,b) \in \mathcal{X} \times \mathcal{Y}.$$

- $I(P,Q)$ is used to denote the mutual information between random variable $(X,Y) \sim (P,[PQ])$ coupled via $P_{Y|X}(y|x) = Q(y|x)$,

- $D(P_1 \parallel P_2)$ is used to denote the Kullback Leibler divergence between distribution P_1 and P_2. The conditional Kullback Leibler divergence is also denoted by

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

- The Bhatcharrya distance between P_1, P_2 is also denoted by

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

- $y_j^n := [y_j, \ldots, y_n]$ is a vector of length $n-j+1$ and we simply use y^n instead of y_1^n.

- The indicator function of event A is denoted by $\mathbb{1}_A$.
- We also use the notation $a_n \doteq e^{nb}$ when

$$\lim_{n\to\infty} \frac{\log a_n}{n} = b.$$

- $[m : n]$ for $m \leq n$ is used to denote the set $\{m, m+1, \ldots, n\}, m, n \in \mathbb{R}$. We also use $[m] := [1 : m]$.
- When all elements of the random vector X^n are generated i.i.d according to distribution P, we denote it as $X^n \overset{\text{i.i.d}}{\sim} P$.
- $\lfloor x \rfloor_r$ is used to denote the remainder of x divided by r.
- We use the notation $[a]^+$ to represent

$$[a]^+ := \begin{cases} a, & a > 0 \\ 0, & a \leq 0 \end{cases}.$$

- We use standard big-O notation and we write

$$f(n) = o(g(n)) : \lim_{n\to\infty} \frac{f(n)}{g(n)} = 0,$$
$$f(n) = Og(n)) : \lim \sup_{n\to\infty} \frac{|f(n)|}{g(n)} < \infty.$$

- We write $f(n) = \mathrm{poly}(n)$ when $\exists 0 < k < \infty$ such that $f(n) = O(n^k)$.

CHAPTER 2

ASYNCHRONOUS CAPACITY OF A SINGLE BURSTY USER

Parts of this chapter has been previously published in [2].

It is widely believed that Machine-type Communications and Internet of Things are going to be the next dominant paradigm in wireless technology. The traffic pattern imposed by the devices within these networks have unique features different from the ones in human-type communication networks. The communications that take place within these networks are often sporadic and bursty, but must nonetheless be reliably detected and decoded. For example, each sensor node may want to transmit a signal to the base station only when some incident has taken place.

In this chapter, we consider the problem of both detecting and decoding asynchronous data bursts of a single user. In conventional methods the user transmits a pilot signal at the beginning of each data burst to notify the decoder of the upcoming data; the decoding phase may be performed using any synchronized decoding method. The loss in this approach is negligible when synchronization is done once and the cost of acquiring synchronization is absorbed into the lengthy data stream that follows. For sparse / bursty transmission, as in the problem considered here, this approach is not suitable as the training based schemes are known to be sub-optimal [42]. In this work we do not enforce the usage of pilot symbols, and the codebook serves the dual purpose of synchronization and data transfer. This Chapter's

central goal is to characterize the trade-off between the reliable transmission rate between one transmitter and one receiver, the burstiness of that transmitter, and the level of asynchronism.

In section 2.1 we introduce the slotted bursty and strongly asynchronous channel model with fixed number of transmissions and derive its capacity region. We also find an equivalent capacity region expression for the special case with zero rate (synchronization only). In section 2.2 we introduce a model for slotted bursty and strongly asynchronous channel with random number of transmissions and find its capacity region. Section 4.2 concludes this chapter.

2.1 System model for fixed number of transmissions and main results

We consider a discrete memoryless channel with transition probability matrix $Q(y|x)$ defined over all (x, y) in the finite input and output alphabets $(\mathcal{X}, \mathcal{Y})$. We also define a noise symbol $\star \in \mathcal{X}$ for which $Q_\star(y) > 0,\ \forall y \in \mathcal{Y}$.

Let M be the number of messages, A be the number of blocks, and K be the number of transmissions. An (M, A, K, n, ϵ) code for the *slotted bursty and strongly asynchronous* discrete memoryless channel with transition probability matrix $Q(y|x)$ *with fixed number of transmissions* consists of:

- A message set $[M]$, from which messages are selected uniformly at random.
- Encoding functions $f_i : [M] \to \mathcal{X}^n,\ i \in [A]$, where we define $x_i^n(m) := f_i(m)$. The transmitter chooses uniformly at random one set of K blocks for transmission out of the $\binom{A}{K}$ possible ones, and a set of K messages from M^K possible ones, also uniformly at random, and sends $x_{\nu_i}^n(m_i)$ in block ν_i for $i \in [K]$ and $\star^n$ in every other block. We denote the chosen blocks and messages as $((\nu_1, m_1), \ldots, (\nu_K, m_K))$.

- A destination decoder function

$$g(\mathcal{Y}^{nA}) = ((\widehat{\nu_1}, \widehat{m_1}), \ldots, (\widehat{\nu_K}, \widehat{m_K}))\,,$$

such that the average probability of error associated to it, given by

$$P_e^{(n)} := \frac{1}{M^K \binom{A}{K}} \sum_{(\nu_1,m_1),\ldots,(\nu_K,m_K)} \mathbb{P}[g(\mathcal{Y}^{nA}) \neq ((\nu_1, m_1), \ldots, (\nu_K, m_K))\,|H_{((\nu_1,m_1),\ldots,(\nu_K,m_K))}],$$

satisfies $P_e^{(n)} \leq \epsilon$, where $H_{((\nu_1,m_1),\ldots,(\nu_K,m_K))}$ is the hypothesis that user transmits message m_i at block ν_i with the codebook $x_{\nu_i}^n(m_i)$, for all $i \in [K]$.

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 ϵ_n going to zero as n goes to infinity. The capacity region is the set of all possible achievable (R, α, ν) triplets.

We now introduce our main result. In Theorem 1 we show that an exponential number of transmissions for a single user is possible at the expense of a reduced rate and/or reduced asynchronous window length compared to the case of only one transmission $K_n = 1$ (or more generally $\nu = 0$).

Theorem 1. *Achievable and impermissible regions for the capacity region of a slotted bursty and strongly asynchronous discrete memoryless channel with transition probability matrix* $Q(y|x)$ *are given by*

$$\mathcal{R}^{\text{in}} := \bigcup_{\lambda\in[0,1],P\in\mathcal{P}_{\mathcal{X}}} \left\{ \begin{array}{c} \nu \leq \alpha \\ \alpha + R < D(Q_\lambda \parallel Q_\star|P) \\ \nu < D(Q_\lambda \parallel Q|P) \\ R < I(P,Q) \end{array} \right\}, \tag{2.1}$$

and

$$\mathcal{R}^{\text{out}} := \bigcup_{\lambda\in[0,1],P\in\mathcal{P}_{\mathcal{X}}} \left\{ \{\nu > \alpha\} \cup \left\{ \begin{array}{c} \alpha > D([PQ_\lambda] \parallel Q_\star) + [I(P,Q_\lambda) - R]^+ \\ \nu > D(Q_\lambda \parallel Q|P) \end{array} \right\} \cup \{R > I(P,Q)\} \right\}, \tag{2.2}$$

where

$$Q_\lambda(.|x) := \frac{Q_x^\lambda(.)Q_\star^{1-\lambda}(.)}{\sum_{y'\in\mathcal{Y}} Q_x^\lambda(y')Q_\star^{1-\lambda}(y')}. \tag{2.3}$$

Proof. **Achievability.** *Codebook generation.* The user generates A_n constant composition codebooks, of rate R and blocklength n, by selecting each message's codeword uniformly and independently from the P-type set of sequences in $\mathcal{X}^n$, one codebook for each available block.

Decoder. We perform a two-stage decoding. First, the decoder finds the location of the transmitted codewords (first stage, the synchronization stage) and it decodes the messages

(second stage, the decoding stage). The probability of error for this two-stage decoder is given by

$$P_e^{(n)} \leq \mathbb{P}[\text{synchronization error}] + \mathbb{P}[\text{decoding error}|\text{no synchronization error}].$$

For the first stage, fix

$$T : -D(Q_\star \parallel Q|P) \leq T \leq D(Q \parallel Q_\star|P),$$

which can be changed for different trade-off points. At each block $j \in [A_n]$, if there exists *any* message $m \in [M_n]$ such that the Log Likelihood Ratio (LLR)

$$L\left(y_j^n, x_j^n(m)\right) := \frac{1}{n} \log \frac{Q\left(y_j^n|x_j^n(m)\right)}{Q_{\star^n}(y_j^n)} \geq T, \tag{2.4}$$

declare a codeword transmission block and a noise block otherwise. Given the hypothesis $H_1 := H_{((1,1),\ldots,(K_n,1))}$ the probability of the synchronization error in the first stage is given by

$$
\begin{aligned}
&\mathbb{P}\left[\text{synch error}|H_1\right] \\
&\leq \mathbb{P}\left[\bigcup_{j=1}^{K_n}\bigcap_{m=1}^{M_n} L\left(Y_j^n, x_j^n(m)\right) < T|H_1\right] \\
&\quad + \mathbb{P}\left[\bigcup_{j=K_n+1}^{A_n}\bigcup_{m=1}^{M_n} L\left(Y_j^n, x_j^n(m)\right) \geq T|H_1\right] \\
&\leq \sum_{j=1}^{K_n} \mathbb{P}\left[L\left(Y_j^n, x_j^n(1)\right) < T|H_1\right] + e^{nR} \sum_{j=K_n+1}^{A_n} \mathbb{P}\left[L\left(Y_j^n, x_j^n(1)\right) \geq T|H_1\right] \\
&\leq e^{n\nu} \sum_{\substack{\widehat{Q}: \\ D(\widehat{Q}\|Q_\star|P)-D(\widehat{Q}\|Q|P)<T}} \mathbb{P}\left[Y^n \in T_{\widehat{Q}}\left(x^n(1)\right)|H_1\right] \\
&\quad + e^{n(R+\alpha)} \sum_{\substack{\widehat{Q}: \\ D(\widehat{Q}\|Q_\star|P)-D(\widehat{Q}\|Q|P)\geq T}} \mathbb{P}\left[Y^n \in T_{\widehat{Q}}(x^n(1))|H_1\right] \qquad (2.5) \\
&\leq e^{n\nu} e^{-nD(Q_\lambda\|Q|P)} + e^{n(\alpha+R)} e^{-nD(Q_\lambda\|Q_\star|P)}, \qquad (2.6)
\end{aligned}
$$

where Q_λ is defined in (Equation 2.3) and

$$\lambda : D(Q_\lambda \parallel Q_\star|P) - D(Q_\lambda \parallel Q|P) = T.$$

The expression in (Equation 2.6) is the result of finding the minimum exponent in (Equation 2.5) using the Lagrangian method as in [51, Sec. 11.7].

By (Equation 2.6), the probability of error in the synchronization goes to zero as n goes to infinity when

$$\nu < D(Q_\lambda \parallel Q|P), \tag{2.7a}$$

$$\alpha + R < D(Q_\lambda \parallel Q_\star|P). \tag{2.7b}$$

Conditioning on the 'no synchronization error' and having found all K_n 'not noisy' blocks, we can use a Maximum Likelihood (ML) decoder for random constant composition codes, introduced and analyzed in [52], on the super-block of length nK_n to distinguish among e^{nK_nR} different message combinations. If $R < I(P, Q)$, the probability of the error of the second stage also vanishes as n goes to infinity.

Converse. The main technical difficulty and innovation in the proof relies on analyzing the probability of error in a ML decoder. In this regard, we boil down the problem to finding an exponentially decaying 'lower' bounds on the probability of the missed detection (where the likelihood ratio defined in (Equation 2.4) of an active block is less than a threshold) and false alarm (where the likelihood ratio defined in (Equation 2.4) of an idle block is larger than the threshold) error events. By the type counting argument and the fact that we have polynomially many types in blocklength at the expense of a small reduce in rate [53] we can restrict our attention to constant composition codes. In other words, we assume the use of codewords $x_i^n(.)$ with constant compositions P_i in each block $i \in [A_n]$. Given the hypothesis $H_1 := H_{((1,1)\dots(K_n,1))}$,

with a ML decoder (which achieves the minimum average probability of error) and for any $T \in \mathbb{R}$, the error events are given by

$$\{\text{error}|H_1\} = \bigcup_{\substack{((l_1,\tilde{m}_1)\ldots,(l_{K_n},\tilde{m}_{K_n})) \\ \neq((1,m_1),\ldots,(K_n,m_{K_n}))}} \left\{ \sum_{i=1}^{K_n} L\left(Y_i^n, x_i^n(m_i)\right) \leq \sum_{i=1}^{K_n} L\left(Y_{l_i}^n, x_{l_i}^n(\tilde{m}_i)\right) \right\}, \tag{2.8}$$

where (Equation 2.8) the union of the events that the sum of the LLRs of the true hypothesis $((1, m_1), \ldots, (K_n, m_{K_n}))$ is less than the sum of the LLRs of the wrong hypothesis

$$((l_1, \tilde{m}_1) \ldots, (l_{K_n}, \tilde{m}_{K_n})) \neq ((1, m_1), \ldots, (K_n, m_{K_n})),$$

where wrong means that we have at least one decoding or one synchronization error. We now focus our attention on a subset of these events which have a *single* synchronization error. i.e.,

$$\{\text{error}|H_1\} \supseteq \bigcup_{\substack{i\in[K_n] \\ j\in[K_n+1:A_n] \\ m\in[M_n]}} \left\{ L\left(Y_i^n, x_i^n(m_i)\right) \leq L\left(Y_j^n, x_j^n(m)\right) \right\} \tag{2.9}$$

$$\supseteq \left\{ \bigcup_{i\in[K_n]} \{L\left(Y_i^n, x_i^n(m_i)\right) \leq T\} \right\} \bigcap \left\{ \bigcup_{\substack{j\in[K_n+1:A_n] \\ m\in[M_n]}} \left\{ L\left(Y_j^n, x_j^n(m)\right) \geq T \right\} \right\}. \tag{2.10}$$

In other words, (Equation 2.9) is the union over the events that *(any message, any noisy block)* is selected instead of *one* of the *(correct message, correct block)*s; with the underlying assumption that the rest of the blocks are chosen correctly. We also further restrict

$$T \in [-D(Q_\star \parallel Q|P_{i^\star}), D(Q \parallel Q_\star|P_{i^\star})],$$

where $i^\star$ is chosen such that

$$i^\star := \arg \max_{\substack{i,\lambda_i: \\ D(Q_{\lambda_i}\|Q_\star|P_i)-D(Q_{\lambda_i}\|Q|P_i)=T}} D(Q_{\lambda_i} \parallel Q|P_i). \tag{2.11}$$

The reason for this choice of i^* will be become clear later (see (Equation 2.13) and (Equation 2.14)). By (Equation 2.10) we have

$$\begin{aligned}
&\mathbb{P}\left[\text{error}\middle|H_1\right] \\
&\geq \mathbb{P}\left[\bigcup_{i\in[K_n]} L\left(Y_i^n, x_i^n(m)\right) \leq T|H_1\right] \cdot \mathbb{P}\left[\bigcup_{\substack{j\in[K_n+1:A_n] \\ m\in[M_n]}} L\left(Y_j^n, x_j^n(m)\right) \geq T|H_1\right] && (2.12)\\
&\geq \left(1 - e^{-n\left[\nu - D\left(Q_{\lambda_{i^\star}}\|Q|P_{i^\star}\right)\right]}\right) && (2.13)\\
&\quad \cdot \left(1 - e^{-n\left[\alpha + R\mathbb{1}_{\left\{R<I\left(P,Q_{\lambda_{i^*}}\right)\right\}} - D\left(Q_{\lambda_{i^*}}\|Q_\star|P_{i^*}\right)\right]}\right), && (2.14)
\end{aligned}$$

where (Equation 2.12) is due to the independence of $Y_j^n, j \in [A_n]$ and where (Equation 2.13) and (Equation 2.14) are proved in Appendix A and B, respectively.

The lower bound on the probability of error given in (Equation 2.13) and (Equation 2.14), would be bounded away from zero if

$$\nu > D\left(Q_{\lambda_{i^\star}} \parallel Q|P_{i^\star}\right),$$

$$\alpha + R\mathbb{1}_{\{R<I(P,Q_{\lambda_{i^*}})\}} > D\left(Q_{\lambda_{i^\star}} \parallel Q_\star|P_{i^\star}\right) = I(P, Q_{\lambda_{i^\star}}) + D([P_{i^*}Q_{\lambda_{i^\star}}] \parallel Q_\star),$$

which can be equivalently be written as

$$\nu > D\left(Q_{\lambda_{i^\star}} \parallel Q|P_{i^\star}\right), \tag{2.16a}$$

$$\alpha > D([P_{i^*}Q_{\lambda_{i^\star}}] \parallel Q_\star) + \left[I(P, Q_{\lambda_{i^\star}}) - R\right]^+, \tag{2.16b}$$

and hence this region is impermissible.

Any asynchronous channel can be reduced to a synchronous channel by providing the decoder with side information about the transmission time. Hence, the same bound on the rate of a synchronous channel, i.e. $R < I(P_{i^\star}, Q)$ also applies to the asynchronous channel. By the symmetry of the hypothesis, the same lower bound on probability of error holds for the average probability of error and hence we retrieve the bounds given in (Equation 2.2). □

Remark 1. *We note that for all those* $\lambda : R < I(P, Q_\lambda)$*, the achievability and the converse bounds match.*

The main novelty in this problem is to find exponentially decaying upper and lower bounds on the probability of error. The achievability scheme analysis is easier as we can easily pose it

as a hypothesis testing problem. However, in the converse, we have to deal with the optimal ML decoder. As a first step in reducing the complexity of the ML decoder, we considered a set of error events with single synchronization errors (which we believe is the major error set and many other events are its subsets). Next, we had to find the probability that the LLRs of the active blocks are smaller than a threshold. This again, would be easy to calculate for a single LLR; its probability is a function of the (imaginary) channel Q_λ defined in (Equation 2.3). However, we have to deal with unions of such events as in (Equation 2.12). Calculation of these unions is also easy for $\nu = 0$. In this case the optimal $\lambda = 1$ and hence $Q_{\lambda=1} = Q$ and one can leverage the fact that the probability of decoding error for channel Q is small to transform the union into a summation. If however $\nu \neq 0$ and hence $\lambda \neq 1$, probability of error for channel Q_λ (for the same code as channel Q) would be dependent on the rate R. Transformation of a union to a summation is not straightforward anymore and hence we had to provide several additional steps (in Appendix B and C) to do so.

For a fixed λ, a comparison between the bounds given by (Equation 2.16b) and (Equation 2.7b) is shown in Fig. Figure 1. It is easy to see that the bounds given in (Equation 2.1) and (Equation 2.2) will coincide (i.e., complement one another) for the case $\nu = 0$ ($\lambda = 1$) and retrieve the capacity region previously derived in [20].

Remark 2. *It is worth noting that the region specified in Figure 1, need not be convex since* α *is a channel parameter and can not be chosen by user.*

We now concentrate our attention to the synchronization case only.

Remark 3. *By specializing Theorem 1 for* $R = 0$*, we can see that* $\mathcal{R}^{\text{in}}|_{R=0} = \mathcal{R}^{\text{out}}|_{R=0} = \mathcal{R}|_{R=0}$.

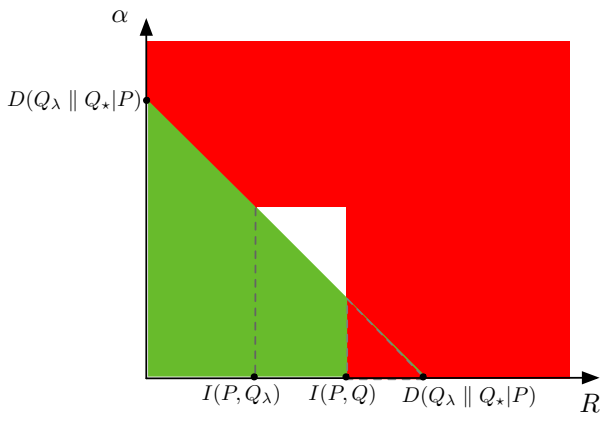


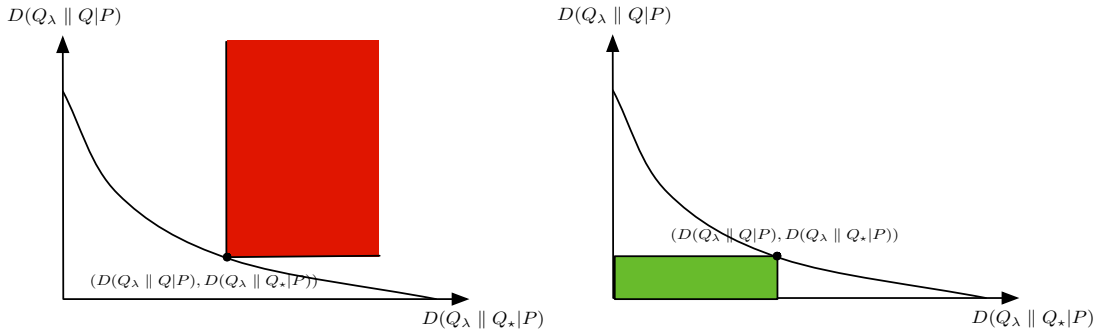
Figure 1. Comparison of Impermissible region given in (Equation 2.16b) (red region) and achievable region given by (Equation 2.7b) (green region) for fixed λ.

