Distributed Coding in A Multiple Access Environment

Yanru Tang[1], Faeze Heydaryan[2] and Jie Luo[3]

[1]*Electrical and Computer Engineering Department, Colorado State University; yrtang@colostate.edu*
[2]*Electrical and Computer Engineering Department, Colorado State University; faeze66@colostate.edu*
[3]*Electrical and Computer Engineering Department, Colorado State University; rockey@colostate.edu*

ABSTRACT

With the fast expansion of communication networks and the increasing dynamic of wireless communication activities, a significant proportion of messages in wireless networks are being transmitted using distributed protocols that feature opportunistic channel access without full user coordination. This challenges the basic assumption of long message transmissions among coordinated users in classical channel coding theory. In this monograph, we introduce channel coding theorems for the distributed communication model where users choose their channel codes individually. We show that, although reliable message recovery is not always guaranteed in distributed communication systems, the notion of fundamental limit still exists, and can indeed be viewed as an extension to its classical correspondence.

Due to historical priority of developing wireline networks, network architectures tend to achieve system modularity by compromising communication and energy efficiency. Such a

choice is reasonable for wireline systems but can be disastrous for wireless radio networks. Therefore, to reduce efficiency loss, large scale communication networks often adopt wireless communication only at the last hop. Because of such a special structure, architectural inefficiency in wireless part of the network can be mitigated by enhancing the interface between the physical and the data link layers. The enhanced interface, to be proposed, provides each link layer user with multiple transmission options, and supports efficient distributed networking by enabling advanced communication adaptation at the data link layer. In this monograph, we focus on the introduction of distributed channel coding theory, which serves as the physical layer foundation for the enhanced physical-link layer interface. Nevertheless, early research results at the data link layer for the enhanced interface are also presented and discussed.

1

Introduction

A fundamental challenge in wireless networking is to efficiently share the open wireless channel among highly dynamic users. Classical information theory [20] and network theory [11] both have been investigating this key topic for half a century, but from two different angles and along two separate paths that have not yet converged [22].

Because wireless medium often needs to be shared among devices with tight bandwidth and power budgets, communication efficiency is a central concern in wireless systems. Classical information theory [20], particularly channel coding theory, addresses the "efficiency" concern by characterizing the fundamental performance limitation of a wireless channel, and this consequently provides design guidance for wireless systems to achieve or to approach the theoretical efficiency limits. However, information theory was originally developed in an environment when major wireless applications, such as mobile telephony and TV broadcast, only involved transmitting long messages to or from a small number of structured users. To achieve optimal efficiency, channel coding theory suggests that users in a communication party should jointly choose their channel codes, which includes the joint optimization of communication parameters such as information rate and transmission power [29][21][20]. This is termed the "coordinated" communication model in this monograph. Classical channel coding theory assumes that, so long as the messages are long enough and their corresponding coding schemes are optimized, overhead and possible inefficiency in coordinating the communication party should be negligible.

Wireless devices nowadays are often connected into communication networks which typically involve large numbers of users and a wide range of network functions. Modularized architecture is a crucial requirement for developing such large complex network systems [11]. Classical network theory addresses the "modularity" concern by proposing layered network architectures such as the open systems interconnection (OSI) model and its variations [96][69]. By partitioning communication functions into abstraction layers with clearly defined interfaces, OSI model allows system design and optimization to be focused on one or a small number of neighboring layers without the worry of how the outcome can fit into the general system. However, modularity usually does not come without a cost, and compromising low priority resources is a natural choice for achieving system modularity. Classical network theory was originally developed in an environment when the key demand was to connect computers to build the wireline internet infrastructure. For wireline systems, bandwidth of a network cable and communication power of a computer are relatively abundant. Consequently, classical network theory emphasizes the support of a wide range of communication functions in the design of layering interfaces and network protocols, but pays relatively less attention to the impact that the design proposals can have on communication efficiency of the involved systems.

With the computing power of mobile devices and wireless sensors exceeding previous generation large computers, the demand of wireless networking applications is increasing at a dramatic pace. However, developments of advanced wireless networks still suffer from the lack of a theoretical foundation that addresses both concerns of "efficiency" and "modularity" simultaneously. Because classical information theory and network theory each only emphasizes one aspect of the concerns and ignores the other one that is equally important, the need of a unification of the two classical frameworks should be quite apparent [22].

Indeed, such a vision has been recognized for decades, as witnessed by a long list of publications ranging from cross layer utility optimizations [76][93][27] to understanding networking phenomena from information theoretic perspectives [31][60][5][56], from the milestone results on wireless network scaling law [38][39][92], to the celebrated development of fountain channel coding [15][54][74][70][88], and to the historical discovery of network coding [2][52][49][91][43]. These results investigated efficiency problems in various layers of the network architecture from different perspectives. However, not all the problems are specific to wireless networks and therefore are not necessarily among the list of pressing concerns due to the increasing demand of wireless networking. Most of the research results mentioned above also did not suggest explicit architectural revisions to address the corresponding efficiency problems.

The viewpoint that we are going to introduce in this monograph is unique in the following senses. The associated architectural problem lies in the physical and the data link layers. It is an efficiency bottleneck, but only for wireless part of the networks. Furthermore, the research investigations to be presented are motivated and centered around a particular proposal of interface enhancement between the physical and the data link layers. The proposal was originally suggested in [58][87] and then in [55], but has never been thoroughly presented and explained. Therefore, this monograph serves as the first rigorous, in a relative sense, introduction of the research vision and the corresponding research results.

1.1 The Single-hop Cellular Structure

Direct extensions of classical information theoretic and network theoretic frameworks to wireless networking have their own inherited challenges at the bottom two layers, especially when there is a lack of balanced respect to the efficiency and the modularity concerns. Understanding these challenges is essential for identifying the missing pieces needed for the potential unification of the classical frameworks.

On one hand, channel coding theory provides design guidance by characterizing performance limitations such as channel capacity of a communication system. While such efforts have been highly successful in single user [71][72][81] and structured multiuser systems such as multiple access [1][53][90][94] and broadcast systems [19][9][10][30][89], the picture does not look so bright when it comes to a general multiuser network. Deriving channel capacity or capacity region of a general multiuser system is often extremely challenging. Even if one can be confident about solving the capacity problems, an equally important concern is the assumption of the coordinated communication model which has infiltrated into many aspects of the channel coding problem formulations [22]. More specifically, because a wireless network often involves a significant number of users with dynamic short message transmissions, the assumption that all users can be fully coordinated with a negligible overhead is no longer justified in such an environment. Performance limitations obtained in classical channel coding theory provide little guidance to the design and optimizations of distributed and partially distributed communication systems, which are commonly seen in wireless networks [11].

On the other hand, while extending the existing network architecture to wireless systems appeared to be more practical, not all extensions can stand the test of time. With revisions to handle wireless-specific problems such as the hidden and the exposed nodes problems [7], wireless devices can be effectively connected to carry out networking functions. Such extension enabled the exponential growth of Wi-Fi networks [77], which belong to the class of single-hop wireless networks in the sense that either the transmitter or the receiver in each transmission is directly connected to a wireline network. In Wi-Fi networks, wireless routers and client devices are often organized into a cellular-type structure with each micro cell being managed by one router and with interference between different cells well controlled via channel or space separations. By scheduling communication activities within each cell, and exploiting multiple access, broadcast and multiple antenna communication techniques, communication efficiency can be managed at an acceptable level. However, when it comes to multi-hop wireless networks, such as multi-hop bluetooth networks [61] and WiMax networks [3], the stories are quite different. While wireless devices can be connected effectively, most of the proposed multi-hop wireless networks failed to become popular mainly due to their low communication efficiency. Although it is well known that the throughput of wireless systems often does not scale well [38][39][92], the fact that only Wi-Fi-type networks can sustain an acceptable level of efficiency is primarily due to the architectural design details that intentionally or unintentionally compromised bandwidth and energy efficiency of many of the wireless systems.

Because of the difficulties in extending classical theoretical frameworks, major network systems tend to use wireline networks as their backbone and to use wireless links only at the last hop. Wireless devices are often organized into a cellular-type structure to best exploit operational guidance from both classical information theory and classical network theory. In this monograph, we term this special structure the "single-hop cellular structure", as illustrated in Figure 1.1. There have

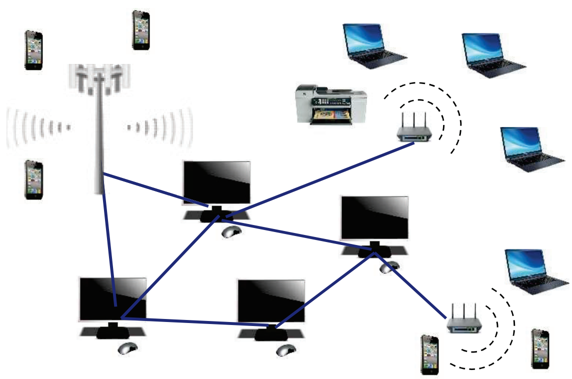

Figure 1.1: The single-hop cellular structure.

been continuous demands and research efforts to extend wireless systems beyond the single-hop cellular structure [22][37]. However, most of these efforts face a clear dilemma. That is, while the inefficiency of the current network architecture limited its capability in supporting complex wireless network structures, a complete redesign of the network architecture is also in lack of a strong incentive because the current architecture does work reasonably well for the wireline part of the networks. This dilemma does not necessarily imply that an ultimate unification of the classical theories will not happen. It does however suggest that consummation of the classical frameworks should be carried out in well motivated steps.

In the rest of the monograph, we will only consider wireless networks with the single-hop cellular structure due to its dominance in current wireless systems. Because a wireless channel usually has a much lower capacity than a wireline cable, with the objective of addressing the throughput bottleneck, we also choose to focus on the bottom two layers of the network, i.e., the physical and the data link layers. Note that once a data packet travels one hop into the wireline network, bandwidth and energy efficiency is no longer the primary concern, and hence research challenges at the higher layers become fundamentally different. Nevertheless, even with just two layers and a special network structure, the necessity of unifying information theory and network theory for wireless systems is still quite convincing.

1.2 The Missing Support of Distributed Communication

Data networks often have large numbers of bursty short messages that need to be disseminated in a timely manner [11][22]. Coordinating all users in a communication party in such an environment can be infeasible or expensive in terms of overhead. A significant proportion of the messages in current wireless networks are therefore transmitted using distributed communication protocols, where an individual user can adjust its communication parameters, such as a transmission/idling decision, without sharing such a decision with other users including its targeted receiver [58]. Such a communication model is incompatible with the joint coding design assumption of the classical channel coding theory. Distributed communication can also cause key issues that do not appear in a coordinated communication system. For example, without full user coordination, data packets transmitted from multiple users can experience collision at their receivers [60]. Collision detection and collision resolution therefore are core problems at the physical and the data link layers [11]. However, these problems are completely ignored in classical channel coding theory [20].

One may think that classical network theory and current network architecture provide reasonable support for distributed communication and networking at the bottom two layers. Unfortunately, this is true only for wireline systems when communication efficiency is not a key concern. Current layering architecture assumes that a link layer user can only determine whether a packet should be transmitted or not [11]. Other communication details are handled at the physical layer. In distributed communication when physical layer does not have full capability of joint channel code optimization, data link layer has to get involved into communication adaptation. A simple example is the collision resolution protocols such as the exponential backoff-based DCF protocol in IEEE 802.11 [12]. However, with each link layer user only having binary transmission/idling options, advanced wireless capabilities such as rate, power and antenna beam adaptations all become irrelevant at the data link layer. This can lead to a quite significant efficiency reduction in the throughput performance of a wireless system.

For example, let us consider a multiple access system with K homogenous users and a single receiver. Assume unit channel gain from each user to the receiver, and additive Gaussian noise with zero mean and variance N_0. Assume that each user has a transmission power of P. From classical channel coding theory [20], we know that, if each user encodes its own message at a rate of $\frac{1}{2}\log_2\left(1+\frac{P}{N_0}\right)$ bits/symbol, then reliable message recovery is only possible if the users transmit sequentially. Sum rate of the system therefore is upper bounded by the single user channel capacity of $C_1 = \frac{1}{2}\log_2\left(1+\frac{P}{N_0}\right)$ bits/symbol, irrespective of the user number K. Alternatively, if users transmit in parallel with an individual rate of $\frac{1}{2K}\log_2\left(1+\frac{KP}{N_0}\right)$, then sum rate of the system can approach the sum channel capacity of $C_K = \frac{1}{2}\log_2\left(1+\frac{KP}{N_0}\right)$ bits/symbol, which grows unboundedly in K. A similar conclusion applies to the same system with a distributed communication model as well. Assume that each user has bursty short messages and cannot afford the overhead of joint coding optimization. If each message is encoded at a rate only slightly less than C_1[1], then sum rate of the system

[1]Note that the rate needs to be smaller than C_1 in order to support reliable decoding with a finite codeword length [28].

is upper bounded by C_1 bits/symbol. Alternatively, if messages arrive with a statistics such that on average $\tilde{K}$ users should have messages to transmit at any moment, then it is generally beneficial to encode each message at a rate close to $\frac{1}{2\tilde{K}} \log_2 \left(1 + \frac{\tilde{K}P}{N_0}\right)$ to support parallel transmissions from up to $\tilde{K}$ users. However, because traffic statistics is unknown at the design stage of a protocol and may also vary in time, in the case of distributed communication, maintaining a high throughput efficiency requires users to have reasonable flexibility of adapting their communication parameters, such as communication rate, at the data link layer. Such a capability is not supported by the physical-link layer interface in the current network architecture.

1.3 An Enhanced Physical-Link Layer Interface

The nature of distributed communication implies that communication parameters cannot be jointly and fully optimized at the physical layer. However, system traffic at the data link layer may still be more or less stationary. To improve communication efficiency, data link layer should exploit advanced wireless capabilities to adapt transmission schemes accordingly, and this needs to be done under the constraint of maintaining a layered (or modularized) network architecture.

To achieve such an objective, we propose an enhancement to the physical-link layer interface [55]. In the enhanced interface, each link layer user can be equipped with multiple transmission options as opposed to the binary transmission/idling options. Different transmission options may correspond to different communication settings such as different power, rates or antenna beams. We generally assume that each link layer user should have a handful of possibly device-dependent transmission options. To maintain the layered architecture, under the distributed communication model, we assume that link layer protocol should inform the physical layer whether a message needs to be transmitted, and if so, which transmission option should be used. Such decisions are not controlled or optimized at the physical layer. We assume that a physical layer receiver should decode the message only if a pre-determined error probability threshold can be met [11][55]. Otherwise the receiver should report collision to the data link layer. At the data link layer, we assume that a user can only choose from the list of provided transmission options, as opposed to being able to adapt the communication parameters arbitrarily.

While the interface enhancement appears to be minor, it involves key research questions whose answers cannot be found in the classical frameworks. At the physical layer, due to possible lack of user coordination, reliable message delivery cannot always be guaranteed. However, it is a fundamental requirement in the layered architecture that any message forwarded to the data link layer must be reliable [11]. Furthermore, because transmission decisions are made at the data link layer, i.e., they are not controlled by a physical layer protocol, any assumption of such a control, such as information rate optimization, may not be valid in physical layer channel coding. With these constraints, whether the notion of fundamental limit still exists for a distributed communication system is a key question that needs to be answered. In Sections 2 and 3 of this monograph, we will show that not only the notion of channel capacity still exists for a distributed system, it can indeed be viewed as an extension to the corresponding result in classical channel coding theory. Meanwhile, at the data link layer when a user is equipped with multiple transmission options, one needs to understand how packet transmission schemes should be adapted in response to the events of transmission success and packet collision. In existing link layer protocols, when only a single transmission option (plus an idling option) is available, a common practice in response to packet collision is to reduce the packet transmission probability of each user [42][11][12]. From classical channel coding theory, we know that a more efficient approach could be reducing the communication rate of each user [20]. However, while transmission options with different power and rate combinations may be available, there is no guarantee that the ideal option should be on the list. Furthermore, different link layer networks may also have different utility optimization objectives. Whether a general link layer distributed medium access control framework exists to optimize transmission schemes under these constraints is an important question that needs to be answered. Although we are not yet able to provide rigorous answers to this question, in Section 4, we present early research results to show that a stochastic approximation framework could be a good starting point to investigate the corresponding link layer problems.

2

Channel Capacity in Distributed Communication

Classical channel coding theory assumes the coordinated communication model [29][21][20]. Each transmitter in a communication party should be backlogged with an infinite reservoir of traffic. To achieve reliable communication, transmitters should jointly determine their codebooks which set values to communication parameters such as transmission power and information rates. Codebook information should be shared with other transmitters and receiver before the encoded symbols are transmitted to the receiver continuously over a long time interval. Channel capacity and channel coding theorems are proved using the standard argument of jointly typical sequences by taking the codeword length to infinity [71][72].

In distributed communication, however, users need to prepare their coding schemes with incomplete and maybe highly limited information about the communication channel and the communication environment. Therefore, instead of choosing one code, each user should prepare an ensemble of codes corresponding to different communication settings such as different power and rate combinations [58][87][55]. The ensemble can be shared with other users, for example by specifying it in the physical layer protocol. When messages become available, depending on decisions from the data link layer protocol, each physical layer transmitter chooses a particular code to encode its message and to send the corresponding codeword through the channel. The coding choice may be unknown to other users including the receiver. Because users are not fully coordinated, reliable message delivery may not always be supported by the channel. It is the receiver's responsibility to detect whether reliable decoding can be achieved or not[1]. If the answer is positive, the receiver should decode the messages of interest. Otherwise, the receiver should report collision. Although the receiver does not always output decoded messages, expected outcomes at the receiver under different system-wise communication settings should still be clearly defined. Therefore, one can generalize the definition of "communication error" from the classical meaning of erroneous decoding to the generalized meaning that output of the receiver is different from the expected outcome [55]. With such an extended definition, it is then possible to characterize asymptotic performance limitation of the system, in the sense of diminishing error probability when taking the codeword length to infinity, or to investigate performance tradeoff bounds under the assumption of a finite codeword length [58][87][55].

In this section, we will introduce coding theorems for the distributed communication model. We will start by maintaining the assumption that codeword length of each user can still be taken to infinity and will postpone the discussions on finite codeword length to Section 3. Throughout the section, we only present results for channels with finite input and output alphabets. The results can be easily extended to channels with continuous input and output alphabets using the same approach for similar extensions in classical channel coding theory [29].

2.1 Distributed Single User Communication

Let us first consider a single user communication system over a discrete-time memoryless channel. The channel is modeled by a conditional distribution function $P_{Y|X}$ where $X \in \mathcal{X}$ is the channel input symbol

[1] An alternative assumption that the transmitters should tell the receiver about their transmission/idling status was investigated in [63].

and $Y \in \mathcal{Y}$ is the channel output symbol. The sets $\mathcal{X}$ and $\mathcal{Y}$ are the finite input and output alphabets. We assume that time is slotted with each slot equaling N symbol durations, which is also the length of a packet or a codeword. Unless otherwise specified, we will confine our focus on block channel codes of length N that represents coding within each packet or each time slot[2]. Throughout this section, we assume that the communication channel is time-invariant. We assume that channel input alphabet $\mathcal{X}$ should be known at the transmitter. Channel distribution function $P_{Y|X}$ should be known at the receiver but *unknown* at the transmitter.

We assume that the physical layer transmitter is equipped with an ensemble of random block codes [28] each corresponding to a transmission/idling option at the data link layer. The ensemble is denoted by $\mathcal{G}^{(N)} = \{g_1, g_2, \dots, g_M\}$ with a finite cardinality $|\mathcal{G}^{(N)}| = M$. Coding scheme of the distributed system is described as follows. Note that the description is extended from a mathematical definition of random coding introduced in [70]. For $g \in \mathcal{G}^{(N)}$, let $\mathcal{L}_g = \{\mathcal{C}_{g\theta} : \theta \in \Theta^{(N)}\}$ be a library of codebooks, indexed by a set $\Theta^{(N)}$. Each codebook contains $\lfloor e^{Nr_g} \rfloor$ codewords of length N, where r_g is a pre-determined parameter termed the "communication rate" of the corresponding code[3]. Denote by $[\mathcal{C}_{g\theta}(w)]_j$ the jth symbol of the codeword corresponding to message w in codebook $\mathcal{C}_{g\theta}$. At the beginning of a time slot, the transmitter randomly generates a codebook index θ according to a distribution $\gamma^{(N)}$. The distribution $\gamma^{(N)}$ and the codebooks $\mathcal{C}_{g\theta}$, $\forall g \in \mathcal{G}^{(N)}$, are chosen such that random variables $X_{gwj} : \theta \to [\mathcal{C}_{g\theta}(w)]_j$, $\forall j, w$ and for each g, are i.i.d. according to a pre-determined input distribution P_{gX}. Note that input distributions of different codes (i.e., corresponding to different g) may be different. We assume that the ensemble of code libraries and the value of θ should both be known at the receiver. That is, the receiver knows the randomly generated codebook of each code. This can be achieved by sharing the random codebook generation algorithm with the receiver.

[2]Note that such a focus does not prevent the implementation of other coding schemes across multiple packets.

[3]Recall that each codeword represents one single packet, as opposed to multiple packets, that can be possibly chosen for transmission.

For example, suppose codebooks of the user are generated using a pseudo random sequence generator seeded by time and its user identity. The system can specify the pseudo random algorithm in a physical layer protocol and share it with the receiver. So long as clocks of the user and the receiver are synchronized, the receiver only needs the user identity information to generate the random codebooks [55].

At the beginning of a time slot, depending on a link layer decision, the transmitter chooses a code $g \in \mathcal{G}^{(N)}$ from the ensemble. We assume that such a decision is made *arbitrarily* in the sense that it is not controlled by the physical layer, and therefore even statistical information of the coding choice may be unknown at the physical layer. More importantly, we assume that g is unknown to the receiver. Given code index g, the transmitter then uses the random coding scheme to map a message w into a codeword of N channel input symbols $X_g^{(N)}(w) = [X_{gw1}, X_{gw2}, \dots, X_{gwN}]$, and sends the symbols through the channel. Upon receiving the channel output symbols $Y^{(N)} = [Y_1, Y_2, \dots, Y_N]$, the receiver either outputs an estimated message and code index pair $(\hat{w}, \hat{g})$, or reports "collision".

Note that in the above description, a random block code g is characterized by its communication rate r_g and its input distribution P_{gX}. With an abuse of the notation, we also regard $g = (r_g, P_{gX})$ as a variable representing a rate and distribution pair, which is not a function of the codeword length N. We will use "code space" to refer to the space of g, which is the space of rate and distribution pairs.

We assume that the receiver should choose a region R, termed the "operation region", in the code space. Note that while R itself is not a function of the codeword length, for a given codeword length N, R represents a subset of the random block codes whose rate and distribution pairs belong to the specified region in the code space. Because the channel is unknown at the transmitter especially when code ensemble $\mathcal{G}^{(N)}$ is designed, not all codes in the ensemble can support reliable message delivery over the channel. We assume that, the receiver should "intend" to decode the message if $g \in R$ and to report collision if $g \notin R$. Note that the receiver does not know g and will need to make a judgement on whether to decode the message or not based only on the channel output sequence $Y^{(N)}$. Given the operation region R, and conditioned on g being the actual code used for transmission, communication error probability as a function of g for codeword length N is defined as follows[4].

$$P_e^{(N)}(g) = \begin{cases} \max_w Pr\{(\hat{w},\hat{g}) \neq (w,g) | (w,g)\}, & \forall g \in R \\ \max_w 1 - Pr\left\{ \begin{array}{l} \text{“collision” or} \\ (\hat{w},\hat{g}) = (w,g) \end{array} \middle| (w,g) \right\}, & \forall g \notin R \end{cases}. \tag{2.1}$$

Definition 2.1. We say that an operation region R is asymptotically achievable for the discrete memoryless channel $P_{Y|X}$, if for all finite M and all code ensemble $\mathcal{G} = \{g_1, \dots, g_M\}$, decoding algorithms can be designed for the sequence of random code ensembles $\mathcal{G}^{(N)} = \mathcal{G}$ to achieve

$$\lim_{N\to\infty} P_e^{(N)}(g_i) = 0, \qquad \forall i \in \{1, \dots, M\}. \tag{2.2}$$

The following property is implied by the achievability definition.

Theorem 2.1. For a single user discrete memoryless channel $P_{Y|X}$ with finite input and output alphabets, if an operation region R is asymptotically achievable, then any subset of the region $\tilde{R} \subseteq R$ is also asymptotically achievable.

The following theorem gives the maximum achievable region of the single user communication channel.

Theorem 2.2. For a single user discrete memoryless channel $P_{Y|X}$ with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, the following region C_d in the code space is asymptotically achievable

$$C_d = \{g | g = (r_g, P_{gX}), r_g < I_g(X;Y)\}, \tag{2.3}$$

where $I_g(X;Y)$ denotes the mutual information between X and Y with respect to joint distribution $P_{XY} = P_{Y|X}P_{gX}$.

[4] In error probability definition (2.1), we regard correct message output as one of the acceptable outcomes for $g \notin R$. Compared with an alternative definition that only accepts collision report for $g \notin R$, using definition (2.1) does not change the results of Theorems 2.1 and 2.2. This is because the receiver can always choose to report collision for $g \notin R$ when g can be detected reliably.

Let $C_d^c = \{g|g = (r_g, P_{gX}), r_g \leq I_g(X;Y)\}$ be the closure of C_d. The achievable region $\mathcal{C}_d$ is maximum in the sense that for any region R that is asymptotically achievable, we must have $R \subseteq C_d^c$.