It can be easily seen in Fig. Figure 2 that by taking the union over $\lambda \in [0, 1]$, the achievability and converse regions match for $R = 0$.

In the following example, we consider a Binary Symmetric Channel and plot its achievable region.

Example 1. *To illustrate the capacity region in Theorem 1, we consider a Binary Symmetric Channel (BSC)* Q *with cross over probability* δ *as it is shown in Fig. Figure 3. We also assume* $\star = 0$. *For the channel* Q_λ *in* (Equation 2.3) *we have*

$$Q_\lambda(0|0) = 1 - \delta,$$
$$\epsilon_\lambda := Q_\lambda(0|1) = \frac{\delta^\lambda (1-\delta)^{(1-\lambda)}}{\delta^\lambda (1-\delta)^{(1-\lambda)} + (1-\delta)^\lambda \delta^{(1-\delta)}}.$$



(a) Impermissible region (b) Achievable region

Figure 2. The union of the regions over different values of λ will result in matching achievability and converse bounds for $R = 0$.

By changing $p = \mathbb{P}[X = 0] \in [0, \frac{1}{2}]$ *and* $\lambda \in [0, 1]$*, we obtain the achievability region shown in Fig. 4(a). In addition, the (optimal) trade off for* $(R, \alpha, \nu = 0)$ *can be seen in Fig. 4(b) which resembles the one in [22, Fig. 1]. The trade off between* (α, ν) *can be seen in Fig. 4(c) which has the curvature we expect to see, like the one in Fig. Figure 10 in the Appendix.*

Theorem 2 provides another form for the trade-off between $(R = 0, \alpha, \nu)$ which implies that using a repetition pattern for synchronization pattern is optimal.

Theorem 2. *For* $R = 0$*, the capacity region* $\mathcal{R}|_{R=0}$ *in Remark 3 is equivalent to*

$$\mathcal{R}^{synch} := \bigcup_{x \in \mathcal{X}, \lambda \in [0,1]} \left\{ \begin{array}{c} \nu < \alpha \\ \alpha < D(Q_\lambda \parallel Q_\star) \\ \nu < D(Q_\lambda \parallel Q_x) \end{array} \right\}. \tag{2.17}$$

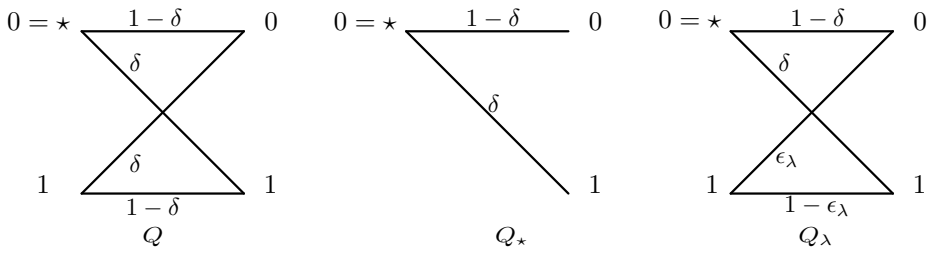


Figure 3. Strongly synchronous binary symmetric channel

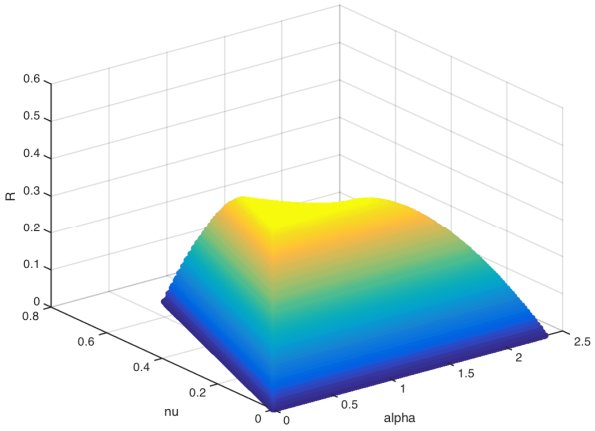
Proof. $\mathcal{R}^{\text{synch}} \subseteq \mathcal{R}|_{R=0}$ is trivial since we can restrict the set of distributions $P \in \mathcal{P}_{\mathcal{X}}$ in $\mathcal{R}|_{R=0}$ to the distributions with weight one on a single symbol x and zero weight on all other symbols.

We also prove $\mathcal{R}|_{R=0} \subseteq \mathcal{R}^{\text{synch}}$ by contradiction and by means of the following Lemma proved in Appendix D.

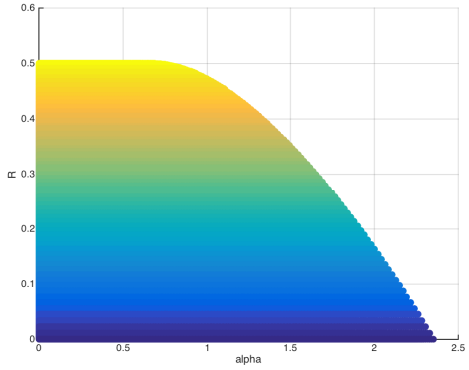
Lemma 3. *The curve* $(D(Q_\lambda \parallel Q_\star|P), D(Q_\lambda \parallel Q|P))$ *characterized by* $\lambda \in [0,1]$ *is the lower envelope of the set of curves*

$$\bigcup_{x \in \mathcal{X}} \left\{ (D(Q_{\lambda_x} \parallel Q_\star|P), D(Q_{\lambda_x} \parallel Q|P)) \right\},$$

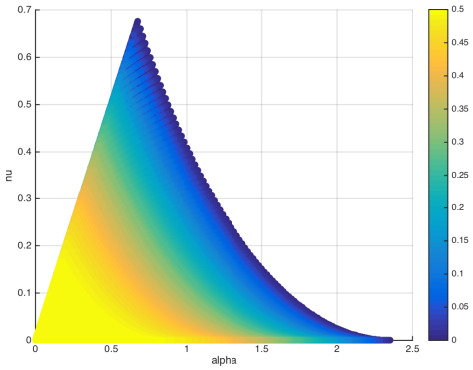
which are each characterized by $\lambda_x \in [0,1]$.



(a) (R, α, ν) trade-off



(b) (R, α) trade-off for $\nu = 0$



(c) (α, ν) trade-off for different rates, specified by the color

Figure 4. Achievability bound on capacity region of slotted bursty and strongly asynchronous BSC with fixed number of transmissions with cross over probability $\delta = 0.11$.

We continue the proof by assuming $\mathcal{R}|_{R=0} \not\subseteq \mathcal{R}^{\text{synch}}$. Then there exists an element

$$(r_1, r_2, 0) = (D(Q_\lambda \parallel Q|P), D(Q_\lambda \parallel Q|P), 0) \in \mathcal{R}|_{R=0},$$
$$(r_1, r_2) \notin \mathcal{R}^{\text{synch}},$$

that is, which lies above all the $\{D(Q_\lambda \parallel Q_\star), D(Q_\lambda \parallel Q_x)\}$ curves for all $x \in \mathcal{X}$. Hence, for any $x \in \mathcal{X}$, there exists a λ_x such that

$$r_1 = D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q_\star),$$
$$r_2 = D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q_x).$$

As a result

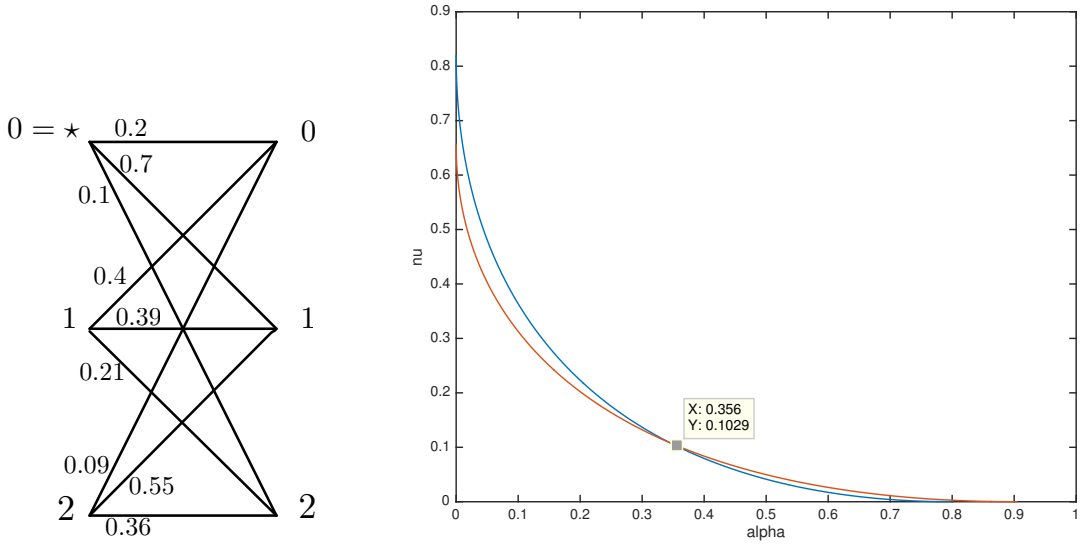
$$D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q_\star|P),$$
$$D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q|P),$$

which contradicts Lemma 3 that $(D(Q_\lambda \parallel Q|P), D(Q_\lambda \parallel Q|P))$ is the lower envelope of the set of $\bigcup_{x\in\mathcal{X}} \{(D(Q_{\lambda_x} \parallel Q_\star|P), D(Q_{\lambda_x} \parallel Q|P))\}$ curves and hence the initial assumption that $\mathcal{R}|_{R=0} \not\subseteq \mathcal{R}^{\text{synch}}$ is not feasible. □

Note that by adapting the achievability scheme to synchronize only $(R = 0)$, we do not need a different synchronization pattern for each block. Using the same synchronization pattern in

every block suffices to drive the probability of error in the synchronization stage to zero and since it matches the converse, it is optimal.

Theorem 2 also implies that depending on the value of α and ν, using a repetition synchronization pattern with a single symbol is optimal. This symbol may change depending on the considered value of α and ν. For the ternary channel in Fig. 5(a), for example, the resulting curves by using symbol $x = 1$ and $x = 2$ are shown in Fig. 5(b). As it is clear, for the regime $\alpha > 0.356$, symbol $x = 1$ has to be used whereas in the regime $\alpha \leq 0.356$ symbol $x = 2$ has to be used in the synchronization pattern.



(a) (b)

Figure 5. Channel with different synchronization pattern symbols for different (α, ν) regimes.

2.2 System model for random transmissions and main results

We consider again a discrete memoryless channel with transition probability matrix $Q(y|x)$ defined over all (x, y) in the finite input and output alphabets $(\mathcal{X}, \mathcal{Y})$. We also define a noise symbol $\star \in \mathcal{X}$ for which $Q_\star(y) > 0,\ \forall y \in \mathcal{Y}$.

An (M, A, p, n, ϵ) code for the *slotted bursty and strongly asynchronous* discrete memoryless channel with transition probability matrix $Q(y|x)$ *with random access* is defined as follows.

- A message set $[M]$, from which messages are selected uniformly at random.
- Encoding functions $f_i : [M] \to \mathcal{X}^n,\ i \in [A]$, where we define $x_i^n(m) := f_i(m)$. For each block $i \in [A]$, the transmitter chooses a message among M possible ones and transmit $x_i^n(m_i)$ through the channel with probability p or remains idle and transmits $\star^n$ with probability $1 - p$.
- A destination decoder function

$$g(\mathcal{Y}^{nA}) = \left((\widehat{\nu}_1, \widehat{m_1}), \ldots, (\widehat{\nu_{\hat{k}}}, \widehat{m_{\hat{k}}})\right),$$

such that the average probability of error associated to it, given by

$$P_e^{(n)} := \sum_{k=1}^{A} \sum_{(\nu_1, m_1), \ldots, (\nu_k, m_k)} \frac{1}{M^k} p^k (1-p)^{A-k} \mathbb{P}[g(y^{nA}) \neq ((\nu_1, m_1), \ldots, (m_k, \nu_k))\, | H_{((\nu_1, m_1), \ldots, (\nu_k, m_k))}],$$

satisfies $P_e^{(n)} \leq \epsilon$, where $H_{((\nu_1, m_1), \ldots, (\nu_k, m_k))}$ is the hypothesis that user transmits message m_i at block ν_i with the codebook $x_{\nu_i}^n$, for all $i \in [k]$.

A tuple (R, α, β) is said to be achievable if there exists a sequence of codes $(e^{nR}, e^{n\alpha}, e^{-n\beta}, n, \epsilon_n)$ with $\lim_{n\to\infty} \epsilon_n = 0$. The capacity region is the set of all possible achievable (R, α, β) triplets.

Theorem 4. *Achievable and impermissible regions for the capacity region of a slotted bursty and strongly asynchronous random access channel with transition probability matrix* $Q(y|x)$ *are given by*

$$\mathcal{R}^{\text{in}} := \bigcup_{\lambda\in[0,1], P\in\mathcal{P}_{\mathcal{X}}} \left\{ \begin{array}{ll} \alpha + R & < D(Q_\lambda \parallel Q_\star|P) \\ \alpha - \beta & < D(Q_\lambda \parallel Q|P) \\ R & < I(P, Q) \end{array} \right\}, \tag{2.18}$$

and

$$\mathcal{R}^{\text{out}} := \bigcup_{\lambda\in[0,1], P\in\mathcal{P}_{\mathcal{X}}} \left\{ \left\{ \begin{array}{ll} \alpha & > D(Q_\lambda \parallel Q_\star|P) + [I(P, Q_\lambda) - R]^+ \\ \alpha - \beta & > D(Q_\lambda \parallel Q|P) \end{array} \right\} \cup \{R > I(P, Q)\} \right\}. \tag{2.19}$$

Proof. **Achievability.** The encoder and decoder are the same as the one given for the achievability proof of Theorem 1, except that the number of active blocks is not fixed. We denote $p_n := e^{-n\beta}$ and $\hat{H}_k$ to be the hypothesis that the user is active in k blocks. By the

symmetry of the probability of error among hypotheses with the same number of occupied blocks, we can write

$$\begin{aligned} P_e^{(n)} &= \sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1-p_n)^{A_n-k}\ \mathbb{P}[\text{Error}|\hat{H}_k] \\ &\leq \sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1-p_n)^{A_n-k} \mathbb{P}[\text{Synchronization error}|\hat{H}_k] \\ &\quad + \sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1-p_n)^{A_n-k} \mathbb{P}[\text{Decoding error}|\hat{H}_k, \text{No synchronization error}]. \end{aligned} \tag{2.20}$$

With similar steps as those in the proof of Theorem 1, we obtain

$$\mathbb{P}[\text{synchronization error}|\hat{H}_k] \leq k\ e^{-nD(Q_\lambda \| Q|P)} + e^{nR}\left(e^{n\alpha} - k\right) e^{-nD(Q_\lambda \| Q_\star|P)}, \tag{2.21}$$

where

$$\lambda : D(Q_\lambda \parallel Q_\star|P) - D(Q_\lambda \parallel Q|P) = T.$$

By (Equation 2.21), we can upper bound (Equation 2.20) as

$$\begin{aligned} &\sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1-p_n)^{(A_n-k)} \mathbb{P}[\text{synchronization error}|\hat{H}_k] \\ &\leq e^{n\alpha} e^{-n\beta} e^{-nD(Q_\lambda \| Q|P)} + e^{n(\alpha+R)} e^{-nD(Q_\lambda \| Q_\star|P)}, \end{aligned}$$

which goes to zero for

$$\alpha - \beta < D(Q_\lambda \parallel Q|P),$$

$$\alpha + R < D(Q_\lambda \parallel Q_\star|P).$$

For the decoding stage, with the same strategy as the one in Theorem 1 we obtain the third bound in (Equation 2.18).

Converse. The converse argument is also similar to the converse proof of Theorem 1. It can be shown that

$$\mathbb{P}\left[\text{error}\middle|\widehat{H}_k\right] \geq \left(1 - \frac{e^{D\left(Q_{\lambda_{i^\star}} \| Q|P_{i^\star}\right)}}{k}\right) \cdot \left(1 - \frac{e^{-n\left[R\mathbb{1}_{\left\{R<I(P,Q_{\lambda_{i^*}})\right\}} - D\left(Q_{\lambda_{i^\star}} \| Q_\star|P_{i^\star}\right)\right]}}{A_n - k}\right).$$

Hence

$$\begin{aligned}\mathbb{P}[\text{error}] &\geq \sum_{k=1}^{A_n-1} \binom{A_n}{k} (e^{-n\beta})^k (1-e^{-n\beta})^{A_n-k} \left(1 - \frac{e^{nD_1}}{k}\right) \left(1 - \frac{e^{n\left(D_2 - R\mathbb{1}_{\left\{R<I(P,Q_{\lambda_{i^*}})\right\}}\right)}}{A_n - k}\right) \\ &\geq 1 - (1-e^{-n\beta})^{A_n} - e^{-n\beta A_n} - \frac{2e^{nD_1}}{e^{-n\beta}e^{n\alpha}} - \frac{2e^{n\left(D_2 - R\mathbb{1}_{\left\{R<I(P,Q_{\lambda_{i^*}})\right\}}\right)}}{(1-e^{-n\beta})e^{n\alpha}}, \end{aligned} \tag{2.22}$$

where

$$D_1 := D\left(Q_{\lambda_{i^\star}} \parallel Q|P_{i^\star}\right),$$

$$D_2 := D\left(Q_{\lambda_{i^\star}} \parallel Q_\star|P_{i^\star}\right),$$

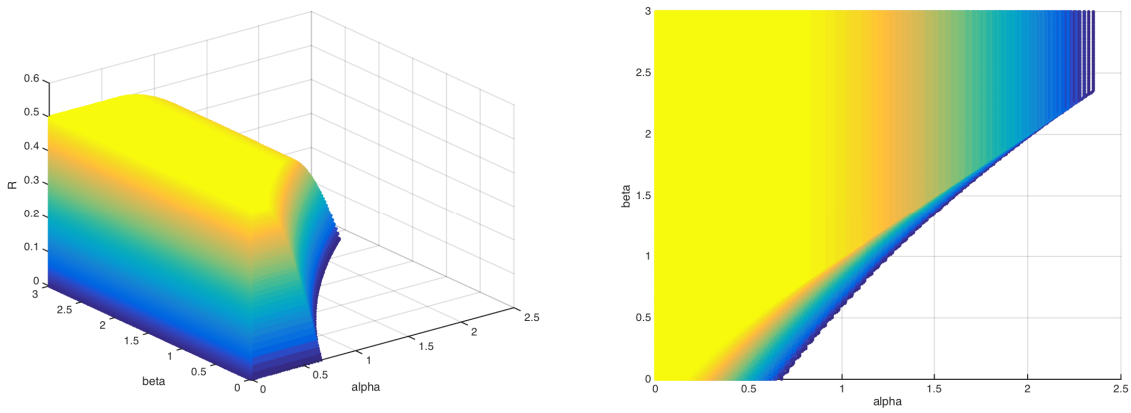
and where (Equation 2.22) is proved in Appendix E. This retrieves the first two bounds in (Equation 2.19). The third bound in (Equation 2.19) is by the usual bound on the reliable rate of a synchronous channel. □

It is easy to see that (Equation 2.18) and (Equation 2.19) match for the cases that $R = 0$ or $\beta = \alpha$. The latter case corresponds to $\lambda = 1$.

Example 2. *We consider the same BSC channel defined in Example 1 and illustrate its achievability region for the slotted bursty and strongly asynchronous channel with random access in Fig. 6(a). For values of* $\beta > D(Q \parallel Q_\star|P) = 2.3527$*, the achievable region is similar to the to the case* $\beta = 2.3527$ *and the surface remains unchanged. This is also apparent in Fig. 6(b) where the trade-off between* (α, β) *is depicted. This is in fact obvious in Theorem 4 since for values of* $\beta > D(Q \parallel Q_\star|P)$ *the achievability* (Equation 2.18) *and converse bound* (Equation 2.19) *match and are equal to the capacity region for one for only one transmission as the one in [22, Fig. 1].*

2.3 Conclusion

In this Chapter we studied a slotted bursty and strongly asynchronous discrete memoryless channel where a user transmits a randomly selected message among $M_n = e^{nR}$ messages in



(a) (R, α, β) trade-off (b) (α, β) trade-off for different rates, specified by colors

Figure 6. Capacity region of slotted bursty and strongly asynchronous BSC with random access with cross over probability $\delta = 0.11$.

each one of the $K_n = e^{n\nu}$ randomly selected blocks of the available $A_n = e^{n\alpha}$ blocks. We derive the upper and lower bounds on the trade-off among (R, α, ν) by finding achievability and converse bounds where we analyze an optimal Maximum Likelihood decoder in the converse. For the case that the number of transmissions of the user is not fixed and the user may access the channel with probability $e^{-n\beta}$, we again provide upper and lower bounds on the trade-off between (R, α, β).

CHAPTER 3

IDENTIFICATION OF A MASSIVE NUMBER OF DISTRIBUTIONS

Parts of this chapter has been previously published in [3].

Hypothesis testing is a classical problem in statistics where in its simplest form one has to make a decision in favor of one of the two possible hypothesis based on some observations. More specifically, given a random observation vector, one seeks to identify the distribution from a given set of distributions that generated it. Pioneering work in classical hypothesis testing include the proof of the optimality of likelihood ratio tests under certain criteria in the Neyman-Pearson Theorem [54]. Derivation of error exponents of different error types and their trade-offs for binary and M-ary hypothesis testing in [55] and [56] and the analysis of sequential hypothesis testing in [57].

One of the main areas in hypothesis testing is the identification and ranking problems. The classical identification problem is consist of a finite number of distinct sources, each generating 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.

In here, we assume A sequences of length n are generated i.i.d according to A distinct distributions; in particular random vectors $X_i^n \overset{\text{i.i.d}}{\sim} P_{\sigma_i}, i \in [A]$, for some unknown permutation σ of the distributions. The goal is to reliably identify the permutation σ with vanishing error probability as n goes to infinity from an observation of $[X_1^n, \ldots, X_A^n]$. A motivation is the identification of users using only channel output sequences, without the use of pilot / explicit

identification signals [1]. In this scenario, the problem's difficulty increases with the number of users. In addition, in modeling the systems with a massive number of users (such as the Internet of Things), it may be reasonable to assume that the number of users grow with the transmission blocklength [1], [16], and that the user's identities must be distinguished from the received data. As the result, it is useful to understand exactly how the number of distributions affects the system performance, in particular for the case that the cardinality of the distributions grows with the blocklength. Notice that in this scenario, the number of hypothesis, would be doubly exponential in blocklength and the analysis of the optimal decoder becomes much harder than the classical (with constant number of distributions) identification problems.

In this chapter, we consider the identification problem for the case that the number of distributions grow with the observation blocklength n as motivated by the massive user identification problem in the Internet of Things paradigm. The key novel element in this work consist of analyzing and reducing the complexity of the optimal maximum likelihood decoder, with double exponential number of hypothesis, using a graph theoretic result.

We first introduce a set of special notations. In section 3.2, we introduce the problem formulation and the main result of this chapter. The main theorem proof consist of achievability and converse bounds.

3.1 Special notation

The special notation used in this chapter is as follows.

- We use S_n, where $|S_n| = n!$, to denote the set of all possible permutations of a set of n elements.

- For a permutation $\sigma \in S_n$, σ_i denotes the i-th element of the permutation.
- $K_k\left(a_1, \ldots, a_{\binom{k}{2}}\right)$ is the complete graph with k nodes with edge index $i \in [\binom{k}{2}]$ and edge weights $a_i,\ i \in [\binom{k}{2}]$. We may drop the edge argument and simply write K_k when the edge specification is not needed.
- A cycle c of length r in K_k may be interchangeably defined by a vector of vertices as $c^{(v)} = [v_1, \ldots, v_r]$ or by a set of edges $c^{(e)} = \{a_1, \ldots, a_r\}$ where a_i is the edge between $(v_i, v_{i+1}), \forall i \in [r-1]$ and a_r is that between (v_r, v_1). With this notation, $c^{(v)}(i)$ is then used to indicate the i-th vertex of the cycle c.
- $C_k^{(r)}$ is used to denote the set of all cycles of length r in the complete graph $K_k\left(a_1, \ldots, a_{\binom{k}{2}}\right)$.
- The cycle gain, denoted by $G(c)$, for cycle $c = \{a_1, \ldots, a_r\} \in C_k^{(r)}$ is the product of the edge weights within the cycle c, i.e.,

$$G(c) = \prod_{i=1}^{r} a_i,\ \forall a_i \in c.$$

3.2 Problem formulation

Let $P := \{P_1, \ldots, P_A\}, P_i \in \mathcal{P}_{\mathcal{X}}, \forall i \in [A]$ consist of A distinct distributions and also let Σ be uniformly distributed over S_A, the set of permutations of A elements. In addition, assume that we have A independent random vectors $\{X_1^n, X_2^n, \ldots, X_A^n\}$ of length n each. For σ, a realization of Σ, assign the distribution $P_{\sigma_i}^n$ to the random vector $X_i^n, \forall i \in [A]$. After observing a sample $x^{nA} = [x_1^n, \ldots, x_A^n]$ of the random vector $X^{nA} = [X_1^n, \ldots, X_A^n]$, we would like to identify $P_{\sigma_i}, \forall i \in$

$[A]$. More specifically, we are interested in finding a permutation $\hat{\sigma} : \mathcal{X}^{nA} \to S_A$ to indicate that $X_i^n \overset{\text{i.i.d}}{\sim} P_{\hat{\sigma}_i},\ \forall i \in [A]$. Let $\hat{\Sigma} = \hat{\sigma}(X^{nA})$.

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

$$\begin{aligned} P_e^{(n)} &= \mathbb{P}\left[\hat{\Sigma} \neq \Sigma\right] \\ &= \frac{1}{(A)!} \sum_{\sigma \in S_A} \mathbb{P}\left[\hat{\Sigma} \neq \sigma | X_i^n \overset{\text{i.i.d}}{\sim} P_{\sigma_i}, \forall i \in [A]\right] \\ &= \mathbb{P}\left[\hat{\Sigma} \neq [A] \middle| X_i^n \overset{\text{i.i.d}}{\sim} P_i, \forall i \in [A]\right]. \end{aligned}$$

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

3.2.1 Condition for Identifiability

In Theorem 5 we characterize the relation between the number of distributions and the pairwise distance of the distributions for reliable identification.

Theorem 5. *A sequence of distributions* $P = \{P_1, \ldots, P_{A_n}\}$ *is identifiable iff*

$$\lim_{n\to\infty} \sum_{1 \le i < j \le A_n} e^{-2nB(P_i, P_j)} = 0. \tag{3.1}$$

Proof. The rest of this section contains the proof. To prove Theorem 5, we provide upper and lower bounds on the probability of error in the following subsections.

3.2.2 Upper bound on the probability of identification error

We use the optimal Maximum Likelihood (ML) decoder, which minimizes the average probability of error, given by

$$\hat{\sigma}(x_1^n, \ldots, x_{A_n}^n) := \arg \max_{\sigma \in S_{A_n}} \sum_{i=1}^{A_n} \log \left(P_{\sigma_i} \left(x_i^n \right) \right), \tag{3.2}$$

where $P_{\sigma_i}\left(x_i^n\right) = \prod_{t=1}^{n} P_{\sigma_i}\left(x_{i,t}\right)$. The average probability of error associated with the ML decoder can also be written as

$$\begin{aligned}
P_e^{(n)} &= \mathbb{P}\left[\hat{\Sigma} \neq [A_n] \middle| \widehat{H}\right] \\
&= \mathbb{P}\left[\bigcup_{\hat{\sigma} \neq [A_n]} \hat{\Sigma} = \hat{\sigma} \middle| \widehat{H}\right] \\
&= \mathbb{P}\left[\bigcup_{r=2}^{A_n} \bigcup_{\left\{\sum_{i=1}^{A_n} \mathbb{1}_{\{\hat{\sigma}_i \neq i\}} = r\right\}}^{\hat{\sigma}:} \hat{\Sigma} = \hat{\sigma} \middle| \widehat{H}\right] \qquad (3.3) \\
&= \mathbb{P}\left[\bigcup_{r=2}^{A_n} \bigcup_{\left\{\sum_{i=1}^{A_n} \mathbb{1}_{\{\hat{\sigma}_i \neq i\}} = r\right\}}^{\hat{\sigma}:} \sum_{i=1}^{A_n} \log \frac{P_{\hat{\sigma}_i}\left(X_i^n\right)}{P_i\left(X_i^n\right)} \geq 0 \middle| \widehat{H}\right] \qquad (3.4)
\end{aligned}$$

where $\widehat{H} := \left\{ X_i^n \overset{\text{i.i.d}}{\sim} P_i, \forall i \in [A_n] \right\}$ and where (Equation 3.3) is due to the requirement that each sequence is distributed according to a distinct distribution and hence the number of incorrect distributions ranges from $[2 : A_n]$. In order to avoid considering the same set of error events multiple times, we incorporate a graph theoretic interpretation of $\left\{ \sum_{i=1}^{A_n} \mathbb{1}_{\{\hat{\Sigma}_i \neq i\}} = r \right\}$

in (Equation 3.4) which is used to denote the fact that we have identified r distributions incorrectly. Consider the two sequences $[i_1, \ldots, i_r]$ and $[\hat{\sigma}_{i_1}, \ldots, \hat{\sigma}_{i_r}]$ for which we have

$$\left\{ \sum_{i=1}^{A_n} \mathbb{1}_{\{\hat{\sigma}_i \neq i\}} = \sum_{j=1}^{r} \mathbb{1}_{\{\hat{\sigma}_{i_j} \neq i_j\}} = r \right\}.$$

These two sequences in (Equation 3.4) in fact indicate the event that we have (incorrectly) identified $X_{i_j}^n \overset{\text{i.i.d}}{\sim} P_{\hat{\sigma}_{i_j}}$ instead of the (true) distribution $X_{i_j}^n \overset{\text{i.i.d}}{\sim} P_{i_j}, \forall j \in [r]$. For a complete graph K_{A_n}, the set of edges between $((i_1, \hat{\sigma}_{i_1}), \ldots, (i_r, \hat{\sigma}_{i_r}))$ in K_{A_n} would produce a single cycle of length r or a set of disjoint cycles with total length r. However, we should note that in the latter case where the sequence of edges construct a set of (lets say of size L) disjoint cycles (each with some length $\tilde{r}_l$ for $\tilde{r}_l < r$ such that $\sum_{l=1}^{L} \tilde{r}_l = r$), then those cycles and their corresponding sequences are already taken into account in the (union of) set of $\tilde{r}_l$ error events.

As an example, assume $A_n = 4$ and consider the error event

$$\log \frac{P_2(X_1^n)}{P_1(X_1^n)} + \log \frac{P_1(X_2^n)}{P_2(X_2^n)} + \log \frac{P_4(X_3^n)}{P_3(X_3^n)} + \log \frac{P_3(X_4^n)}{P_4(X_4^n)} \geq 0,$$

which corresponds to the (error) event of choosing $[\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \hat{\sigma}_4] = [2, 1, 4, 3]$ over $[1, 2, 3, 4]$ with $r = 4$ errors. In the graph representation, this gives two cycles of length 2 each, which correspond to

$$\left\{ \log \frac{P_2(X_1^n)}{P_1(X_1^n)} + \log \frac{P_1(X_2^n)}{P_2(X_2^n)} \geq 0 \right\} \cap \left\{ \log \frac{P_4(X_3^n)}{P_3(X_3^n)} + \log \frac{P_3(X_4^n)}{P_4(X_4^n)} \geq 0 \right\},$$

and are already accounted for in the events

$$\{[\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \hat{\sigma}_4] = [2, 1, 3, 4]\} \cup \{[\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \hat{\sigma}_4] = [1, 2, 4, 3]\}$$

with $r = 2$.

As the result, in order to avoid double counting, in evaluating (Equation 3.4) for each r we should only consider the sets of sequences which produce a *single* cycle of length r.

Before proceeding further, we define the edge weights for a complete weighted graph

$$K_{A_n}(a_{(1,2)}, \dots a_{(K_n,1)}).$$

In particular, we define $a_{(i,j)} := e^{-nB(P_i,P_j)}$ to be the edge weight between vertices (i, j) in the complete graph K_{A_n} shown in Fig. Figure 7.

Hence, we can upper bound the probability of error in (Equation 3.4) as

$$\begin{aligned} P_e^{(n)} &\leq \sum_{r=2}^{A_n} \sum_{c \in C_{A_n}^{(r)}} \mathbb{P}\left[\sum_{i=1}^{r} \log \frac{P_{\lfloor c^{(\nu)}(i+1) \rfloor_r}\left(X^n_{c^{(\nu)}(i)}\right)}{P_{c^{(\nu)}(i)}\left(X^n_{c^{(\nu)}(i)}\right)} \geq 0 | \widehat{H}\right] \\ &\leq \sum_{r=2}^{A_n} \sum_{c \in C_{A_n}^{(r)}} e^{-n \sum_{i=1}^{r} B(P_{c^{(\nu)}(i)}, P_{c^{(\nu)}(\lfloor i+1 \rfloor_r)})} \qquad (3.5) \\ &= \sum_{r=2}^{A_n} \sum_{c \in C_{A_n}^{(r)}} G(c), \qquad (3.6) \end{aligned}$$

where r enumerates the number of incorrect matchings and where $c(i)$ is the i-th vertex in the cycle c. In (Equation 3.6), we have leveraged the fact that $e^{-nB(P_i,P_j)}$ is the edge weight between

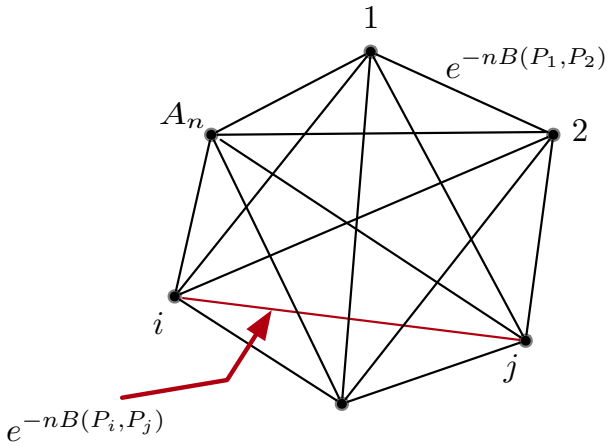


Figure 7. Complete graph K_{A_n} with edge weight $e^{-nB(P_i,P_j)}$ for every pair of vertices $i \neq j \in [K_n]$.

vertices (i, j) in the complete graph K_{A_n} and hence $G(c) = e^{-n\sum_{i=1}^{r} B(P_{c^{(v)}(i)}, P_{c^{(v)}(\lfloor i+1 \rfloor_r)})}$ is the gain of cycle c. The inequality in (Equation 3.5) is by

$$\begin{aligned}
&\mathbb{P}\left[\sum_{i=1}^{r} \log \frac{P_{\lfloor c^{(v)}(i+1) \rfloor_r}\left(X^n_{c^{(v)}(i)}\right)}{P_{c^{(v)}(i)}\left(X^n_{c^{(v)}(i)}\right)} \geq 0 | \widehat{H}\right] \\
&\leq \exp\left\{ n \inf_t \log \mathbb{E}\left[\prod_{i=1}^{r}\left(\frac{P_{c^{(v)}(\lfloor i+1 \rfloor_r)}\left(X^n_{c(i)}\right)}{P_{c^{(v)}(i)}\left(X^n_{c(i)}\right)}\right)^t\right]\right\} \\
&\leq \exp\left\{ n \sum_{i=1}^{r} \log \mathbb{E}\left[\left(\frac{P_{c^{(v)}(\lfloor i+1 \rfloor_r)}\left(X^n_{c(i)}\right)}{P_{c^{(v)}(i)}\left(X^n_{c(i)}\right)}\right)^{1/2}\right]\right\} \qquad (3.7) \\
&= \exp\left\{ -n \sum_{i=1}^{r} B(P_{c^{(v)}(i)}, P_{c^{(v)}(\lfloor i+1 \rfloor_r)})\right\}.
\end{aligned}$$

The fact that we used $t = 1/2$ in (Equation 3.7) 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 see later, the rest of the error events are dominated by the set of only 2 incorrectly identified distributions. This can be seen as follows for $X_1^n \overset{\text{i.i.d}}{\sim} P_1, X_2^n \overset{\text{i.i.d}}{\sim} P_2$

$$\begin{aligned}
&\mathbb{P}\left[\log \frac{P_1(X_2^n)}{P_2(X_2^n)} + \log \frac{P_2(X_1^n)}{P_1(X_1^n)} \geq 0\right] \\
&= \sum_{\substack{\hat{P}_1, \hat{P_2}: \\ \sum_{x\in\mathcal{X}} \hat{P}_1(x) \log \frac{P_2(x)}{P_1(x)} + \\ \hat{P_2}(y) \log \frac{P_1(x)}{P_2(x)} \geq 0}} \exp\left\{nD\left(\hat{P}_1 \parallel P_1\right) - nD\left(\hat{P}_2 \parallel P_2\right)\right\} \\
&\doteq e^{-nD\left(\tilde{P}\|P_1\right) - nD\left(\tilde{P}\|P_2\right)} = e^{-2nB(P_1,P_2)},
\end{aligned} \tag{3.8}$$

where $\tilde{P}$ in the first equality in (Equation 3.8), 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 (Equation 3.8) is proved.

In order to calculate the expression in (Equation 3.6), we use the following graph theoretic Lemma, the proof of which is given in the Appendix F.

Lemma 6. *In a complete graph* $K_k(a_1, \ldots, a_{n_k})$ *and for the set of cycles of length* $r, \mathcal{C}_k^{(r)} = \{c_1, \ldots c_{N_{r,k}}\}$ *we have*

$$\frac{1}{N_{r,k}} \left(G(c_1) + \ldots G(c_{N_{r,k}})\right) \leq \left(\frac{a_1^2 + \ldots + a_{n_k}^2}{n_k}\right)^{\frac{r}{2}}$$

where $N_{r,k}, n_k$ *are the number of cycles of length* r *and the number of edges in the complete graph* K_k*, respectively.*

By Lemma 6 and (Equation 3.6) we prove in Appendix G that the upper bound on the probability of error $P_e^{(n)}$ goes to zero if

$$\lim_{n\to\infty} \sum_{1\le i<j\le A_n} e^{-2nB(P_i,P_j)} = 0. \tag{3.9}$$

As a result of Lemma 6, it can be seen from (Equation G.1) that the sum of probabilities that $r \geq 3$ 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 the error events with two wrong distributions.

3.2.3 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 (Equation 3.4), we only consider the set of error events with only two incorrect distributions, i.e., the set of events with $r = 2$. In this case we have

$$\begin{aligned} P_e^{(n)} &\geq \mathbb{P}\left[\bigcup_{1\le i<j\le A_n} \log\frac{P_i(X_j^n)}{P_j(X_j^n)} + \log\frac{P_j(X_i^n)}{P_i(X_i^n)} \geq 0 | \widehat{H}\right] \\ &\geq \frac{\left(\sum_{1\le i<j\le A_n} \mathbb{P}\left[\xi_{i,j}\right]\right)^2}{\sum_{\substack{(i,j),(j,k)\\(i,j)\neq(l,k)\\i\neq j, l\neq k}} \mathbb{P}[\xi_{i,j}, \xi_{k,l}]}, \end{aligned} \tag{3.10}$$

where (Equation 3.10) is by [58] and where

$$\xi_{i,j} := \left\{ \log \frac{P_i}{P_j}(X_j^n) + \log \frac{P_j}{P_i}(X_i^n) \geq 0 | \widehat{H} \right\}. \tag{3.11}$$

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

$$P_e^{(n)} \geq \frac{\left(\sum_{1 \leq i < j \leq A_n} e^{-2nB(P_i,P_j)} \right)^2}{\sum\limits_{i,j,k} e^{-nB(P_i,P_j)-nB(P_i,P_k)-nB(P_k,P_j)} + \left(\sum\limits_{i,j} e^{-2nB(P_i,P_j)} \right)^2} \tag{3.12}$$

$$\geq \frac{\left(\sum_{i,j} e^{-2nB(P_i,P_j)} \right)^2}{8 \left(\sum\limits_{1 \leq i < j \leq A_n} e^{-2nB(P_i,P_j)} \right)^{\frac{3}{2}} + \left(\sum\limits_{1 \leq i < j \leq A_n} e^{-2nB(P_i,P_j)} \right)^2} \tag{3.13}$$

$$= \frac{\sqrt{\sum_{1 \leq i < j \leq A_n} e^{-2nB(P_i,P_j)}}}{8 + \sqrt{\sum_{1 \leq i < j \leq A_n} e^{-2nB(P_i,P_j)}}}, \tag{3.14}$$

where (Equation 3.13) is by Lemma 6. As it can be seen from (Equation 3.14), if

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

the probability of error is bounded away from zero. As a result, we have to have

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

which also matches our upper bound on the probability of error in (Equation G.2). □

As it is clear from the result of Theorem 5, when A_n is a constant or grows polynomially with n, the sequence of distributions in P are always identifiable and the probability of error in the identification problem decays to zero as the blocklength n goes to infinity. The interesting aspect of Theorem 5 is in the regime that A_n increases exponentially with the blocklength.

3.3 Conclusion

In this Chapter, we generalized the identification problem to the case that the number of distributions grows with the blocklength n. By analyzing the optimal Maximum Likelihood decoder we found matching upper and lower bounds on the probability of identification error. We also derived and used some graph theoretic results in order to calculate the probability of error in the Maximum Likelihood decoder. This result characterizes the relation between the number of distributions and the pairwise distance of the distributions for reliable identification. Having proved the criterion for identifiability of massive number of distributions in Theorem 5, we move on to the SAS-MAC problem. We use the result of Theorem 5 to identify the massive number of users, by their induced probability distribution at the receiver.

CHAPTER 4

STRONGLY ASYNCHRONOUS SLOTTED MASSIVE ACCESS CHANNEL

Parts of this chapter has been previously published in [1].