Although Theorem 2.2 is implied by Theorem 2.4, we still provide a full proof in Appendix A.1 for the sake of easy understanding. The proof also serves as a reference for understanding the more notation-complicated proof of Theorem 2.4.

We define C_d as the "distributed capacity" of the discrete time memoryless channel $P_{Y|X}$. C_d coincides with the classical Shannon capacity C of the same channel [71][20] in the sense that we can regard Shannon capacity $C = \{r|r < \max_{P_X} I(X;Y)\}$ as the set of information rates achievable in the Shannon sense and write its closure as

$$C^c = \{r|\exists g \in C_d^c, r_g = r\}. \tag{2.4}$$

Also note that the same capacity region C_d can have different meanings in different communication models. In coordinated communication, on one hand, the transmitter and the receiver can achieve reliable message delivery at any rate r inside the capacity region using a random block code $g \in C_d$ with $r_g = r$ [20]. In distributed communication, on the other hand, the transmitter chooses an arbitrary code g to transmit its message. While coding choice is not shared with the receiver, if $g \in C_d$, the receiver will be able to decode the message and to detect g reliably, while if $g \notin C_d$, the receiver will be able to report collision reliably[5].

2.2 Distributed Multiple Access Communication

Next, let us consider a time-slotted multiple access system with K transmitters and one receiver. Each time slot equals the length of N symbols. We use bold font variable to represent a vector whose entries are the corresponding variables of all users. We model the discrete-time memoryless multiple access channel by a conditional distribution $P_{Y|\boldsymbol{X}}$, where $\boldsymbol{X} = [X_1, \ldots, X_K]^T \in \boldsymbol{\mathcal{X}}$ is the channel input symbol vector of all users with $\boldsymbol{\mathcal{X}}$ being the vector of finite input alphabets,

[5] According to Corollary 2.7, the receiver should be able to ensure collision report for $g \notin C_d$.

and $Y \in \mathcal{Y}$ is the channel output symbol with $\mathcal{Y}$ being the finite output alphabet. We assume that channel input alphabet $\mathcal{X}_k$ is known to user k, $k \in \{1, \ldots, K\}$. Channel distribution function $P_{Y|\boldsymbol{X}}$ is known at the receiver. Whether the channel is known at the transmitters or not is irrelevant to the results we are going to present.

Assume that each physical layer transmitter is equipped with an ensemble of M random block codes [28] each corresponding to a transmission/idling option at the data link layer. Let $\boldsymbol{\mathcal{G}}^{(N)} = \left[\mathcal{G}_1^{(N)}, \ldots, \mathcal{G}_K^{(N)}\right]^T$ denote the vector of code ensembles of the users, where $\mathcal{G}_k^{(N)} = \{g_{k1}, \ldots, g_{kM}\}$ is the ensemble of random block codes of user k, $k \in \{1, \ldots, K\}$[6]. We use $\boldsymbol{g}$ to represent a possible code vector chosen by the users. $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}^{(N)}$ if for all $k \in \{1, \ldots, K\}$, the kth entry g_k of $\boldsymbol{g}$ satisfies $g_k \in \mathcal{G}_k^{(N)}$. We also use $\boldsymbol{g} = (\boldsymbol{r_g}, \boldsymbol{P_{gX}})$ to represent the pair of rate and distribution vectors corresponding to the random code vector. For any code ensemble vector $\boldsymbol{\mathcal{G}}^{(N)}$, we use $\boldsymbol{\mathcal{G}}$ to represent the corresponding ensemble in the code space.

At the beginning of each time slot, a random codebook is generated for each code and for each user. We assume that the receiver knows the randomly generated codebooks as this can be achieved by sharing the codebook generation algorithms with the receiver without requiring much online information exchange [55]. Depending on decisions from the link layer protocol, each transmitter chooses a code from its code ensemble to map a message to a codeword, and then sends the codeword through the channel. Coding choices of the users are denoted by a code index vector $\boldsymbol{g}$. At the physical layer, the value of $\boldsymbol{g}$ is regarded as "arbitrary". We assume that $\boldsymbol{g}$ is unknown to the receiver. Given code index vector $\boldsymbol{g}$, the transmitters then map a message vector $\boldsymbol{w}$ into N vectors of channel input symbols, denoted by $\boldsymbol{X}_{\boldsymbol{g}}^{(N)}(\boldsymbol{w})$, and send them through the multiple access channel. Upon receiving the channel output symbols $Y^{(N)} = [Y_1, Y_2, \ldots, Y_N]$, the receiver outputs an estimated message and code index vector pair $(\hat{\boldsymbol{w}}, \hat{\boldsymbol{g}})$ if a pre-determined error probability requirement can be met, otherwise the receiver reports

[6]The assumption that code ensembles of all users should have the same cardinality is made to simplify the notations. All theorems remain valid if cardinalities of code ensembles of different users are different.

collision. Here, we assume that the receiver should either decode the messages of all users, or report collision for all users.

We assume that the receiver should choose an "operation region" $\boldsymbol{R}$ in the code space, or in the space of rate and distribution vector pairs. Because users are not coordinated, reliable message recovery may not be possible for all coding choices. We assume that the receiver should "intend" to decode the messages if $\boldsymbol{g} \in \boldsymbol{R}$, and to report collision if $\boldsymbol{g} \notin \boldsymbol{R}$. As before, the receiver needs to make a decision on decoding or collision report without the knowledge of the actual code index vector $\boldsymbol{g}$. Given the operation region $\boldsymbol{R}$ and conditioned on $\boldsymbol{g}$ being the actual code index vector, communication error probability as a function of $\boldsymbol{g}$ for codeword length N is defined as follows[7].

$$P_e^{(N)}(\boldsymbol{g}) = \begin{cases} \max_{\boldsymbol{w}} Pr\{(\hat{\boldsymbol{w}}, \hat{\boldsymbol{g}}) \neq (\boldsymbol{w}, \boldsymbol{g}) | (\boldsymbol{w}, \boldsymbol{g})\}, & \forall \boldsymbol{g} \in \boldsymbol{R} \\ \max_{\boldsymbol{w}} 1 - Pr\left\{ \begin{array}{c} \text{"collision" or} \\ (\hat{\boldsymbol{w}}, \hat{\boldsymbol{g}}) = (\boldsymbol{w}, \boldsymbol{g}) \end{array} \middle| (\boldsymbol{w}, \boldsymbol{g}) \right\}, & \forall \boldsymbol{g} \notin \boldsymbol{R} \end{cases}. \tag{2.5}$$

Definition 2.2. We say that an operation region $\boldsymbol{R}$ is asymptotically achievable for the multiple access channel $P_{Y|\boldsymbol{X}}$, if for all finite M and all code ensemble vectors $\boldsymbol{\mathcal{G}}$, decoding algorithms can be designed for the sequence of random code ensembles $\boldsymbol{\mathcal{G}}^{(N)} = \boldsymbol{\mathcal{G}}$ to achieve

$$\lim_{N \to \infty} P_e^{(N)}(\boldsymbol{g}) = 0, \qquad \forall \boldsymbol{g} \in \boldsymbol{\mathcal{G}}. \tag{2.6}$$

Similar to Theorem 2.1, the following property is implied by the achievability definition.

Theorem 2.3. For a discrete memoryless multiple access channel $P_{Y|\boldsymbol{X}}$ with finite input and output alphabets, if an operation region $\boldsymbol{R}$ is asymptotically achievable, then any of its subsets $\tilde{\boldsymbol{R}} \subseteq \boldsymbol{R}$ is also asymptotically achievable.

The next theorem gives the maximum achievable region of the multiple access communication channel.

[7] As in the previous section, whether we regard correct decoding as an acceptable outcome for $\boldsymbol{g} \notin \boldsymbol{R}$ or not does not change the conclusions of the Theorems in this section, because the receiver can always choose to report collision for $\boldsymbol{g} \notin \boldsymbol{R}$ in cases when $\boldsymbol{g}$ can be detected reliably.

Theorem 2.4. For a discrete-time memoryless multiple access channel $P_{Y|\boldsymbol{X}}$ with finite input and output alphabets, the following region $\boldsymbol{C}_d$ in the code space is asymptotically achievable

$$\boldsymbol{C}_d = \left\{ \boldsymbol{g} \,\middle|\, \boldsymbol{g} = (\boldsymbol{r}_{\boldsymbol{g}}, \boldsymbol{P}_{\boldsymbol{g}\boldsymbol{X}}), \forall S \subseteq \{1, \ldots, K\}, \sum_{k \in S} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_S; Y | \boldsymbol{X}_{\bar{S}}) \right\}, \tag{2.7}$$

where $\boldsymbol{X}_S$ denotes the channel input symbols for users in S, $\boldsymbol{X}_{\bar{S}}$ denotes the channel input symbols for users not in S, and $I_{\boldsymbol{g}}(\boldsymbol{X}_S; Y|\boldsymbol{X}_{\bar{S}})$ denotes the mutual information between $\boldsymbol{X}_S$ and Y given $\boldsymbol{X}_{\bar{S}}$ with respect to joint distribution $P_{\boldsymbol{X}Y} = P_{Y|\boldsymbol{X}} \prod_{k=1}^{K} P_{g_k X_k}$.

Let $\boldsymbol{C}_d^c$ be the closure of $\boldsymbol{C}_d$. The achievable region $\boldsymbol{C}_d$ is maximum in the sense that for any region $\boldsymbol{R}$ that is asymptotically achievable, we must have $\boldsymbol{R} \subseteq \boldsymbol{C}_d^c$.

Theorem 2.4 is implied by Corollary 2.8.

We define $\boldsymbol{C}_d$ as the "distributed capacity" of the multiple access channel $P_{Y|\boldsymbol{X}}$. Note that $\boldsymbol{C}_d$ coincides with the classical Shannon capacity $\boldsymbol{C}$ of the same channel [1][53] in the following sense

$$\boldsymbol{C}^c = \text{convex hull}\left(\{\boldsymbol{r} | \exists \boldsymbol{g} \in \boldsymbol{C}_d^c, \boldsymbol{r}_{\boldsymbol{g}} = \boldsymbol{r}\}\right), \tag{2.8}$$

where $\boldsymbol{C}^c$ and $\boldsymbol{C}_d^c$ are the closures of $\boldsymbol{C}$ and $\boldsymbol{C}_d$, respectively. We also regard $\boldsymbol{C}_d$ as an extension of the Shannon capacity in a sense similar to the one explained at the end of Section 2.1.

Note that, similar to classical channel coding theory, Theorem 2.4 holds even if the input and output alphabets of the channel are continuous [29]. Also similar to classical channel coding theory, one can pose a constraint in the code space to limit the coding choices of the users. In this case the result of Theorem 2.4 can be used to obtain the constrained distributed capacity of the channel. We give two examples originally presented in [58] to illustrate these cases.

Example 2.1. Let us consider a K-user multiple access system over a discrete-time memoryless channel with additive Gaussian noise. The channel is modeled by $Y = \sum_{k=1}^{K} X_k + V$, where V is the Gaussian noise with zero mean and variance N_0.

Let us pose a constraint that all random coding options for user k, $k \in \{1, \ldots, K\}$, must have Gaussian input distribution with zero mean and variance P_k. Let $\boldsymbol{r}$ be the rate vector of the users. Constrained distributed capacity of the multiple access channel is given by the following region in the space of the rate vectors.

$$\boldsymbol{C}_d = \left\{ \boldsymbol{r_g} \middle| \forall S \subseteq \{1, \ldots, K\}, \sum_{k \in S} r_{g_k} < \frac{1}{2} \log \left(1 + \frac{\sum_{k \in S} P_k}{N_0} \right) \right\}, \quad (2.9)$$

The constrained distributed channel capacity is identical to the Shannon channel capacity given in [90][20].

When $K = 2$ and $P_1 = P_2 = 5N_0$, the capacity region is a pentagon illustrated in Figure 2.1, where the rates are measured in nats/symbol.

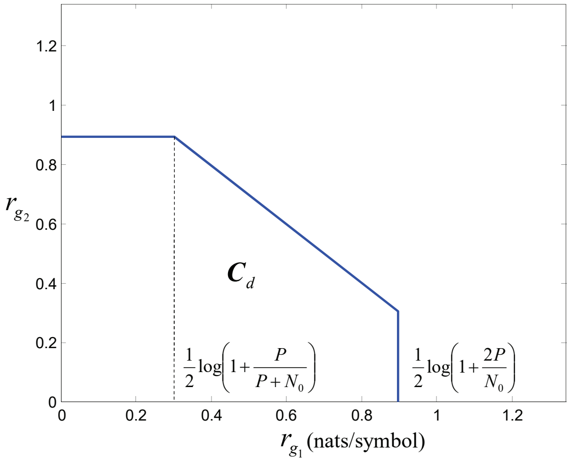

Figure 2.1: Distributed capacity of a two user Gaussian channel.

Example 2.2. Next, let us consider a K-user distributed multiple access system over a symbol collision channel. Let n be a non-negative integer. We model the channel as follows. Let the input alphabet of any user k be given by $\mathcal{X}_k = \{0, 1, \ldots, 2^n\}$, where 0 represents an idling symbol. The channel output alphabet is given by $\mathcal{Y} = \{0, 1, \ldots, 2^n, c\}$, where c

represents a collision symbol. If all users idle, the receiver receives an idling symbol $Y = 0$. If only one user, say user k, transmits a nonzero symbol X_k, the receiver receives $Y = X_k$. If multiple users transmit nonzero symbols, the receiver receives a collision symbol $Y = c$. Assume the constraint that, for any input distribution, probabilities of transmitting different nonzero symbols must be identical. Consequently, input distribution $P_{g_k X_k}$ can be characterized by a single parameter p_{g_k}, which is the probability that any particular symbol in the transmitted codeword takes a nonzero value. $P_{g_k X_k}$ is therefore given by

$$P_{g_k X_k} = \begin{cases} 1 - p_{g_k} & X_k = 0 \\ \frac{1}{2^n} p_{g_k} & X_k = j \in \{1, \dots, 2^n\} \end{cases} . \tag{2.10}$$

In other words, each code vector $\boldsymbol{g} = (\boldsymbol{r_g}, \boldsymbol{p_g})$ can be mapped to a point in the space of rate vector and probability vector pairs.

We define the distributed rate capacity region of the system as $\boldsymbol{C}_d^{(\boldsymbol{r})} = \{\boldsymbol{r} | \exists \boldsymbol{g} \in \boldsymbol{C}_d \text{ with } \boldsymbol{r} = \boldsymbol{r_g}\}$, where the rates are measured in bits/symbol. By following the proof of [58, Proposition 1], for any user subset $S \subseteq \{1, \dots, K\}$, we have,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} I_{\boldsymbol{g}}(\boldsymbol{X}_S; Y | \boldsymbol{X}_{\bar{S}}) &= \lim_{n\to\infty} \frac{1}{n} H_{\boldsymbol{g}}(Y | \boldsymbol{X}_{\bar{S}} = \boldsymbol{0}) Pr\{\boldsymbol{X}_{\bar{S}} = \boldsymbol{0}\} \\ = Pr\{Y \notin \{0, c\}\} \prod_{j \in \bar{S}} (1 - p_j) &= \sum_{k \in S} p_k \prod_{j \in \bar{S}} (1 - p_j). \end{aligned} \tag{2.11}$$

Therefore, the normalized distributed rate capacity region possesses the following limiting behavior.

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{C}_d^{(\boldsymbol{r})} = \left\{ \boldsymbol{\lambda} \middle| \begin{array}{l} \exists 0 \le \boldsymbol{p} \le 1, \forall k \in \{1, \dots, K\}, \\ \lambda_k \le p_k \prod_{j \neq k} (1 - p_j) \end{array} \right\}. \tag{2.12}$$

When $K = 2$, (2.12) gives the throughput, stability and the information capacity region of the collision channel [67][44], as illustrated in Figure 2.2. In this case, the region is non-convex[8], and can be written in a simple form of $\{\boldsymbol{\lambda} | \sqrt{\lambda_1} + \sqrt{\lambda_2} \le 1\}$ [67][58].

[8]But it is coordinate convex [58].

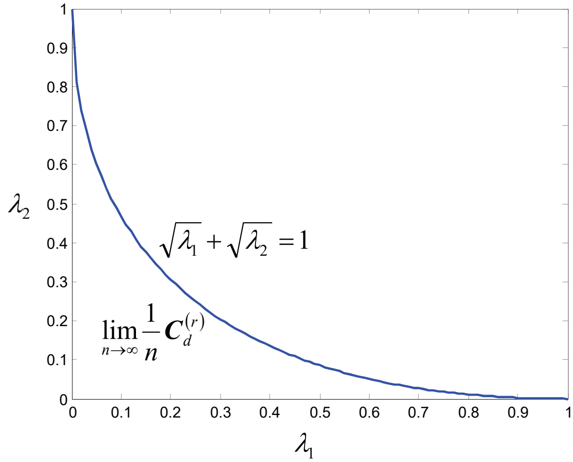

Figure 2.2: Throughput region of a two user collision channel.

2.3 Distributed Communication with Single User Decoding

In the previous section, we assumed that the receiver should either decode the messages of all users or to report collision for all users. In this section, we will extend the model further to assume that each transmitter has its own associated receiver[9] [58]. A receiver is only interested in decoding the message of its associated transmitter even though it can choose to decode the messages of other users if necessary.

Let us consider the system introduced in Section 2.2 with the assumption that the receiver is now only associated to user 1. That is, the receiver is only interested in decoding the message of user 1. We still assume that the randomly generated codebooks of all users are known at the receiver, and therefore, the receiver can choose to decode the messages of any other user if necessary. Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code index vector chosen by the transmitters. Upon observing the channel output symbols $Y^{(N)}$, the receiver outputs an estimated message and code index pair $(\hat{w}_1, \hat{g}_1)$ for user 1 if

[9]This does not prevent a system from associating the same receiver to multiple transmitters.

a pre-determined error probability requirement can be met, otherwise the receiver reports collision for user 1.

We assume that the receiver should choose an "operation region" $\boldsymbol{R}_1$ in the space of code index vectors[10]. The receiver "intends" to decode the message of user 1 if the actual code index vector $\boldsymbol{g} \in \boldsymbol{R}_1$ is inside the operation region, and to report collision for user 1 if $\boldsymbol{g} \notin \boldsymbol{R}_1$. Given the operation region $\boldsymbol{R}_1$ and conditioned on $\boldsymbol{g}$ being the actual code index vector, communication error probability as a function of $\boldsymbol{g}$ for codeword length N is defined as follows.

$$P_e^{(N)}(\boldsymbol{g}) = \begin{cases} \max_{\boldsymbol{w}} Pr\{(\hat{w}_1, \hat{g}_1) \neq (w_1, g_1) | (\boldsymbol{w}, \boldsymbol{g})\}, & \forall \boldsymbol{g} \in \boldsymbol{R}_1 \\ \max_{\boldsymbol{w}} 1 - Pr\left\{ \begin{array}{l} \text{"collision" or} \\ (\hat{w}_1, \hat{g}_1) = (w_1, g_1) \end{array} \middle| (\boldsymbol{w}, \boldsymbol{g}) \right\}, & \forall \boldsymbol{g} \notin \boldsymbol{R}_1 \end{cases}. \tag{2.13}$$

Definition 2.3. We say that an operation region $\boldsymbol{R}_1$ is asymptotically achievable for the multiple access channel $P_{Y|\boldsymbol{X}}$ for user 1, if for all finite M and all code ensemble vectors $\boldsymbol{\mathcal{G}}$, decoding algorithms can be designed for the sequence of random code ensembles $\boldsymbol{\mathcal{G}}^{(N)} = \boldsymbol{\mathcal{G}}$ to achieve

$$\lim_{N\to\infty} P_e^{(N)}(\boldsymbol{g}) = 0, \qquad \forall \boldsymbol{g} \in \boldsymbol{\mathcal{G}}. \tag{2.14}$$

Theorem 2.5. For a discrete-time memoryless multiple access channel $P_{Y|\boldsymbol{X}}$ with finite input and output alphabets, if an operation region $\boldsymbol{R}_1$ is asymptotically achievable for user 1, then any subset $\tilde{\boldsymbol{R}}_1 \subseteq \boldsymbol{R}_1$ is also asymptotically achievable for user 1.

Unlike similar theorems presented in the previous two sections, the conclusion of Theorem 2.5 does depend on the error probability definition (2.13), where we regard correct message decoding as an acceptable outcome for $\boldsymbol{g} \notin \boldsymbol{R}_1$. Because the receiver does not always decode the messages of users other than user 1, the receiver may not be able to

[10] Although the receiver is only interested in decoding the message of user 1, the operation region is still defined in the space of code index vectors involving the code indices of all users. This is because decodability of the message of user 1 depends on the transmission status of other users in the system.

correctly detect the part of the code index vector $\boldsymbol{g}$ corresponding to the un-decoded users. Therefore, the receiver may not be able to tell whether the actual code index vector $\boldsymbol{g}$ satisfies $\boldsymbol{g} \in \boldsymbol{R}_1$ or not. With the error probability definition (2.13), correct detection of the full code index vector is not required. That is, so long as the receiver does not output an erroneous message for user 1, whether the receiver guarantees collision report for $\boldsymbol{g} \notin \boldsymbol{R}_1$ or not is not a concern to the system.

Alternatively, suppose we only accept collision report for $\boldsymbol{g} \notin \boldsymbol{R}_1$, and define the error probability as

$$P_e^{(N)}(\boldsymbol{g}) = \begin{cases} \max_{\boldsymbol{w}} Pr\{(\hat{w}_1, \hat{g}_1) \neq (w_1, g_1)|(\boldsymbol{w}, \boldsymbol{g})\}, & \forall \boldsymbol{g} \in \boldsymbol{R}_1 \\ \max_{\boldsymbol{w}} 1 - Pr\left\{\text{“collision”}\right|(\boldsymbol{w}, \boldsymbol{g})\}, & \forall \boldsymbol{g} \notin \boldsymbol{R}_1 \end{cases}. \tag{2.15}$$

Consequently, the receiver will have to detect whether $\boldsymbol{g} \in \boldsymbol{R}_1$ or $\boldsymbol{g} \notin \boldsymbol{R}_1$. Depending on the feasibility of such a detection task, Theorem 2.5 may no longer hold. That is, even if a region $\boldsymbol{R}_1$ is asymptotically achievable for user 1, there may exist a subset $\tilde{\boldsymbol{R}}_1 \subseteq \boldsymbol{R}_1$ that is not asymptotically achievable for user 1. A simple example of such a situation is illustrated below.

Example 2.3. Consider a distributed multiple access system with two users. Let X_1, X_2 be the channel input symbols of the two users, and let Y be the channel output symbol, all having finite alphabets. Assume that input symbol of user 2 has no impact to the channel output. That is, the channel satisfies $P(Y|X_1, X_2) = P(Y|X_1)$. With the error probability definition of (2.15), according to Theorem 2.2, the following region $\boldsymbol{R}_1 = \left\{ \boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \middle| r_{g_1} < I_{g_1}(X_1; Y) \right\}$ is asymptotically achievable for user 1. However, the following subset $\tilde{\boldsymbol{R}}_1 = \left\{ \boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \middle| r_{g_1} < I_{g_1}(X_1; Y), r_{g_2} < 0.5 \right\}$ with $\tilde{\boldsymbol{R}}_1 \subseteq \boldsymbol{R}_1$ is not achievable for user 1. This is because the receiver has no capability of detecting the communication rate of user 2, and therefore cannot tell whether $r_{g_2} < 0.5$ is true or false (or equivalently, whether or not $\boldsymbol{g} \in \tilde{\boldsymbol{R}}_1$).

Let us come back to the error probability definition of (2.13). The following theorem gives the maximum achievable region of the multiple access channel for user 1 [58].

Theorem 2.6. For a discrete memoryless multiple access channel $P_{Y|\boldsymbol{X}}$ with finite input and output alphabets, the following region $\boldsymbol{C}_{d1}$ in the code space is asymptotically achievable for user 1.

$$\boldsymbol{C}_{d1} = \left\{ \boldsymbol{g} \;\middle|\; \begin{array}{l} \boldsymbol{g} = (\boldsymbol{r}_{\boldsymbol{g}}, \boldsymbol{P}_{\boldsymbol{g}\boldsymbol{X}}), \forall S \subseteq \{1, \dots, K\}, 1 \in S, \exists \tilde{S} \subseteq S, 1 \in \tilde{S}, \\ \text{such that, } \sum_{k \in \tilde{S}} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}}; Y | \boldsymbol{X}_{\bar{S}}) \end{array} \right\}. \tag{2.16}$$

where $I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}}; Y|\boldsymbol{X}_{\bar{S}})$ denotes the mutual information between $\boldsymbol{X}_{\tilde{S}}$ and Y given $\boldsymbol{X}_{\bar{S}}$ with respect to joint distribution corresponding to code index vector $\boldsymbol{g}$, i.e., $P_{\boldsymbol{X}Y} = P_{Y|\boldsymbol{X}} \prod_{k=1}^{K} P_{g_k X_k}$.

The achievable region $\boldsymbol{C}_{d1}$ is maximum in the sense that for any region $\boldsymbol{R}_1$ that is asymptotically achievable for user 1, we must have $\boldsymbol{R}_1 \subseteq \boldsymbol{C}_{d1}^c$, where $\boldsymbol{C}_{d1}^c$ is the closure of $\boldsymbol{C}_{d1}$.