One assumption used in much of the network information theoretic analysis is that the number of users, even though large, is a constant. More specifically, in conventional studies, the number of users is let to go to infinity only after performing asymptotic analysis on probability of error for large coding blocklengths. By allowing the number of users and the blocklength to be functions of one another and to go to infinity simultaneously, lots of previously developed tools in analysis of probability of error fails. Another unrealistic information theoretic assumption is that of block and symbol synchronism between transmitters and receivers. It is often assumed that the receiver knows the starting point of the codeword and is only concerned about correctly decoding the received symbols. In practice, this synchronization is usually achieved by transmitting a pilot signal before the data signal. The pilot signal does not carry any information and has negligible impact (as the blocklength goes to infinity) on the achievable rates for streams of data where synchronization is done once, or for finitely many users. In large decentralized networks, however, the use of pilot signals for synchronization may cause an unacceptable amount of overhead. This motivates us to find the channel capacity of a network consisting of massive number of users without an a priori assumption of synchronization.

In this work the usage of pilot symbols is not assumed, and the codebook may in theory serve the purpose of synchronization as well as of data transfer. For example, one could imagine that the codebook is sufficiently different in idle and busy time blocks so as to achieve synchronization at the receiver.

In addition, we do not assume preambles in our codebook design to distinguish the users. Each user's codebook should be sufficiently different so as to allow the receiver to identify both the message and its transmitter. These tasks become harder to achieve as the length of the possible transmission window as well as the number of users increases, in particular if they increase exponentially with the blocklength.

We first mention the special notation convention that we use within this chapter and after introducing our channel model we state our main results.

4.0.1 Special Notation.

- A vector of length n with a subscript refers to $Y_s^n := [Y_{(s-1)n+1}, \ldots, Y_{sn}]$; the subscript could also indicate a user; its meaning should be clear from the context.
- For the SAS-MAC we use the shorthand notation

$$Q_S(y|x_S) := Q(y|x_S, \star_{S^c}), \forall S \subseteq [K], \tag{4.1}$$

to indicate that the users indexed by S transmit x_i, and users indexed by $S^c := [K] \setminus S$ transmit their idle symbol.

- When $|S| = 1$, we use

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

and when $|S| = 0$, we use

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

4.1 System Model

The discrete memoryless *classical MAC* with K-user, denoted as $(\mathcal{X}_1 \times \ldots \times \mathcal{X}_K, Q(.|.), \mathcal{Y})$, consists of K+1 finite sets $(\mathcal{X}_1, \ldots, \mathcal{X}_K, \mathcal{Y})$ and a collection of conditional distributions $Q(y|x_1, \ldots, x_K)$ on $\mathcal{Y}$, one for each input $(x_1, \ldots, x_K)$. This MAC is memoryless since we assume

$$Q(y^n|x_1^n, \ldots, x_K^n) = \prod_{t=1}^{n} Q(y_t|x_{1,t}, \ldots, x_{K,t}), \forall n \in \mathbb{N}.$$

Let M be the number of messages, A be the number of blocks, and K be the number of users. An (M, A, K, n, ϵ) code for the *asynchronous* multiple access channel consists of:

- A message set $[M]$, for each user $i \in [K]$, from which messages are chosen uniformly at random.
- An encoding function $f_i : [M] \to \mathcal{X}_i^n$, for each user $i \in [K]$. We define $x_i^n(m) := f_i(m)$; Each user choses a message $m_i \in [M]$, to convey to the receiver, and a block index $t_i \in [A]$, both uniformly at random and independent of one another. It then transmits $[\star_i^{n(t_i-1)} x_i^n(m_i) \star_i^{n(A-t_i)}] \in \mathcal{X}_i^{nA}$, where $\star_i \in \mathcal{X}_i$ is the designated 'idle' symbol for user i.

- A destination decoding function

$$g(\mathcal{Y}^{nA}) = \left((\hat{t}_1, \hat{m}_1), \ldots, (\hat{t}_K, \hat{m}_K)\right),$$

such that its associated average probability of error, $P_e^{(n)}$, satisfies

$$\epsilon \geq p_e^{(n)} := \frac{1}{A^K M^K} \sum_{(t_1,m_1),\ldots,(t_K,m_K)} \mathbb{P}\left[g(Y^n) \neq ((t_1, m_1), \ldots, (t_K, m_K)) \,|H_{((t_1,m_1),\ldots,(t_K,m_K))}\right].$$

The hypothesis that user $i, i \in [K]$ has chosen message m_i and block t_i is denoted by

$$H_{((t_1,m_1),\ldots,(t_K,m_K))}.$$

For the *SAS-MAC with asynchronism level* α, a code is defined over an asynchronous MAC channel with A increasing exponentially with blocklength n as $A_n = e^{n\alpha}$. For both exponential number users $K_n : \log(K_n) = O(n)$, and subexponential number of users $K_n : \log(K_n) = o(n)$, a tuple $(R_1, \ldots, R_{K_n}, \alpha)$ is said to be achievable if there exists a sequence of codes $(e^{nR}, e^{n\alpha}, e^{n\nu}, n, \epsilon_n)$ with ϵ_n vanishing to zero as n goes to infinity.

The capacity region is the closure of all such achievable tuples.

We now consider the performance of the SAS-MAC for three different scalings of the number of users.

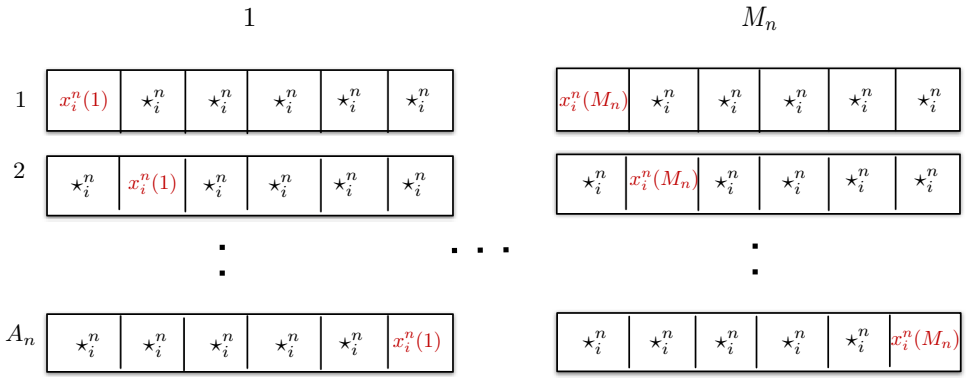


Figure 8. Extended codebook.

4.1.1 Exponential regime: case $\log(K_n) = n\nu : \nu > \alpha$

The following theorem provides a converse bound on ν. More specifically, it provides an upper bound for feasible values of ν.

Theorem 7. *For a SAS-MAC with asynchronism level* α *and* $\log(K_n) = n\nu : \nu > \alpha$*, reliable synchronization is not possible, i.e., even with* $M_n = 1, \forall i \in [K_n]$*, one has strictly positive error probabilities* $P_e^{(n)} > 0$.

Proof. 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 nA_n constructed such that $\forall m_i \in [M_n]$ and $\forall t_i \in [A_n]$

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

as depicted in Fig. Figure 8. By using Fano's inequality, i.e., $H(X_1^{nA_n}, \ldots, X_{K_n}^{nA_n} | Y^{nA_n}) \leq n\epsilon_n$: $\epsilon_n \to 0$ as n goes to infinity, for any codebook of length nA_n we have

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

where $\frac{\log\left(1+\frac{R}{\alpha}\right)}{n} \geq 0$ and $\frac{\log(1+\frac{\epsilon_n}{e^{n\alpha}|\mathcal{Y}|})}{n} \geq 0$ vanish as n goes to infinity. This implies that $\nu \leq \alpha$ is a necessary condition for reliable communications. In other words, for $\nu > \alpha$ not even synchronization, i.e., $M_n = 1, \forall i \in [K_n]$, is possible. □

4.1.2 Sub-exponential regime: case $\log(K_n) = o(n)$

For $\nu < \frac{\alpha}{2}$ the probability δ_n that more than one user transmits a codeword in each block is:

$$\delta_n = 1 - \frac{A_n(A_n - 1)\ldots(A_n - K_n + 1)}{A_n^{K_n}}, \tag{4.2}$$

which goes to zero as n goes to infinity for $\nu < \frac{\alpha}{2}$. Hence one may analyze the probability of error conditioned on the fact that users are transmitting in different blocks, i.e., no collision. This

assumption reduces the number of different hypotheses for each block that must be considered in the error analysis.

Theorem 8. *For a SAS-MAC with asynchronous level* α *and* $\log(K_n) = o(n)$*, the capacity region is the Cartesian product of the corresponding strong asynchronous point-to-point capacities given by*

$$R < \max_{P_i : D([P_i Q_i] \| Q_\star) > \alpha} I(P_i, Q_i), \forall i \in [K_n]. \tag{4.3}$$

Proof. Each user generates an i.i.d. random codebook according to the distribution P_i on $\mathcal{X}, \forall i \in [K_n]$. The decoder uses the following 'slot by slot' strong typicality decoder: for every block $s \in [A]$ it finds the empirical distribution of the output sequence Y_s^n and codeword $x_i^n(m_i)$ for every $m_i \in [e^{nR}]$, $i \in [K_n]$; it announces that m_i was the sent codeword in block s if m_i is the unique message index such that $(x_i^n(m_i), Y_s^n) \in T_\epsilon^n(P_i Q_i)$; if no codeword passes the test, the decoder declares that no user was active on block s and moves forward to block $s + 1$; if more than one codeword passes the test, the decoder picks one uniformly at random and moves forward to block $s + 1$.

Assuming no collision, since all hypotheses are equally likely, and by averaging over all random codes C, we have that the average probability of error is the same as that obtained by conditioning over $H^{(1)} := H_{(1,1)\dots(K_n,1)}$. By the union bound we can write

$$
\begin{aligned}
P_e^{(n)} &\leq \mathbb{P}\left[\text{error}|H^{(1)}\right] + \delta_n \\
&\leq \delta_n + \sum_{i\in[K_n]} \mathbb{P}\left[(x_i^n(1), Y_i^n) \notin T_\epsilon^n(P_iQ_i)|H^{(1)}\right] && (4.4)\\
&+ \sum_{i\in[K_n]} \sum_{m_i\in[2:e^{nR}]} \mathbb{P}\left[(x_i^n(m), Y_i^n) \in T_\epsilon^n(P_iQ_i)|H^{(1)}\right] && (4.5)\\
&+ \sum_{s\in[K_n+1:A]} \sum_{i\in[K_n]} \sum_{m_i\in[e^{nR}]} \mathbb{P}\left[(x_i^n(m_i), Y_s^n) \in T_\epsilon^n(P_iQ_i)|H^{(1)}\right] && (4.6)\\
&+ \sum_{i\in[K_n]} \sum_{\substack{j\in[K_n]\\ j\neq i}} \sum_{m_j\in[e^{nR}]} \mathbb{P}\left[(x_j^n(m_j), Y_i^n) \in T_\epsilon^n(P_jQ_j)|H^{(1)}\right] && (4.7)\\
&\leq \delta_n + \sum_{i\in[K_n]} e^{-n\epsilon^2} + \sum_{i\in[K_n]} e^{-n(I(P_i,Q_i)-R)} && (4.8)\\
&+ \sum_{i\in[K_n]} e^{-n(I(P_i,Q_i)+D([P_iQ_i]\|Q_\star)-\alpha-R)} && (4.9)\\
&+ \sum_{i\in[K_n]} \sum_{\substack{j\in[K_n]\\ j\neq i}} e^{-n(I(P_j,Q_j)+D([P_jQ_j]\|[P_iQ_i])-R)} && (4.10)
\end{aligned}
$$

where δ_n is the probability of collision given in (Equation 4.2) which goes to zero as n goes to infinity, the term (Equation 4.4) is the probability that the true codeword is not typical with its corresponding output, the term in (Equation 4.5) is the probability of classical synchronous point-to-point error, the term in (Equation 4.6) is the probability that a noise block, or a block where no user was active, mimics any of the codewords, and finally the term in (Equation 4.7)

is the probability that users are confused with one another. The bound in (Equation 4.8) is due to the typicality decoder and those in (Equation 4.9) and (Equation 4.10) are proved in Appendix I. All together, by assuming K_n to be sub-exponential in n, so that $K_n e^{-n\epsilon^2} \to 0$, we get

$$R < I(P_j, Q_j), \tag{4.11}$$

$$R + \alpha < I(P_j, Q_j) + D([P_j Q_j] \parallel Q_\star), \tag{4.12}$$

$$R < I(P_j, Q_j) + D([P_j Q_j] \parallel [P_i Q_i]), \forall i \neq j, \tag{4.13}$$

where the bound in (Equation 4.13) is redundant due the more restrictive bound in (Equation 4.11). The achievable rates obtained above match the converse bound given by the Cartesian product of the corresponding point-to-point capacities in (Equation 4.3). Finally, (Equation 4.11) – (Equation 4.12) are equivalent to (Equation 4.3) as proven in [22] and hence the theorem is proved. □

Remark 4. *As it can be seen in* (Equation 4.8)*, this typicality decoder imposes the condition* $e^{n\nu} e^{-n\epsilon^2} \to 0$ *for exponential number of users which corresponds to* $\nu = 0$ *when* $\epsilon \to 0$*. This begs the question of whether indeed it is possible to support exponentially many users* ($\nu > 0$)*. Next subsections affirmatively answer this question. Also note that in calculating the per-user probability of error, all the summations over users* $i \in [K_n]$ *in* (Equation 4.8)*,* (Equation 4.9)*,* (Equation 4.10) *are eliminated and hence would relax the requirement* $\nu = 0$ *for this typicality decoder.*

4.1.3 Exponential regime: case $\log(K_n) = n\nu : 0 < \nu < \frac{\alpha}{2}$

Now we investigate a SAS-MAC with an exponential number of users. This regime is the hardest to deal with as the typicality decoder seems to fail as the number of users grow exponentially fast. For example, if we apply the previously introduced strong typicality decoder to this case, error events as in (Equation 4.4) would restrict ν to be zero. In the following Subsections, we investigate achievability results for different assumptions on the users channels.

4.1.4 Users with Identical Channels

The following theorem is an achievable region for the SAS-MAC for the case that different users have identical channels toward the base station when they are the sole active user. In this scenario, users' identification and decoding can be merged together.

Theorem 9. *For a SAS-MAC with asynchronous level α, occupancy level ν and rate R, assume that $Q_{\{i\}}(y|x) = Q(y|x)$ (recall definition* (Equation 4.1)*) for all users. Then, the following (R, α, ν) region is achievable*

$$\bigcup_{\substack{P\in\mathcal{P}_{\mathcal{X}}\\ \lambda\in[0,1]}} \left\{ \begin{array}{cc} \nu & < \frac{\alpha}{2} \\ \nu & < D(Q_\lambda \parallel Q|P) \\ \alpha + R + \nu & < D(Q_\lambda \parallel Q_\star|P) \\ R + \nu & < I(P, Q) \end{array} \right\}, \tag{4.14}$$

where

$$Q_\lambda(y|x) := \frac{Q(y|x)^\lambda Q_\star(y)^{1-\lambda}}{\sum_{y' \in \mathcal{Y}} Q(y'|x)^\lambda Q_\star(y')^{1-\lambda}}, \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{4.15}$$

Proof. Before starting the proof, we note that for $\nu < \frac{\alpha}{2}$ (first bound in (Equation 4.14)), with probability approaching to one as the blocklength n grows to infinity, the users transmit in distinct blocks. Hence, in analyzing the joint probability of error of our achievability scheme, we can safely condition on the hypothesis that users do not collide. The probability of error given the hypothesis that collision has occurred, which may be large, is then multiplied by the probability of collision and hence is vanishing as the blocklength goes to infinity, regardless of the achievability scheme. The probability of error for this two-stage decoder can be decomposed as

$$\mathbb{P}\left[\text{Error}\right] = \mathbb{P}\left[\text{Synchronization error}\right] \tag{4.16}$$

$$+ \mathbb{P}\left[\text{Decoding error}|\text{No synchronization error}\right]. \tag{4.17}$$

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 (Equation 1.4)). The codeword of user $i \in [K_n]$ for message $m \in [M_n]$ is denoted as $x_i^n(m)$.

Probability of error analysis

A two-stage decoder is used, to first synchronize and then decode (which also identifies the user's identities) the users' messages. We now introduce the two stages and bound the probability of error for each stage.

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

$$T \in [-D(Q_\star \parallel Q|P), D(Q \parallel Q_\star|P)]. \tag{4.18}$$

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

$$L(x_i^n(m), y_j^n) := \frac{1}{n} \log \frac{Q(y_j^n|x_i^n(m))}{Q_\star(y_j^n)} \geq T, \tag{4.19}$$

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

$$H^{(1)} := H_{((1,1),(2,1),\ldots,(K_n,1))}, \tag{4.20}$$

be the hypothesis that user $i \in [K_n]$ is active in block i and sends message $m_i = 1$. The average probability of synchronization error, averaged over the different hypotheses, is upper bounded by

$$\mathbb{P}\left[\text{Synchronization error}\right] = \mathbb{P}\left[\text{Synchronization error}|H^{(1)}\right] \tag{4.21}$$

$$\leq \sum_{j=1}^{K_n} \mathbb{P}\left[\bigcap_{i=1}^{K_n}\bigcap_{m=1}^{M_n} L(x_i^n(m), Y_j^n) < T|H^{(1)}\right] + \sum_{j=K_n+1}^{A_n} \mathbb{P}\left[\bigcup_{i=1}^{K_n}\bigcup_{m=1}^{M_n} L(x_i^n(m), Y_j^n) \geq T|H^{(1)}\right] \tag{4.22}$$

$$\leq \sum_{j=1}^{K_n} \mathbb{P}\left[L(x_j^n(1), Y_j^n) < T|H^{(1)}\right] + \sum_{j=K_n+1}^{A_n}\sum_{i=1}^{K_n}\sum_{m=1}^{M_n} \mathbb{P}\left[L(x_i^n(m), Y_j^n) \geq T|H^{(1)}\right] \tag{4.23}$$

$$\leq e^{n\nu} e^{-nD(Q_\lambda\|Q|P)} + e^{n(\alpha+\nu+R)} e^{-nD(Q_\lambda\|Q_\star|P)}, \tag{4.24}$$

where (Equation 4.21) is by the symmetry of different hypothesis and (Equation 4.24) can be derived as in [51, Chapter 11]. The upper bound on the probability of error for the synchronization error in (Equation 4.24) vanishes as n goes to infinity if the second and third bound in (Equation 4.14) hold.

Decoding stage. In this stage, by conditioning on no synchronization error, we have a superblock of length nK_n, for which we have to distinguish between $K_n!(M_n)^{K_n} \doteq e^{nK_n(R+\nu)}$ different messages. We note that all the codewords for this superblock also have constant composition P (since they are made by the concatenation of constant composition codewords). We can hence use a Maximum Likelihood (ML) decoder for random constant composition codes, introduced and analyzed in [52], on the super-block of length nK_n to distinguish among

$e^{nK_n(R+\nu)}$ different messages with vanishing probability of error if $R+\nu < I(P,Q)$. This retrieves the last bound in (Equation 4.14).

This concludes the proof. □

4.1.5 Users with Different Choice of Channels

We now move on to a more general case in which we remove the restriction that the channels of all users are the same. Theorem 10 finds an achievable region when we allow the users channels to be chosen from a set of conditional distributions of polynomial size in the blocklength.

Theorem 10. *For a SAS-MAC with asynchronous level* α*, occupancy level* ν *and rate* R*, assume that* $Q_i(y|x) = W_{c(i)}(y|x)$ *is the channel for user* $i \in [K_n]$*, for some* $c(i) \in [S_n]$ *where* $S_n = poly(n)$*. Then, the following region is achievable*

$$\bigcup_{n\geq 1}\bigcup_{\substack{P_j\in\mathcal{P}_{\mathcal{X}}\\ \lambda_j\in[0,1]}}\bigcap_{j\in[S_n]}\left\{\begin{array}{ll} \nu_j & < \frac{\alpha}{2} \\ \nu_j & < D([P_jW_j]_{\lambda_j} \parallel [P_jW_j]) \\ \alpha & < D([P_jW_j]_{\lambda_j} \parallel Q_\star) \\ R+\nu_j & < I(P_j,W_j) \end{array}\right\}, \tag{4.25}$$

where

$$\nu_j := \frac{1}{n}\log(\mathcal{N}_j), \tag{4.26}$$

$$\mathcal{N}_j := \sum_{i=1}^{K_n} \mathbb{1}_{\{Q_i=W_j\}} : \sum_{j=1}^{S_n}\mathcal{N}_j = K_n. \tag{4.27}$$

Proof. Before starting the proof, we should note that with similar arguments as the ones in Theorem 9, by imposing the first bound in (Equation 4.25), different users transmit in distinct blocks with a probability which goes to one as blocklength goes to infinity; thus we can assume no user collision in the following. We now propose a three-stage achievability scheme. The three stages perform the task of synchronization, identification and decoding, respectively. The joint probability of error for this three-stage achievability scheme can be decomposed as

$$\mathbb{P}\left[\text{Error}\right] = \mathbb{P}\left[\text{Synchronization error}\right] \tag{4.28}$$

$$+ \mathbb{P}\left[\text{Identification error}|\text{No synchronization error}\right] \tag{4.29}$$

$$+ \mathbb{P}\left[\text{Decoding error}|\text{No synchronization and No identification error}\right]. \tag{4.30}$$

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 $S_n = \text{poly}(n)$ 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 $c(i) \in [S_n]$ 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 $x_i^n(m)$.