The proof of Theorem 2.6 is presented in Appendix A.2.

In fact, the following theorem shows that $\boldsymbol{C}_{d1}$ is also achievable for user 1 under the alternative error probability definition (2.15), although in this case there is no guarantee that a subset of $\boldsymbol{C}_{d1}$ should also be achievable.

Corollary 2.7. Theorem 2.6 still holds under error probability definition (2.15).

The proof of Corollary 2.7 is presented in Appendix A.3.

The above results can be further extended to the case when the receiver is interested in decoding the messages of a user subset.

Definition 2.4. Let $S_0 \subseteq \{1, \dots, K\}$ be a user subset. We say that an operation region $\boldsymbol{R}_{S_0}$ is asymptotically achievable for the multiple access channel $P_{Y|\boldsymbol{X}}$ for user subset S_0, if $\forall k \in S_0$, $\boldsymbol{R}_{S_0}$ is asymptotically achievable for user k.

The following corollary gives the maximum achievable region for the multiple access channel for user subset S_0.

Corollary 2.8. For a discrete memoryless multiple access channel $P_{Y|\boldsymbol{X}}$ with finite input and output alphabets, let $\boldsymbol{C}_{dk}$ be the maximum achievable region for user k. The expression of $\boldsymbol{C}_{dk}$ can be obtained from (2.16) by replacing user index 1 with user index k. Let $S_0 \subseteq \{1, \dots, K\}$

be a user subset. The maximum achievable region for user subset S_0 is given by

$$\boldsymbol{C}_{dS_0} = \bigcap_{k \in S_0} \boldsymbol{C}_{dk} =$$
$$\left\{ \boldsymbol{g} \,\middle|\, \begin{array}{l} \boldsymbol{g} = (\boldsymbol{r}_{\boldsymbol{g}}, \boldsymbol{P}_{\boldsymbol{g}\boldsymbol{X}}), \forall S \subseteq \{1, \dots, K\}, S \cap S_0 \neq \phi, \exists \tilde{S}, \\ S \cap S_0 \subseteq \tilde{S} \subseteq S, \text{ such that, } \sum_{k \in \tilde{S}} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}}; Y | \boldsymbol{X}_{\bar{S}}) \end{array} \right\}. \tag{2.17}$$

where ϕ is the empty set, and $I_{\boldsymbol{g}}()$ denotes the mutual information function calculated with respect to joint distribution corresponding to code index vector $\boldsymbol{g}$, i.e., $P_{\boldsymbol{X}Y} = P_{Y|\boldsymbol{X}} \prod_{k=1}^{K} P_{g_k X_k}$.

The proof of Corollary 2.8 is presented in Appendix A.4.

Example 2.4. Consider the K-user distributed multiple access system with additive Gaussian noise as introduced in Example 2.1. We still pose the constraint that, for any user k, input distributions of all coding options must be Gaussian with zero mean and variance P_k. If the receiver is only associated to user 1, then according to Theorem 2.6, the maximum achievable region for user 1 is given by

$$\boldsymbol{C}_{d1} = \left\{ \boldsymbol{r}_{\boldsymbol{g}} \,\middle|\, \begin{array}{l} \forall S \subseteq \{1, \dots, K\}, 1 \in S, \exists \tilde{S} \subseteq S, 1 \in \tilde{S}, \\ \text{such that, } \sum_{k \in \tilde{S}} r_{g_k} < \frac{1}{2} \log \left(1 + \frac{\sum_{k \in \tilde{S}} P_k}{\sum_{k \in S \setminus \tilde{S}} P_k + N_0} \right) \end{array} \right\}. \tag{2.18}$$

Similarly, one can also use Theorem 2.6 and Corollary 2.8 to obtain the maximum achievable region for any other user and for any user group.

When $K = 2$ and $P_1 = P_2 = 5N_0$, the maximum achievable region for user 1 and the maximum achievable region for user 2 are illustrated respectively in Figure 2.3, where the rates are measured in nats/symbol. It can be seen that intersection of the two regions equals the constrained channel capacity, i.e. the pentagon region, as illustrated in Figure 2.1.

2.4 Interfering User and Compound Channel

In a distributed system when users access the communication channel opportunistically, a receiver often needs to consider users that can

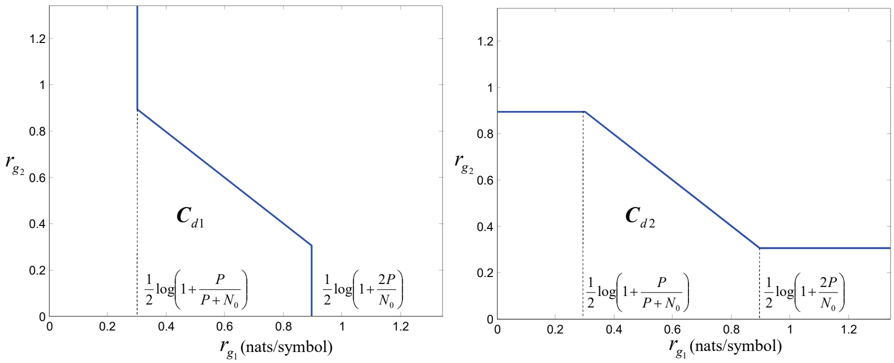

Figure 2.3: Maximum achievable regions for each individual user of a two user Gaussian channel.

potentially be active in the area. The system therefore may involve a significant number of users in its decoding operation. When there is a lack of user coordination and the receiver is not interested in decoding the messages of the remotely located users, it is not always reasonable to assume that codebook information of all users should be known at the receiver [55]. From a complexity perspective, it is also possible that the receiver may choose not to fully process codebook information of all users even if such information is available. Furthermore, without continuous transmission coverage, it is generally difficult for a receiver to keep track of the status of the wireless channels from every transmitter. This is especially true for users whose bursty messages are not decodable at the receiver. Therefore, the assumption that the multiple access channel should be known precisely at the receiver is not always valid. In this section, we extend the system model to include interfering users whose codebook information is not fully known at the receiver, and also to include the case of distributed communication over a compound channel. We investigate the two problems together because they actually fall into the same problem formulation [55].

In order to motivate the general system model, let us first consider the following two examples.

Example 2.5. (**Interfering User**) Let us start with the time-slotted distributed communication system introduced in Section 2.3. We add one more user into the system, indexed as user 0. Assume that user 0 is equipped with a code ensemble $\mathcal{G}_0 = \{g_{01}, \dots, g_{0M}\}$. Each code g_{0i} in the ensemble corresponds to a random block code with rate $r_{g_{0i}}$ and input distribution $P_{g_{0i}X_0}$. The memoryless channel is characterized by a conditional distribution $P_{Y|\boldsymbol{X}}$ with finite input alphabets $\boldsymbol{\mathcal{X}}$ and finite output alphabet $\mathcal{Y}$. We term user 0 an "interfering user" in the following sense. We assume that the receiver only knows the input distribution information of the code ensemble of user 0, but the randomly generated codebooks of user 0 are unknown at the receiver. In other words, it is not possible for the receiver to decode the message of user 0. Denote the actual code index of user 0 by g_0. Given g_0, the channel can be characterized by the following marginal conditional distribution function that only involves the input symbols of users other than user 0, i.e., $\boldsymbol{X}_{\overline{\{0\}}}$.

$$P_{Y|\boldsymbol{X}_{\overline{\{0\}}}}(g_0) = \sum_{X_0 \in \mathcal{X}_0} P_{Y|\boldsymbol{X}} P_{g_0 X_0}. \tag{2.19}$$

Note that the channel function $P_{Y|\boldsymbol{X}_{\overline{\{0\}}}}(g_0)$ is a function of g_0. Because codeword information of user 0 is not available at the receiver, the receiver should only work with the marginal channel function $P_{Y|\boldsymbol{X}_{\overline{\{0\}}}}(g_0)$ in its decoding operation[11].

Example 2.6. (**Compound Channel**) Let us again start with the time-slotted distributed communication system introduced in Section 2.3. We assume that the channel is replaced by a compound multiple access channel with M possible realizations, $P_{Y|\boldsymbol{X}} \in \{P^{(1)}_{Y|\boldsymbol{X}}, P^{(2)}_{Y|\boldsymbol{X}}, \dots, P^{(M)}_{Y|\boldsymbol{X}}\}$. The receiver knows the ensemble of possible channel realizations but does not know the particular realization chosen at the beginning of each time slot. To model such a system, we can introduce a virtual user, indexed as user 0. We assume that the user is equipped with an ensemble of M communication options, denoted by $\mathcal{G}_0 = \{1, \dots, M\}$. At the beginning of each time slot, user 0 chooses a value of its communication parameter $g_0 \in \mathcal{G}_0$ arbitrarily, and this determines the channel

[11] Extending the system model to include multiple interfering users is quite straightforward [55].

realization by setting $P_{Y|\boldsymbol{X}} = P^{(g_0)}_{Y|\boldsymbol{X}}$. We can think the channel model as $P_{Y|\boldsymbol{X}}(g_0) = P^{(g_0)}_{Y|\boldsymbol{X}}$, which is a function of g_0. The channel function and the ensemble $\mathcal{G}_0$ are known at the receiver, but the receiver does not know the value of $g_0 \in \mathcal{G}_0$ chosen at the beginning of each time slot.

Next, we show that the two examples can actually be covered by the same general system model presented below.

Consider a time-slotted distributed communication system with $K+1$ transmitters and one receiver. The transmitters (users) are indexed by $\{0, 1, \ldots, K\}$. We categorize users $1, \ldots, K$ as "regular users" and user 0 as an "interfering user". We assume that each regular user $k \in \{1, \ldots, K\}$ is equipped with an ensemble of random block codes, denoted by $\mathcal{G}_k = \{g_{k1}, \ldots, g_{kM}\}$. Each code $g_{ki} = (r_{g_{ki}}, P_{g_{ki}X_k})$ is characterized by a rate and input distribution pair. At the beginning of each time slot, regular user k chooses a code index $g_k \in \mathcal{G}_k$ arbitrarily, and then maps a message w_k to a codeword sequence $X^{(N)}_{g_k}(w_k)$ and sends the sequence of channel input symbols through the multiple access channel. We assume that the receiver knows the ensemble of randomly generated codebooks of each regular user k, but the receiver does not know the value of code index g_k. For the interfering user 0, we assume that the user has a parameter g_0 which, for convenience, is still termed the code index parameter. User 0 is equipped with an ensemble of communication options, denoted by $\mathcal{G}_0 = \{g_{01}, \ldots, g_{0M}\}$. We assume that, at the beginning of each time slot, the interfering user should choose $g_0 \in \mathcal{G}_0$ arbitrarily, and this affects the communication channel model in a sense explained below. We denote the vector of channel input symbols of the regular users by $\boldsymbol{X}$, and denote the message vector of the regular users by $\boldsymbol{w}$. We use $\boldsymbol{\mathcal{G}}$ to denote the vector of code ensembles of all users including the interfering user. $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$ is used to denote a vector of code indices of all users including the interfering user. Once the regular users send their channel input vector sequence $\boldsymbol{X}^{(N)}_{\boldsymbol{g}}(\boldsymbol{w})$ through the multiple access channel, output sequence $Y^{(N)}$ of the channel is generated according to a conditional distribution function $P_{Y|\boldsymbol{X}}(g_0)$, which is a function of the code index parameter g_0 of the interfering user[12]. We assume that the receiver should know the channel function $P_{Y|\boldsymbol{X}}(g_0)$ and the ensemble $\boldsymbol{\mathcal{G}}$ including $\mathcal{G}_0$, but the receiver does not know the code index vector $\boldsymbol{g}$ including g_0.

We assume that the receiver is associated to user 1. The receiver should output an estimated message and code index pair $(\hat{w}_1, \hat{g}_1)$ for user 1 if a pre-determined error probability threshold can be met, otherwise the receiver should report collision for user 1. Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code index vector pair. The receiver should choose an operation region $\boldsymbol{R}_1$ in the space of code index vectors $\boldsymbol{g}$. The receiver intends to decode the message of user 1 if $\boldsymbol{g} \in \boldsymbol{R}_1$, and intends to report collision for user 1 if $\boldsymbol{g} \notin \boldsymbol{R}_1$.

The following theorem shows that key results presented in Section 2.3 can be extended to distributed communication with the existence of an interfering user.

Theorem 2.9. Let g_0 be the code index of an interfering user. For a discrete memoryless multiple access channel $P_{Y|\boldsymbol{X}}(g_0)$ with finite input and output alphabets, conclusions of Theorems 2.5, 2.6, and Corollaries 2.7, 2.8 still hold, so long as the following extensions are applied to the statements in the theorems, corollaries and in their proofs.

1. Channel input vectors $\boldsymbol{X}$, rate vectors $\boldsymbol{r}_{\boldsymbol{g}}$, input distribution vectors $\boldsymbol{P}_{\boldsymbol{g}\boldsymbol{X}}$ should only contain entries corresponding to the regular users $1, \dots, K$.

2. Code index vectors $\boldsymbol{g} = (\boldsymbol{r}_{\boldsymbol{g}}, \boldsymbol{P}_{\boldsymbol{g}\boldsymbol{X}}, g_0)$ as well as code ensemble vector $\boldsymbol{\mathcal{G}}$ should contain one more entry corresponding to the code index of the interfering user.

3. Given code index vector $\boldsymbol{g}$, mutual information function $I_{\boldsymbol{g}}()$, entropy function $H_{\boldsymbol{g}}()$, and probability function $p_{\boldsymbol{g}}()$ should all be computed with respect to joint distribution $P_{\boldsymbol{X}Y} = P_{Y|\boldsymbol{X}}(g_0) \prod_{k=1}^{K} P_{g_k X_k}$, i.e., with a channel function of $P_{Y|\boldsymbol{X}}(g_0)$.

4. User subsets $S \subseteq \{1, \dots, K\}$ should only contain the regular users. The complement set $\bar{S}$ should be defined as $\bar{S} = \{1, \dots, K\} \setminus S$, i.e., the interference user should be excluded from $\bar{S}$.

[12] Note that such a function can be defined independently from the communication option ensemble $\mathcal{G}_0$ of the interfering user, i.e., the range of g_0 in the definition can extend beyond the ensemble $\mathcal{G}_0$.

5. The maximum number of possible code index vectors should be upper bounded by M^{K+1}.

With the above extensions, if error probability is defined as in (2.13), then any subset of an achievable region should also be achievable, $\boldsymbol{C}_{d1}$ given in (2.16) is the maximum asymptotically achievable region for user 1, $\boldsymbol{C}_{dS_0}$ given in (2.17) is the maximum asymptotically achievable region for user subset $S_0 \subseteq \{1, \ldots, K\}$, and the regions are still achievable if the error probability definition is changed to (2.15).

Theorem 2.9 can be proven by following the proofs of Theorems 2.5, 2.6, and Corollaries 2.7, 2.8 with the extensions listed above.

Example 2.7. Consider a single user communication system over a discrete-time memoryless channel with an unknown channel gain and additive Gaussian noise. The channel is modeled by

$$Y = hX + V, \tag{2.20}$$

where $h \geq 0$ is the unknown channel gain and V is the Gaussian noise with zero mean and variance N_0.

Let us pose the constraint that input distributions of all coding options must be zero mean with variance P. We can formulate the problem by constructing a system with two users. User 1 is the regular user whose coding options $\mathcal{G}_1 = \{r_1, \ldots, r_{M_1}\}$ represent an ensemble of Gaussian random block codes with the same input distribution but with different rates. User 0 is a interfering (or virtual) user whose communication options $\mathcal{G}_1 = \{h_1, \ldots, h_{M_0}\}$ represent the ensemble of compound gains that can possibly be taken by the channel. Consequently, distributed capacity region of the system is simply a region in the space of rate and channel gain pairs given by

$$\boldsymbol{C}_{d1} = \left\{ (r_{g_1}, h) \middle| r_{g_1} < \frac{1}{2} \log \left(1 + \frac{h^2 P}{N_0} \right) \right\}, \tag{2.21}$$

where the rate is measured in nats/symbol. The capacity region is illustrated in Figure 2.4 for $P = 5N_0$.

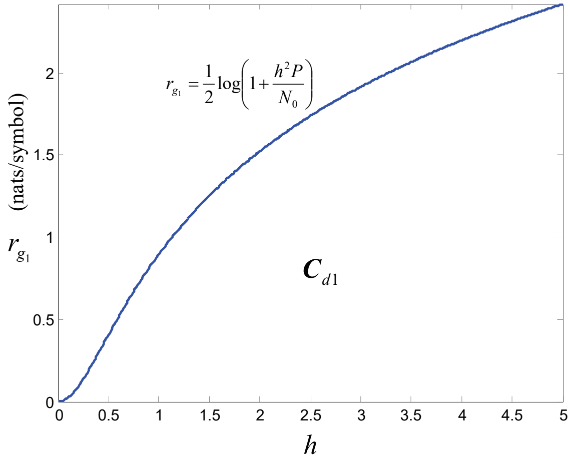

Figure 2.4: Distributed capacity region of a single user system over a Gaussian channel with an unknown channel gain.

3

Performance Bound with A Finite Codeword Length

Distributed communication often involves transmissions of bursty short messages, and therefore asymptotic analysis that takes codeword length to infinity may not be a valid approximation to the practical communication setting [22]. In this section, we extend the coding theorems to distributed communication systems with a finite codeword length [87].

While performance bounds on tradeoffs among decoding error probability, information rate, and codeword length have been extensively investigated in classical channel coding theory [25][28][73][84][26][13][65], the distributed communication model brought several new challenges that must be carefully considered in the formulation and analysis of the corresponding problems [87][55]. First, because data packets in distributed communication are relatively short in length, validity of the obtained tradeoff bounds should not require a large codeword length. Second, each user in a distributed communication system can choose its code from the ensemble arbitrarily. Different coding choices may lead to different types of error events such as decoding error and collision detection error. Therefore, when analyzing error probability performance of a distributed communication system, one may want to assign different weights to the probabilities of different error events. We will see in the section that, distributed communication can also involve tricky detection problems (such as collision detection or compound channel detection) whose performance depends on the decoding algorithm design at the receiver. If the decoding algorithm is not carefully thought out, error performance of the system can be dominated by an ill-posed detection task, which does not necessarily reflect the practical objective of the system design [55].

3.1 The Generalized Error Performance Measure

Let us again start with a single user system that consists of one transmitter and one receiver. The discrete-time memoryless channel is modeled by a conditional distribution function $P_{Y|X}$ with $X \in \mathcal{X}$ being the channel input symbol and $Y \in \mathcal{Y}$ being the channel output symbol. $\mathcal{X}$ and $\mathcal{Y}$ are the finite input and output alphabets, respectively. We use $P(Y|X)$ to denote the conditional probability of channel output symbol Y given channel input symbol X. Time is slotted with each slot equaling the length of N symbol durations, which is also the length of a packet. As before, we focus on channel coding within one time slot or one packet. We assume that the transmitter is equipped with an ensemble of random block codes, denoted by $\mathcal{G} = \{g_1, \dots, g_M\}$, with cardinality $|\mathcal{G}| = M$. Each code $g \in \mathcal{G}$ is characterized by a rate and input distribution pair (r_g, P_{gX}). The probability of a channel input symbol X with respect to input distribution P_{gX} is denoted by $P_g(X)$. At the beginning of each time slot, the transmitter arbitrarily chooses a code index $g \in \mathcal{G}$, and then encodes a message w into a codeword $X_g^{(N)}(w)$ of N symbols and sends the symbols through the channel. We assume that the receiver knows the randomly generated codebooks of all codes. But the receiver does not know the value of code index g. The receiver chooses an operation region $R \subseteq \mathcal{G}$ which is a subset of the code ensemble $\mathcal{G}$. Upon receiving the sequence of channel output symbols $Y^{(N)}$, the receiver intends to output an estimated message and code index pair $(\hat{w}, \hat{g})$ if $g \in R$, and to report collision if $g \notin R$.

Let the actual message and code index pair be (w, g). We define the conditional error probability as a function of g as

$$P_e(g) = \begin{cases} \max_w Pr\{(\hat{w},\hat{g}) \neq (w,g) | (w,g)\}, & \forall g \in R \\ \max_w 1 - Pr\left\{ \begin{array}{c} \text{“collision” or} \\ (\hat{w},\hat{g}) = (w,g) \end{array} \middle| (w,g) \right\}, & \forall g \notin R \end{cases}. \tag{3.1}$$

Note that whether or not we regard correct message decoding as an acceptable outcome for $g \notin R$ does not make any difference in the performance of a single user system. This is because the receiver can always choose to report collision for $g \notin R$ when g can be detected correctly.

Let $\{\alpha_g\}$ be a set of pre-determined weight parameters each being assigned to a code index $g \in \mathcal{G}$. $\{\alpha_g\}$ satisfies

$$\left\{ \alpha_g \middle| \alpha_g \geq 0, \forall g \in \mathcal{G}, \sum_g e^{-N\alpha_g} = 1 \right\}. \tag{3.2}$$

We define the “generalized error performance” (GEP) measure[1] of the system as

$$\text{GEP} = \sum_g P_e(g) e^{-N\alpha_g}. \tag{3.3}$$

GEP defined in (3.3) is the expected communication error probability of the system if code index g is chosen with a presumed probability of $e^{-N\alpha_g}$.

The following theorem gives an achievable bound of the generalized error probability.

Theorem 3.1. Consider the single user distributed communication system described above. There exists a decoding algorithm such that the generalized error performance of the system is upper bounded by

$$\text{GEP} \leq \sum_{g \in R} \left[\sum_{\tilde{g} \in R} \exp(-NE_m(g,\tilde{g})) + 2 \sum_{\tilde{g} \notin R} \exp(-NE_i(g,\tilde{g})) \right], \tag{3.4}$$

where $E_m(g,\tilde{g})$ and $E_i(g,\tilde{g})$ in the above equation are given by

[1] Introduction of the GEP measure was originally inspired by the results presented in [24].

$$E_m(g,\tilde{g}) = \max_{0<\rho\le 1} -\rho r_{\tilde{g}} + \max_{0\le s\le 1} -\log \sum_Y \left(\sum_X P_g(X)[P(Y|X)e^{-\alpha_g}]^{1-s}\right)\left(\sum_X P_{\tilde{g}}(X)[P(Y|X)e^{-\alpha_{\tilde{g}}}]^{\frac{s}{\rho}}\right)^{\rho},$$
$$E_i(g,\tilde{g}) = \max_{0<\rho\le 1} -\rho r_g + \max_{0\le s\le 1-\rho} -\log \sum_Y \left(\sum_X P_g(X)[P(Y|X)e^{-\alpha_g}]^{\frac{s}{s+\rho}}\right)^{s+\rho}\left(\sum_X P_{\tilde{g}}(X)P(Y|X)e^{-\alpha_{\tilde{g}}}\right)^{1-s} \tag{3.5}$$

The proof of Theorem 3.1 is provided in Appendix B.1. Although Theorem 3.1 is implied by Theorem 3.2, because the proofs are quite sophisticated, we provide complete proofs for both theorems and organize them in the same structure so that readers can regard the proof of Theorem 3.1 as a reference to help understanding the more notation-complicated proof of Theorem 3.2.

Note that the meaning of weight parameters $\{\alpha_g\}$ can be interpreted in various ways. For example, suppose the system chooses $\alpha_g = +\infty$ for all $g \notin R$. On one hand, this can be interpreted as that the system does not care about its error performance when $g \notin R$. On the other hand, it can also be interpreted as an assumption that the transmitter should not choose a code index g with $g \notin R$ in the first place.

Furthermore, with appropriate choices of $\{\alpha_g\}$, GEP can also be used as a tool to bound other error probability measures of interest. For example, if one wants to get an achievable bound for $\max_g P_e(g)$, due to the union bound,

$$\max_g P_e(g) \le \sum_g P_e(g) = M \times \text{GEP} \quad \text{with } \alpha_g = \frac{1}{N}\log M. \tag{3.6}$$

An achievable bound on $\max_g P_e(g)$ can be obtained by combining (3.6) with (3.4).

Compared with a coordinated communication model where transmitter should always choose a coding scheme that supports reliable decoding, a distributed communication system needs to prepare for the situation when transmitter chooses a coding option outside the operation region or even the capacity region. We already showed that distributed capacity of a single user channel coincides with the Shannon capacity in a sense explained after Theorem 2.2. It is intuitive to expect that certain price must be paid to support the detection of transmission options outside the operation region in the distributed communication model. In the following example, we show that such a price can indeed be seen in the scaling law of error performance bound with respect to the codeword length.

Example 3.1. Consider a single user communication system over a binary symmetric channel with crossover probability 0.1. Shannon capacity of the channel is given by $C = 0.37$ nats/symbol. Assume that the transmitter has two coding options both with uniform input distribution. Rates of the two options are $r_1 < C = 0.37$ and $r_2 = 0.6 > C$. Assume that the operation region contains only the first coding option, i.e., $R = \{g = 1\}$. Let $e^{-N\alpha_1} = e^{-N\alpha_2} = 0.5$. In the case of a large codeword length, the scaling law of error probability bound (3.4) in the codeword length is determined by the following error exponent,

$$E_d = \min\{E_m(g = 1, \tilde{g} = 1), E_i(g = 1, \tilde{g} = 1)\}. \tag{3.7}$$

If we choose $e^{-N\alpha_1} = 1$, $e^{-N\alpha_2} = 0$ and force the transmitter to use the first coding option, then the corresponding error exponent becomes the classical random coding exponent given as in [28] by

$$E_r = E_m(g = 1, \tilde{g} = 1). \tag{3.8}$$

A comparison of E_d and E_r as a function of r_1 is illustrated in Figure 3.1. It can be seen that the value of E_d is almost half of E_r in this case[2].