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

$$T_{c(i)} \in \left[-D\left(Q_\star \parallel [P_{c(i)}W_{c(i)}]\right), D\left([P_{c(i)}W_{c(i)}] \parallel Q_\star\right)\right], \ i \in [K_n]. \tag{4.31}$$

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

$$L_i(y_j^n) := \frac{1}{n} \log \frac{[P_{c(i)}W_{c(i)}](y_j^n)}{Q_\star(y_j^n)} \geq T_{c(i)}, \tag{4.32}$$

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 (Equation 4.19) by leveraging the constant composition construction of codewords in Theorem 9. In here, we can leverage the i.i.d constructure of the codewords and calculate the probability of error corresponding to (Equation 4.32).

We now find an upper bound on the average probability of error for this scheme over different hypothesis. Before doing so, we should note that by the symmetry of different hypotheses, the average probability of error over different hypothesis is equal to probability of error given the hypothesis that user $i \in [K_n]$ transmits in block i; this hypotheses is denoted by

$$H^{(2)} := H_{((1,.),(2,.),\ldots,(K_n,.))}, \tag{4.33}$$

where $(.)$ is used instead of specifying the messages to emphasize that the decoder finds the location of the users, irrespective of their transmitted messages.

The average probability of synchronization error, averaged over the different hypotheses, is upper bounded by

$$\mathbb{P}\left[\text{Synchronization error}\right] = \mathbb{P}\left[\text{Synchronization error}|H^{(2)}\right] \tag{4.34}$$

$$\leq \mathbb{P}\left[\bigcup_{i=1}^{K_n} L_i(Y_i^n) < T_{c(i)}|H^{(2)}\right] + \mathbb{P}\left[\bigcup_{j=K_n+1}^{A_n}\bigcup_{i=1}^{K_n} L_i(Y_j^n) \geq T_{c(i)}|H^{(2)}\right] \tag{4.35}$$

$$\leq \sum_{i=1}^{K_n} \mathbb{P}\left[L_i(Y_i^n) < T_{c(i)}|H^{(2)}\right] + (A_n - K_n)\mathbb{P}\left[\bigcup_{i=1}^{K_n} L_i(Y^n) \geq T_{c(i)}|H^{(2)}\right] \tag{4.36}$$

$$\leq \sum_{i=1}^{K_n} e^{-nD\left([P_{c(i)}W_{c(i)}]_{\lambda_i}\|[P_{c(i)}W_{c(i)}]\right)} + e^{n\alpha}\sum_{j=1}^{S_n} e^{-nD\left([P_jW_j]_{\lambda_j}\|Q_\star\right)} \tag{4.37}$$

$$= \sum_{j=1}^{S_n} \mathcal{N}_j e^{-nD\left([P_jW_j]_{\lambda_j}\|[P_jW_j]\right)} + e^{n\alpha}e^{-nD\left([P_jW_j]_{\lambda_j}\|Q_\star\right)}, \tag{4.38}$$

where

$$[P_jQ_j]_\lambda(y) := \frac{([P_jQ_j](y))^\lambda (Q_\star(y))^{1-\lambda}}{\sum_{y'\in\mathcal{Y}} ([P_jQ_j](y'))^\lambda (Q_\star(y'))^{1-\lambda}}. \tag{4.39}$$

The probability of error in this stage will decay to zero if for all $j \in [S_n]$

$$\nu_j := \frac{1}{n}\log(\mathcal{N}_j) < D\left([P_jW_j]_{\lambda_j} \parallel [P_jW_j]\right), \tag{4.40}$$

$$\alpha < D\left([P_jW_j]_{\lambda_j} \parallel Q_\star\right). \tag{4.41}$$

This retrieves the second and third bounds in (Equation 4.25).

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

$$[P_{c(i)}Q_{c(i)}](y) := \sum_{x \in \mathcal{X}} P_{c(i)}(x)Q_{c(i)}(y|x). \tag{4.42}$$

We leverage this property and customize the result in Theorem 5 to identify the different distributions of the different users. In Theorem 5, assumed that all the distributions are distinct, while in here, our distributions are not distinct. The only modification that is needed in order to use the result of Theorem 5 is as follows. We need to consider a graph in which the edge between every two similar distributions have 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 goes to zero as blocklength n goes to infinity since

$$\mathbb{P}[\text{Identification error}|\text{No synchronization error}] \leq \sum_{1 \leq i < j \leq S_n} e^{-2nB\left([P_iQ_i],[P_jQ_j]\right)} \to 0, \tag{4.43}$$

and since $S_n = \text{poly}(n)$ by assumption.

Decoding stage. After finding the permutation of users in the active blocks, we can go ahead with the third stage of the achievability scheme to find the transmitted messages of the users. In this stage, we can group the users who have similar channel Q_j to get superblocks of

length $n\mathcal{N}_j, j \in [S_n]$. For each superblock, we have to distinguish $(M_n)^{\mathcal{N}_j}\,(\mathcal{N}_j)^{\mathcal{N}_j} \approx e^{n\mathcal{N}_j(R+\nu_j)}$ different message permutation. By using a typicality decoder, we conclude that the probability of decoding error for each superblock will go exponentially fast (in blocklength) to zero if

$$R + \nu_j < I(P_j, W_j), \forall j \in [S_n]. \tag{4.44}$$

This retrieves the last bound in (Equation 4.25) and concludes the proof. □

Remark 5. *The achievability proof of Theorem 9 relies on constant composition codes whereas the achievability proof of Theorem 10 relies on i.i.d codebook. The reason for these restrictions is that in 10 we also need to distinguish different users and in order to use the result of [3], we focused our attention on i.i.d codebooks.*

4.1.6 Users with no restriction on their channels

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

In our results, we use the following notation. The *Chernoff distance* between two distributions is defined as

$$C(P, Q) := \sup_{0 \le t \le 1} -\log \left(\sum_x P(x)^t Q(x)^{1-t} \right). \tag{4.45}$$

We extend this definition and introduce the quantity

$$C(P_i, Q_i, P_j, Q_j) := \sup_{0 \le t \le 1} \mu_{i,j}(t), \tag{4.46}$$

where

$$\mu_{i,j}(t) := -\log \sum_{x_i, x_j, y} P_i(x_i) P_j(x_j) Q_i(y|x_i)^{1-t} Q_j(y|x_j)^t \tag{4.47}$$

is a concave function of t. We also define

$$C(., Q_\star, P_j, Q_j) := \sup_{0 \le t \le 1} -\log \left(\sum_{x_j, y} P_j(x_j) Q_\star(y)^{1-t} Q_j(y|x_j)^t \right)$$

to address the special case with $i = 0$ and hence all users are idle.

Theorem 11. *For a SAS-MAC with asynchronous level* α*, occupancy level* ν *and rate* R*, the following region is achievable*

$$\bigcup_{n\geq 1}\bigcup_{P_i\in\mathcal{P}_{\mathcal{X}}}\bigcap_{i\in[K_n]}\left\{\begin{array}{cc} \nu & <\frac{\alpha}{2} \\ \nu+R & <B(P_i,Q_i), \\ 2\nu+R & <\inf_{j\neq i}C(P_j,Q_j,P_i,Q_i), \\ \alpha+\nu+R & <C(\,.\,,Q_\star,P_i,Q_i), \end{array}\right\}. \tag{4.48}$$

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:

The receiver uses the following block by block decoder: for each block $s\in[A_n]$, the decoder outputs

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

where $x_0^n=\emptyset$.

We now find an upper bound on probability of error given the hypothesis $H^{(1)}$ in (Equation 4.20) for this decoder as follows

$$
\begin{aligned}
P_e^{(n)} &\leq \sum_{i\in[K_n]} \sum_{m\in[2:M_n]} \mathbb{P}\left[\log \frac{Q_i(Y_i^n \mid x_i^n(m))}{Q_i(Y_i^n \mid x_i^n(1))} > 0 | H^{(1)}\right] \\
&+ \sum_{i\in[K_n]} \sum_{\substack{j\in[0:K_n] \\ j\neq i}} \sum_{m\in[M_n]} \mathbb{P}\left[\log \frac{Q_j(Y_i^n | x_j^n(m))}{Q_i(Y_i^n | x_i^n(1))} > 0 | H^{(1)}\right] \\
&+ \sum_{s\in[K_n+1:A_n]} \sum_{j\in[K_n]} \sum_{m\in[M_n]} \mathbb{P}\left[\log \frac{Q_j(Y_s^n | x_j^n(m))}{Q_\star(Y_s^n)} > 0 | H^{(1)}\right] \\
&\leq \sum_{i\in[K_n]} e^{nR} e^{-n\sup_t - \log \mathbb{E}\left[\left(\frac{Q_i(Y_i|\overline{X}_i)}{Q_i(Y_i|X_i)}\right)^t\right]} \\
&+ \sum_{i\in[K_n]} \sum_{\substack{j\in[0:K_n] \\ j\neq i}} e^{nR} e^{-n\sup_t - \log \mathbb{E}\left[\left(\frac{Q_j(Y_i|X_j)}{Q_i(Y_i|X_i)}\right)^t\right]} \\
&+ e^{n\alpha} \sum_{j\in[K_n]} e^{nR} e^{-n\sup_t - \log \mathbb{E}\left[\left(\frac{Q_j(Y_s|X_j)}{Q_\star(Y_s)}\right)^t\right]},
\end{aligned}
$$

where $P_{X,\overline{X}}(x, x') = P_X(x)P_X(x')$. The last inequality is due to the Chernoff bound. In order for each term in the probability of error upper bound to vanish as n grows to infinity, we find the conditions stated in the theorem. □

Remark 6. *Note that (see Appendix J):*

$$\mathsf{B}(\mathsf{P},\mathsf{Q}) \coloneqq \mathsf{C}(\mathsf{P},\mathsf{Q},\mathsf{P},\mathsf{Q}) = -\log \sum_{x,x',y} \mathsf{P}(x)\mathsf{P}(x')\sqrt{\mathsf{Q}(y|x)\mathsf{Q}(y|x')}, \tag{4.49a}$$

$$\mathsf{C}(\,.\,,\mathsf{Q}_\star,\mathsf{P}_j,\mathsf{Q}_j) \le \mathsf{I}(\mathsf{P}_j,\mathsf{Q}_j) + \mathsf{D}\left([\mathsf{P}_j\mathsf{Q}_j] \parallel \mathsf{Q}_\star\right), \tag{4.49b}$$

$$\mathsf{C}(\mathsf{P}_i,\mathsf{Q}_i,\mathsf{P}_j,\mathsf{Q}_j) \le \mathsf{I}(\mathsf{P}_j,\mathsf{Q}_j) + \mathsf{D}\left(\mathsf{P}_i[\mathsf{P}_j\mathsf{Q}_j] \parallel \mathsf{P}_i\mathsf{Q}_i\right), \tag{4.49c}$$

where, due to symmetry, in $\mathsf{C}(\mathsf{P},\mathsf{Q},\mathsf{P},\mathsf{Q})$ *the supremum is achieved at the midpoint* $t = \frac{1}{2}$*, and hence* $\mathsf{B}(\mathsf{P},\mathsf{Q}) = \mathsf{C}(\mathsf{P},\mathsf{Q},\mathsf{P},\mathsf{Q}) = \mu(\frac{1}{2})$*. The bounds in* (Equation 4.49) *show that in the achievable rates in Theorem 11 are less than the corresponding point-to-point bounds.*

Example 3. *Consider the SAS-MAC with asynchronism level* α*, occupancy level* ν*, and rate* R *with input-output relationship* $Y = \sum_{i\in[K_n]} X_i \oplus Z$ *with* $Z \sim \mathsf{Ber}(\delta)$ *for some* $\delta \in (0, 1/2)$*. In our notation*

$$\mathsf{Q}(y|x) = \mathbb{P}[X_i \oplus Z = y | X_i = x] = \mathbb{P}[Z = x \oplus y] \tag{4.50}$$

$$= \begin{cases} 1-\delta & x \oplus y = 0 \ (\text{i.e., } x = y) \\ \\ \delta & x \oplus y = 1 \ (\text{i.e., } x \neq y) \end{cases}. \tag{4.51}$$

Assume that the input distribution used is $P = \mathrm{Ber}(p)$ *for some* $p \in (0, 1/2)$*. The achievability region of this example, based on Theorem 9, includes the following region*

$$\bigcup_{\substack{p\in[0,\frac{1}{2}]\\ \lambda\in[0,1]}} \left\{ \begin{array}{cc} \nu & < \alpha/2 \\ \nu & < p \cdot d(\epsilon_\lambda \parallel \delta) \\ \alpha + R + \nu & < p \cdot d(\epsilon_\lambda \parallel 1-\delta) \\ R + \nu & < h(p * \delta) - h(p) \end{array} \right\}, \tag{4.52}$$

where

$$\begin{aligned} d(p \parallel q) &:= p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}, \\ \epsilon_\lambda &:= \frac{\delta^\lambda (1-\delta)^{(1-\lambda)}}{\delta^\lambda (1-\delta)^{(1-\lambda)} + (1-\delta)^\lambda \delta^{(1-\delta)}}, \\ p * q &:= p(1-q) + (1-p)q. \end{aligned}$$

Moreover, by assuming $P_i = \mathrm{Ber}(p_i)$ *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 equal to* $t = 1/2$ *and hence the achievability region for this channel based on Theorem 11 is given by*

$$\bigcup_{n\geq 1} \bigcup_{P_i \in \mathcal{P}_\mathcal{X}} \bigcap_{i\in[K_n]} \left\{ \begin{array}{cc} \nu & < \frac{\alpha}{2} \\ \nu + R & < B(P_i, Q) = g(p_i * p_i, \delta), \\ 2\nu + R & < \inf_{j\neq i} C(P_j, Q, P_i, Q) = \inf_{i\neq j} g(p_i * p_j, \delta), \\ \alpha + \nu + R & < C(\,.\,, Q_\star, P_i, Q) = g(p_i, \delta),, \end{array} \right\}$$

where

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

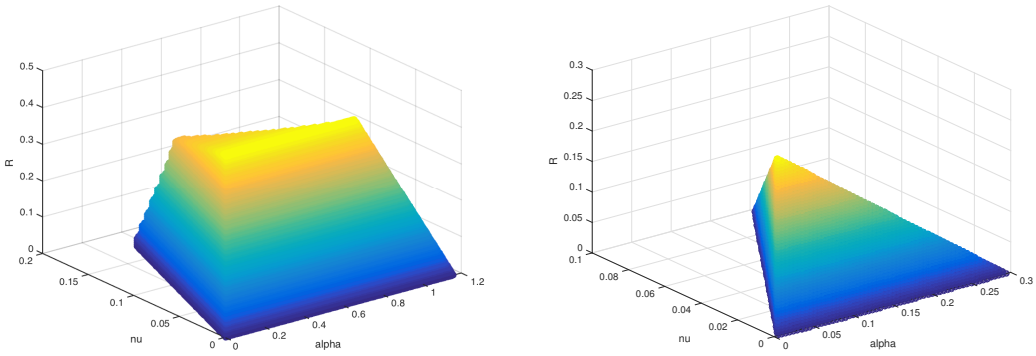
Finally, by symmetry, we can see that the optimal $p_i = \frac{1}{2}, \forall i \in [K_n]$ *and hence* $g(\frac{1}{2},\delta) = -\log\left(1/2+\sqrt{\delta(1-\delta)}\right) > 0$. *So on the BSC(*$\delta$*) strictly positive rates and* ν *are achievable. In this regrard, the region in Theorem 11 reduces to*

$$\alpha+\nu+R < -\log\left(1/2+\sqrt{\delta(1-\delta)}\right). \tag{4.53}$$

The achievable region in (Equation 4.52) *for* (α,ν,R) *is shown in Fig. 9(a). In addition, the achievable region for* (α,ν,R) *with the achievable scheme in Theorem 11 is also plotted in Fig. 9(b) for comparison. Fig. 3 shows that the achievable scheme in Theorem 9 indeed results in a larger achievable region than the one in Theorem 11.*

Note that the fact that the achievability region for Theorem 9 is larger than the achievability region of Theorem 11 for identical channels, is not surprising. In Theorem 9 we separated the synchronization and decoding steps whereas in Theorem 11 we sequentially (in blocks) synchronize and decode at the same time. The sequential decoding step is a restrictive factor in the derived bounds.

Thus far, we have provided two achievability regions for SAS-MAC for the case that different users have identical channels; the case that their channels belong to a set of size that grows polynomially in blocklength, and the case without any restriction on the users' channels.



(a) Achievable region in (Equation 4.52). (b) Achievable region in (Equation 4.53).

Figure 9. Comparison of the achievable region in Theorem 9 and Theorem (11), for the Binary Symmetric Channel with cross over probability $\delta = .11$.

Theorem 12 on the other hand, provides a converse to the capacity region of SAS-MAC that holds for any choice of uesers channels. This theorem can be easily generalized to SAS-MAC as well.

Theorem 12. *For the SAS-MAC with asynchronous level* α*, occupancy level* ν *and rate* R*, such that* $\nu < \alpha/2$*, the following region is impermissible*

$$\bigcup_{n\geq 1}\bigcup_{\substack{i\in[K_n]\\ P_i\in\mathcal{P}_{\mathcal{X}}\\ \lambda_i\in[0,1]}}\left\{\left\{\begin{array}{ll}\nu & > \frac{1}{K_n}\sum_{i=1}^{K_n} D(Q_{i\lambda_i}\parallel Q_i|P_i),\\ \alpha & > \frac{1}{K_n}\sum_{i=1}^{K_n} D(Q_{i\lambda_i}\parallel Q_\star|P_i) - (1-\bar{r}_n)(\nu+R)\end{array}\right\}\bigcup\{R > I(P_i,Q_i)\}\right\}, \quad (4.54)$$

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

$$\bar{r}_n := \inf_T \frac{1}{K_n M_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i\lambda_i}(T \neq i, m|x_i^n(m))). \tag{4.55}$$

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

$$F_n := M_n K_n, \qquad Q_{i,m}^n(y^n) := Q_i(y^n|x_i^n(m)), \tag{4.56}$$

$$\overline{P_{Y^n}}(y^n) := \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m}^n(y^n), \qquad Q_{i,m\lambda_i}^n(y^n) := \frac{(Q_{i,m}^n(y^n))^{\lambda_i} (Q_\star^n(y^n))^{1-\lambda_i}}{\sum_{y^n} \left(Q_{i,m}^n(y^n)\right)^{\lambda_i} (Q_\star^n(y^n))^{1-\lambda_i}}, \tag{4.57}$$

$$\overline{P_{Y^n}^{(\lambda)}}(y^n) := \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m\lambda_i}^n(y^n), \quad \left(\overline{P_{Y^n}}\right)_t (y^n) := \frac{\left(\overline{P_{Y^n}}(y^n)\right)^t (Q_\star^n(y^n))^{1-t}}{\sum_{y^n} \left(\overline{P_{Y^n}}(y^n)\right)^t (Q_\star^n(y^n))^{1-t}}, \tag{4.58}$$

$$Q_\star^n(y^n) := \prod_{i=1}^{n} Q_\star(y_i). \tag{4.59}$$

We use the optimal Maximum Likelihood (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

$$\arg \max_{\substack{(l_1,\ldots,l_{K_n}) \\ l_i \neq l_j, \forall i \neq j \\ l_i \in [A_n], i \in [K_n]}} \sum_{i=1}^{K_n} \log \frac{\overline{P_{Y^n}}(Y_{l_i}^n)}{Q_\star^n(Y_{l_i}^n)}. \tag{4.60}$$

Given the hypothesis that the users are active in blocks $[K_n]$, denoted by $H^{(2)}$ in (Equation 4.33), the corresponding error events to the ML decoder are given by

$$\begin{aligned}\left\{\text{error}|H^{(2)}\right\} &= \bigcup_{\substack{(l_1,\ldots,l_{K_n})\\ \neq(1,\ldots,K_n)}} \left\{\sum_{i=1}^{K_n} \log \frac{\overline{P_{Y^n}}(Y^n_{l_i})}{Q^n_\star(Y^n_{l_i})} > \sum_{i=1}^{K_n} \log \frac{\overline{P_{Y^n}}(Y^n_i)}{Q^n_\star(Y^n_i)}\right\} \\ &\supseteq \bigcup_{\substack{i\in[K_n]\\ j\in[K_n+1:A_n]}} \left\{\log \frac{\overline{P_{Y^n}}(Y^n_j)}{Q^n_\star(Y^n_j)} \geq \log \frac{\overline{P_{Y^n}}(Y^n_i)}{Q^n_\star(Y^n_i)}\right\} \\ &\supseteq \left\{\bigcup_{j\in[K_n+1:A_n]} \log \frac{\overline{P_{Y^n}}(Y^n_j)}{Q^n_\star(Y^n_j)} \geq T\right\} \bigcap \left\{\bigcup_{i\in[K_n]} T \geq \log \frac{\overline{P_{Y^n}}(Y^n_i)}{Q^n_\star(Y^n_i)}\right\}, \end{aligned} \tag{4.61}$$

for any $T \in \mathbb{R}$. We restrict ourselves to a subset of T's and we take T to be

$$T := \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \left(D\left(Q^n_{i,m\lambda_i} \parallel Q^n_\star\right) - D\left(Q^n_{i,m\lambda_i} \parallel \overline{P_{Y^n}}\right)\right), \tag{4.62}$$