3.2 Decoder of A User Group

Next, we will extend the result to a multiple access system with $K + 1$ users, indexed by $\{0, 1, \ldots, K\}$. Let $D \subseteq \{0, \ldots, K\}$ be a user subset. We assume that users in D are regular users, while users not

[2] The situation can certainly vary depending on the coding parameters of the system.

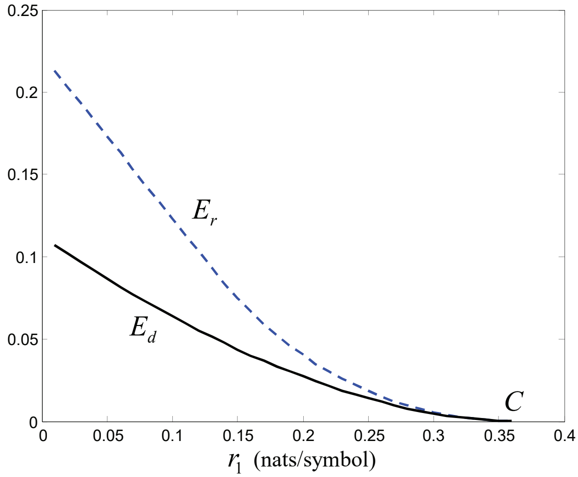

Figure 3.1: Comparison of achievable distributed communication error exponent and the classical random coding exponent.

in D are interfering users. We assume that each user, say user k, is equipped with an ensemble of M communication options, denoted by $\mathcal{G}_k = \{g_{k1}, \ldots, g_{kM}\}$. For a regular user $k \in D$, $\mathcal{G}_k$ represents an ensemble of random block codes, with each code $g_k \in \mathcal{G}_k$ being characterized by a rate and input distribution pair $(r_{g_k}, P_{g_k X_k})$, where X_k is the channel input symbol of user k with a finite alphabet $\mathcal{X}_k$. The probability of a channel input symbol X_k with respect to input distribution $P_{g_k X_k}$ is denoted by $P_{g_k}(X_k)$. For an interfering user $k \notin D$, $\mathcal{G}_k$ may or may not represent a code ensemble. However, for convenience, we still call $\mathcal{G}_k$ the code ensemble and $g_k \in \mathcal{G}_k$ the code index of user k, irrespective of whether k is a regular user or an interfering user. We use $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{g}$ to denote the vectors of code ensembles and code indices of the users, respectively. We assume that $\boldsymbol{\mathcal{G}}$ should be known at the receiver, but the receiver does not know the code index vector $\boldsymbol{g}$ chosen at the beginning of each time slot. We also assume that the randomly generated codebooks of the regular users should be known at the receiver. For the interfering users, we assume that either there is no message to decode, or the receiver chooses not to decode their messages.

Let $\boldsymbol{X}_D$ denote the vectors of channel input symbols of the regular users. Let Y be the channel output symbol with a finite alphabet $\mathcal{Y}$. The discrete-time memoryless channel is characterized by a conditional distribution function $P_{Y|\boldsymbol{X}_D}(\boldsymbol{g}_{\bar{D}})$, where $\boldsymbol{g}_{\bar{D}}$ is the vector of code indices of the interfering users. In other words, channel statistics experienced by the regular users depends on the code indices of the interfering users. We use $P(Y|\boldsymbol{X}_D, \boldsymbol{g}_{\bar{D}})$ to denote the probability of channel output Y given input vector $\boldsymbol{X}_D$ and with code index vector of the interfering user being $\boldsymbol{g}_{\bar{D}}$.

At the beginning of each time slot, users choose their code index vector $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$, the regular users encode their messages $\boldsymbol{w}_D$ into a sequence of N channel input symbol vectors $\boldsymbol{X}_D^{(N)}$, and send them through the channel. Upon receiving the channel output sequence $Y^{(N)}$, the receiver either outputs an estimated message vector and code index vector pair $(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D)$ for the regular users, or reports collision for all regular users. We assume that the receiver should choose an operation region $\boldsymbol{R}_D$ in the space of the code index vectors. The receiver intends to decode the messages of the regular users if $\boldsymbol{g} \in \boldsymbol{R}_D$. The receiver intends to accept collision report for all regular users if $\boldsymbol{g} \notin \boldsymbol{R}_D$.

Let the actual messages of the regular users be $\boldsymbol{w}_D$, and the actual code indices of all users be $\boldsymbol{g}$. We define the conditional error probability as a function of $\boldsymbol{g}$ as

$$P_e(\boldsymbol{g}) = \begin{cases} \max_{\boldsymbol{w}_D} Pr\{(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D) \neq (\boldsymbol{w}_D, \boldsymbol{g}_D) | (\boldsymbol{w}_D, \boldsymbol{g})\}, & \forall \boldsymbol{g} \in \boldsymbol{R}_D \\ \max_{\boldsymbol{w}_D} 1 - Pr\left\{ \begin{array}{l} \text{“collision” or} \\ (\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D) = (\boldsymbol{w}_D, \boldsymbol{g}_D) \end{array} \middle| (\boldsymbol{w}_D, \boldsymbol{g}) \right\}, & \forall \boldsymbol{g} \notin \boldsymbol{R}_D \end{cases} . \tag{3.9}$$

As before, we still regard correct message decoding as an acceptable output for $\boldsymbol{g} \notin \boldsymbol{R}_D$, which means that collision detection for code index vectors outside the operation region is not enforced.

Let $\{\alpha_{\boldsymbol{g}}\}$ be a set of pre-determined weight parameters each being assigned to a code index vector $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$. $\{\alpha_{\boldsymbol{g}}\}$ satisfies

$$\left\{ \alpha_{\boldsymbol{g}} \middle| \alpha_{\boldsymbol{g}} \geq 0, \forall \boldsymbol{g} \in \boldsymbol{\mathcal{G}}, \sum_{\boldsymbol{g}} e^{-N\alpha_{\boldsymbol{g}}} = 1 \right\}. \tag{3.10}$$

We define the "generalized error performance" measure of the system as

$$\mathrm{GEP}_D = \sum_{\boldsymbol{g}} P_e(\boldsymbol{g}) e^{-N\alpha_{\boldsymbol{g}}}. \tag{3.11}$$

An achievable bound of the generalized error probability of the multiple access system is given in the following theorem.

Theorem 3.2. Consider the distributed multiple access system described above. There exists a decoding algorithm such that the generalized error performance defined in (3.11) is upper bounded by

$$\begin{aligned} \mathrm{GEP}_D \leq \sum_{\boldsymbol{g} \in \boldsymbol{R}_D} \sum_{S \subset D} \Bigg[& \sum_{\tilde{\boldsymbol{g}} \in \boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S = \boldsymbol{g}_S} \exp(-N E_{mD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)) \\ & + 2 \sum_{\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S = \boldsymbol{g}_S} \exp(-N E_{iD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)) \Bigg], \end{aligned} \tag{3.12}$$

where $E_{mD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)$ and $E_{iD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)$ for $S \subset D$ in the above equation are given by

$$\begin{aligned} & E_{mD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S) = \\ & \quad \max_{0<\rho\leq 1} -\rho \sum_{k \in D \setminus S} r_{\tilde{g}_k} + \max_{0 \leq s \leq 1} -\log \sum_{Y} \sum_{\boldsymbol{X}_S} \prod_{k \in S} P_{g_k}(X_k) \\ & \quad \times \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y|\boldsymbol{X}_D, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{1-s} \prod_{k \in D \setminus S} P_{g_k}(X_k) \right) \\ & \quad \times \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y|\boldsymbol{X}_D, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \right]^{\frac{s}{\rho}} \prod_{k \in D \setminus S} P_{\tilde{g}_k}(X_k) \right)^{\rho}, \\ & E_{iD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S) = \\ & \quad \max_{0<\rho\leq 1} -\rho \sum_{k \in D \setminus S} r_{g_k} + \max_{0 \leq s \leq 1-\rho} -\log \sum_{Y} \sum_{\boldsymbol{X}_S} \prod_{k \in S} P_{g_k}(X_k) \\ & \quad \times \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y|\boldsymbol{X}_D, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{\frac{s}{s+\rho}} \prod_{k \in D \setminus S} P_{g_k}(X_k) \right)^{s+\rho} \end{aligned}$$

$$\times \left(\sum_{\boldsymbol{X}_{D \setminus S}} P(Y|\boldsymbol{X}_D, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \prod_{k \in D \setminus S} P_{\tilde{g}_k}(X_k) \right)^{1-s}. \qquad (3.13)$$

Theorem 3.2 is implied by Theorem 3.3 by setting $\widehat{\boldsymbol{R}}_D = \bar{\boldsymbol{R}}_D$, which leads to $E_{iD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, D) = \infty$ in (3.18).

While the bound given in (3.12) and (3.13) are quite sophisticated in appearance, it may not always be difficult to evaluate especially when the number of regular users $|D|$ is relatively small. This will be explained further in Section 3.5.

3.3 Operation Region and Operation Margin

In the previous section, when the actual code index vector $\boldsymbol{g} \notin \boldsymbol{R}_D$ is outside the operation region, we still regard correct message decoding as one of the acceptable outputs at the receiver. Consequently, when the receiver forwards a decoded message to the data link layer, it is not always possible for the link layer to tell whether $\boldsymbol{g}$ is inside or outside the operation region. The lack of collision report enforcement may cause problems at the data link layer because many link layer protocols regard collision report as a key guidance for adapting packet transmission probabilities of the users [42][12]. As we introduced before, instead of using error probability definition (3.9), one can regard collision report as the only acceptable output for $\boldsymbol{g} \notin \boldsymbol{R}_D$ and adopt the following error probability definition

$$P_e(\boldsymbol{g}) = \begin{cases} \max_{\boldsymbol{w}_D} Pr\{(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D) \neq (\boldsymbol{w}_D, \boldsymbol{g}_D) | (\boldsymbol{w}_D, \boldsymbol{g})\}, & \forall \boldsymbol{g} \in \boldsymbol{R}_D \\ \max_{\boldsymbol{w}_D} 1 - Pr\{\text{“collision”} | (\boldsymbol{w}_D, \boldsymbol{g})\}, & \forall \boldsymbol{g} \notin \boldsymbol{R}_D \end{cases}. \qquad (3.14)$$

Nevertheless, we show in the following example that both (3.9) and (3.14) could be bad choices for dealing with collision detection.

Example 3.2. Let us consider a single user distributed communication system over a compound binary symmetric channel. The channel model is given by

$$Y = \begin{cases} X & \text{with probability } 1-p \\ \bar{X} & \text{with probability } p \end{cases}. \tag{3.15}$$

where $X, Y \in \{0,1\}$ are the input and output symbols and $0 \leq p \leq 1$ is the crossover probability that can take 9 possible values $p \in \{0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23\}$. Shannon capacities of the channel corresponding to the possible p values are given by $\{0.58,\ 0.56, 0.54, 0.53, 0.51, 0.50, 0.49, 0.47, 0.46\}$ nats/symbol. Suppose that the transmitter only has a unique coding option, which features uniform input distribution with a rate of $r = 0.50$ nats/symbol.

According to the problem formulation introduced in Section 2.4, the system model should contain a regular user with a single coding option and an interfering user with 9 communication options corresponding to the possible channel realizations. It is easy to see that any operation region chosen by the receiver must take the form of $R = \{p|p \leq p_0\}$ for certain p_0. Let us take $p_0 = 0.17$ and $R = \{p|p \leq 0.17\}$ for example. If one chooses error probability definition (3.9) that does not enforce collision report for $p \notin R$, then the upper layer won't know much about the channel if the receiver outputs a decoded message. Alternatively, if one chooses error probability definition (3.14) that does enforce collision report for $p \notin R$, then it implies that the receiver must distinguish channel realization $p = 0.17 \in R$ from $p = 0.18 \notin R$. Note that they correspond to channel capacities 0.54 and 0.53, both are reasonably larger than the rate $r = 0.50$ of the regular user. Even though message decoding can enjoy a low error probability when $p \in \{0.17, 0.18\}$, error probability on distinguishing these two channel realizations can be quite high. Consequently, error performance of the system will be dominated by the difficulty of an unreasonable channel detection task, which may not reflect the design objective of the system.

Alternatively, we can partition the values of p into three sets. Let $R = \{p|p = 0.15, 0.16, 0.17\}$ be termed the operation region. For $p \in R$, only correct message decoding is regarded as the acceptable output. Let $\widehat{R} = \{p|p = 0.18, 0.19\}$ be termed the "operation margin". For $p \in \widehat{R}$, both correct message decoding and collision report are regarded as acceptable outputs. Then for $p \notin R \cup \widehat{R}$, only collision report is regarded as the acceptable output. With such a choice, channel detection task is relaxed to the requirement of distinguishing channels $p \leq 0.17$ from $p \geq 0.20$, which is relatively less challenging. In the meantime, because a message decoding implies that $p \leq 0.19$ and a collision report implies that $p > 0.17$, output at the receiver does carry certain channel information that can be used to guide transmission adaptation at the data link layer.

The above example showed that, in addition to the determination of an operation region, introducing an operation margin and only enforcing collision report outside the operation region and the operation margin can effectively avoid an ill-posed collision detection problem at the receiver. With such an understanding, we extend the result of the previous section as follows.

Consider the multiple access system with $K+1$ users introduced in Section 3.2, where users in $D \subseteq \{0,\ldots,K\}$ are regular users and users not in D are interfering users. Assume that in addition to an operation region $\boldsymbol{R}_D$, the receiver also chooses an "operation margin" $\widehat{\boldsymbol{R}}_D$ with $\widehat{\boldsymbol{R}}_D \cap \boldsymbol{R}_D = \phi$. Let the actual messages of the regular users be $\boldsymbol{w}_D$, and the actual code indices of all users be $\boldsymbol{g}$. We define the conditional error probability as a function of $\boldsymbol{g}$ as

$$P_e(\boldsymbol{g}) = \begin{cases} \max_{\boldsymbol{w}_D} Pr\{(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D) \neq (\boldsymbol{w}_D, \boldsymbol{g}_D) | (\boldsymbol{w}_D, \boldsymbol{g})\}, & \forall \boldsymbol{g} \in \boldsymbol{R}_D \\ \max_{\boldsymbol{w}_D} 1 - Pr\left\{ \begin{array}{l} \text{“collision” or} \\ (\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D) = (\boldsymbol{w}_D, \boldsymbol{g}_D) \end{array} \middle| (\boldsymbol{w}_D, \boldsymbol{g}) \right\}, & \forall \boldsymbol{g} \in \widehat{\boldsymbol{R}}_D \\ \max_{\boldsymbol{w}_D} 1 - Pr\left\{\text{“collision”} | (\boldsymbol{w}_D, \boldsymbol{g})\right\}, & \forall \boldsymbol{g} \notin \boldsymbol{R}_D \cup \widehat{\boldsymbol{R}}_D \end{cases} \tag{3.16}$$

Note that error probability definition (3.16) is general in the sense that it covers error probability definition (3.9) as a special case with $\widehat{\boldsymbol{R}}_D = \bar{\boldsymbol{R}}_D$ and also covers error probability definition (3.14) as a special case with $\widehat{\boldsymbol{R}}_D = \phi$.

Let $\{\alpha_{\boldsymbol{g}}\}$ be a set of pre-determined weight parameters satisfying (3.10) each being assigned to a code index vector $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$. We maintain the definition of the "generalized error performance" measure as in (3.11). With the operation margin added, the following theorem gives the corresponding achievable bound of the generalized error probability of the multiple access system.

Theorem 3.3. Consider the distributed multiple access system described above. There exists a decoding algorithm such that the generalized error performance defined in (3.11) is upper bounded by

$$\begin{aligned} \mathrm{GEP}_D \leq \sum_{\boldsymbol{g}\in\boldsymbol{R}_D} \Bigg\{ \sum_{S\subset D} \Bigg[\sum_{\tilde{\boldsymbol{g}}\in\boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S=\boldsymbol{g}_S} \exp(-NE_{mD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S)) \\ +2\sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S=\boldsymbol{g}_S} \exp(-NE_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S)) \Bigg] \\ +2\sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}_D\cup\widehat{\boldsymbol{R}}_D, \tilde{\boldsymbol{g}}_D=\boldsymbol{g}_D} \exp(-NE_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},D)) \Bigg\}, \end{aligned} \tag{3.17}$$

where $E_{mD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S)$, $E_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S)$ for $S \subset D$ and $E_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},D)$ in the above equation are given by

$$\begin{aligned} &E_{mD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S) = \\ &\quad \max_{0<\rho\leq 1} -\rho\sum_{k\in D\setminus S} r_{\tilde{g}_k} + \max_{0\leq s\leq 1} -\log\sum_Y\sum_{\boldsymbol{X}_S}\prod_{k\in S} P_{g_k}(X_k) \\ &\quad \times\left(\sum_{\boldsymbol{X}_{D\setminus S}} [P(Y|\boldsymbol{X}_D,\boldsymbol{g}_{\bar{D}})e^{-\alpha_{\boldsymbol{g}}}]^{1-s}\prod_{k\in D\setminus S} P_{g_k}(X_k)\right) \\ &\quad \times\left(\sum_{\boldsymbol{X}_{D\setminus S}} [P(Y|\boldsymbol{X}_D,\tilde{\boldsymbol{g}}_{\bar{D}})e^{-\alpha_{\tilde{\boldsymbol{g}}}}]^{\frac{s}{\rho}}\prod_{k\in D\setminus S} P_{\tilde{g}_k}(X_k)\right)^{\rho}, \\ &E_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S) = \\ &\quad \max_{0<\rho\leq 1} -\rho\sum_{k\in D\setminus S} r_{g_k} + \max_{0\leq s\leq 1-\rho} -\log\sum_Y\sum_{\boldsymbol{X}_S}\prod_{k\in S} P_{g_k}(X_k) \\ &\quad \times\left(\sum_{\boldsymbol{X}_{D\setminus S}} [P(Y|\boldsymbol{X}_D,\boldsymbol{g}_{\bar{D}})e^{-\alpha_{\boldsymbol{g}}}]^{\frac{s}{s+\rho}}\prod_{k\in D\setminus S} P_{g_k}(X_k)\right)^{s+\rho} \end{aligned}$$

$$\times \left(\sum_{\boldsymbol{X}_{D \setminus S}} P(Y|\boldsymbol{X}_D, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \prod_{k \in D \setminus S} P_{\tilde{g}_k}(X_k) \right)^{1-s}.$$

$$E_{iD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, D) = \max_{0 \le s \le 1} -\log \sum_{Y} \sum_{\boldsymbol{X}_D} \prod_{k \in D} P_{g_k}(X_k)$$

$$\left[P(Y|\boldsymbol{X}_D, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}}\right]^s \left[P(Y|\boldsymbol{X}_D, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}}\right]^{1-s}. \tag{3.18}$$

The proof of Theorem 3.3 is provided in Appendix B.2.

We generally expect that the number of regular users $|D|$ should be small and therefore the error performance bound given in the theorem is not necessarily difficult to evaluate. Furthermore, because the purpose of choosing an operation margin $\widehat{\boldsymbol{R}}_D$ is to avoid an ill-posed detection problem, given a particular system, finding an appropriate choice of $\widehat{\boldsymbol{R}}_D$ should not be difficult.

3.4 Single User Decoding

With the preparations of the previous three sections, in this section, we consider the general case of a multiple access system with $K+1$ users, indexed by $\{0, 1, \ldots, K\}$. Let user 0 be an interfering user, and let other users be regular users. We assume that the receiver is associated to user 1. The receiver is only interested in decoding the message of user 1 but it can choose to decode the messages of other regular users if necessary. As before, we assume that each user, say user k, is equipped with an ensemble of M communication options, denoted by $\mathcal{G}_k = \{g_{k1}, \ldots, g_{kM}\}$. For a regular user $k \neq 0$, $\mathcal{G}_k$ represents an ensemble of random block codes, with each code $g_k \in \mathcal{G}_k$ being characterized by a rate and input distribution pair $(r_{g_k}, P_{g_k X_k})$, where X_k is the channel input symbol of user k with a finite alphabet $\mathcal{X}_k$. We use $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{g}$ to denote the vectors of code ensembles and code indices of all users, respectively. We assume that $\boldsymbol{\mathcal{G}}$ and the randomly generated codebooks of the regular users should be known at the receiver, but the receiver does not know the code index vector $\boldsymbol{g}$ chosen at the beginning of each time slot.

Let $\boldsymbol{X}$ denote the vectors of channel input symbols of the regular users. Let Y be the channel output symbol with a finite alphabet $\mathcal{Y}$.

The discrete-time memoryless channel is characterized by a conditional distribution function $P_{Y|\boldsymbol{X}}(g_0)$, where g_0 is the code index (or communication option) of the interfering user. We use $P_{g_k}(X_k)$ to denote the probability of a channel input symbol X_k with respect to input distribution $P_{g_k X_k}$, and use $P(Y|\boldsymbol{X}, g_0)$ to denote the probability of channel output Y given input vector $\boldsymbol{X}$ and with code index of the interfering user being g_0.

At the beginning of each time slot, users choose their code index vector $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$, the regular users encode their messages $\boldsymbol{w}$ into a sequence of N channel input symbol vectors $\boldsymbol{X}^{(N)}$, and send them through the channel. Upon receiving the channel output sequence $Y^{(N)}$, the receiver either outputs an estimated message and code index pair $(\hat{w}_1, \hat{g}_1)$ for user 1, or reports collision for user 1. We assume that the receiver should choose an operation region $\boldsymbol{R}_1$ and an operation margin $\widehat{\boldsymbol{R}}_1$ with $\boldsymbol{R}_1 \cap \widehat{\boldsymbol{R}}_1 = \phi$ in the space of the code index vectors. The receiver intends to decode the message of user 1 if $\boldsymbol{g} \in \boldsymbol{R}_1$. The receiver intends to enforce collision report for user 1 if $\boldsymbol{g} \notin \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1$.

Let the actual messages of the regular users be $\boldsymbol{w}$, and the actual code indices of all users be $\boldsymbol{g}$. We define the conditional error probability as a function of $\boldsymbol{g}$ as

$$P_e(\boldsymbol{g}) = \left\{ \begin{array}{ll} \max_{\boldsymbol{w}} Pr\{(\hat{w}_1, \hat{g}_1) \neq (w_1, g_1) | (w_1, g_1)\}, & \forall \boldsymbol{g} \in \boldsymbol{R}_1 \\ \max_{\boldsymbol{w}} 1 - Pr\left\{ \begin{array}{l} \text{“collision” or} \\ (\hat{w}_1, \hat{g}_1) = (w_1, g_1) \end{array} \middle| (\boldsymbol{w}, \boldsymbol{g}) \right\}, & \forall \boldsymbol{g} \in \widehat{\boldsymbol{R}}_1 \\ \max_{\boldsymbol{w}} 1 - Pr\left\{\text{“collision”} | (\boldsymbol{w}, \boldsymbol{g})\right\}, & \forall \boldsymbol{g} \notin \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1 \end{array} \right. \tag{3.19}$$

Note that we only enforce collision report for $\boldsymbol{g} \notin \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1$ and regard both collision report and correct message decoding as acceptable outputs for $\boldsymbol{g} \in \widehat{\boldsymbol{R}}_1$. As explained in the previous section, the operation margin should be chosen appropriately to avoid any ill-posed collision detection problem.

Let $\{\alpha_{\boldsymbol{g}}\}$ be a set of pre-determined weight parameters satisfying

(3.10) each being assigned to a code index vector $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$. We define the "generalized error performance" measure of the system as

$$\text{GEP} = \sum_{\boldsymbol{g}} P_e(\boldsymbol{g}) e^{-N\alpha_{\boldsymbol{g}}}. \tag{3.20}$$

To obtain an achievable bound of the generalized error probability, we will first need to define a special decoder termed a $(D, \boldsymbol{R}_D)$ decoder. While the receiver is only interested in decoding the message of user 1, the receiver unavoidably faces the decision whether the messages of some other regular users should also be jointly decoded. Let $D \subseteq \{1, 2, \dots, K\}$ be a user subset with $1 \in D$. Let $\boldsymbol{R}_D \subseteq \boldsymbol{R}_1$ be an operation region. A $(D, \boldsymbol{R}_D)$ decoder represents the decision that the receiver decides to decode messages of all and only the regular users in D. To do so, the receiver uses an operation region of $\boldsymbol{R}_D$ and an operation margin of $\widehat{\boldsymbol{R}}_D = \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1 \setminus \boldsymbol{R}_D$. Based on the error probability definition of (3.16), and the same set of weight parameters $\{\alpha_{\boldsymbol{g}}\}$, let generalized error probability of the $(D, \boldsymbol{R}_D)$ decoder be denoted by GEP_D as given in (3.11).

An achievable bound of the generalized error probability of the multiple access system with the receiver being associated only to user 1 is given in the following theorem.

Theorem 3.4. Consider the distributed multiple access system described above. Assume that the receiver is only interested in decoding the message of user 1. Let $\boldsymbol{R}_1$ be the operation region, $\widehat{\boldsymbol{R}}_1$ be the operation margin, and $\{\alpha_{\boldsymbol{g}}\}$ be the set of weight parameters. Let σ be a partition of the operation region $\boldsymbol{R}_1$, as described below

$$\boldsymbol{R}_1 = \bigcup_{D, D \subseteq \{1,\dots,K\}, 1 \in D} \boldsymbol{R}_D, \qquad \boldsymbol{R}_{D'} \cap \boldsymbol{R}_D = \phi,$$
$$\forall D, D' \subseteq \{1, \dots, K\}, D' \neq D, 1 \in D, D'. \tag{3.21}$$

There exists a decoding algorithm such that the generalized error performance defined in (3.20) is upper bounded by

$$\text{GEP} \leq \min_{\sigma} \sum_{D, D \subseteq \{1,\dots,K\}, 1 \in D} \text{GEP}_D, \tag{3.22}$$

where GEP_D represents the generalized error probability of the $(D, \boldsymbol{R}_D)$ decoder with receiver decoding the messages of all and only the users in D, with the operation region being $\boldsymbol{R}_D$ and the operation margin being $\widehat{\boldsymbol{R}}_D = \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1 \setminus \boldsymbol{R}_D$. Note that GEP_D in (3.22) can be further bounded by (3.17).

The proof of Theorem 3.4 is provided in Appendix B.3.

Note that Theorem 3.4 did not provide an explicit algorithm to calculate the partition that minimizes either $\sum_{D, D \subseteq \{1,\ldots,K\}, 1 \in D} \mathrm{GEP}_D$ or its upper bound obtained from (3.17). To find the optimal partition, one may need to compute every single term on the right hand sides of (3.22), (3.17) and (3.18) for all code index vectors and for all user subsets. Complexity of such calculations can clearly be a concern. While a detailed complexity analysis is beyond the scope of this monograph, some of the key points are discussed in the following section.

3.5 Complexity and Code Index Detection

Because error performance bounds presented in the previous sections are quite sophisticated, it is natural to wonder whether the associated complexity of these results is always too high to be practical. In this section, we share some thoughts about such a concern. We begin the discussion by presenting a basic distributed communication scheme that is in close proximity to the ones widely seen in wireless networks such as Wi-Fi networks today [45].

A basic scheme: Consider a time-slotted distributed communication system where each user in the system has only a single transmission option plus an idling option. Each user either transmits a data packet or idles in a time slot. At the receiver when decoding the message of a particular user, the receiver regards signals from all other users as interference. In other words, there is no joint message decoding of multiple users.

Let us look at the system from the perspective of a particular user, say user 1. Assume that the receiver is associated to user 1. The basic distributed communication scheme illustrated above falls into the problem formulations of Sections 3.2, 3.3, and 3.4 with each user having two coding options, i.e., one transmission option and one idling option, and with user 1 being the only regular user in the system[3]. This indicates that certain systems modeled in the previous sections can indeed have a complexity low enough to support immediate practical implementation. Therefore, extensions of the basic scheme, especially those involving only a handful of coding options for each user and the joint decoding of a small number of users, do not necessarily have a high computational complexity.

Due to lack of user coordination, receiver in a distributed communication system often needs to prepare for users that can potentially be active in the area. The number of users involved in the decoding operation of a receiver can therefore significantly exceed the number of active users. Consequently, when investigating computational complexity of the system, complexity scaling law with respect to the user number should be an important factor. Furthermore, while packets in a distributed system should be relatively short in length, they may not be short enough to completely relieve the classical coding complexity concern. Complexity scaling law with respect to the codeword length may remain a factor although its significance in the overall complexity tradeoff may be different from that of a classical system.

Because distributed channel coding essentially combines a message recovery task with a code index detection task at the receiver, a simple way to avoid calculating the likelihood of too many codewords in channel decoding is to first let the receiver detect the code index vector of the users using only the distribution information of channel input and output symbols [55]. The receiver can then process the codewords corresponding to the detected code indices. Note that, similar to the discussions presented in Section 3.3, without decoding the codewords, a receiver may not have the capability to detect the code index vector precisely. For example, in a multiple access system with homogenous users, the receiver may be able to tell how many users are transmitting. But without message decoding, the receiver won't be able to detect the identities of the active users. Therefore, an appropriate expectation is to let the receiver detect whether the code index vector should

[3]Note that the impact of multiple interfering users can be modeled using one interfering user only.

belong to a chosen subset or not. We formulate the detection problem as follows.

Consider a distributed multiple access system with K users being indexed by $\{1, \ldots, K\}$. We do not distinguish regular users and interfering users because the receiver does not decode any codeword in this step of the operation. We assume that each user, say user k, should be equipped with an ensemble of M coding or communication options, denoted by $\mathcal{G}_k = \{g_{k1}, \ldots, g_{kM}\}$. Let $\boldsymbol{g}$ and $\boldsymbol{\mathcal{G}}$ be the code index vector and the code ensemble vector of the users, respectively. $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$ if $g_k \in \mathcal{G}_k$ for all $k = 1, \ldots, K$. Given the code index vector $\boldsymbol{g}$, we denote the conditional distribution of channel output symbol $Y \in \mathcal{Y}$ by $P(Y|\boldsymbol{g})$.

Let $\boldsymbol{R}, \widehat{\boldsymbol{R}} \subseteq \boldsymbol{\mathcal{G}}$ with $\boldsymbol{R} \cap \widehat{\boldsymbol{R}} = \phi$ be two code index vector subsets. We assume that the receiver intends to report a "true/neutral/false" decision for the code index vector detection problem. The receiver intends to output "true" if $\boldsymbol{g} \in \boldsymbol{R}$ and to output "false" if $\boldsymbol{g} \notin \boldsymbol{R} \cup \widehat{\boldsymbol{R}}$. The receiver intends to accept all outputs if $\boldsymbol{g} \in \widehat{\boldsymbol{R}}$. For convenience, we still term $\boldsymbol{R}$ the operation region and $\widehat{\boldsymbol{R}}$ the operation margin. As explained in Section 3.3, $\widehat{\boldsymbol{R}}$ is designed to avoid an ill-posed detection problem. That is, with the operation margin, the receiver only intends to use a "true" output to claim that $\boldsymbol{g} \in \boldsymbol{R} \cup \widehat{\boldsymbol{R}}$, and to use a "false" output to claim that $\boldsymbol{g} \notin \boldsymbol{R}$.

Given code index vector $\boldsymbol{g}$, we define the conditional error probability of the code index detection problem as follows.

$$P_e(\boldsymbol{g}) = \begin{cases} Pr\{\text{false}|\boldsymbol{g}\}, & \forall \boldsymbol{g} \in \boldsymbol{R} \\ Pr\left\{\text{true}|\boldsymbol{g}\right\}, & \forall \boldsymbol{g} \notin \boldsymbol{R} \cup \widehat{\boldsymbol{R}} \end{cases}. \tag{3.23}$$

Let $\{\alpha_{\boldsymbol{g}}\}$ be a set of weight parameters satisfying (3.10) each being assigned to a code index vector $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$. We define the "generalized error performance" measure as in (3.20). An achievable bound of the generalized error performance is given in the following theorem.

Theorem 3.5. Consider the distributed code index vector detection problem described above. Let $\boldsymbol{R}$ be the operation region, $\widehat{\boldsymbol{R}}$ be the operation margin, and $\{\alpha_{\boldsymbol{g}}\}$ be the set of weight parameters. There exists a decoding algorithm such that the generalized error performance defined in (3.20) is upper bounded by

$$\mathrm{GEP} \leq \sum_{\boldsymbol{g}\in\boldsymbol{R}} \sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}\cup\widehat{\boldsymbol{R}}} \left[\exp(-NE_m(\boldsymbol{g},\tilde{\boldsymbol{g}})) + \exp(-NE_m(\tilde{\boldsymbol{g}},\boldsymbol{g}))\right], \quad (3.24)$$

where $E_m(\boldsymbol{g},\tilde{\boldsymbol{g}})$ in the above equation is given by

$$E_m(\boldsymbol{g},\tilde{\boldsymbol{g}}) = \max_{s\geq 0} -\log\left(\sum_{Y}[P(Y|\boldsymbol{g})e^{-\alpha_{\boldsymbol{g}}}]^{(1-s)}[P(Y|\tilde{\boldsymbol{g}})e^{-\alpha_{\tilde{\boldsymbol{g}}}}]^{s}\right). \quad (3.25)$$

The proof of Theorem 3.5 is given in Appendix B.4.

4

An Enhanced Physical-Link Layer Interface

Existing network architecture only allows a data link layer user to determine whether a packet should be transmitted or not [11]. In distributed networking when communication optimization cannot be done fully at the physical layer, such a single transmission option (or binary transmitting/idling options) significantly limited the capability of exploiting advanced wireless adaptations such as power, rate, and antenna adjustments at the data link layer [55]. As introduced in Section 1.3, this architectural inefficiency can be mitigated by enhancing the physical-link layer interface to equip each link layer user with multiple transmission options. Different transmission options can correspond to different communication settings such as different rate and power combinations. Distributed channel coding theorems introduced in Sections 2 and 3 provided a basic physical layer foundation to support such an interface enhancement in terms of allowing each physical layer transmitter to prepare an ensemble of codes corresponding to the link layer transmission/idling options, and giving a link layer protocol full control of transmission decision without sharing the decision with other users or with the receiver. In this section, we discuss the support of the interface enhancement at the data link layer. Note that, while navigating through the provided transmission options enables certain capability of advanced communication adaptation, due to the layering architecture (or more precisely, the modularity requirement), a link layer protocol is bounded with the provided transmission options and can only construct its transmission scheme within this constraint to optimize a network utility. Therefore, the key question is, for data link layer users in a distributed network, whether there exists a general framework to efficiently exploit an arbitrary and often limited set of provided transmission options to optimize a chosen network utility. Unfortunately, a rigorous answer to this question is not yet available. In the following, we present some early research results to shed light on this link layer problem.

Distributed adaptive medium access control (MAC) protocols can be categorized into splitting algorithms [32][16][33][75][66][95][34] and back-off approaches [83][42][47][12][17]. In splitting algorithms such as the FCFS algorithm [32], under the assumption that noiseless channel feedback is instantly available, users maintain a common virtual interval of their random identity values. The interval is partitioned and ordered, which determines the transmission schedule of the users, according to a sequence of channel feedback messages. While splitting algorithms can often achieve a relatively high system throughput, their function depends on the assumptions of instant availability of channel feedback and correct reception of feedback sequence. Both of the two conditions, unfortunately, can be violated in a wireless environment. Theoretical analysis of a splitting algorithm, taking into account the wireless-related factors such as channel fading, measurement noise, feedback error, and transmission delay, can be extremely challenging. Analysis of the back-off algorithms, on the other hand, has proven to be more trackable [42][41][12]. In back-off algorithms such as the 802.11 DCF protocol [12], conditioned on packet availability, each user should transmit with a particular probability. In most cases [42][17], a user should decrease its transmission probability in response to a packet collision (or transmission failure) event, and increase its transmission probability in response to a transmission success event. Distributed probability adaptation in a back-off algorithm often falls into the framework of stochastic approximation [42][41], whose theoretical analysis enjoys a rigorous set of mathematical and statistical tools developed in the literature [68][48][46][50][14]. Practical back-off algorithms can also be analyzed using Markov models [12]. Most of the existing analysis of the splitting and the back-off algorithms either assume a throughput optimization objective and/or a simple collision channel model. While there has been no analytical framework that can deal with the optimization of a general network utility with a general channel model, the interesting topic of how collision resolution algorithms should be revised to work with wireless-related physical layer properties, such as capture effect and multipacket reception, has attracted significant research efforts in the literature [62][18][35][36][51][17].

In the rest of this section, we will introduce a distributed MAC framework abstracted from the back-off algorithms. In order to maintain a relatively simple and trackable investigation, we choose to focus on distributed link-layer multiple access networking with an unknown number of homogeneous users, and also assume that all users should have saturated message queues. Motivations of such a focus are explained as follows. First, the assumption of saturated message queues is introduced to avoid the complication of random message arrivals. While bursty message arrival is rather an important character of distributed network systems [11][22], it is known to create coupling between transmission activities of the users [82][67], and such coupling often leads to open research problems in throughput and stability analysis [4][78][59][56] of systems with a relatively small number of users [6][40]. Results obtained with the assumption of saturated message queues can often serve as achievable bounds to the corresponding results for systems with random message arrivals [41][56]. Second, because each user only interacts with the receiver, the assumption of multiple access networking with homogeneous users mainly represents the communication environment envisioned by each link layer user[1]. In other words, without further knowledge about the actual networking environment, a link

[1]Note that the assumption of user symmetry is also reflected in many existing channel models such as the collision channel model [11] and the multipacket reception channel model [35][36].

layer protocol should be designed to help a user to get a fair share of the multiple access channel under the assumption of user homogeneity. Early research investigation aims at achieving such a design objective in the assumed networking environment. Understanding the behavior of the link layer algorithm in a general networking environment is a future research task that is beyond the scope of this monograph. Finally, because users in a distributed network often access the channel opportunistically, it is difficult to know how many users are actually active [41]. We assume that the homogeneous users in a distributed multiple access network should be able to calculate its optimal transmission scheme if the user number is known, but we would like to develop distributed algorithms to lead the system to a desired operation point without the knowledge of the actual user number[2]. Note that, rather than developing a practical MAC protocol, our primary objective is to obtain useful insights about distributed medium access control through the analysis of systems with/without the enhanced physical-link layer interface.

4.1 A Stochastic Approximation Framework

Consider a time-slotted distributed multiple access network with a memoryless channel and K homogeneous users. The length of each time slot equals the transmission duration of one packet. We assume that neither the users nor the receiver should know the user number K. Each user, say user k, is equipped with M transmission options plus an idling option, denoted by $\mathcal{G}_k = \{g_{k0}, g_{k1}, \ldots, g_{kM}\}$ with g_{k0} being the idling option. These options correspond to the code ensemble $\mathcal{G}_k$ prepared by the physical layer transmitter of user k, as explained in Sections 2 and 3. We assume that all users are backlogged with saturated message queues. At the beginning of each time slot t, according to an associated probability vector, each user either idles or randomly chooses a transmission option to send a message. Transmission decisions of the users are made individually. The decisions are not shared among the

[2] In back-off algorithms, the necessity of probability adaptation generally implies the assumption that the number of active users is unknown to the system.

users or with the receiver. The M-length probability vector associated to user k in time slot t is denoted by $\boldsymbol{p}_k(t) = p_k(t)\boldsymbol{d}_k(t)$, where $p_k(t)$ is termed the "transmission probability" of user k, and $\boldsymbol{d}_k(t)$, termed the "transmission direction" vector of user k, is an M-length probability vector whose entries $d_{km}(t)$, for $1 \leq m \leq M$, satisfy $d_{km}(t) \geq 0$ and $\sum_{m=1}^{M} d_{km}(t) = 1$.

At the end of each time slot t, based upon available channel feedback, each user k calculates a target probability vector $\tilde{\boldsymbol{p}}_k(t) = \tilde{p}_k(t)\tilde{\boldsymbol{d}}_k(t)$. User k then updates its transmission probability vector by

$$\boldsymbol{p}_k(t+1) = (1-\alpha(t))\boldsymbol{p}_k(t) + \alpha(t)\tilde{\boldsymbol{p}}_k(t) = \boldsymbol{p}_k(t) + \alpha(t)(\tilde{\boldsymbol{p}}_k(t) - \boldsymbol{p}_k(t)), \tag{4.1}$$

where $\alpha(t) > 0$ is a step size parameter of time slot t. Let $\boldsymbol{P}(t) = [\boldsymbol{p}_1^T(t), \boldsymbol{p}_2^T(t), \ldots, \boldsymbol{p}_K^T(t)]^T$ be a vector of length MK that consists of the transmission probability vectors of all users in time slot t. Let $\tilde{\boldsymbol{P}}(t) = [\tilde{\boldsymbol{p}}_1^T(t), \tilde{\boldsymbol{p}}_2^T(t), \ldots, \tilde{\boldsymbol{p}}_K^T(t)]^T$ be the corresponding target vector. $\boldsymbol{P}(t)$ is updated by

$$\boldsymbol{P}(t+1) = \boldsymbol{P}(t) + \alpha(t)(\tilde{\boldsymbol{P}}(t) - \boldsymbol{P}(t)). \tag{4.2}$$

Note that (4.2) falls into the framework of stochastic approximation algorithms [68][48][46], where the actual target transmission probability vector $\tilde{\boldsymbol{P}}(t)$ is often calculated based upon noisy estimates of certain system variables.

Define $\hat{\boldsymbol{P}}(t) = [\hat{\boldsymbol{p}}_1^T(t), \hat{\boldsymbol{p}}_2^T(t), \ldots, \hat{\boldsymbol{p}}_K^T(t)]^T$ as the "theoretical value" of $\tilde{\boldsymbol{P}}(t)$ when there is no measurement noise and no feedback error in time slot t, with $\hat{\boldsymbol{p}}_k(t)$ being the corresponding theoretical value of $\tilde{\boldsymbol{p}}_k(t)$, for $1 \leq k \leq K$. Let $E_t[\tilde{\boldsymbol{P}}(t)]$ be the expectation of $\tilde{\boldsymbol{P}}(t)$ conditioned on system state at the beginning of time slot t. Let us express $E_t[\tilde{\boldsymbol{P}}(t)]$ as follows

$$E_t[\tilde{\boldsymbol{P}}(t)] = \hat{\boldsymbol{P}}(t) + \boldsymbol{G}(t) = \hat{\boldsymbol{P}}(\boldsymbol{P}(t)) + \boldsymbol{G}(\boldsymbol{P}(t)), \tag{4.3}$$

where $\boldsymbol{G}(t) = E_t[\tilde{\boldsymbol{P}}(t)] - \hat{\boldsymbol{P}}(t)$ is defined as the bias term in the target probability vector calculation. Given the communication channel, both $\hat{\boldsymbol{P}}(t)$ and $\boldsymbol{G}(t)$ are functions of $\boldsymbol{P}(t)$, which is the transmission probability vector in time slot t.

Next, we present two conditions that are typically required for the convergence of a stochastic approximation algorithm [68][48][46].

Condition 1. (Mean and Bias) There exists a constant $K_m > 0$ and a bounding sequence $0 \leq \beta(t) \leq 1$, such that

$$\|\boldsymbol{G}(\boldsymbol{P}(t))\| \leq K_m \beta(t). \tag{4.4}$$

Furthermore, we assume that $\beta(t)$ should be controllable in the sense that one can design protocols to ensure $\beta(t) \leq \epsilon$ for any chosen $\epsilon > 0$ and for large enough t.

Condition 2. (Lipschitz Continuity) There exists a constant $K_l > 0$, such that

$$\|\hat{\boldsymbol{P}}(\boldsymbol{P}_a) - \hat{\boldsymbol{P}}(\boldsymbol{P}_b)\| \leq K_l \|\boldsymbol{P}_a - \boldsymbol{P}_b\|, \qquad \text{for all } \boldsymbol{P}_a, \boldsymbol{P}_b. \tag{4.5}$$

Under these conditions, according to stochastic approximation theory [46][50][14], if the step size sequence $\alpha(t)$ and the bounding sequence $\beta(t)$ are small enough, trajectory of the transmission probability vector $\boldsymbol{P}(t)$ under distributed adaptation given in (4.2) can be approximated by the following associated ordinary differential equation (ODE),

$$\frac{d\boldsymbol{P}(t)}{dt} = -[\boldsymbol{P}(t) - \hat{\boldsymbol{P}}(t)], \tag{4.6}$$

where, with an abuse of notation, we also used t to denote the continuous time variable. Because all entries of $\boldsymbol{P}(t)$ and $\hat{\boldsymbol{P}}(t)$ stay in the range of $[0, 1]$, any equilibrium $\boldsymbol{P}$ of the associated ODE given in (4.6) must satisfy

$$\boldsymbol{P} = \hat{\boldsymbol{P}}(\boldsymbol{P}). \tag{4.7}$$

Suppose that the solution to (4.7), which is also the equilibrium of (4.6), is unique at $\boldsymbol{P}^* = [\boldsymbol{p}_1^{*T}, \ldots, \boldsymbol{p}_K^{*T}]^T$. According to stochastic approximation theory, if the step size sequence $\alpha(t)$ and the bounding sequence $\beta(t)$ are small in value, we have the following convergence results.

Theorem 4.1. For distributed transmission probability adaptation given in (4.2), assume that the associated ODE given in (4.6) has a unique stable equilibrium at $\boldsymbol{P}^*$. Suppose that $\alpha(t)$ and $\beta(t)$ satisfy the following conditions

$$\sum_{t=0}^{\infty} \alpha(t) = \infty, \sum_{t=0}^{\infty} \alpha(t)^2 < \infty, \sum_{t=0}^{\infty} \alpha(t)\beta(t) < \infty. \tag{4.8}$$

Under Conditions 1 and 2, $\boldsymbol{P}(t)$ converges to $\boldsymbol{P}^*$ with probability one.

Theorem 4.1 is implied by [50, Theorem 4.3].

Theorem 4.2. For distributed transmission probability adaptation given in (4.2), assume that the associated ODE given in (4.6) has a unique stable equilibrium at $\boldsymbol{P}^*$. Let Conditions 1 and 2 be met. Then for any $\epsilon > 0$, there exists a constant $K_w > 0$, such that, for any $0 < \underline{\alpha} < \overline{\alpha} < 1$ satisfying the following constraint

$$\exists T_0 \geq 0, \underline{\alpha} \leq \alpha(t) \leq \overline{\alpha}, \beta(t) \leq \sqrt{\overline{\alpha}}, \forall t \geq T_0, \tag{4.9}$$

$\boldsymbol{P}(t)$ converges to $\boldsymbol{P}^*$ in the following sense

$$\limsup_{t \to \infty} Pr\left\{ \|\boldsymbol{P}(t) - \boldsymbol{P}^*\| \geq \epsilon \right\} < K_w \overline{\alpha}. \tag{4.10}$$

Theorem 4.2 can be obtained by following the proof of [14, Theorem 2.3] with minor revisions.

Note that, for simplicity, our system model assumed the same step size sequence $\alpha(t)$ and the same bounding sequence $\beta(t)$ for all users. We also assumed that all users should update their transmission probability vectors (synchronously) in each time slot. However, by following the literature of stochastic approximation theory [46][50], it is easy to show that different users can use different step sizes and bounding sequences, and can also adapt their probability vectors asynchronously. So long as the step sizes and bounding sequences of all users satisfy the same constraints given in (4.8) and (4.9), and they also update their probability vectors frequently enough, then convergence results stated in Theorems 4.1 and 4.2 should remain valid.

With convergence of the probability vectors guaranteed by Theorems 4.1 and 4.2, the key objective of the system design is to develop distributed MAC algorithms to satisfy Conditions 1 and 2 and to place the unique equilibrium of the associated ODE at the desired point. Unfortunately, achieving such an objective is not always easy especially when the enhanced physical-link layer interface is introduced. Because users are homogeneous, due to symmetry, if an equilibrium of the system is unique, transmission probability vectors of the users at the equilibrium must be identical. We choose to enforce such a property by guaranteeing that all users should obtain the same target transmission probability vector in each time slot. The corresponding part of the system design is introduced below.

In each time slot, we assume that there is a set of V virtual packets being transmitted through the channel. The virtual packet set remains the same over different time slots. Each virtual packet in the set is an assumed packet whose coding parameters are known to the users and to the receiver, but it is not physically transmitted in the system, i.e., the packet is "virtual". We assume that, without knowing the transmission status of the users, the receiver can detect whether the reception of each virtual packet should be regarded as successful or not, and therefore can estimate the success probability of each virtual packet. For example, suppose that the link layer channel is a collision channel, and a virtual packet has the same coding parameters of a real packet. Then, the virtual packet reception should be regarded as successful if and only if no real packet is transmitted in a time slot. Success probability of the virtual packet in this case equals the idling probability of the collision channel. For another example, if all packets including the virtual packets are encoded using random block codes, given the physical layer channel model, reception of each virtual packet corresponds to a detection task that judges whether or not code index vector of the real users should belong to a specific operation region. Such detection tasks and their performance bounds have been extensively discussed in the previous sections.

Let $\boldsymbol{q}_v(t)$ be a V-length vector whose entry $q_{vi}(t)$, for $1 \leq i \leq V$, is the success probability of the ith virtual packet in time slot t. We assume that the receiver should measure and feed the estimated $\boldsymbol{q}_v(t)$ back to all users (transmitters). Upon receiving the estimated $\boldsymbol{q}_v(t)$, each user should calculate the M-length target transmission probability vector as the same function of the $\boldsymbol{q}_v(t)$ estimate. Denote the theoretical target probability vector by $\hat{\boldsymbol{p}}(\boldsymbol{q}_v(t))$. The target transmission probability vectors of all users is given by $\hat{\boldsymbol{P}}(t) = \mathbf{1} \otimes \hat{\boldsymbol{p}}(\boldsymbol{q}_v(t))$, where $\mathbf{1}$ denotes a K-length vector of all 1's and $\otimes$ represents the Kronecker product. Consequently, according to (4.6), if $\boldsymbol{P}^*$ is an equilibrium of the system, we must have $\boldsymbol{P}^* = \mathbf{1} \otimes \boldsymbol{p}^*$. Because $\boldsymbol{q}_v$ is a function of the transmission probability vectors, we must have $\boldsymbol{P}^* = \mathbf{1} \otimes \boldsymbol{p}^* = \mathbf{1} \otimes \hat{\boldsymbol{p}}(\boldsymbol{p}^*)$, where $\hat{\boldsymbol{p}}(\boldsymbol{p}^*)$

denotes the theoretical target probability vector of the users given that all users have the same transmission probability vector $\boldsymbol{p}^*$.