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

We also find the following lower bounds, which are proved in the Appendix K,

$$Q^n_\star\left[\log \frac{\overline{P_{Y^n}}}{Q^n_\star}(Y^n) \geq T\right] \geq e^{-\frac{n}{K_n}\left(\sum_{i=1}^{K_n} D(Q_{i\lambda_i}\|Q_\star|P_i) - (R+\nu)(1-\bar{r}_n) + \frac{h(\bar{r}_n)}{n}\right)}, \tag{4.63}$$

$$\overline{P_{Y^n}}\left[\log \frac{\overline{P_{Y^n}}}{Q^n_\star}(Y^n) \leq T\right] \geq e^{-\frac{n}{K_n}\sum_{i=1}^{K_n} D(Q_{i\lambda_i}\|Q_i|P_i)}. \tag{4.64}$$

By using the inequalities in (Equation 4.63) and (Equation 4.64), we find a lower bound on the probability of (Equation 4.61) as follows:

$$\mathbb{P}\left[\bigcup_{j\in[K_n+1:A_n]} \log\frac{\overline{P_{Y^n}}(Y_j^n)}{Q_\star^n(Y_j^n)} \geq T \cap \bigcup_{i\in[K_n]} T \geq \log\frac{\overline{P_{Y^n}}(Y_i^n)}{Q_\star^n(Y_i^n)}|H^{(2)}\right] \tag{4.65}$$

$$= \mathbb{P}\left[\bigcup_{j\in[K_n+1:A_n]} \log\frac{\overline{P_{Y^n}}(Y_j^n)}{Q_\star^n(Y_j^n)} \geq T|H^{(2)}\right] \mathbb{P}\left[\bigcup_{i\in[K_n]} T \geq \log\frac{\overline{P_{Y^n}}(Y_i^n)}{Q_\star^n(Y_i^n)}|H^{(2)}\right] \tag{4.66}$$

$$=: \mathbb{P}\left[Z_1 \geq 1\right]\mathbb{P}[Z_2 \geq 1] \tag{4.67}$$

$$\geq \left(1-\frac{\mathrm{Var}[Z_1]}{\mathbb{E}^2[Z_1]}\right)\left(1-\frac{\mathrm{Var}[Z_2]}{\mathbb{E}^2[Z_2]}\right) \tag{4.68}$$

$$\geq \left(1-\frac{1}{\sum_{j=K_n+1}^{A_n}\mathbb{P}[\xi_j=1]}\right)\left(1-\frac{1}{\sum_{i=1}^{K_n}\mathbb{P}[\zeta_i=1]}\right) \tag{4.69}$$

$$\geq \left(1-e^{-n\alpha+\frac{n}{K_n}\left(\sum_{i=1}^{K_n} D(Q_{i\lambda_i}\|Q_\star|P_i)-(R+\nu)(1-\bar{\tau}_n)+\frac{h(\bar{\tau}_n)}{n}\right)}\right)\left(1-e^{-n\nu+\frac{n}{K_n}\sum_{i=1}^{K_n} D(Q_{i\lambda_i}\|Q_i|P_i)}\right), \tag{4.70}$$

where (Equation 4.66) 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 Equation 4.68 by Chebyshev's inequality, where we have defined

$$Z_1 := \sum_{j=K_n+1}^{A_n} \xi_j, \qquad \xi_j := \mathrm{Ber}\left(Q_\star^n\left[\log\frac{\overline{P_{Y^n}}}{Q_\star^n}(Y_j^n) \geq T\right]\right), \tag{4.71}$$

$$Z_2 := \sum_{i=1}^{K_n} \zeta_i, \qquad \zeta_i := \mathrm{Ber}\left(\overline{P_{Y^n}}\left[\log\frac{\overline{P_{Y^n}}}{Q_\star^n}(Y_i^n) \leq T\right]\right), \tag{4.72}$$

where $\{\xi_j, \zeta_i\}$ are mutually independent. We can see from (Equation 4.70) that if

$$\nu > \frac{1}{K_n} \sum_{i=1}^{K_n} D(Q_{i\lambda_i} \parallel Q_i | P_i), \tag{4.73}$$

$$\alpha > \frac{1}{K_n} \sum_{i=1}^{K_n} D(Q_{i\lambda_i} \parallel Q_\star | P_i) - (1 - \bar{r}_n)(\nu + R), \tag{4.74}$$

then the probability of error is strictly bounded away from zero and hence it is impermissible. Moreover, the usual converse bound on the rate of a synchronous channel also applies to any asynchronous channel and hence the region where $R > I(P, Q)$ is also impermissible. This concludes the proof. □

It should be noted that even though the expression (Equation 4.54) involves a union over all blocklengths n, in order to compute this bound, 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). However, since we still have exponential (in blocklength n) number of users, and in theory we have to optimize all of their distributions, we need to take the union with respect to all blocklengths.

4.2 Discussion and conclusion

In this chapter we studied a Strongly Asynchronous and Slotted Massive Access Channel (SAS-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 global probability of error in which we required all users messages and identities to be jointly correctly decoded. Our results are focused on the region $\nu < \frac{\alpha}{2}$, where the probability

of user collisions in vanishing. We proved in Theorem **??** that for the region $\nu > \alpha$, even synchronization is not possible. Hence, we would like to take this chance to discuss some of the difficulties that one may face in analyzing the region $\frac{\alpha}{2} \leq \nu \leq \alpha$.

As we have mentioned before, for the region $\nu < \frac{\alpha}{2}$, with probability $\frac{\binom{A_n}{K_n}}{(A_n)^{K_n}}$ which approaches to one as blocklength n goes to infinity, the users transmit in distinct blocks. Hence, in analyzing the probability of error of our achievable schemes, we could safely condition on the hypothesis that users are not colliding. For the region $\frac{\alpha}{2} \leq \nu \leq \alpha$, we lose this simplifying assumption. In particular, based on Lemma 13 (proved in the Appendix L), for the region $\frac{\alpha}{2} \leq \nu \leq \alpha$, the probability of every arrangement of users is itself vanishing in the blocklength.

Lemma 13. *For the region $\frac{\alpha}{2} \leq \nu \leq \alpha$ 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 13, one needs to propose an achievable scheme that accounts for every possible arrangement and collision of users and drives the probability of error in all (or most) of the hypothesis to zero. It is also worth noting that the number of possible hypotheses is doubly exponential in the blocklength. Finally, it is worth emphasizing the reason why the authors in [59] can get to $\nu \leq \alpha$. In [59] the authors require the recovery of the messages of a *large fraction* of users and they also require the per-user probability of error to be vanishing. To prove whether or not strictly positive (R, α, ν) are possible in the region $\frac{\alpha}{2} \leq \nu \leq \alpha$, with vanishing global probability of error, is an open problem.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTION

In this thesis, motivated by the new emerging Internet of Things paradigm, we investigate a novel Strongly Asynchronous Slotted Massive Access Channel model. In this model the users are bursty and transmit sporadically within a large time window. Moreover, we allow the number users to grow exponentially with the transmission blocklength. In this model, many conventionally developed information theoretic tools (such as a priori synchronization assumption and typicality arguments) fail. Hence the study of this model requires new technical tools.

In this work we first introduced two models to capture the sporadic transmission of messages of a single bursty user in a network. In both models we assumed that the user may transmit in an asynchronous window $A_n = e^{n\alpha}$ blocks of n channel uses each. In the first model the user transmits a randomly selected message among $M_n = e^{nR}$ messages in each one of the $K_n = e^{n\nu}$ randomly selected blocks of the available $A_n = e^{n\alpha}$ blocks where as in the second model the user may transmit with each block with probability $p_n = e^{-n\beta}$. In the former model we studied the tradeoff between (R, α, ν) and in the latter we investigated the tradeoff between (R, α, β).

Moreover, in conventional information theoretic studies of multiuser networks, the number of users (even for many users) is assumed to be fixed and the capacity results are derived by letting the codeword blocklength n to go to infinity. In an internet of things paradigm however, we have a massive deployment of devices with sensing facilities and the number these

interconnected devices may be even comparable with the coding blocklength. As a result, in the second part of our thesis, we allow the number of users within a network to grow with the coding blocklength. More specifically, each user can transmit a random message among $M_n = e^{nR}$ possible ones only once at random among a window of length $A_n = e^{n\alpha}$. In addition, we also allow the number of users to grow exponentially with blocklength such that we have $K_n = e^{n\nu}$ users within a network. We also require the receiver to distinguish noise from the codewords and the users identity without the use of pilot signals.

What renders the single user version of the problem – a single user transmitting multiple times rather than multiple users transmitting once each – more tractable is that one is guaranteed that in each block there is at most one transmitted message and we do not need to detect the user's identity.

There are still several unexplored paths in this topic, some of which we present here.

5.0.1 SAS-MAC analysis for the regime $\nu \geq \frac{\alpha}{2}$

In this thesis, we restricted ourselves to the regime $\nu < \frac{\alpha}{2}$. By doing so, we could easily assume that in each block we have at most one active user with high probability. For the regime $\nu \geq \frac{\alpha}{2}$ this is no longer true. In this regime, arbitrary subsets of users may be active (where probability of each hypothesis is itself vanishing with blocklength) and one should develop new achievability/converse results do deal with this problem.

5.0.2 Finite blocklength analysis

One of the main reasons for allowing the number of users to grow with the blocklength is that the number of users can be comparable with the coding blocklenth. The asymptotic

analysis of probability of error is a good first step in finding the information theoretic limits of this channel model. However, infinite coding blocklength is not be accommodated and hence the analysis would be more realistic if done for the regime of finite blocklength.

5.0.3 Removing the slot synchronism assumption

In our analysis, we constraint the users to be block synchornous. i.e., their transmission times (even though random) was an integer multiple of the blocklength n and hence users where either non colliding or completely colliding. In a more realistic scenario, there is no coordination between users and their transmission times can start any time within the asynchronous window. It is interesting to understand how this relaxing assumption can change the analysis.

5.0.4 Massive identification problem with a subset of distributions

In our massive identification problem, we assumed that *all* of the distributions will generate a distinct i.i.d sequence. The main problem was to find the permutation of the distributions that generated the samples. A more general case would be to allow only a subset of the distributions to generate i.i.d samples (with an exponential number of distribution in blocklength in total). The main question to answer is to find the subset and to find the permutation of this subset.

APPENDICES

Appendix A

Proof of (Equation 2.13)

The main trick in the proof of (Equation 2.13) is to find an equivalent event and lower bound the probability of that event instead. In this regard we have

$$\mathbb{P}\left[\bigcup_{i\in[K_n]}\frac{1}{n}\log\frac{Q(Y_i^n|x_i^n(m_i))}{Q_{\star^n}(Y_i^n)}\leq T\right] \tag{A.1}$$

$$=\mathbb{P}[Z_1\geq 1] \tag{A.2}$$

$$\geq 1-\frac{\mathrm{Var}[Z_1]}{\mathbb{E}^2[Z_1]}=1-\frac{\sum_{i=1}^{K_n}p_i(1-p_i)}{\left(\sum_{i=1}^{K_n}P_i\right)^2}\geq 1-\frac{1}{\sum_{i=1}^{K_n}p_i} \tag{A.3}$$

$$\geq 1-e^{-n\left(\nu-D\left(Q_{\lambda_i^\star}\|Q|P_{i^\star}\right)\right)}, \tag{A.4}$$

where we define

$$Z_1:=\sum_{i=1}^{K_n}\xi_i,\qquad \xi_i\sim\text{Bernoulli}(p_i),$$

$$p_i:=Q_{x_i^n(m_i)}\left[\frac{1}{n}\log\frac{Q(Y_i^n|x_i^n(m_i))}{Q_{\star^n}(Y_i^n)}\leq T\right]$$

$$p_i\geq Q_{x_i^n(m_i)}\left[Y_i^n\in T_{Q_{\lambda_i}}(x_i^n(m_i))\right]=e^{-nD(Q_{\lambda_i}\|Q|P_i)}. \tag{A.5}$$

The equality in (Equation A.2) is due to the equivalence of the events to the ones in (Equation A.1) and (Equation A.3) is by Chebyshev's inequlity. The inequality in (Equation A.4) is by the

Appendix A (Continued)

choice of $\mathfrak{i}^\star$ in (Equation 2.11) and finally (Equation A.5) is true because of the special choice of $\mathsf{T} = \mathsf{D}(\mathsf{Q}_{\lambda_\mathfrak{i}} \parallel \mathsf{Q}_\star|\mathsf{P}_\mathfrak{i}) - \mathsf{D}(\mathsf{Q}_{\lambda_\mathfrak{i}} \parallel \mathsf{Q}|\mathsf{P}_\mathfrak{i})$.

Appendix B

Proof of (Equation 2.14)

To find a lower bound on the term in (Equation 2.12), we proceed as before by writing

$$\mathbb{P}\left[\bigcup_{j\in[K_n+1:A_n]}\bigcup_{m\in[M_n]}\frac{1}{n}\log\frac{Q\left(Y_j^n|x_{i^\star}^n(m)\right)}{Q_{\star^n}(Y_j^n)}\geq T\right] \tag{B.1}$$

$$=\mathbb{P}\left[Z_2\geq 1\right] \tag{B.2}$$

$$\geq 1-\frac{\mathrm{Var}[Z_2]}{\mathbb{E}^2[Z_2]}=1-\frac{\sum_{j=K_n+1}^{A_n}q_j(1-q_j)}{\left(\sum_{j=K_n+1}^{A_n}q_j\right)^2}\geq 1-\frac{1}{\sum_{j=K_n+1}^{A_n}q_j} \tag{B.3}$$

$$\geq 1-\exp\left\{-n\left(\alpha+R\mathbb{1}_{\{R<I(P,Q_{\lambda_{i^*}})\}}-D(Q_{\lambda_{i^*}}\parallel Q_\star|P_{i^*})\right)\right\},$$

where we have defined

$$Z_2:=\sum_{j\in[K_n+1:A_n]}\zeta_j,\qquad \zeta_j\sim\text{Bernoulli}(q_j),$$

$$q_j:=Q_{\star^n}\left[\bigcup_{m\in[M_n]}\frac{1}{n}\log\frac{Q(Y_j^n|x_j^n(m))}{Q_{\star^n}(Y_j^n)}\geq T\right], \tag{B.4}$$

$$q_j\geq\exp\left\{n\left(R\mathbb{1}_{\left\{R<I(P,Q_{\lambda_j})\right\}}-D(Q_{\lambda_j}\parallel Q_\star|P_j)\right)\right\}. \tag{B.5}$$

The equality in (Equation B.2) is true because the two events in the probabilities are the same and the first inequality in (Equation B.3) is by Chebyshev inequality. The inequality

Appendix B (Continued)

in (Equation B.5) is proved in Appendix C. We should note that $\zeta_j, j \in [K_n + 1 : A_n]$, are independent since $Y_j^n, j \in [K_n + 1 : A_n]$ are independent.

Appendix C

Proof of (Equation B.5)

We first define a new typical set $T^{\delta}_{Q^n_{\lambda+\epsilon}}$ as follows.

Definition 7. *For ϵ and δ define*

$$T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n) := \left\{ y^n : \sum_{a,b} \frac{1}{n}\mathcal{N}(a,b|x^n,y^n) \log \frac{Q(b|a)}{Q_\star(b)} \geq T, \right.$$
$$\left. \left| \frac{1}{n}\mathcal{N}(a,b|x^n,y^n) - P(a)Q_{\lambda+\epsilon}(b|a) \right| < \delta, \forall (a,b) \in \mathcal{X}\times\mathcal{Y} \right\}.$$

The new constraint $\sum_{a,b} \frac{1}{n}\mathcal{N}(a,b|x^n,y^n) \log \frac{Q(b|a)}{Q_\star(b)} \geq T$ that we included in the typical set definition ensures that all the sequences y^n that belong to $T^{\delta}_{Q^n_{\lambda+\epsilon}}$ will also satisfy $\frac{1}{n}\log \frac{Q(y^n|x^n)}{Q_{\star^n}(y^n)} \geq T$. In addition, define

$$\Delta := \sum_{a,b} P(a)Q_{\lambda+\epsilon}(b|a) \log \frac{Q(b|a)}{Q_\star(b)} - T,$$

where $\Delta > 0$ since $T = \sum_{a,b} P(a)Q_\lambda(b|a) \log \frac{Q(b|a)}{Q_\star(b)}$ is decreasing in λ [60]. By the Law of Large Numbers

$$Q^n_{\lambda+\epsilon}\left[\left| \frac{1}{n}\mathcal{N}(a,b|x^n,Y^n) - P(a)Q_{\lambda+\epsilon}(b|a) \right| > \delta | x^n \right] \to 0$$

Appendix C (Continued)

and

$$Q_{\lambda+\epsilon}^{n}\left[\sum_{a,b}\frac{1}{n}\mathcal{N}(a,b|x^{n},Y^{n})\log\frac{Q(b|a)}{Q_{\star}(b)}\geq T|x^{n}\right]\rightarrow 0$$

and hence for any $\delta_1 > 0$ there exists n_1 such that for all $n \geq n_1$ we have

$$Q_{\lambda+\epsilon}^{n}\left[T_{Q_{\lambda+\epsilon}}^{\delta}(x^{n})|x^{n}\right] > 1-\delta_1. \tag{C.1}$$

Moreover, assume that $D_{Q_{\lambda+\epsilon}^{n}}(m)$ is the optimal (and disjoint) decoding region for message m, whose codeword is passed through the channel $Q_{\lambda+\epsilon}^{n}$. We also denote the average probability of decoding error associated with channel $Q_{\lambda+\epsilon}^{n}$ to be

$$P_e^{(n)}(Q_{\lambda+\epsilon}) := \frac{1}{e^{nR}}\sum_{m=1}^{e^{nR}}\sum_{y^{n}\in D_{Q_{\lambda+\epsilon}^{n}}^{c}(m)} Q_{\lambda+\epsilon}^{n}(y^{n}).$$

Now, if we drop half of the codewords in $(x^{n}(1),\ldots,x^{n}(M_n))$ with the largest probability of the error, the remaining half must must all satisfy

$$Q_{\lambda+\epsilon}^{n}\left[Y^{n}\notin D_{Q_{\lambda+\epsilon}^{n}}(m)|x^{n}(m)\right] < 2P_e^{(n)}(Q_{\lambda+\epsilon}); \tag{C.2}$$

otherwise, the average probability of error for the decoding regions $D_{Q_{\lambda+\epsilon}^{n}}(m)$ will be larger than $P_e^{(n)}(Q_{\lambda+\epsilon})$ and we reach a contradiction. Henceforth we restrict our attention to this half of the codebook (which without loss of generality we assume is the first $\frac{M_n}{2}$ codewords).