We want to point out that, the introduction of virtual packets and the assumption of feeding back $\boldsymbol{q}_v(t)$ to the transmitters are rather rare both in the literature of MAC algorithms and in practical MAC protocols. The key purpose of such a system design is to feed back a measure that is common to all users. This enables users to calculate the same target transmission probability vector and consequently guarantees that transmission probability vectors of all users at any system equilibrium must be identical. As we will see in the following sections, such a property can significantly simplify the design and analysis of the distributed MAC algorithm. If a user only knows the success/failure status of its own packets on the other hand, as commonly assumed in existing MAC algorithms, then guaranteeing identical transmission probability vector at the equilibrium under our problem formulation can become a challenging task.

In a practical system, the measurement of $\boldsymbol{q}_v(t)$ is likely to experience measurement noise and feedback error. We assume that, if users keep $\boldsymbol{P}$ at a constant vector, and $\boldsymbol{q}_v$ is measured over an interval of Q time slots, then the measurement should converge to its true value with probability one as Q is taken to infinity. Other than this assumption, system noise is not involved in the discussions of the design objectives, i.e., to meet Conditions 1 and 2 and to place the unique equilibrium of the associated ODE at the desired point. Therefore, in the following sections, we assume that $\boldsymbol{q}_v(t)$ can be measured precisely at the receiver and be fed back to the users. This leads to $\tilde{\boldsymbol{P}}(t) = \hat{\boldsymbol{P}}(t) = \mathbf{1} \otimes \hat{\boldsymbol{p}}(t)$. To simplify the notation, we will also skip time index t in the rest of the discussions.

4.2 Single Transmission Option

Let us first consider the simple case of classical physical-link layer interface, where each user only has a single transmission option plus an idling option. Each user, say user k, should maintain a scalar transmission probability parameter p_k, which specifies the probability at which user k transmits a packet in a time slot. Transmission probabilities of all users are listed in a K-length vector $\boldsymbol{p}$. In this section, we show that, with a general channel model and without knowing the user number K, a distributed MAC algorithm can be designed to lead the system to a unique equilibrium that is not far from optimal with respect to a chosen symmetric network utility.

Given the physical layer channel and the provided transmission options, we specify the link layer multiple access channel using two sets of channel parameters. Define $\{C_{rj}\}$ for $j \geq 0$ as the "real channel parameter set", where C_{rj} is the conditional success probability of a real packet should it be transmitted in parallel with j other real packets. We also assume that there is a single virtual packet being transmitted in each time slot. Virtual packets transmitted in different time slots are identical. Given coding parameters of the virtual packet, we define $\{C_{vj}\}$ for $j \geq 0$ as the "virtual channel parameter set", where C_{vj} is the success probability of the virtual packet should it be transmitted in parallel with j real packets. Assume $C_{vj} \geq C_{v(j+1)}$ for all $j \geq 0$, which means that, if the number of parallel real packet transmissions increases, the virtual packet should have a non-increasing chance of getting through the channel. Let $\epsilon_v > 0$ be a pre-determined small constant. Define J_{ϵ_v} as the minimum integer such that $C_{vJ_{\epsilon_v}}$ is strictly larger than $C_{v(J_{\epsilon_v}+1)} + \epsilon_v$, i.e.,

$$J_{\epsilon_v} = \arg\min_{j} C_{vj} > C_{v(j+1)} + \epsilon_v. \tag{4.11}$$

We assume that both the real and the virtual channel parameter sets should be known at the users and at the receiver. Note that, while $\{C_{rj}\}$ has nothing to do with the virtual packet, $\{C_{vj}\}$ is dependent on the coding parameters of the virtual packet.

We assume that users intend to maximize a symmetric network utility, denoted by $U(K, p, \{C_{rj}\})$. Under the assumption that all users should transmit with the same probability, i.e., $\boldsymbol{p} = p\mathbf{1}$, system utility is a function of the unknown user number K, the common transmission probability p, and the real channel parameter set $\{C_{rj}\}$. For example, if we choose sum throughput of the system as the utility function,

$U(K, p, \{C_{rj}\})$ should be given by

$$U(K, p, \{C_{rj}\}) = K \sum_{j=0}^{K-1} \binom{K-1}{j} p^{j+1}(1-p)^{K-1-j} C_{rj}. \quad (4.12)$$

For most of the utility functions of interest, such as the sum throughput function given above, an asymptotically optimal solution should roughly keep the expected load of the channel at a constant [35][36]. Therefore, if p_K^* is the optimum transmission probability for user number K, we should have $\lim_{K\to\infty} Kp_K^* = x^*$, where $x^* > 0$ is obtained from the following asymptotic utility optimization.

$$x^* = \arg\max_x \lim_{K\to\infty} U\left(K, \frac{x}{K}, \{C_{rj}\}\right). \quad (4.13)$$

Note that virtual packet is not involved in the calculation of x^*.

Without knowing the actual user number K, we will show next that it is possible to set the system equilibrium at $\boldsymbol{p}^* = \min\{p_{\max}, \frac{x^*}{K+b}\}\mathbf{1}$, where $b \geq 1$ is a pre-determined design parameter, and $p_{\max}$ is defined as

$$p_{\max} = \min\left\{1, \frac{x^*}{J_{\epsilon_v} + b}\right\}. \quad (4.14)$$

We will also show later that, when the optimum transmission probabilities satisfies $\lim_{K\to\infty} Kp_K^* = x^*$, setting the equilibrium at $\boldsymbol{p}^* = \min\{p_{\max}, \frac{x^*}{K+b}\}\mathbf{1}$ is often not far from optimal even when the user number is small.

We intend to design a distributed MAC algorithm to maximize $U(K, p, \{C_{rj}\})$ by maintaining channel contention at a desired level. Let q_v denote the success probability of the virtual packet, measured at the receiver. We term q_v the "channel contention measure" because it is a measurement of the contention level of the system. Note that $q_v(\boldsymbol{p}, K)$ is a function of user number K and the transmission probability vector $\boldsymbol{p}$. Because $q_v(\boldsymbol{p}, K)$ equals the summation of a finite number of polynomial terms, $q_v(\boldsymbol{p}, K)$ should be Lipschitz continuous in $\boldsymbol{p}$ for any finite K. When the transmission probabilities of all users are equal, i.e., $\boldsymbol{p} = p\mathbf{1}$, we also write success probability of the virtual packet as

$$q_v(p, K) = \sum_{j=0}^{K} \binom{K}{j} p^j (1-p)^{K-j} C_{vj}, \quad (4.15)$$

where $\{C_{vj}\}$ is the set of virtual channel parameters. We assume that, upon obtaining q_v from the receiver, each user should first obtain a user number estimate, denoted by $\hat{K}$, and then set the corresponding transmission probability target at $\tilde{p} = \hat{p} = \min\left\{p_{\max}, \frac{x^*}{\hat{K}+b}\right\}$, where $x^* > 0$ is obtained from (4.13). We will show that, for any $x^* > 0$, without knowing K, one can always choose an appropriate b and design a distributed MAC algorithm to ensure system convergence to the desired equilibrium of $\boldsymbol{p}^* = \min\{p_{\max}, \frac{x^*}{K+b}\}\mathbf{1}$.

Convergence of the MAC algorithm to be introduced depends on two key monotonicity properties presented below. First, the following theorem shows that, given user number K, $q_v(p, K)$ is non-increasing in p.

Theorem 4.3. Under the assumption that $C_{vj} \geq C_{v(j+1)}$ for all $j \geq 0$, $q_v(p, K)$ given in (4.15) satisfies $\frac{\partial q_v(p,K)}{\partial p} \leq 0$. Furthermore, $\frac{\partial q_v(p,K)}{\partial p} < 0$ holds with strict inequality for $K > J_{\epsilon_v}$ and $p \in (0, 1)$.

The proof of Theorem 4.3 is given in Appendix C.1.

Given that $\hat{p} = \frac{x^*}{\hat{K}+b}$. Let $N = \lfloor \hat{K} \rfloor$ be the largest integer below $\hat{K}$. We define a continuous function $q_v^*(\hat{p})$, which can also be viewed as a function of $\hat{K}$, as follows

$$q_v^*(\hat{p}) = \frac{\hat{p} - p_{N+1}}{p_N - p_{N+1}} q_N(\hat{p}) + \frac{p_N - \hat{p}}{p_N - p_{N+1}} q_{N+1}(\hat{p}), \tag{4.16}$$

where $p_N = \min\left\{p_{\max}, \frac{x^*}{N+b}\right\}$, $p_{N+1} = \min\left\{p_{\max}, \frac{x^*}{N+1+b}\right\}$, and

$$q_N(p) = \sum_{j=0}^{N} \binom{N}{j} p^j (1-p)^{N-j} C_{vj},$$
$$q_{N+1}(p) = \sum_{j=0}^{N+1} \binom{N+1}{j} p^j (1-p)^{N+1-j} C_{vj}. \tag{4.17}$$

We term $q_v^*(\hat{p})$ the "theoretical channel contention measure" because it serves as a reference to the theoretical contention level of the system in the following sense. If user number of the system indeed equals $K = \hat{K}$ with $\hat{K} \geq J_{\epsilon_v}$, then $q_v^*(\hat{p})$ defined in (4.16) equals the actual channel contention measure at the desired equilibrium $\boldsymbol{p}^* = \frac{x^*}{K+b}\mathbf{1} = \frac{x^*}{\hat{K}+b}\mathbf{1}$ when all users transmit with the same probability $\hat{p} = \frac{x^*}{\hat{K}+b}$.

The following theorem gives the second monotonicity property, which shows that, given an arbitrary $x^* > 0$, with an appropriate choice of b, $q_v^*(\hat{p})$ is non-decreasing in $\hat{p}$.

Theorem 4.4. Let $x^* > 0$. Let $b \geq \max\{1, x^* - \gamma_{\epsilon_v}\}$, with γ_{ϵ_v} being defined as

$$\gamma_{\epsilon_v} = \min_{N, N \geq J_{\epsilon_v}, N \geq x^* - b} \frac{\sum_{j=0}^{N} j\binom{N}{j}\left(\frac{p_{N+1}}{1-p_{N+1}}\right)^j (C_{vj} - C_{v(j+1)})}{\sum_{j=0}^{N} \binom{N}{j}\left(\frac{p_{N+1}}{1-p_{N+1}}\right)^j (C_{vj} - C_{v(j+1)})}. \quad (4.18)$$

Then $q_v^*(\hat{p})$ defined in (4.16) is non-decreasing in $\hat{p}$. Furthermore, if $b > \max\{1, x^* - \gamma_{\epsilon_v}\}$ holds with strict inequality, then $q_v^*(\hat{p})$ is strictly increasing in $\hat{p}$ for $\hat{p} \in (0, p_{\max})$.

The proof of Theorem 4.4 is given in Appendix C.2.

Note that, if ϵ_v is small enough to satisfy $C_{vj} = C_{v(j+1)}$ for all $j < J_{\epsilon_v}$, then we have $\gamma_{\epsilon_v} = J_{\epsilon_v}$. Otherwise, $\gamma_{\epsilon_v} \leq J_{\epsilon_v}$ is generally true.

With the two key monotonicity properties, we are now ready to propose the following distributed MAC algorithm.

Distributed MAC algorithm:

1) Initialize the transmission probabilities of all users. Let the transmission probability of user k be denoted by p_k.

2) Let $Q \geq 1$ be a pre-determined integer. Over an interval of Q time slots, the receiver measures the success probability of a virtual packet, denoted by q_v, and feeds q_v back to all transmitters.

3) Upon receiving q_v, each user (transmitter) derives a transmission probability target $\hat{p}$ by solving the following equation

$$q_v^*(\hat{p}) = q_v. \quad (4.19)$$

If a $\hat{p} \in [0, p_{\max}]$ satisfying (4.19) cannot be found, each user sets $\hat{p}$ at $\hat{p} = p_{\max}$ when $q_v > q_v^*(p_{\max})$, or at $\hat{p} = 0$ when $q_v < q_v^*(0)$.

4) Each user, say user k, then updates its transmission probability by

$$p_k = (1 - \alpha)p_k + \alpha\hat{p}, \tag{4.20}$$

where $\alpha > 0$ is the step size parameter for user k.

5) The process is repeated from Step 2 till probabilities of all users converge.

Convergence of the proposed MAC algorithm is stated in the following theorem.

Theorem 4.5. Consider the K-user distributed multiple access network presented in this section. Given $x^* > 0$ and $\epsilon_v > 0$. Suppose that b is chosen to satisfy $b > \max\{1, x^* - \gamma_{\epsilon_v}\}$ where γ_{ϵ_v} is defined in (4.18). With the proposed MAC algorithm, associated ODE of the system given in (4.6) has a unique equilibrium at $\boldsymbol{p}^* = \min\{p_{\max}, \frac{x^*}{K+b}\}\mathbf{1}$. Furthermore, probability target $\hat{p}(\boldsymbol{p})$ as a function of the transmission probability vector $\boldsymbol{p}$ satisfies Conditions 1 and 2. Consequently, the distributed probability adaptation converges to the equilibrium in the sense specified in Theorems 4.1 and 4.2.

The proof of Theorem 4.5 is given in Appendix C.3.

In the above analysis, we did not pose any design constraint on the coding parameters of the virtual packet. Convergence of the distributed MAC algorithm is guaranteed so long as parameter b is chosen to satisfy $b > \max\{1, x^* - \gamma_{\epsilon_v}\}$, where $\gamma_{\epsilon_v} = J_{\epsilon_v}$ if ϵ_v is small enough. However, one should note that optimality of the MAC algorithm can be affected by the value of b and J_{ϵ_v}. Both b and J_{ϵ_v} are determined by the virtual channel parameter set $\{C_{vj}\}$ which is dependent on the virtual packet design. Assume that setting the transmission probabilities of all users at $p = \min\left\{1, \frac{x^*}{K}\right\}$ is an ideal choice for optimizing the chosen utility, which is indeed the case for sum throughput optimization over a collision channel [41][35]. Because the proposed MAC algorithm sets system equilibrium at $\boldsymbol{p}^* = \min\{p_{\max}, \frac{x^*}{K+b}\}\mathbf{1}$, there are two optimality concerns. On one hand, for a large user number K, it is a general preference that one should design the virtual packet to allow a relatively small value of b, which implies that γ_{ϵ_v} and J_{ϵ_v} should not be much smaller than x^*. On the other hand, for a small user number K, one should also design the virtual packet to support a J_{ϵ_v} value not much larger than x^*, so that $p_{\max} = \min\{1, \frac{x^*}{J_{\epsilon_v}+b}\}$ can be as close to 1 as possible. Considering both optimality concerns, a general guideline is to design coding parameters of the virtual packets such that J_{ϵ_v} (and γ_{ϵ_v}) should be slightly smaller than x^*, and b should be close to 1.

Example 4.1. Consider distributed multiple access networking over a multi-packet reception channel. We assume that all packets should be received successfully if the number of users transmitting in parallel is no more than $\hat{M} = 5$. Otherwise, the receiver should report collision to all users. The real channel parameter set $\{C_{rj}\}$ in this case is given by $C_{rj} = 1$ for $j < 5$ and $C_{rj} = 0$ for $j \geq 5$. We assume that users intend to optimize the symmetric throughput of the system. Consequently, if the user number equals K and all users transmit with an identical probability of p, system utility $U(K, p, \{C_{rj}\})$ is given by

$$U(K, p, \{C_{rj}\}) = \sum_{j=0}^{\min\{K-1,4\}} K \binom{K-1}{j} p^{j+1}(1-p)^{K-1-j}. \qquad (4.21)$$

Let $U_{\text{opt}}(K)$ be the optimal sum throughput of the system under the assumption that K is known.

$$U_{\text{opt}}(K) = \max_p U(K, p, \{C_{rj}\}). \qquad (4.22)$$

From the asymptotic utility optimization given in (4.13), we obtain $x^* = 3.64$. According to the design guideline presented above, we assume that a virtual packet should be equivalent to the combination of 2 real packets. Consequently, the virtual channel parameter set $\{C_{vj}\}$ is given by $C_{vj} = 1$ for $j < 4$ and $C_{vj} = 0$ for $j \geq 4$. Choose $\epsilon_v = 0.01$, which implies $\gamma_{\epsilon_v} = J_{\epsilon_v} = 3$, and hence we can set $b = 1.01 > \max\{1, x^* - \gamma_{\epsilon_v}\}$. We use $U^*(K)$ to denote the sum throughput of the system when transmission probabilities of all users are set at $p = \min\{p_{\max}, \frac{x^*}{K+b}\}$, where $p_{\max} = \min\{1, \frac{x^*}{J_{\epsilon_v}+b}\}$.

Figure 4.1 illustrated the two utility values, $U_{\text{opt}}(K)$ and $U^*(K)$, as functions of user number K. It can be seen that $U^*(K)$ is reasonably close to $U_{\text{opt}}(K)$ when user number K is not close to $\hat{M}$. Note that

Full text available at: http://dx.doi.org/10.1561/1300000063

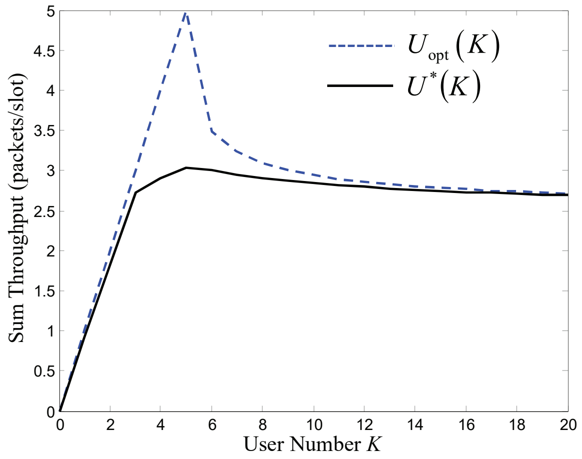

Figure 4.1: Sum throughput of the system as functions of the user number.

$U_{\text{opt}}(K)$ is not necessarily achievable without the knowledge of user number K.

Example 4.2. In this example, we consider distributed multiple access networking over a simple fading channel. Assume that the system has $K = 8$ users and one receiver. In each time slot, with a probability of 0.3, the channel can support no more than $\hat{M}_1 = 4$ parallel real packet transmissions, and with a probability of 0.7, the channel can support no more than $\hat{M}_2 = 6$ parallel real packet transmissions[3]. The real channel parameter set $\{C_{rj}\}$ in this case is given by $C_{rj} = 1$ for $j < 4$, $C_{rj} = 0.7$ for $4 \leq j < 6$, and $C_{rj} = 0$ for $j \geq 6$. Assume that users intend to optimize the symmetric system throughput weighted by a transmission energy cost of $E = 0.3$. If user number equals K and all users transmit with a probability of p, system utility $U(K, p, \{C_{rj}\})$ is given by

$$U(K, p, \{C_{rj}\}) = \sum_{j=0}^{K-1} K \binom{K-1}{j} p^{j+1}(1-p)^{K-1-j} C_{rj} - EKp. \quad (4.23)$$

[3]Such a channel can appear if there is an interfering user that transmits a packet with a probability of 0.3 in each time slot. One packet from the interfering user is equivalent to the combination of two packets from a regular user.

Consequently, x^* can be obtained from the asymptotic utility optimization (4.13) as $x^* = 3.29$.

Assume that a virtual packet should have the same coding parameters as those of a real packet. Consequently, the virtual channel parameter set $\{C_{vj}\}$ is identical to the real channel parameter set, i.e., $C_{vj} = C_{rj}$ for all $j \geq 0$. With $\epsilon_v = 0.01$, we have $\gamma_{\epsilon_v} = J_{\epsilon_v} = 3$ and hence we can set $b = 1.01 > \max\{1, x^* - \gamma_{\epsilon_v}\}$.

We initialize the transmission probabilities of all users at 0. In each time slot, a channel state flag is randomly generated to indicate whether the channel can support the parallel transmissions of no more than $\hat{M}_1 = 4$ packets or $\hat{M}_2 = 6$ packets. Each user also randomly determines whether a packet should be transmitted according to its own transmission probability parameter. Consequently, whether a real packet and the virtual packet can go through the channel successfully or not is determined using the corresponding channel model. We use the following exponential moving average approach to measure q_v[4], which is the success probability of the virtual packet. q_v is initialized at $q_v = 1$. In each time slot, if the virtual packet can be received successfully, an indicator variable I_v is set at $I_v = 1$. If the virtual packet reception fails, we set $I_v = 0$. Success probability of the virtual packet is then updated by $q_v = (1 - \frac{1}{300})q_v + \frac{1}{300}I_v$. The rest of the probability updates proceeds according to the distributed MAC algorithm introduced before with a constant step size of $\alpha = 0.05$. Convergence behavior in system utility is illustrated in Figure 4.2, where system utility is measured using the same exponential moving average approach as the measurement of q_v except that initial value of the utility is set at 0. Two reference values are shown in the figure. $U_{\text{opt}}(K)$ is the optimal utility as defined in (4.22), while $U^*(K)$ is the theoretical utility at the designed equilibrium.

[4]While this approach is different from the one proposed in the distributed MAC algorithm, simulations show that an exponential averaging measurement of q_v can often lead system to the designed equilibrium in a relatively small number of time slots.

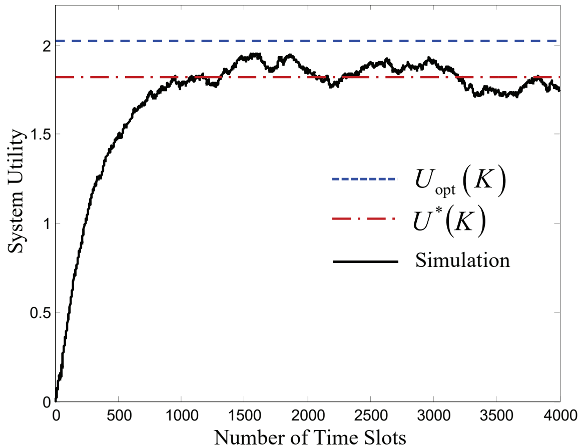

Figure 4.2: Convergence in sum utility of a $K = 8$ user multiple access network over a simple fading channel.

4.3 Multiple Transmission Options, Single Virtual Packet

In this section, we consider the case when each user is equipped with M transmission options plus an idling option. Each user, say user k, should maintain an M-length transmission probability vector $\boldsymbol{p}_k = p_k\boldsymbol{d}_k$, where p_k is the transmission probability and $\boldsymbol{d}_k$ is the transmission direction vector. Transmission probability vectors of all users are listed in an MK-length vector $\boldsymbol{P} = [\boldsymbol{p}_1^T, \dots, \boldsymbol{p}_K^T]^T$. As in the previous section, with a general channel model and without knowing the user number K, the objective is to design a distributed MAC algorithm to lead the system to a unique equilibrium that maximizes a chosen symmetric network utility.

Given the physical layer channel and the transmission options, we specify the link layer multiple access channel using two sets of channel parameter functions. Assume that all users have the same transmission direction vector $\boldsymbol{d}$. Define $\{C_{rij}(\boldsymbol{d})\}$ for $1 \le i \le M$ and $j \ge 0$ as the "real channel parameter function set", where $C_{rij}(\boldsymbol{d})$ is the conditional success probability of the ith real packet, should it be transmitted in parallel with other j real packets. Because each packet can be generated from a randomly chosen transmission option, $C_{rij}(\boldsymbol{d})$ is a function of $\boldsymbol{d}$. We still assume that there is a single virtual packet being transmitted in each time slot. Virtual packets being transmitted in different time slots are identical. Given coding parameters of the virtual packet, under the assumption that all users should have the same transmission probability vector $\boldsymbol{p} = p\boldsymbol{d}$, we define $\{C_{vj}(\boldsymbol{d})\}$ as the "virtual channel parameter function set", where $C_{vj}(\boldsymbol{d})$ is the success probability of the virtual packet, should it be transmitted in parallel with j real packets. We assume that $C_{vj}(\boldsymbol{d}) \geq C_{v(j+1)}(\boldsymbol{d})$ should hold for all $j \geq 0$ and for all $\boldsymbol{d}$. That is, under the same transmission direction vector $\boldsymbol{d}$, if the number of parallel real packet transmissions increases, the chance of a virtual packet getting through the channel should not increase. Let $\epsilon_v > 0$ be a pre-determined small constant. We define $J_{\epsilon_v}(\boldsymbol{d})$ as the minimum integer such that $C_{vJ_{\epsilon_v}}(\boldsymbol{d})$ is ϵ_v larger than $C_{v(J_{\epsilon_v}+1)}(\boldsymbol{d})$, i.e.,

$$J_{\epsilon_v}(\boldsymbol{d}) = \arg\min_{j} C_{vj}(\boldsymbol{d}) > C_{v(j+1)}(\boldsymbol{d}) + \epsilon_v. \tag{4.24}$$

Both the real and the virtual channel parameter function sets are assumed to be known at the transmitters and at the receiver.