Appendix C (Continued)

As the result for the channel $Q^n_{\lambda+\epsilon}$ and its optimal decoding regions $D_{Q^n_{\lambda+\epsilon}}(m)$, by (Equation C.1) and (Equation C.2) we have

$$Q^n_{\lambda+\epsilon}\left[T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n(m)) \cap D_{Q^n_{\lambda+\epsilon}}(m)|x^n(m)\right] \geq 1-\delta_1-2P_e^{(n)}(Q_{\lambda+\epsilon}). \tag{C.3}$$

In addition, we can conclude from [22, Lemma 10] that for any two distributions P_1^n, P_2^n and any event A such that

$$P_1^n(A) \geq \alpha,$$

we have

$$P_2^n(A) \geq \beta_\alpha(P_1^n, P_2^n) \geq \frac{\alpha}{2}\exp\{-nD(P_1 \parallel P_2)\}. \tag{C.4}$$

Appendix C (Continued)

In case the lower bound given in (Equation C.3), i.e. $1-\delta_1-P_e^{(n)}(Q_{\lambda+\epsilon})$, is positive (which we discuss shortly) and by (Equation C.4) we can write

$$\begin{aligned}
&Q_{\star^n}\left[\bigcup_{m\in[M_n]}\frac{1}{n}\log\frac{Q(Y_i^n|x^n(m))}{Q_\star^n(Y_i^n)}\geq T\right]\geq Q_{\star^n}\left[\bigcup_{m\in[\frac{M_n}{2}]}T^\delta_{Q^n_{\lambda+\epsilon}}(x^n(m))\right]\\
&\geq Q_{\star^n}\left[\bigcup_{m\in[\frac{M_n}{2}]}T^\delta_{Q^n_{\lambda+\epsilon}}(x^n(m))\cap D_{Q^n_{\lambda+\epsilon}}(m)\right]\\
&=\sum_{m=1}^{\frac{M_n}{2}}Q_{\star^n}\left[T^\delta_{Q^n_{\lambda+\epsilon}}(x^n(m))\cap D_{Q^n_{\lambda+\epsilon}}(m)\right]\\
&\geq\sum_{m=1}^{\frac{M_n}{2}}\frac{1-\delta_1-2P_e^{(n)}(Q_{\lambda+\epsilon})}{2}e^{-nD(Q_{\lambda+\epsilon}\|Q_\star|P)}\\
&\doteq e^{nR}e^{-nD(Q_{\lambda+\epsilon}\|Q_\star|P)}.
\end{aligned}\tag{C.5}$$

In addition, due to continuity of the divergence, as ϵ vanishes to zero, we have

$$D(Q_{\lambda+\epsilon}\parallel Q_\star|P)\to D(Q_\lambda\parallel Q_\star|P).$$

We now disccus the cases that $1-\delta_1-P_e^{(n)}(Q_{\lambda+\epsilon})$ is positive. A sufficient condition for $1-\delta_1-P_e^{(n)}(Q_{\lambda+\epsilon})$ to be positive is that $P_e^{(n)}(Q_{\lambda+\epsilon})$ be vanishing as n goes to infinity. This would be true if

$$R<I(P,Q_{\lambda+\epsilon}).$$

Appendix C (Continued)

If, on the other hand $R \geq I(P, Q_{\lambda+\epsilon})$, we still can lower bound (Equation B.4) by

$$\begin{aligned} Q_{\star^n}\left[\bigcup_{m\in[M_n]} \frac{1}{n}\log\frac{Q(Y_j^n|x^n(m))}{Q_{\star^n}(Y_j^n)} \geq T\right] &\geq Q_{\star^n}\left[\frac{1}{n}\log\frac{Q(Y_j^n|x^n(1))}{Q_{\star^n}(Y_j^n)} \geq T\right] \\ &\geq Q_{\star^n}\left[Y_j^n \in T^{\delta}_{Q_\lambda}(x^n(1))\right] \\ &\geq e^{-nD(Q_\lambda\|Q_\star|P)}. \end{aligned}$$

Appendix D

Proof of Lemma 3

We provide the proof for a binary alphabet $\mathcal{X} = \{a, b\}$ in a proof by contradiction. The proof for the general $|\mathcal{X}| > 2$ is a straightforward generalization. For $x = a, b$ define

$$E_0^{(x)}(\lambda_x) := D\left(Q_{\lambda_x} \parallel Q_\star\right),$$

$$E_1^{(x)}(\lambda_x) := D\left(Q_{\lambda_x} \parallel Q_x\right).$$

Assume that the claim of the Lemma 3 is not valid and hence there exists $(\lambda_a, \lambda_b, \widetilde{\lambda}) \in [0, 1]^3$ such that

$$D(Q_{\lambda_x} \parallel Q|P) < D(Q_{\widetilde{\lambda}} \parallel Q|P),$$

$$D(Q_{\lambda_x} \parallel Q_\star|P) < D(Q_{\widetilde{\lambda}} \parallel Q_\star|P),$$

or equivalently

$$\rho E_1^{(a)}(\lambda_a) + \bar{\rho} E_1^{(b)}(\lambda_b) < \rho E_1^{(a)}(\widetilde{\lambda}) + \bar{\rho} E_1^{(b)}(\widetilde{\lambda}), \tag{D.1a}$$

$$\rho E_0^{(a)}(\lambda_a) + \bar{\rho} E_0^{(b)}(\lambda_b) < \rho E_0^{(a)}(\widetilde{\lambda}) + \bar{\rho} E_0^{(b)}(\widetilde{\lambda}), \tag{D.1b}$$

Appendix D (Continued)

where $\rho := \mathbb{P}(x = a)$ and $\bar{\rho} = 1 - \rho = \mathbb{P}(x = b)$. By [60, Theorem 2] we can exclude the cases where $\lambda_a, \lambda_b < \widetilde{\lambda}$ and $\lambda_a, \lambda_b > \widetilde{\lambda}$ and assume $\lambda_a < \widetilde{\lambda} < \lambda_b$, which implies

$$E_1^{(x)}(\lambda_a) > E_1^{(x)}(\widetilde{\lambda}) > E_1^{(x)}(\lambda_b),$$
$$E_0^{(x)}(\lambda_a) < E_0^{(x)}(\widetilde{\lambda}) < E_0^{(x)}(\lambda_b),$$

for $x \in \{a, b\}$. Hence, by rearranging (Equation D.1) and by dividing the two equations, we get

$$\frac{\left(E_1^{(a)}(\lambda_a) - E_1^{(a)}(\widetilde{\lambda})\right)}{\left(E_0^{(a)}(\lambda_a) - E_0^{(a)}(\widetilde{\lambda})\right)} > \frac{\left(E_1^{(b)}(\widetilde{\lambda}) - E_1^{(b)}(\lambda_b)\right)}{\left(E_0^{(b)}(\widetilde{\lambda}) - E_0^{(b)}(\lambda_b)\right)}. \tag{D.2}$$

Note since the $\left(E_0^{(x)}(\lambda), E_1^{(x)}(\lambda)\right)$ curve is convex and strictly decreasing, we have

$$\frac{\partial E_1^{(a)}\left(E_0^{(a)}(\lambda)\right)}{\partial \lambda}\Big|_{\lambda=\widetilde{\lambda}} \geq \frac{\left(E_1^{(a)}(\lambda_a) - E_1^{(a)}(\lambda)\right)}{\left(E_0^{(a)}(\lambda_a) - E_0^{(a)}(\lambda)\right)}, \tag{D.3}$$
$$\frac{\left(E_1^{(b)}(\lambda) - E_1^{(b)}(\lambda_b)\right)}{\left(E_0^{(b)}(\lambda) - E_0^{(b)}(\lambda_b)\right)} \geq \frac{\partial E_1^{(b)}\left(E_0^{(b)}(\lambda)\right)}{\partial \lambda}\Big|_{\lambda=\widetilde{\lambda}}, \tag{D.4}$$

where $\frac{\partial E_1^{(x)}\left(E_0^{(x)}(\lambda)\right)}{\partial \lambda}$ is the slope of the $\left(E_0^{(x)}(\lambda), E_1^{(x)}(\lambda)\right)$, which can be visually seen in Fig. Figure 10. However, according to [60, Theorem 6], the slope of the $\left(E_0^{(x)}(\lambda), E_1^{(x)}(\lambda)\right)$ curve at $\lambda = \widetilde{\lambda}$ is equal to $\frac{\widetilde{\lambda}-1}{\widetilde{\lambda}}$ and is independent of x.

Putting (Equation D.2), (Equation D.3) and (Equation D.4) together, we reach a contradiction and the proof is complete.

Appendix D (Continued)

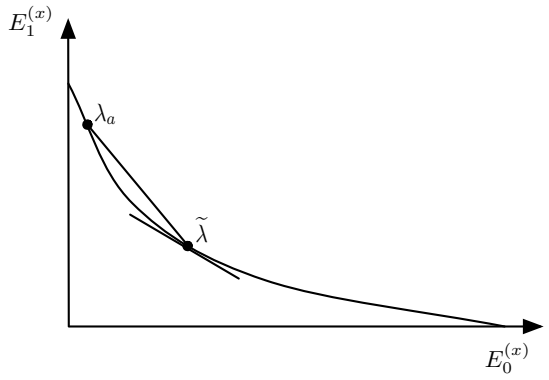


Figure 10. Slope at $\lambda = \widetilde{\lambda}$ is larger than the slope of the line between λ_a and $\widetilde{\lambda}$.

Appendix E

Proof of (Equation 2.22)

Note that

$$\begin{aligned}\sum_{k=1}^{A_n-1}\binom{A_n}{k}p^k(1-p)^{A_n-k}\frac{1}{k} &= \frac{1}{A_n+1}\sum_{k=1}^{A_n-1}\binom{A_n+1}{k+1}p^k(1-p)^{A_n-k}\frac{k+1}{k}\\ &\le \frac{2}{A_n+1}\sum_{k=1}^{A_n-1}\binom{A_n+1}{k+1}p^k(1-p)^{A_n-k}\\ &\le \frac{2}{p(A_n+1)}\sum_{j=0}^{A_n+1}\binom{A_n+1}{j}p^j(1-p)^{A_n+1-j}\\ &= \frac{2}{p(A_n+1)} \le \frac{2}{pA_n},\end{aligned}$$

and similarly

$$\sum_{k=1}^{A_n-1}\binom{A_n}{k}p^k(1-p)^{A_n-k}\frac{1}{A_n-k} \le \frac{2}{(1-p)A_n}.$$

Appendix F

Proof of Lemma 6

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

$$r(n_k)^{\frac{r}{2}-1}\left(G(c_1)+\ldots G(c_{N_{r,k}})\right) \leq \frac{N_{r,k}r}{n_k}\left(a_1^2+\ldots+{a_{n_k}}^2\right)^{\frac{r}{2}}. \tag{F.1}$$

We may drop the subscripts and use $N := N_{r,k}$ and $n := n_k$ in the following for notational ease. Our goal is to expand the right hand side (RHS) of (Equation F.1) such that all elements have coefficient 1. Then, we parse these elements into N different groups (details will be provided later) such that using the AM-GM inequality (i.e., $n\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \leq \sum_{i=1}^{n} a_i$) on each group, we get one of the N terms on the LHS of (Equation F.1). Before stating the rigorous proof, we provide an example of this strategy for the graph with $k = 4$ vertices shown in Fig. Figure 11. In this example, we consider the Lemma for $r = 4$ cycles (for which $N = 3$).

Appendix F (Continued)

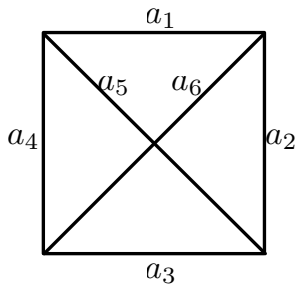


Figure 11. A complete graph with 4 vertices

We may expand the RHS in (Equation F.1) as

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

$$\begin{aligned}
\Theta_1=&\{a_1^4+a_2^4+a_3^4+a_4^4+a_1^2a_3^2+a_1^2a_3^2+a_2^2a_4^2+a_2^2a_4^2\\
&+a_1^2a_2^2+a_1^2a_2^2+a_1^2a_2^2+a_1^2a_2^2+a_1^2a_4^2+a_1^2a_4^2+a_1^2a_4^2+a_1^2a_4^2\\
&+a_2^2a_3^2+a_2^2a_3^2+a_2^2a_3^2+a_2^2a_3^2+a_3^2a_4^2+a_3^2a_4^2+a_3^2a_4^2+a_3^2a_4^2\}\\
\Theta_2=&\{a_1^4+a_6^4+a_3^4+a_5^4+a_5^2a_6^2+a_5^2a_6^2+a_1^2a_3^2+a_1^2a_3^2\\
&+a_1^2a_6^2+a_1^2a_6^2+a_1^2a_6^2+a_1^2a_6^2+a_1^2a_5^2+a_1^2a_5^2+a_1^2a_5^2+a_1^2a_5^2\\
&+a_3^2a_6^2+a_3^2a_6^2+a_3^2a_6^2+a_3^2a_6^2+a_3^2a_5^2+a_3^2a_5^2+a_3^2a_5^2+a_3^2a_5^2\}\\
\Theta_3=&\{a_4^4+a_5^4+a_2^4+a_6^4+a_5^2a_6^2+a_5^2a_6^2+a_2^2a_4^2+a_2^2a_4^2\\
&+a_4^2a_5^2+a_4^2a_5^2+a_4^2a_5^2+a_4^2a_5^2+a_4^2a_6^2+a_4^2a_6^2+a_4^2a_6^2+a_4^2a_6^2\\
&+a_2^2a_5^2+a_2^2a_5^2+a_2^2a_5^2+a_2^2a_5^2+a_2^2a_6^2+a_2^2a_6^2+a_2^2a_6^2+a_2^2a_6^2\}.
\end{aligned}$$

Appendix F (Continued)

It can be easily seen that if we use the AM-GM inequality on Θ_1, Θ_2 and Θ_3, we can get the lower bound equal to $24(a_1a_2a_3a_4), 24(a_1a_6a_3a_5)$ and $24(a_4a_5a_2a_6)$, respectively where $rn^{\frac{r}{2}-1} = 24$ and hence (Equation F.1) holds in this example.

We proceed to prove Lemma 6 for arbitrary k and (even) $r \geq 2$. We propose the following scheme to group the elements on the RHS of (Equation F.1) and then we prove that this grouping indeed leads to the claimed inequality in the Lemma.

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 (Equation F.1). In this regard, we consider all possible subsets of the edges of cycle c_i with $1 : \frac{r}{2}$ 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 (Equation F.1) 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 (Equation F.1). 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 $\frac{1}{l}$ 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 (so that we can assign the same amount of elements to other cycles with similar edges).

We state some facts, which can be easily verified:

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

Appendix F (Continued)

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

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

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

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

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

$$\Theta_i \geq r n^{\frac{r}{2}-1} \left(a_{i_1}^{n_{i_1}} \times \ldots \times a_{i_r}^{n_{1_r}}\right)^{\left(\frac{1}{r n^{\frac{r}{2}-1}}\right)},$$
$$\Theta_j \geq r n^{\frac{r}{2}-1} \left(a_{j_1}^{n_{j_1}} \times \ldots \times a_{j_r}^{n_{j_r}}\right)^{\left(\frac{1}{r n^{\frac{r}{2}-1}}\right)},$$

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

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

Appendix F (Continued)

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.

$$\Theta_i \geq r n^{\frac{r}{2}-1} \left(a_{i_1}^{q_i} \times \ldots \times a_{i_r}^{q_i} \right)^{\left(\frac{1}{r n^{\frac{r}{2}-1}} \right)}. \tag{F.3}$$

It can be seen from (Equation F.2) and (Equation F.3) that all the elements of all groups have the same power $n_{i_t} = p, \forall i \in [N], t \in [r]$. i.e.,

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

Since each element on the RHS of (Equation F.1) 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 (Equation F.1)) 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

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

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

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

and the Lemma is proved for even r.

Appendix F (Continued)

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 (Equation F.1) 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 (Equation F.1) is at most the multiplication of rm different a_i's and on the LHS of (Equation F.1), 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, we can now continue the proof with the same strategy as even values of r for the odd values of r.

Appendix G

Proof of (Equation 3.9)

By Lemma 6 and (Equation 3.6) we can write

$$\begin{aligned} P_e^{(n)} &\le \sum_{r=2}^{A_n} \sum_{c \in C_{A_n}^{(r)}} G(c) \\ &\le \sum_{r=2}^{A_n} \frac{N_{r,A_n}}{(n_{A_n})^{\frac{r}{2}}} \left(a_1^2 + \ldots + a_{n_{A_n}}^2\right)^{r/2} \\ &\le \sum_{r=2}^{A_n} 4^r \left(\sum_{1 \le i < j \le A_n} e^{-2nB(P_i,P_j)} \right)^{r/2} \qquad \text{(G.1)} \\ &\le \frac{16 \left(\sum_{1 \le i < j \le A_n} e^{-2nB(P_i,P_j)} \right)}{1 - 4\sqrt{\sum_{1 \le i < j \le A_n} e^{-2nB(P_i,P_j)}}}, \qquad \text{(G.2)} \end{aligned}$$

where (Equation G.1) is by Fact 1 (see Appendix F) and

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

As the result, (Equation G.2) will go to zero as n goes to infinity if

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

Appendix H

Proof of (Equation 3.12)

We upper bound the denominator of (Equation 3.10) by

$$\begin{aligned}
\mathbb{P}[\xi_{i,j},\xi_{i,k}] &= \mathbb{P}\left[\log\frac{P_i(X_j^n)}{P_j(X_j^n)}+\log\frac{P_j(X_i^n)}{P_i(X_i^n)}\geq 0\cap\ \log\frac{P_i(X_k^n)}{P_k(X_k^n)}+\log\frac{P_k(X_i^n)}{P_i(X_i^n)}\geq 0\right]\\
&\leq \mathbb{P}\left[\log\frac{P_i(X_j^n)}{P_j(X_j^n)}+\log\frac{P_j(X_i^n)}{P_i(X_i^n)}+\log\frac{P_i(X_k^n)}{P_k(X_k^n)}+\log\frac{P_k(X_i^n)}{P_i(X_i^n)}\geq 0\right]\\
&\leq \exp\left\{n\inf_t\log\left(\mathbb{E}\left[\left(\frac{P_i(X_j^n)}{P_j(X_j^n)}\cdot\frac{P_j(X_i^n)}{P_i(X_i^n)}\cdot\frac{P_i(X_k^n)}{P_k(X_k^n)}\cdot\frac{P_k(X_i^n)}{P_i(X_i^n)}\right)^t\right]\right)\right\}\\
&\leq \exp\left\{n\log\mathbb{E}\left[\left(\frac{P_i(X_j^n)}{P_j(X_j^n)}\cdot\frac{P_j(X_i^n)}{P_i(X_i^n)}\cdot\frac{P_i(X_k^n)}{P_k(X_k^n)}\cdot\frac{P_k(X_i^n)}{P_i(X_i^n)}\right)^{\frac{1}{2}}\right]\right\}\\
&= \exp\{-nB(P_i,P_j)-nB(P_j,P_k)-nB(P_i,P_k)\}. \qquad \text{(H.1)}
\end{aligned}$$

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

Appendix I

Proof of (Equation 4.10)

For any joint empirical distribution J defined on $\mathcal{X}_j \times \mathcal{Y}, 1 \leq i \leq K_n$

$$\begin{aligned}
\mathbb{P}[(x_j^n(m_j), Y_i^n) \in T_\epsilon^n(P_j Q_j) | H^{(1)}] &\leq \sum_{J: J \in T_\epsilon^n(P_j Q_j)} \mathbb{P}\left[\widehat{P}_{(x_j^n(m_j), Y_i^n)} = J | H^{(1)}\right] \\
&\overset{(a)}{=} \sum_{J: J \in T_\epsilon^n(P_j Q_j)} e^{-nD(J \| P_j [P_i Q_i])} \\
&\leq \sum_{J: J \in T_\epsilon^n(P_j Q_j)} e^{-n\left(D(P_j Q_j \| P_j [P_i Q_i]) - \delta_\epsilon\right)} \\
&\overset{(b)}{\leq} \mathrm{poly}(n)\, e^{-n\left(D(P_j Q_j \| P_j [P_i Q_i]) - \delta_\epsilon\right)} \\
&= \mathrm{poly}(n)\, e^{-n\left(I(P_j, Q_j) + D([P_j Q_j] \| [P_i Q_i]) - \delta_\epsilon\right)},
\end{aligned}$$

where δ_ϵ can be made arbitrary small with the choice ϵ. Equality in (a) is due to [53, Lemma 2.6] and (b) is by [53, Lemma 2.2]. With similar reasoning, (Equation 4.9) can be proved.

Appendix J

Proof of (Equation 4.49b)

We find an upper bound on $C(\,.\,, Q_\star, P_j, Q_j)$ by noting that $\mu_{0,j}(t)$ in (Equation 4.47) is concave in t with $\mu_{0,j}(1) = 0$ and

$$\frac{\partial \mu_{0,j}(t)}{\partial t} \,|_{t=1} = -I(P_j, Q_j) - D([P_j Q_j] \parallel Q_\star) \leq 0.$$

Hence $\mu_{0,j}(t)$ is always less than $(I(P_j, Q_j) + D([P_j Q_j] \parallel Q_\star))(1-t)$ and that for $0 \leq t \leq 1$ it is always less than $I(P_j, Q_j) + D([P_j Q_j] \parallel Q_\star)$.

The inequality in (Equation 4.49c) follows similarly.

Appendix K

Proof of (Equation 4.63) **and** (Equation 4.64)

Before moving on to calculation of lower bounds on (Equation 4.63) and (Equation 4.64), we note that at the expense of a small decrease in the rate, [22, Eq. 137], we may further restrict our attention to constant composition codewords. Henceforth, we assume that the composition of the codewords for user $i, i \in [K_n]$ is given by $P_i, i \in [K_n]$. Moreover, to make this thesis self-contained, we restate the following Lemmas that we use in the rest of the proof.