We assume that users intend to maximize a symmetric network utility, denoted by $U(K, p\boldsymbol{d}, \{C_{rij}(\boldsymbol{d})\})$. Under the assumption that all users should have the same transmission probability vector $\boldsymbol{p}$, system utility is a function of the unknown user number K, the common transmission probability vector $\boldsymbol{p} = p\boldsymbol{d}$, and the real channel parameter function set $\{C_{rij}(\boldsymbol{d})\}$. For example, if the ith real packet has a communication rate of r_i (in units/time slot), and we choose sum throughput of the system as the utility function, then $U(K, \boldsymbol{p}, \{C_{rij}(\boldsymbol{d})\})$ should be given by

$$U(K, \boldsymbol{p}, \{C_{rij}(\boldsymbol{d})\}) = K \sum_{i=1}^{M} d_i r_i \sum_{j=0}^{K-1} \binom{K-1}{j} p^{j+1}(1-p)^{K-1-j} C_{rij}(\boldsymbol{d}) \tag{4.25}$$

We intend to design a distributed MAC algorithm to maximize $U(K, \boldsymbol{p}, \{C_{rij}(\boldsymbol{d})\})$ by maintaining channel contention at a desired level. Let q_v denote the success probability of the virtual packet. As before, we term q_v the "channel contention measure" because it is used to measure the contention level of the system. $q_v(\boldsymbol{P}, K)$ is a function of the user number K and the MK-length transmission probability vector $\boldsymbol{P}$. Because $q_v(\boldsymbol{P}, K)$ equals the summation of a finite number of polynomial terms, it should be Lipschitz continuous in $\boldsymbol{P}$ for any finite K. When all users have the same transmission probability vector $\boldsymbol{p} = p\boldsymbol{d}$, we also write q_v as

$$q_v(\boldsymbol{p}, K) = \sum_{j=0}^{K} \binom{K}{j} p^j (1-p)^{K-j} C_{vj}(\boldsymbol{d}). \tag{4.26}$$

Upon obtaining q_v from the receiver, we assume that each user should first derive a user number estimate $\hat{K}$ by comparing q_v with a "theoretical channel contention measure" $q_v^*(\hat{K})$, which is a function of $\hat{K}$. A user should then set its transmission probability vector target $\hat{\boldsymbol{p}}$ according to a designed theoretical vector parameter function $\boldsymbol{p}(\hat{K})$. To understand key properties that the $\boldsymbol{p}(\hat{K})$ function should possess, let us first take a look at the following simple example.

Example 4.3. Consider a distributed multiple access network with K homogeneous users. Assume that each user has two transmission options plus an idling option. The two transmission options are labeled as the "high rate" option and the "low rate" option, respectively. If users transmit with the low rate option only, then the channel can support the parallel transmissions of no more than 12 packets. Assume that a high rate packet is equivalent to the combination of 4 low rate packets. Therefore, if n_l low rate packets and n_h high rate packets are transmitted in parallel, the packets can be received successfully if and only if $n_l + 4n_h \leq 12$. Assume that users intend to optimize the sum throughput of the network, and transmission probability vectors of the users should be identical at the equilibrium. When all users have the same probability vector $\boldsymbol{p} = [p_h, p_l]^T$, system utility, denoted by $U(K, \boldsymbol{p})$ as a function of K and $\boldsymbol{p}$, is given by

$$U(K, \boldsymbol{p}) = \sum_{\substack{n_h \geq 0, n_l \geq 0, \\ n_h + n_l \leq K-1, \\ 4(n_h+1) + n_l \leq 12}} 4K \binom{K-1}{n_h, n_l} p_h^{n_h+1} p_l^{n_l} (1 - p_h - p_l)^{K-1-n_h-n_l}$$

$$+ \sum_{\substack{n_h \geq 0, n_l \geq 0, \\ n_h + n_l \leq K-1, \\ 4n_h + n_l + 1 \leq 12}} K \binom{K-1}{n_h, n_l} p_h^{n_h} p_l^{n_l+1} (1 - p_h - p_l)^{K-1-n_h-n_l}. \tag{4.27}$$

Given user number K, let $\boldsymbol{p}^* = \arg\max_{\boldsymbol{p}} U(K, \boldsymbol{p})$ be the optimal transmission probability vector. p_h^* and p_l^* as functions of user number K are illustrated in Figure 4.3. We can see that, if we write $\boldsymbol{p}^* = p^* \boldsymbol{d}^*$,

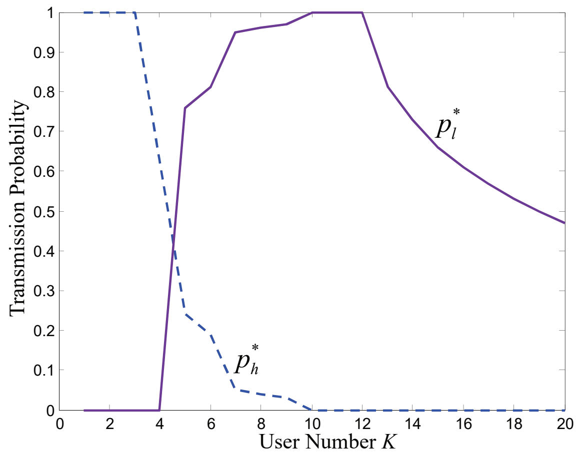

Figure 4.3: Optimal transmission probabilities of a K-user multiple access system with each user having two transmission options.

then we have $\boldsymbol{d}^* = [1, 0]^T$ for $K \leq 4$, and $\boldsymbol{d}^* = [0, 1]^T$ for $K \geq 10$. $\boldsymbol{d}^*$ transits from $[1, 0]^T$ to $[0, 1]^T$ in the region of $4 \leq K \leq 10$.

According to the above observation, we assume that the vector parameter function $\boldsymbol{p}(\hat{K})$ to be designed should possess the following properties termed the "Head and Tail Condition".

Condition 3. (Head and Tail) Let $\epsilon_v > 0$ be a pre-determined constant. Let J_{ϵ_v} be defined in (4.24). There exist two integer-valued constants $0 < \underline{K} \leq \overline{K}$, such that,

1) $\boldsymbol{d}(\hat{K}) = \boldsymbol{d}(\underline{K})$, for all $\hat{K} \leq \underline{K}$, $\underline{K} \geq J_{\epsilon_v}(\boldsymbol{d}(\underline{K}))$.

2) $\boldsymbol{d}(\hat{K}) = \boldsymbol{d}(\overline{K})$, for all $\hat{K} \geq \overline{K}$, $\overline{K} > J_{\epsilon_v}(\boldsymbol{d}(\overline{K}))$.

Condition 3 indicates that, when $\hat{K} \leq \underline{K}$ is small enough or when $\hat{K} \geq \overline{K}$ is large enough, $\boldsymbol{d}(\hat{K})$ should stop changing in $\hat{K}$. In these two regimes, the system with multiple transmission options becomes equivalent to a system with a single transmission option. The virtual channel parameter set of the equivalent system is given by $\{C_{vj}\} = \{C_{vj}(\boldsymbol{d})\}$. Calculation of the real channel parameter set of the equivalent system, on the other hand, depends on the chosen utility function. If the utility function is the sum throughput given in (4.25) for example, the equivalent real channel parameter set $\{C_{rj}\}$ should be obtained by $C_{rj} = \sum_{i=1}^{M} d_i r_i C_{rij}(\boldsymbol{d})$, for $j \geq 0$. We assume that core parameter functions of the MAC algorithm, i.e., the theoretical channel contention measure $q_v^*(\hat{K})$ and the probability target function $p(\hat{K})$, should be designed according to the guideline given in Section 4.2, for $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$. We choose not to repeat the corresponding details in this section.

Let us temporarily assume that the vector parameter function $\boldsymbol{p}(\hat{K})$ has been determined completely, not just for $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$, but also for $\underline{K} < \hat{K} < \overline{K}$. To present the distributed MAC algorithm, we need to define the theoretical channel contention measure $q_v^*(\hat{K})$ as follows. Let $N = \lfloor \hat{K} \rfloor$ be the largest integer below $\hat{K}$. For $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$, $q_v^*(\hat{K})$ is defined by

$$q_v^*(\hat{K}) = \frac{p(\hat{K}) - p(N+1)}{p(N) - p(N+1)} q_v(\boldsymbol{p}(\hat{K}), N) + \frac{p(N) - p(\hat{K})}{p(N) - p(N+1)} q_v(\boldsymbol{p}(\hat{K}), N+1), \tag{4.28}$$

which is consistent with (4.16). For $\underline{K} \leq \hat{K} \leq \overline{K}$, $q_v^*(\hat{K})$ is defined by

$$q_v^*(\hat{K}) = (N + 1 - \hat{K}) q_v(\boldsymbol{p}(\hat{K}), N) + (\hat{K} - N) q_v(\boldsymbol{p}(\hat{K}), N+1). \tag{4.29}$$

In other words, if $\hat{K}$ is integer-valued, $q_v^*(\hat{K}) = q_v(\boldsymbol{p}(\hat{K}), \hat{K})$ equals the channel contention measure when all users have the same transmission probability vector $\boldsymbol{p}(\hat{K})$ and the user number equals $K = \hat{K}$. If $\hat{K}$ is not integer-valued, on the other hand, $q_v^*(\hat{K})$ is a linear interpolation between $q_v(\boldsymbol{p}(\hat{K}), \lfloor \hat{K} \rfloor)$ and $q_v(\boldsymbol{p}(\hat{K}), \lfloor \hat{K} \rfloor + 1)$. Note that the interpolation approach used for $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$ is different from the one used for $\underline{K} \leq \hat{K} \leq \overline{K}$.

Next, we present the distributed MAC algorithm below.

Distributed MAC algorithm:

1) Initialize the transmission probability vectors of all users. Let the transmission probability vector of user k be denoted by $\boldsymbol{p}_k$.

2) Let $Q \geq 1$ be a pre-determined integer. Over an interval of Q time slots, the receiver measures (or estimates) the success probability of the virtual packet, denoted by q_v, and feeds q_v back to all transmitters.

3) Upon receiving q_v, each user (transmitter) derives a user number estimate $\hat{K}$ by solving the following equation

$$q_v^*(\hat{K}) = q_v. \tag{4.30}$$

If a $\hat{K}$ satisfying (4.30) cannot be found, user k sets $\hat{K} = J_{\epsilon_v}(\boldsymbol{d}(\underline{K}))$ if $q_v > q_v^*(J_{\epsilon_v}(\boldsymbol{d}(\underline{K})))$, and sets $\hat{K} = \infty$ otherwise.

4) Each user, say user k, then updates its transmission probability vector by

$$\boldsymbol{p}_k = (1-\alpha)\boldsymbol{p}_k + \alpha \boldsymbol{p}(\hat{K}), \tag{4.31}$$

where $\alpha > 0$ is the step size parameter for user k.

5) The process is repeated from Step 2 till transmission probability vectors of all users converge.

We intend to design the distributed MAC algorithm with the following convergence property. If $K \geq J_{\epsilon_v}(\boldsymbol{d}(\underline{K}))$, we intend to have $\hat{K} = K$ at the equilibrium, while if $K < J_{\epsilon_v}(\boldsymbol{d}(\underline{K}))$, we intend to have $\hat{K} = J_{\epsilon_v}(\boldsymbol{d}(\underline{K}))$ at the equilibrium. In order to ensure convergence of the proposed MAC algorithm, we require that the vector parameter function $\boldsymbol{p}(\hat{K})$ and the corresponding theoretical channel contention measure $q_v^*(\hat{K})$ should satisfy the following "Monotonicty and Gradient Condition" for $\underline{K} \leq \hat{K} \leq \overline{K}$.

Condition 4. (Monotonicity and Gradient) For $\underline{K} \leq \hat{K} \leq \overline{K}$,

1) $\boldsymbol{p}(\hat{K}) = p(\hat{K})\boldsymbol{d}(\hat{K})$ should be Lipschitz continuous in $\hat{K}$, i.e., there exists a constant $K_g > 0$, such that for all $\hat{K}_a, \hat{K}_b \in [\underline{K}, \overline{K}]$, we have
$$\|\boldsymbol{p}(\hat{K}_a) - \boldsymbol{p}(\hat{K}_b)\| \leq K_g|\hat{K}_a - \hat{K}_b|. \tag{4.32}$$

2) $q_v^*(\hat{K})$ should be continuous and be strictly decreasing in $\hat{K}$. There exists a positive constant $\epsilon_q > 0$, such that for all $\hat{K}_a, \hat{K}_b \in [\underline{K}, \overline{K}]$, we have
$$|q_v^*(\hat{K}_a) - q_v^*(\hat{K}_b)| \geq \epsilon_q|\hat{K}_a - \hat{K}_b|. \tag{4.33}$$

3) There exists a constant $\epsilon_v > 0$, such that $\hat{K} > J_{\epsilon_v}(\boldsymbol{d}(\hat{K}))$ should be satisfied for all $\hat{K} \in [\underline{K}, \overline{K}]$.

4) There exist constants $0 < \underline{p} < \overline{p} < 1$, such that $\underline{p} \leq p(\hat{K}) \leq \overline{p}$ should be satisfied for all $\hat{K} \in [\underline{K}, \overline{K}]$.

As a special case, it can be verified that, if one fix $\boldsymbol{d}(\hat{K}) = \boldsymbol{d}(\underline{K}) = \boldsymbol{d}(\overline{K})$ and design $p(\hat{K})$ according to the guideline given in Section 4.2, the resulting $\boldsymbol{p}(\hat{K})$ and $q_v^*(\hat{K})$ functions do satisfy the Monotonicity and Gradient Condition for $\underline{K} \leq \hat{K} \leq \overline{K}$.

Convergence of the distributed MAC algorithm is stated in the following theorem.

Theorem 4.6. Consider a multiple access system with K users adopting the proposed distributed MAC algorithm to update their transmission probability vectors. Under Condition 3, let $\boldsymbol{p}(\hat{K})$ and $q_v^*(\hat{K})$ be designed for $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$ according to the guideline given in Section 4.2. Let $\boldsymbol{p}(\hat{K})$ and $q_v^*(\hat{K})$ be designed to satisfy Condition 4 for $\underline{K} \leq \hat{K} \leq \overline{K}$. Then, associated ODE of the system given in (4.6) has a unique equilibrium at $\boldsymbol{P}^* = \mathbf{1} \otimes \boldsymbol{p}(K)$. The probability target $\hat{\boldsymbol{p}}(\boldsymbol{P})$ as a function of the transmission probability vector $\boldsymbol{P}$ satisfies Conditions 1 and 2. Consequently, the distributed probability vector adaptation converges to the unique equilibrium in the sense specified in Theorems 4.1 and 4.2.

Theorem 4.6 is implied by Theorem 4.8.

Note that, in the Monotonicity and Gradient Condition 4, while we still require $q_v^*(\hat{K})$ to be strictly decreasing in $\hat{K}$, being different from the single option case, we no longer require $q_v(\boldsymbol{p}(\hat{K}), K)$ to be strictly increasing in $\hat{K}$ for a given K. Also being different from the single option case where the $\boldsymbol{p}(\hat{K})$ function is completely specified in a closed form, Condition 4 did not explain how $\boldsymbol{p}(\hat{K})$ should be designed to satisfy the conditions.

Next, we will show that, so long as one can manually design $\boldsymbol{p}(\hat{K})$ for a set of chosen points with integer-valued $\hat{K}$ to satisfy a set of "Pinpoints Condition", then there is a simple approach to complete the $\boldsymbol{p}(\hat{K})$ function for $\underline{K} \leq \hat{K} \leq \overline{K}$ to satisfy Condition 4.

Condition 5. (Pinpoints) Let $\underline{K} = \hat{K}_0 < \hat{K}_1 < \ldots < \hat{K}_L = \overline{K}$ be a collection of integer-valued points. For $i = 1, \ldots, L$, and $0 \leq \lambda < 1$, define

$$\begin{aligned} \hat{K}_{i\lambda} &= (1-\lambda)\hat{K}_{i-1} + \lambda\hat{K}_i \\ \boldsymbol{d}_{i\lambda} &= (1-\lambda)\boldsymbol{d}(\hat{K}_{i-1}) + \lambda\boldsymbol{d}(\hat{K}_i) \\ q_{vi\lambda}^* &= (1-\lambda)q_v^*(\hat{K}_{i-1}) + \lambda q_v^*(\hat{K}_i). \end{aligned} \tag{4.34}$$

We have the following conditions.

1) There exists a positive constant $\epsilon_q > 0$, such that, for all $i = 1, \ldots, L$, $q_v^*(\hat{K}_{i-1}) - q_v^*(\hat{K}_i) \geq \epsilon_q$.

2) There exists a constant $\epsilon_v > 0$, such that for all $i = 1, \ldots, L$ and $0 \leq \lambda < 1$, $\hat{K}_{i\lambda} > J_{\epsilon_v}(\boldsymbol{d}_{i\lambda})$, where $J_{\epsilon_v}(\boldsymbol{d}_{i\lambda})$ is defined in (4.24).

3) There exist $0 < \underline{p} < \overline{p} < 1$, such that $\underline{p} \leq p(\hat{K}_i) \leq \overline{p}$ should be satisfied for all $i = 1, \ldots, L$.

3) Extend the definition of $q_v(\boldsymbol{p}, \hat{K})$ to non-integer-valued $\hat{K}$ as

$$\begin{aligned} q_v(\boldsymbol{p}, \hat{K}) = (\lfloor \hat{K} \rfloor + 1 - \hat{K}) q_v(\boldsymbol{p}, \lfloor \hat{K} \rfloor) \\ +(\hat{K} - \lfloor \hat{K} \rfloor) q_v(\boldsymbol{p}, \lfloor \hat{K} \rfloor + 1). \end{aligned} \tag{4.35}$$

The following inequality should be satisfied for all $i = 1, \ldots, L$ and for all $0 \leq \lambda < 1$.

$$q_v\left(\overline{p}\boldsymbol{d}_{i\lambda}, \hat{K}_{i\lambda}\right) \leq q_{vi\lambda}^* \leq q_v\left(\underline{p}\boldsymbol{d}_{i\lambda}, \hat{K}_{i\lambda}\right). \tag{4.36}$$

The next "Interpolation Approach" shows that, so long as $\boldsymbol{p}(\hat{K})$ is designed for the pinpoints, it is easy to complete the whole $\boldsymbol{p}(\hat{K})$ function for $\underline{K} \le \hat{K} \le \overline{K}$.

Interpolation Approach: Assume that $\boldsymbol{p}(\hat{K})$ is designed for a given set of pinpoints $\{\hat{K}_i\}$, $i = 0, \dots, L$, with $\hat{K}_0 = \underline{K} < \hat{K}_1, < \dots < \hat{K}_L = \overline{K}$, to satisfy Condition 5. For $i = 1, \dots, L$ and $0 \le \lambda < 1$, let $\hat{K}_{i\lambda}$, $\boldsymbol{d}_{i\lambda}$ and $q^*_{vi\lambda}$ be defined in (4.34). Let $q_v(\boldsymbol{p}, \hat{K})$ be defined in (4.35). We choose $p(\hat{K}_{i\lambda})$ to satisfy

$$q_v(p(\hat{K}_{i\lambda})\boldsymbol{d}(\hat{K}_{i\lambda}), \hat{K}_{i\lambda}) = q^*_{vi\lambda}. \tag{4.37}$$

Consequently, $\boldsymbol{p}(\hat{K}_{i\lambda})$ is designed as $\boldsymbol{p}(\hat{K}_{i\lambda}) = p(\hat{K}_{i\lambda})\boldsymbol{d}_{i\lambda}$.

Note that according to (4.36), a solution of $\underline{p} \le p(\hat{K}_{i\lambda}) \le \overline{p}$ satisfying (4.37) must exist. Effectiveness of the Interpolation Approach is stated in the following theorem.

Theorem 4.7. Assume that $\boldsymbol{p}(\hat{K})$ is designed for a set of $L+1$ pinpoints $\{\hat{K}_i\}$, $i = 0, \dots, L$, with $\hat{K}_0 = \underline{K} < \hat{K}_1, < \dots < \hat{K}_L = \overline{K}$, to satisfy Condition 5. After completing the function using the Interpolation Approach, $\boldsymbol{p}(\hat{K})$ and $q^*_v(\hat{K})$ functions satisfy the Monotonicity and Gradient Condition 4 for $\underline{K} \le \hat{K} \le \overline{K}$.

Theorem 4.7 is implied by Theorem 4.9.

Note that, in the single transmission option case discussed in Section 4.2, $p(\hat{K})$ is specified in a closed form with a small number of design parameters. Monotonicity property of $q^*_v(\hat{K})$ is proven theoretically. With multiple transmission options, however, such a direct-design approach faces a key challenge. Due to generality of the system model, when $\boldsymbol{d}(\hat{K})$ changes in $\hat{K}$ and consequently affects the channel parameters, it is often difficult to theoretically characterize its impact on the $q^*_v(\hat{K})$ function. Alternatively, we switched to a search-assisted approach to first manually design $\boldsymbol{p}(\hat{K})$ for a set of pinpoints to satisfy Condition 5, and then to use the Interpolation Approach to complete the $\boldsymbol{p}(\hat{K})$ function. Note that the Interpolation Approach only ensures convergence of the proposed MAC algorithm. It pays no attention to the optimality, in terms of the utility value, of the design outcome. Therefore, one often needs to carefully adjust the design of the pinpoints to direct the $\boldsymbol{p}(\hat{K})$ function toward a near optimal solution.

Example 4.4. Let us use the system introduced in Example 4.3 to illustrate the design procedure of the $\boldsymbol{p}(\hat{K})$ function. First, we consider the "Head" and the "Tail" regimes when $\hat{K}$ is either small or large in value. We will add subscript "H" to parameters of the "Head" regime, and add subscript "T" to parameters of the "Tail" regime. Without specifying the values of $\underline{K}$ and $\overline{K}$, we first determine the optimal transmission directions in these two regimes, $\boldsymbol{d}_H = [1, 0]^T$ and $\boldsymbol{d}_T = [0, 1]^T$. In other words, users should only use the high rate option in the "Head" regime and only use the low rate option in the "Tail" regime. In the "Head" regime, the channel can support the parallel transmissions of no more than 3 high rate packets. The real channel parameter set for the equivalent single option system is given by $\{C_{rj}\}_H$ with $C_{rj} = 1$ for $j \le 2$ and $C_{rj} = 0$ otherwise. By following the single option system design guideline, we get $x_H^* = \arg\max_x (x + x^2 + \frac{x^3}{2})e^{-x} = 2.27$. We design the virtual packet to be equivalent to a real high rate packet. Consequently, virtual channel parameter set for the equivalent single option system is given by $\{C_{vj}\}_H = \{C_{rj}\}_H$. With $\epsilon_v = 0.01$, we get $\gamma_{\epsilon_v H} = J_{\epsilon_v H} = 2$, and $b_H = 1.01$. In the "Tail" regime, on the other hand, the channel can support the parallel transmissions of no more than 12 low rate packets. The real channel parameter set for the equivalent system is given by $\{C_{rj}\}_T$ with $C_{rj} = 1$ for $j \le 11$ and $C_{rj} = 0$ otherwise. This yields $x_T^* = \arg\max_x \sum_{i=0}^{11} \frac{x^{i+1}}{i!} e^{-x} = 8.82$. Because we already chose the virtual packet to be equivalent to a high rate real packet, virtual channel parameter set in this case is given by $\{C_{vj}\}_T$ with $C_{vj} = 1$ for $j \le 8$ and $C_{vj} = 0$ otherwise. Therefore, with $\epsilon_v = 0.01$, we have $\gamma_{\epsilon_v T} = J_{\epsilon_v T} = 8$, and luckily, this supports $b_T = 1.01$.

To determine the values of $\underline{K}$ and $\overline{K}$, we compare the following two schemes. In the first "high rate option only" scheme, we fix $\boldsymbol{d}(\hat{K})$ at $[1, 0]^T$ for all $\hat{K}$, and set $p(\hat{K}) = \min\left\{p_{\max H}, \frac{x_H^*}{\hat{K} + b_H}\right\}$, where $p_{\max H} = \frac{x_H^*}{J_{\epsilon_v H} + b_H}$. In the second "low rate option only" scheme, we fix $\boldsymbol{d}(\hat{K})$ at $[0, 1]^T$ for all $\hat{K}$, and set $p(\hat{K}) = \min\left\{p_{\max T}, \frac{x_T^*}{\hat{K} + b_T}\right\}$, where $p_{\max T} = \frac{x_T^*}{J_{\epsilon_v T} + b_T}$. By comparing the utility values and the theoretical channel contention measures of the two schemes, we choose $\underline{K} = 4$ and $\overline{K} = 10$.

Now consider the design conditions for $\underline{K} \le \hat{K} \le \overline{K}$. For transmission directions $\boldsymbol{d}$ with $d_1 > 0$, we generally have $J_{\epsilon_v} = 2$. Therefore, so long as $\boldsymbol{d}(\hat{K})$ does not transit too quickly to $[0,1]^T$, the condition of $\hat{K} > J_{\epsilon_v}(\boldsymbol{d}(\hat{K}))$ should hold true. Consequently, only two other key conditions need to be satisfied. The first condition is that $q_v^*(\hat{K})$ of the selected pinpoints must be strictly decreasing in $\hat{K}$. The second condition is that $p(\hat{K})$ found in the Interpolation Approach should be bounded away from 0 and 1. From the optimal scheme, we can see that $\boldsymbol{d}(\hat{K})$ should transit toward $[0,1]^T$ faster than a linearly transition from $\underline{K}$ to $\overline{K}$.