Lemma 14 (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:*

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

Lemma 15 (Fano)**.** *Let* F *be an arbitrary set of size* N*. For* $\bar{P} = \frac{\sum_{\theta \in F} P_\theta}{N}$ *we have*

$$\frac{1}{N} \sum_{\theta \in F} D\left(P_\theta \parallel \bar{P}\right) \geq (1 - \bar{r}) \log\left(N(1 - \bar{r})\right) + \bar{r} \log\left(\frac{N\bar{r}}{N-1}\right), \tag{K.2}$$

Appendix K (Continued)

where

$$\bar{r} := \inf_{T} \frac{1}{N} \sum_{\theta \in \mathbb{F}} P_\theta \{T \neq \theta\} \tag{K.3}$$

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

We now continue with the proof of (Equation 4.63). Using the Chernoff bound we can write

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

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

$$\begin{aligned}
A(t) &:= \frac{t}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m\lambda_i}^n \parallel Q_\star^n\right) - D\left(Q_{i,m\lambda_i}^n \parallel \overline{P_{Y^n}}\right) - \log \mathbb{E}_{Q_\star^n} \left[\left(\frac{\overline{P_{Y^n}}}{Q_\star^n}(Y^n) \right)^t \right] \\
&= \frac{t}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m\lambda_i}^n(y^n) \log \frac{\overline{P_{Y^n}}(y^n)}{Q_\star^n(y^n)} - \log \sum_{y^n} \left(\overline{P_{Y^n}}(y^n)\right)^t \left(Q_\star^n(y^n)\right)^{1-t} \\
&= \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m\lambda_i}^n(y^n) \log \frac{\left(\overline{P_{Y^n}}(y^n)\right)^t \left(Q_\star^n(y^n)\right)^{1-t}}{\overline{\sum_{y^n} \left(\overline{P_{Y^n}}(y^n)\right)^t \left(Q_\star^n(y^n)\right)^{1-t}}} \\
&= \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m\lambda_i}^n(y^n) \log \frac{\left(\overline{P_{Y^n}}\right)_t(y^n)}{Q_\star^n(y^n)} \\
&= \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m\lambda_i}^n \parallel Q_\star^n\right) - D\left(Q_{i,m\lambda_i}^n \parallel \left(\overline{P_{Y^n}}\right)_t\right) \\
&= \frac{1}{K_n} \sum_{i=1}^{K_n} nD\left(Q_{i\lambda_i} \parallel Q_\star | P_i\right) - \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m\lambda_i}^n \parallel \left(\overline{P_{Y^n}}\right)_t\right),
\end{aligned} \tag{K.4}$$

Appendix K (Continued)

where (Equation K.4) is the result of constant composition structure of the codewords. As the result

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

Moreover,

$$\begin{aligned}
\inf_t \frac{1}{F_n}\sum_{i=1}^{K_n}\sum_{m=1}^{M_n} D\left(Q^n_{i,m\lambda_i} \parallel \left(\overline{P_{Y^n}}\right)_t\right) &= \inf_t \frac{1}{F_n}\sum_{i=1}^{K_n}\sum_{m=1}^{M_n}\sum_{y^n} Q^n_{i,m\lambda_i}(y^n)\log\frac{Q^n_{i,m\lambda_i}(y^n)}{\left(\overline{P_{Y^n}}\right)_t(y^n)} \\
&= \inf_t \frac{1}{F_n}\sum_{i=1}^{K_n}\sum_{m=1}^{M_n}\sum_{y^n} Q^n_{i,m\lambda_i}(y^n)\log\frac{Q^n_{i,m\lambda_i}(y^n)}{\left(\overline{P_{Y^n}}\right)_t(y^n)}\cdot\frac{\overline{P^{(\lambda)}_{Y^n}}(y^n)}{\overline{P^{(\lambda)}_{Y^n}}(y^n)} \\
&= \inf_t \frac{1}{F_n}\sum_{i=1}^{K_n}\sum_{m=1}^{M_n} D\left(Q^n_{i,m\lambda_i} \parallel \overline{P^{(\lambda)}_{Y^n}}\right) + D\left(\overline{P^{(\lambda)}_{Y^n}} \parallel \left(\overline{P_{Y^n}}\right)_t\right) \\
&\geq \frac{1}{F_n}\sum_{i=1}^{K_n}\sum_{m=1}^{M_n} D\left(Q^n_{i,m\lambda_i} \parallel \overline{P^{(\lambda)}_{Y^n}}\right).
\end{aligned}$$

Note that $\overline{P^{(\lambda)}_{Y^n}}$ is the average of $Q^n_{i,m\lambda_i}$ over different m, i's and hence based on Lemma 15, we have

$$\begin{aligned}
\frac{1}{F_n}\sum_{i=1}^{K_n}\sum_{m=1}^{M_n} D\left(Q^n_{i,m} \parallel \overline{P^{(\lambda)}_{Y^n}}\right) &\geq (1-\bar{r}_n)\log\left(F_n(1-\bar{r}_n)\right) + \bar{r}_n\log\left(\frac{F_n\bar{r}_n}{F_n-1}\right) \\
&\approx (1-\bar{r}_n)\log F_n - h(\bar{r}_n),
\end{aligned}$$

Appendix K (Continued)

where $h(.)$ is the binary entropy function. As the result

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

Now we continue with the proof of (Equation 4.64). Again, using the Chernoff bound we have

$$\begin{aligned}
&\overline{P_{Y^n}} \left[\log \frac{P_{Y^n}}{Q_\star^n}(Y^n) \leq \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m\lambda_i}^n \parallel Q_\star^n\right) - D\left(Q_{i,m\lambda_i}^n \parallel \overline{P_{Y^n}}\right) \right] \\
&= \overline{P_{Y^n}} \left[\log \frac{Q_\star^n}{P_{Y^n}}(Y^n) \geq \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m\lambda_i}^n \parallel \overline{P_{Y^n}}\right) - D\left(Q_{i,m\lambda_i}^n \parallel Q_\star^n\right) \right] \doteq e^{-\sup_t B(t)},
\end{aligned}$$

where

$$\begin{aligned}
\sup_t B(t) &:= \sup_t \frac{t}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} \sum_{y^n} Q_{i,m\lambda_i}^n(y^n) \log \frac{Q_\star^n(y^n)\overline{P_{Y^n}}(y^n)}{\overline{P_{Y^n}}(y^n)} - \log \sum_{y^n} \left(Q_\star^n(y^n)\right)^t \left(\overline{P_{Y^n}}(y^n)\right)^{1-t} \\
&= \sup_t \sum_{y^n} \overline{P_{Y^n}^{(\lambda)}}(y^n) \log \frac{\left(\overline{P_{Y^n}}\right)_{1-t}(y^n)}{\overline{P_{Y^n}}(y^n)} \\
&= \sup_t D\left(\overline{P_{Y^n}^{(\lambda)}} \parallel \overline{P_{Y^n}}\right) - D\left(\overline{P_{Y^n}^{(\lambda)}} \parallel \left(\overline{P_{Y^n}}\right)_{1-t}\right) \\
&\leq D\left(\overline{P_{Y^n}^{(\lambda)}} \parallel \overline{P_{Y^n}}\right) = \sum_{y^n} \left(\frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m\lambda_i}^n(y^n) \right) \log \left(\frac{\frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m\lambda_i}^n(y^n)}{\frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m}^n(y^n)} \right) \\
&\leq \sum_{y^n} \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} Q_{i,m\lambda_i}^n(y^n) \log \frac{Q_{i,m\lambda_i}^n(y^n)}{Q_{i,m}^n(y^n)} = \frac{1}{F_n} \sum_{i=1}^{K_n} \sum_{m=1}^{M_n} D\left(Q_{i,m\lambda_i}^n \parallel Q_{i,m}^n\right) \qquad \text{(K.5)} \\
&= \frac{1}{K_n} \sum_{i=1}^{K_n} nD\left(Q_{i\lambda_i} \parallel Q | P_i\right), \qquad \text{(K.6)}
\end{aligned}$$

and where the inequality in (Equation K.5) is by Log-Sum inequality.

Appendix L

Proof of Lemma 13

We will prove the Lemma by contradiction.

Define

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

Assume that the arrangement with highest probability (lets call it $\mathcal{A}$) has at least two blocks, say blocks $1, 2$, for which $\mathsf{t_1} - \mathsf{t_2} > 1$. 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

$$\begin{aligned}\mathbb{P}(\mathcal{A}) &\propto \binom{\mathsf{K_n}}{\mathsf{t_1}}\binom{\mathsf{K_n}-\mathsf{t_1}}{\mathsf{t_2}} = \frac{\mathsf{K_n}!}{\mathsf{t_1}!(\mathsf{K_n}-\mathsf{t_1})!}\frac{(\mathsf{K_n}-\mathsf{t_1})!}{\mathsf{t_2}!(\mathsf{K_n}-\mathsf{t_1}-\mathsf{t_2})!} \\ &= \frac{\mathsf{K_n}!}{\mathsf{t_1}!\mathsf{t_2}!(\mathsf{K_n}-\mathsf{t_1}-\mathsf{t_2})!}.\end{aligned}$$

Appendix L (Continued)

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. Probability of this new arrangement is proportional to

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

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 user within them in the arrangement with the highest probability.

Appendix M

IEEE POLICY ON THESES AND DISSERTATIONS

This Appendix includes the copyright permission granted from the IEEE to use published work in thesis. The following statement has been copied from the CopyRight Clearance Center (Rightslink).

"The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1. The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2. Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3. In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use

Appendix M (Continued)

of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to `http://www.ieee.org/publications_standards/publications/rights/rights_link.html` to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation."

CITED LITERATURE

1. Shahi, S., Tuninetti, D., and Devroye, N.: On the capacity of strong asynchronous multiple access channels with a large number of users. In IEEE International Symposium on Information Theory (ISIT), pages 1486–1490, July 2016.

2. Shahi, S., Tuninetti, D., and Devroye, N.: On the capacity of the slotted strongly asynchronous channel with a bursty user. In 2017 IEEE Information Theory Workshop (ITW), pages 91–95, Nov 2017.

3. Shahi, S., Tuninetti, D., and Devroye, N.: On identifying a massive number of distributions. In IEEE International Symposium on Information Theory (ISIT), June 2018.

4. Gartner press release: Gartner says 8.4 billion connected "things" will be in use in 2017, up 31 percent from 2016. http://www.gartner.com/newsroom/id/3598917, February 17 2017.

5. Shafiq, M. Z., Ji, L., Liu, A. X., Pang, J., and Wang, J.: A first look at cellular machine-to-machine traffic: large scale measurement and characterization. ACM SIGMETRICS Performance Evaluation Review, 40(1):65–76, 2012.

6. Dawy, Z., Saad, W., Ghosh, A., Andrews, J. G., and Yaacoub, E.: Toward massive machine type cellular communications. IEEE Wireless Communications, 24(1):120–128, February 2017.

7. Shariatmadari, H., Ratasuk, R., Iraji, S., Laya, A., Taleb, T., Jäntti, R., and Ghosh, A.: Machine-type communications: current status and future perspectives toward 5g systems. IEEE Communications Magazine, 53(9):10–17, 2015.

8. Berger, T., Mehravari, N., Towsley, D., and Wolf, J.: Random multiple-access communication and group testing. IEEE Transactions on Communications, 32(7):769–779, 1984.

9. Wainwright, M. J.: Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. IEEE Transactions on Information Theory, 55(12):5728–5741, 2009.

CITED LITERATURE (Continued)

10. Chen, Z., Sohrabi, F., and Yu, W.: Sparse activity detection for massive connectivity. arXiv preprint arXiv:1801.05873, 2018.

11. Inan, H. A., Kairouz, P., and Ozgur, A.: Sparse group testing codes for low-energy massive random access. In Communication, Control, and Computing (Allerton), 2017 55th Annual Allerton Conference on, pages 658–665. IEEE, 2017.

12. Pratas, N. K., Pattathil, S., Stefanović, Č., and Popovski, P.: Massive machine-type communication (mmtc) access with integrated authentication. In Communications (ICC), 2017 IEEE International Conference on, pages 1–6. IEEE, 2017.

13. Ali, M. S., Hossain, E., and Kim, D. I.: Lte/lte-a random access for massive machine-type communications in smart cities. IEEE Communications Magazine, 55(1):76–83, 2017.

14. Ordentlich, O. and Polyanskiy, Y.: Low complexity schemes for the random access Gaussian channel. In Information Theory (ISIT), 2017 IEEE International Symposium on, pages 2528–2532. IEEE, 2017.

15. Khaleghi, E., Adjih, C., Alloum, A., and Muhlethaler, P.: Near-far effect on coded slotted aloha. In PIMRC 2017-IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications-Workshop WS-07 on" The Internet of Things (IoT), the Road Ahead: Applications, Challenges, and Solutions", 2017.

16. Chen, X., Chen, T.-Y., and Guo, D.: Capacity of Gaussian many-access channels. IEEE Transactions on Information Theory, 63(6):3516–3539, 2017.

17. Polyanskiy, Y.: A perspective on massive random-access. In 2017 IEEE International Symposium on Information Theory (ISIT), pages 2523–2527, June 2017.

18. Effros, M., Kostina, V., and Yavas, R. C.: Random access channel coding in the finite blocklength regime. arXiv preprint arXiv:1801.09018, 2018.

19. Chen, T.-Y., Chen, X., and Guo, D.: Many-broadcast channels: Definition and capacity in the degraded case. In Information Theory (ISIT), 2014 IEEE International Symposium on, pages 2569–2573. IEEE, 2014.

20. Chandar, V., Tchamkerten, A., and Tse, D.: Asynchronous capacity per unit cost. IEEE Transactions on Information Theory, 59(3):1213–1226, March 2013.

CITED LITERATURE (Continued)

21. Chandar, V., Tchamkerten, A., and Wornell, G.: Optimal sequential frame synchronization. IEEE Transactions on Information Theory, 54(8):3725–3728, Aug 2008.

22. Polyanskiy, Y.: Asynchronous communication: Exact synchronization, universality, and dispersion. IEEE Transactions on Information Theory, 59(3):1256–1270, March 2013.

23. Chen, X., Chen, T. Y., and Guo, D.: Capacity of Gaussian many-access channels. IEEE Transactions on Information Theory, 63(6):3516–3539, June 2017.

24. Liu, L. and Yu, W.: Massive connectivity with massive MIMO-part i: Device activity detection and channel estimation. arXiv preprint arXiv:1706.06438, 2017.

25. Ahlswede, R. and Wegener, I.: Search problems. John Wiley & Sons, Inc., 1987.

26. Kiefer, J. and Sobel, M.: Sequential identification and ranking procedures, with special reference to Koopman-Darmois populations. University of Chicago Press, 1968.

27. Ahlswede, R. and Haroutunian, E.: On logarithmically asymptotically optimal testing of hypotheses and identification. In General Theory of Information Transfer and Combinatorics, pages 553–571. Springer, 2006.

28. Haroutunian, E. and Hakobyan, P.: Multiple hypotheses lao testing for many independent objects. Scholarly Research Exchange, 2009, 2009.

29. Haroutunian, E. and Hakobyan, P.: Multiple objects: error exponents in hypotheses testing and identification. In Information Theory, Combinatorics, and Search Theory, pages 313–345. Springer, 2013.

30. Haroutunian, E. A., Haroutunian, M. E., Harutyunyan, A. N., et al.: Reliability criteria in information theory and in statistical hypothesis testing. Foundations and Trends® in Communications and Information Theory, 4(2–3):97–263, 2008.

31. Unnikrishnan, J.: Asymptotically optimal matching of multiple sequences to source distributions and training sequences. IEEE Transactions on Information Theory, 61(1):452–468, 2015.

32. Zou, S., Liang, Y., Poor, H. V., and Shi, X.: Nonparametric detection of anomalous data streams. IEEE Transactions on Signal Processing, 65(21):5785–5797, 2017.

CITED LITERATURE (Continued)

33. Li, Y., Nitinawarat, S., and Veeravalli, V. V.: Universal outlier hypothesis testing. IEEE Transactions on Information Theory, 60(7):4066–4082, 2014.

34. Li, Y., Nitinawarat, S., Su, Y., and Veeravalli, V. V.: Universal outlier hypothesis testing: Application to anomaly detection. In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on, pages 5595–5599. IEEE, 2015.

35. Cohen, K. and Zhao, Q.: Quickest anomaly detection: A case of active hypothesis testing. In Information Theory and Applications Workshop (ITA), 2014, pages 1–5. IEEE, 2014.

36. Naini, F. M., Unnikrishnan, J., Thiran, P., and Vetterli, M.: Where you are is who you are: User identification by matching statistics. IEEE Transactions on Information Forensics and Security, 11(2):358–372, 2016.

37. Cover, T., McEliece, R., and Posner, E. C.: Asynchronous multiple-access channel capacity. Information Theory, IEEE Transactions on, 27(4):409–413, Jul 1981.

38. Hui, J. and Humblet, P.: The capacity region of the totally asynchronous multiple-access channel. Information Theory, IEEE Transactions on, 31(2):207–216, Mar 1985.

39. Tchamkerten, A., Chandar, V., and Wornell, G. W.: Communication under strong asynchronism. IEEE Transactions on Information Theory, 55(10):4508–4528, 2009.

40. Chandar, V., Tchamkerten, A., and Tse, D.: Asynchronous capacity per unit cost. In Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, pages 280–284, June 2010.

41. Tchamkerten, A., Chandar, V., and Caire, G.: Energy and sampling constrained asynchronous communication. In Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on, pages 2518–2522, July 2013.

42. Tchamkerten, A., Chandar, V., and Wornell, G. W.: Asynchronous communication: Capacity bounds and suboptimality of training. IEEE Transactions on Information Theory, 59(3):1227–1255, March 2013.

43. Yıldırım, S. A., Martinez, A., and i Fàbregas, A. G.: Achievable rates and exponents for asynchronous communication with ml decoding. In Information Theory (ISIT), 2015 IEEE International Symposium on, pages 96–100. IEEE, 2015.

CITED LITERATURE (Continued)

44. Chandar, V. and Tchamkerten, A.: Asynchronous capacity per unit cost under a receiver sampling constraint. In Information Theory (ISIT), 2015 IEEE International Symposium on, pages 511–515. IEEE, 2015.

45. Chandar, V. and Tchamkerten, A.: Sampling constrained asynchronous communication: How to sleep efficiently. IEEE Transactions on Information Theory, 64(3):1867–1878, 2018.

46. Li, L. and Tchamkerten, A.: Second order asymptotics for communication under strong asynchronism. arXiv preprint arXiv:1710.07025, 2017.

47. Weinberger, N. and Merhav, N.: Codeword or noise? exact random coding exponents for joint detection and decoding. IEEE Transactions on Information Theory, 60(9):5077–5094, 2014.

48. Weinberger, N. and Merhav, N.: Channel detection in coded communication. IEEE Transactions on Information Theory, 63(10):6364–6392, 2017.

49. Merhav, N.: Asymptotically optimal decision rules for joint detection and source coding. IEEE Transactions on Information Theory, 60(11):6787–6795, 2014.

50. Chen, X. and Guo, D.: Many-access channels: The gaussian case with random user activities. In Information Theory (ISIT), 2014 IEEE International Symposium on, pages 3127–3131, June 2014.

51. Cover, T. M. and Thomas, J. A.: Elements of information theory. John Wiley & Sons, 2012.

52. Moulin, P.: The log-volume of optimal constant-composition codes for memoryless channels, within o (1) bits. In Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, pages 826–830. IEEE, 2012.

53. Csiszar, I. and Körner, J.: Information theory: coding theorems for discrete memoryless systems. Cambridge University Press, 2011.

54. Neyman, J. and Pearson, E. S.: On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231:289–337, 1933.

CITED LITERATURE (Continued)

55. Blahut, R.: Hypothesis testing and information theory. IEEE Transactions on Information Theory, 20(4):405–417, 1974.

56. Tuncel, E.: Extensions of error exponent analysis in hypothesis testing. In IEEE International Symposium on Information Theory (ISIT), 2005.

57. Wald, A.: Sequential tests of statistical hypotheses. The Annals of Mathematical Statistics, 16(2):117–186, 1945.

58. Chung, K. L. and Erdos, P.: On the application of the borel-cantelli lemma. Transactions of the American Mathematical Society, 72(1):179–186, 1952.

59. Chandar, V. and Tchamkerten, A.: A note on bursty mac. 2015.

60. Blahut, R.: Hypothesis testing and information theory. IEEE Transactions on Information Theory, 20(4):405–417, Jul 1974.

VITA

Name: Sara Shahi

Education: Ph.D in Electrical and Computer Engineering, University of Illinois at Chicago, 2018

Dual B.Sc in Electrical Engineering and in Biomedical Engineering, Amirkabir University of Technology, 2013

Teaching: Teaching Assistant for Introduction to Electrical Engineering, University of Illinois at Chicago, 2013-2014

Teaching Assistant for Communications I, Amirkabir University of Technology, 2012

Publication: **Sara Shahi**, Daniela Tuninetti, Natasha Devroye, "On the Capacity of the AWGN Channel with Additive Radar Interference", IEEE Transactions on Communications, vol. 66, no. 2, pp. 629-643, Feb. 2018.

VITA (Continued)

Sara Shahi, Daniela Tuninetti, Natasha Devroye, "On the Capacity of the Slotted Strongly Asynchronous Channel with a Bursty User", submitted to IEEE Transactions on Information Theory.

Sara Shahi, Daniela Tuninetti, Natasha Devroye, "The Strongly Asynchronous Massive Access Channel", submitted to IEEE Transactions on Information Theory.

Sara Shahi, Daniela Tuninetti, Natasha Devroye, "On Identifying a Massive Number of Distributions", IEEE International Symposium on Information Theory (ISIT), 2018.

Naruporn Nartasilpa, **Sara Shahi**, Ahmad Salim, Daniela Tuninetti, Natasha Devroye, David Zilz, Mark Bell, "Let's Share CommRad: Co-existing Communications and Radar Systems", IEEE Radar Conference (RadarConf18), Oklahoma City, OK, 2018, pp. 1278-1283.

Sara Shahi, Daniela Tuninetti, Natasha Devroye, "On the Capacity of the Slotted Strongly Asynchronous Channel with a Bursty User", Information Theory Workshop (ITW), Kaohsiung, Taiwan, Nov 2017.

VITA (Continued)

Sara Shahi, Daniela Tuninetti, Natasha Devroye, "On the Capacity of the AWGN Channel with Additive Radar Interference", 54th annual Allerton conference on Communication, Control and Computing, Monticello, IL, Sep 2016.

Sara Shahi, Daniela Tuninetti, Natasha Devroye, "On the Capacity of Strong Asynchronous Multiple Access Channels with a Large Number of Users", In IEEE International Symposium on Information Theory (ISIT), pages 1486-1490, July 2016.

Yanying Chen, **Sara Shahi**, Natasha Devroye, "Colour-and-Forward Relaying: What the Destination Needs in the Zero-Error Primitive Relay Channel", 52nd annual Allerton conference on Communication, Control and Computing, Monticello, IL, Oct 2014.