With these considerations, we choose the following 4 pinpoints. At the edge of the "Head" and the "Tail" regimes, we have $\hat{K}_0 = \underline{K} = 4$ with $\boldsymbol{p}(4) = \frac{x_H^*}{\underline{K}+b_H}[1,0]^T$ and $\hat{K}_3 = \overline{K} = 10$ with $\boldsymbol{p}(10) = \frac{x_T^*}{\overline{K}+b_T}[0,1]^T$. For the other two pinpoints, $\hat{K}_1 = 5$ and $\hat{K}_2 = 6$, we set their transmission directions at the corresponding optimal transmission direction vectors, i.e., direction vectors extracted from the optimal $\boldsymbol{p}$ vectors that maximize the sum throughput at $K = 5$ and $K = 6$, respectively. Transmission probabilities of these two pinpoints are chosen such that the resulting $q_v^*(\hat{K})$ equals $\frac{\overline{K}-\hat{K}}{\overline{K}-\underline{K}} q_v^*(\underline{K}) + \frac{\hat{K}-\underline{K}}{\overline{K}-\underline{K}} q_v^*(\overline{K})$. The purpose of including $\hat{K}_1 = 5$ and $\hat{K}_2 = 6$ in the pinpoint set is to force $\boldsymbol{d}(\hat{K})$ to transit quickly toward $[0,1]^T$. The rest of the $\boldsymbol{p}(\hat{K})$ function is completed using the Interpolation Approach. Theoretical channel contention measure $q_v^*(\hat{K})$ of the designed system is illustrated in Figure 4.4 as a function of the user number.

In Figure 4.5, we illustrated the theoretical sum system throughput as a function of user number K for the following four different scenarios: optimum $\boldsymbol{p}(K)$, designed $\boldsymbol{p}(K)$, $\boldsymbol{p}(K)$ from the high rate option only scheme, and $\boldsymbol{p}(K)$ from the low rate option only scheme. Assume that the high rate option only scheme should be reasonably good in the "Head" regime while the low rate option only scheme should be reasonably good in the "Tail" only regime. It can be seen that the designed $\boldsymbol{p}(\hat{K})$ function can help to bridge the two simple schemes and to efficiently exploit the benefit of the two transmission options. Note that the optimal utility illustrated in Figure 4.5 is not necessarily achievable without the knowledge of user number K.

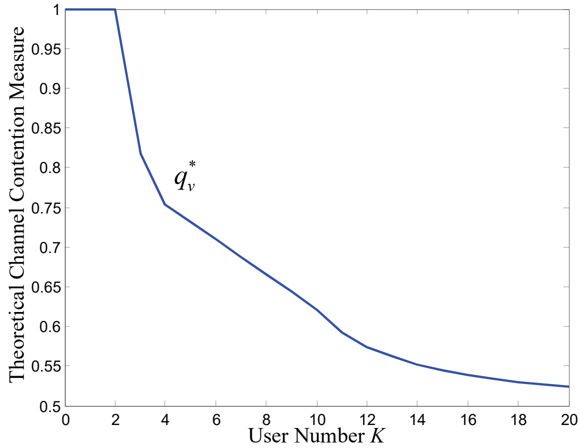

Figure 4.4: Theoretical channel contention measure q_v^* as a function for the user number.

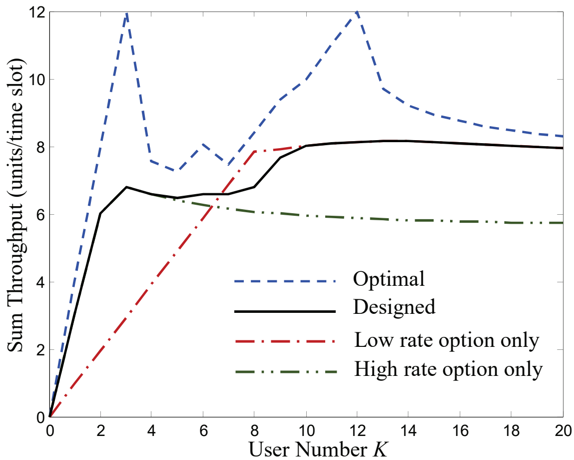

Figure 4.5: Sum throughput of the system as functions of the user number.

Example 4.5. Following Example 4.4, assume that the system has 8 users initially. Transmission probabilities of the users are initialized at $[0, 0]^T$. In each time slot, according to its own transmission probability vector, each user randomly determines whether to transmit a packet or not, and if the answer is positive, which option should be used. The receiver uses the following exponential moving average approach to measure q_v. q_v is initialized at $q_v = 1$. In each time slot, an indicator variable $I_v \in \{0, 1\}$ is used to represent the success/failure status of the virtual packet reception. q_v is then updated as $q_v = (1 - \frac{1}{300})q_v + \frac{1}{300}I_v$, and is fed back to the transmitters at the end of each time slot. Each user then adapts its transmission probability vector according to the proposed MAC algorithm with a constant step size of $\alpha = 0.05$.

We assume that the system experiences three stages. At the beginning in Stage one, the system has 8 users. The system enters Stage two at the 3001th time slot, when 6 more users enter into the system with their transmission probability vectors initialized at $[0, 0]^T$. Then at the 6001th time slot, the system enters Stage three when 8 users exit the system. Convergence behavior in sum throughput of the system is illustrated in Figure 4.6, together with the corresponding optimal throughput and the theoretical throughput at the equilibrium being provided as references. In Figure 4.7, we also illustrates entries of the

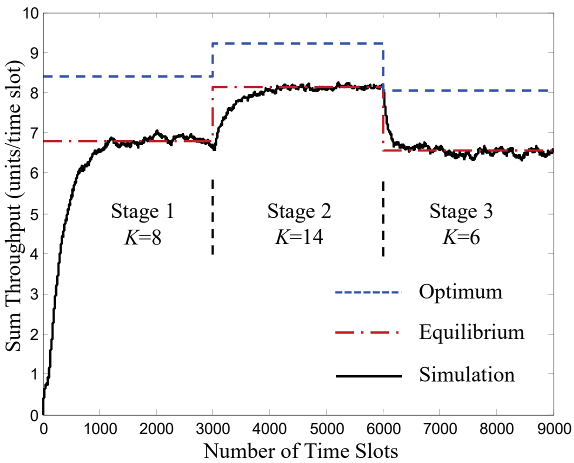

Figure 4.6: Convergence in sum throughput of the system. User number changes from 8 to 14 and then to 6 in three stages.

transmission probability vector target calculated by the users together with the corresponding theoretical values being provided as references.

Note that the simulated probability values presented in the figure are calculated using the same exponential averaging approach explained above.

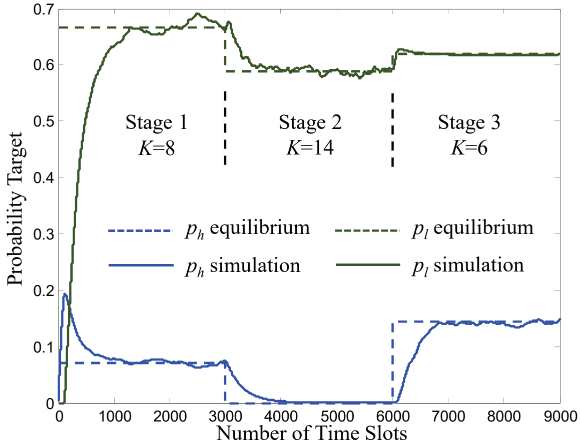

Figure 4.7: Entries of the transmission probability vector taget and their corresponding theoretical values.

Figures 4.6 and 4.7 demonstrated that, with the proposed MAC algorithm and the designed $\boldsymbol{p}(\hat{K})$, $q_v^*(\hat{K})$ functions, users have the capability to quickly adapt to the changes of stages and adjust their transmission probability vectors to the new equilibrium.

According to the Head and Tail Condition 3, the system degrades to an equivalent single option system when $K \leq \underline{K}$ or $K \geq \overline{K}$. In Example 4.4, while $\boldsymbol{d}(\underline{K}) \neq \boldsymbol{d}(\overline{K})$, we found a virtual packet design that supports $b_H = 1.01$ in the "Head" regime and $b_T = 1.01$ in the "Tail" regime. One may think that such a lucky result should not always happen for a general system. Surprisingly, according to our observations, in most of the problems of interest, even if one may not be able to get the ideal result of $b_H = b_T \approx 1$, a single virtual packet can often be designed to support close to ideal performance in both the "Head" and the "Tail" regimes.

4.4 Multiple Transmission Options, Multiple Virtual Packets

Following the system model introduced in Section 4.3, in this section, we assume that there is a set of V virtual packets being transmitted in each time slot. We present such a model extension not only because it enables extra flexibility in system design, but also because obtaining the corresponding technical results is nontrivial.

We assume that virtual packet sets being transmitted in different time slots should be identical. The link layer multiple access channel is still specified using two sets of channel parameter functions. Definition of the "real channel parameter function set" $\{C_{rij}(\boldsymbol{d})\}$ remains the same as in Section 4.3. Given coding parameters of the virtual packets, under the assumption that all users should have the same transmission direction vector $\boldsymbol{d}$, we define $\{C_{vij}(\boldsymbol{d})\}$ as the "virtual channel parameter function set", where $C_{vij}(\boldsymbol{d})$ is the success probability of the ith virtual packet, should it be transmitted in parallel with other j real packets. We assume that $C_{vij}(\boldsymbol{d}) \geq C_{vi(j+1)}(\boldsymbol{d})$ should hold for all $1 \leq i \leq V$, for all $j \geq 0$ and for all $\boldsymbol{d}$. Both the real and the virtual channel parameter function sets are assumed to be known at the transmitters and at the receiver.

Let $\boldsymbol{q}_v$ denote the vector of success probabilities of the virtual packets. We term $\boldsymbol{q}_v(\boldsymbol{P}, K)$ the "channel contention measure vector", which is a function of the user number K and the MK-length transmission probability vector $\boldsymbol{P}$. Because $\boldsymbol{q}_v(\boldsymbol{P}, K)$ equals the summation of a finite number of polynomial terms, it should be Lipschitz continuous in $\boldsymbol{P}$ for any finite K. When all users have the same transmission probability vector $\boldsymbol{p} = p\boldsymbol{d}$, success probability of the ith virtual packet, denoted by q_{vi} for $1 \leq i \leq V$, is also written as

$$q_{vi}(\boldsymbol{p}, K) = \sum_{j=0}^{K} \binom{K}{j} p^j (1-p)^{K-j} C_{vij}(\boldsymbol{d}). \tag{4.38}$$

Let us introduce a new design parameter $\boldsymbol{w}$, termed the "observation vector". $\boldsymbol{w}$ is a V-length vector whose entries satisfy $w_i \geq 0$ for all $1 \leq i \leq V$ and $\sum_{i=1}^{V} w_i = 1$. Upon receiving $\boldsymbol{q}_v$ from the receiver, users calculate the "channel contention measure" q_v as $q_v = \boldsymbol{w}^T \boldsymbol{q}_v$. Note that, given observation vector $\boldsymbol{w}$ and the common transmission direction vector $\boldsymbol{d}$, the system is equivalent to one with a single transmission option. Calculation of the equivalent real channel parameter set $\{C_{rj}\}$, which is dependent on the chosen system utility $U(K, p\boldsymbol{d}, \{C_{rij}(\boldsymbol{d})\})$, remains the same as explained in Section 4.3. The equivalent virtual channel parameter set $\{C_{vj}\}$ is given by $C_{vj} = \sum_{i=1}^{V} w_i C_{vij}(\boldsymbol{d})$. Let $\epsilon_v > 0$ be a pre-determined small constant. We define $J_{\epsilon_v}(\boldsymbol{w}, \boldsymbol{d})$ as the minimum integer such that $C_{vJ_{\epsilon_v}}$ is ϵ_v larger than $C_{v(J_{\epsilon_v}+1)}$, i.e.,

$$J_{\epsilon_v}(\boldsymbol{w}, \boldsymbol{d}) = \arg\min_j \sum_{i=1}^{V} w_i C_{vij}(\boldsymbol{d}) > \sum_{i=1}^{V} w_i C_{vi(j+1)}(\boldsymbol{d}) + \epsilon_v. \quad (4.39)$$

We intend to design two vector parameter functions $\boldsymbol{w}(\hat{K})$ and $\boldsymbol{p}(\hat{K})$, both are functions of the user number estimate $\hat{K}$. As will be explained later, upon receiving $\boldsymbol{q}_v$, a user will use $\boldsymbol{w}(\hat{K})$ and $\boldsymbol{p}(\hat{K})$ to jointly determine a user number estimate $\hat{K}$, and then to set the transmission probability vector target at $\hat{\boldsymbol{p}} = \boldsymbol{p}(\hat{K})$. As in Section 4.3, we assume that the vector parameter functions $\boldsymbol{w}(\hat{K})$ and $\boldsymbol{p}(\hat{K})$ to be designed should satisfy the following "Head and Tail Condition".

Condition 6. (Head and Tail) Let $\epsilon_v > 0$ be a pre-determined constant. Let J_{ϵ_v} be defined in (4.39). There exist two integer-valued constants $0 < \underline{K} \leq \overline{K}$, such that,

1) $\boldsymbol{d}(\hat{K}) = \boldsymbol{d}(\underline{K})$ and $\boldsymbol{w}(\hat{K}) = \boldsymbol{w}(\underline{K})$, for all $\hat{K} \leq \underline{K}$ with $\underline{K} \geq J_{\epsilon_v}(\boldsymbol{w}(\underline{K}), \boldsymbol{d}(\underline{K}))$.

2) $\boldsymbol{d}(\hat{K}) = \boldsymbol{d}(\overline{K})$ and $\boldsymbol{w}(\hat{K}) = \boldsymbol{w}(\overline{K})$ for all $\hat{K} \geq \overline{K}$ with $\overline{K} > J_{\epsilon_v}(\boldsymbol{w}(\overline{K}), \boldsymbol{d}(\overline{K}))$.

Condition 6 indicates that, when $\hat{K} \leq \underline{K}$ or $\hat{K} \geq \overline{K}$, $\boldsymbol{w}(\hat{K})$ and $\boldsymbol{d}(\hat{K})$ should stop changing in $\hat{K}$. As explained in Section 4.3, in these two regimes, the system with multiple transmission options become equivalent to one with a single transmission option. We assume that core parameter functions of the MAC algorithm, i.e., the theoretical channel contention measure $q_v^*(\hat{K})$ and the probability target function $p(\hat{K})$, should be designed according to the guideline given in Section 4.2.

Let us temporarily assume that the vector parameter functions $\boldsymbol{w}(\hat{K})$ and $\boldsymbol{p}(\hat{K})$ have been completely determined for all $\hat{K}$ values. To present the distributed MAC algorithm, we need to define the theoretical channel contention measure $q_v^*(\hat{K})$ as follows. Let $N = \lfloor \hat{K} \rfloor$ be the largest integer below $\hat{K}$. For $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$, $q_v^*(\hat{K})$ is defined by

$$q_v^*(\hat{K}) = \frac{p(\hat{K}) - p(N+1)}{p(N) - p(N+1)} \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), N) + \frac{p(N) - p(\hat{K})}{p(N) - p(N+1)} \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), N+1), \tag{4.40}$$

which is consistent with (4.16). For $\underline{K} \leq \hat{K} \leq \overline{K}$, $q_v^*(\hat{K})$ is defined by

$$q_v^*(\hat{K}) = (N + 1 - \hat{K}) \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), N) + (\hat{K} - N) \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), N+1). \tag{4.41}$$

Next, we present the distributed MAC algorithm below.

Distributed MAC algorithm:

1) Initialize the transmission probability vectors of all users. Let the transmission probability vector of user k be denoted by $\boldsymbol{p}_k$.

2) Let $Q \geq 1$ be a pre-determined integer. Over an interval of Q time slots, the receiver measures (or estimates) the success probabilities of all virtual packets, denoted by $\boldsymbol{q}_v$, and feeds $\boldsymbol{q}_v$ back to all transmitters.

3) Upon receiving $\boldsymbol{q}_v$, each user (transmitter) derives a user number estimate $\hat{K}$ by solving the following equation

$$q_v^*(\hat{K}) = \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v. \tag{4.42}$$

   If a $\hat{K}$ satisfying (4.42) cannot be found, then each user should set $\hat{K} = J_{\epsilon_v}(\boldsymbol{w}(\underline{K}), \boldsymbol{d}(\underline{K}))$ if $\boldsymbol{w}(\underline{K})^T \boldsymbol{q}_v > q_v^*(J_{\epsilon_v}(\boldsymbol{w}(\underline{K}), \boldsymbol{d}(\underline{K})))$, each user should set $\hat{K} = \infty$ otherwise.

4) Each user, say user k, then updates its transmission probability vector by

$$\boldsymbol{p}_k = (1 - \alpha)\boldsymbol{p}_k + \alpha \boldsymbol{p}(\hat{K}), \tag{4.43}$$

   where $\alpha > 0$ is the step size parameter for user k.

5) The process is repeated from Step 2 till transmission probability vectors of all users converge.

We intend to design the distributed MAC algorithm to possess a unique equilibrium at $\hat{K} = \max\{K, J_{\epsilon_v}(\boldsymbol{w}(\underline{K}), \boldsymbol{d}(\underline{K}))\}$. In order to ensure convergence, we require the virtual packet design, the vector parameter function $\boldsymbol{w}(\hat{K})$ and the $q_v^*(\hat{K})$ function should satisfy the following "Majorization Condition".

Condition 7. (Majorization)

1) Channel contention measure vector $\boldsymbol{q}_v$ should satisfy $q_{vi} \leq q_{vj}$ for all $i < j$. This condition can be met, for example, by assuming that all virtual packets should be encoded using random block codes with the same input distribution, but with rate parameters satisfying $r_{vi} \geq r_{vj}$ for all $1 \leq i < j \leq V$.

2) Observation vector function $\boldsymbol{w}(\hat{K})$ should be Lipschitz continuous in $\hat{K}$. There exists a constant $\epsilon_w > 0$, such that $\boldsymbol{w}(\hat{K})$ and $q_v^*(\hat{K})$ should satisfy the following majorization constraint

$$\sum_{i=j}^{V} w_i(\hat{K}_1) - \sum_{i=j}^{V} w_i(\hat{K}_2) \leq (1 - \epsilon_w)[q_v^*(\hat{K}_1) - q_v^*(\hat{K}_2)],$$
$$\forall j \leq V, \text{ and } \forall \hat{K}_1 \leq \hat{K}_2. \qquad (4.44)$$

Note that, because we generally require $q_v^*(\hat{K})$ to be monotonically decreasing in $\hat{K}$, (4.44) can be replaced by the following stronger condition

$$\sum_{i=j}^{V} w_i(\hat{K}_1) \leq \sum_{i=j}^{V} w_i(\hat{K}_2), \qquad \forall j \leq V, \text{ and } \forall \hat{K}_1 \leq \hat{K}_2, \qquad (4.45)$$

which does not involve the evaluation of $q_v^*(\hat{K})$.

Furthermore, we also require that the vector parameter functions $\boldsymbol{p}(\hat{K})$, $\boldsymbol{w}(\hat{K})$, and the corresponding theoretical channel contention measure $q_v^*(\hat{K})$ should satisfy the following "Monotonicty and Gradient Condition" for $\underline{K} \leq \hat{K} \leq \overline{K}$.

Condition 8. (Monotonicity and Gradient) For $\underline{K} \leq \hat{K} \leq \overline{K}$,

1) $\boldsymbol{p}(\hat{K}) = p(\hat{K})\boldsymbol{d}(\hat{K})$ should be Lipschitz continuous in $\hat{K}$, i.e., there exists a constant $K_g > 0$ to satisfy (4.32).

2) $q_v^*(\hat{K})$ should be continuous and be strictly decreasing in $\hat{K}$. There exists a positive constant $\epsilon_q > 0$ to satisfy (4.33).

3) There exists a constant $\epsilon_v > 0$, such that $\hat{K} > J_{\epsilon_v}(\boldsymbol{w}(\hat{K}), \boldsymbol{d}(\hat{K}))$ should be satisfied for all $\hat{K} \in [\underline{K}, \overline{K}]$.

4) There exist constants $0 < \underline{p} < \overline{p} < 1$, such that $\underline{p} \leq p(\hat{K}) \leq \overline{p}$ should be satisfied for all $\hat{K} \in [\underline{K}, \overline{K}]$.

Convergence of the distributed MAC algorithm is stated in the following theorem.

Theorem 4.8. Consider a multiple access system with K users adopting the proposed distributed MAC algorithm to update their transmission probability vectors. Under Condition 6, let $p(\hat{K})$ and $q_v^*(\hat{K})$ be designed for $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$ according to the guideline given in Section 4.2. Let virtual packets, $\boldsymbol{w}(\hat{K})$, $\boldsymbol{p}(\hat{K})$, and $q_v^*(\hat{K})$ be designed to satisfy Conditions 7 and 8. Then, associated ODE of the system given in (4.6) has a unique equilibrium at $\boldsymbol{P}^* = \mathbf{1} \otimes \boldsymbol{p}(K)$. The target probability vector $\hat{\boldsymbol{p}}(\boldsymbol{P})$ as a function of $\boldsymbol{P}$ satisfies Conditions 1 and 2. Consequently, the distributed probability vector adaptation converges to the equilibrium in the sense specified in Theorems 4.1 and 4.2.

The proof of Theorem 4.8 is given in Appendix C.4.

Next, we will show that, so long as one can manually design $\boldsymbol{w}(\hat{K})$ and $\boldsymbol{p}(\hat{K})$ for a set of chosen points with integer-valued $\hat{K}$ to satisfy the following "Pinpoints Condition", then $\boldsymbol{p}(\hat{K})$ can be completed using the "Interpolation Approach" to satisfy Condition 8.

Condition 9. (Pinpoints) Let $\underline{K} = \hat{K}_0 < \hat{K}_1 < \ldots < \hat{K}_L = \overline{K}$ be a set of integer-valued points. For $i = 1, \ldots, L$, and $0 \leq \lambda < 1$, define

$$\begin{aligned}
\hat{K}_{i\lambda} &= (1-\lambda)\hat{K}_{i-1} + \lambda \hat{K}_i \\
\boldsymbol{w}_{i\lambda} &= (1-\lambda)\boldsymbol{w}(\hat{K}_{i-1}) + \lambda \boldsymbol{w}(\hat{K}_i) \\
\boldsymbol{d}_{i\lambda} &= (1-\lambda)\boldsymbol{d}(\hat{K}_{i-1}) + \lambda \boldsymbol{d}(\hat{K}_i)
\end{aligned}$$

$$q^*_{vi\lambda} = (1-\lambda)q^*_v(\hat{K}_{i-1}) + \lambda q^*_v(\hat{K}_i). \tag{4.46}$$

We have the following conditions.

1) There exists a positive constant $\epsilon_q > 0$, such that, for all $i = 1,\ldots,L$, $q^*_v(\hat{K}_{i-1}) - q^*_v(\hat{K}_i) \geq \epsilon_q$.

2) There exists a constant $\epsilon_v > 0$, such that for all $i = 1,\ldots,L$ and $0 \leq \lambda < 1$, $\hat{K}_{i\lambda} > J_{\epsilon_v}(\boldsymbol{w}_{i\lambda}, \boldsymbol{d}_{i\lambda})$, where $J_{\epsilon_v}(\boldsymbol{w}_{i\lambda}, \boldsymbol{d}_{i\lambda})$ is defined in (4.39).

3) There exist $0 < \underline{p} < \overline{p} < 1$, such that $\underline{p} \leq p(\hat{K}_i) \leq \overline{p}$ should be satisfied for all $i = 1,\ldots,L$.

3) Extend the definition of $\boldsymbol{q}_v(\boldsymbol{p}, \hat{K})$ to non-integer-valued $\hat{K}$ as

$$\begin{aligned}\boldsymbol{q}_v(\boldsymbol{p}, \hat{K}) = (\lfloor \hat{K} \rfloor + 1 - \hat{K})\boldsymbol{q}_v(\boldsymbol{p}, \lfloor \hat{K} \rfloor) \\ +(\hat{K} - \lfloor \hat{K} \rfloor)\boldsymbol{q}_v(\boldsymbol{p}, \lfloor \hat{K} \rfloor + 1).\end{aligned} \tag{4.47}$$

The following inequality should be satisfied for all $i = 1,\ldots,L$ and for all $0 \leq \lambda < 1$.

$$\boldsymbol{w}^T_{i\lambda}\boldsymbol{q}_v\left(\overline{p}\boldsymbol{d}_{i\lambda}, \hat{K}_{i\lambda}\right) \leq q^*_{vi\lambda} \leq \boldsymbol{w}^T_{i\lambda}\boldsymbol{q}_v\left(\underline{p}\boldsymbol{d}_{i\lambda}, \hat{K}_{i\lambda}\right). \tag{4.48}$$

Interpolation Approach: Assume that $\boldsymbol{p}(\hat{K})$ is designed for a given set of pinpoints $\{\hat{K}_i\}$, $i = 0,\ldots,L$, with $\hat{K}_0 = \underline{K} < \hat{K}_1, < \ldots < \hat{K}_L = \overline{K}$, to satisfy Conditions 7 and 9. For $i = 1,\ldots,L$ and $0 \leq \lambda < 1$, let $\hat{K}_{i\lambda}$, $\boldsymbol{w}_{i\lambda}$, $\boldsymbol{d}_{i\lambda}$ and $q^*_{vi\lambda}$ be defined in (4.46). Let $\boldsymbol{q}_v(\boldsymbol{p}, \hat{K})$ be defined in (4.47). We choose $p(\hat{K}_{i\lambda})$ to satisfy

$$\boldsymbol{w}^T_{i\lambda}\boldsymbol{q}_v(p(\hat{K}_{i\lambda})\boldsymbol{d}(\hat{K}_{i\lambda}), \hat{K}_{i\lambda}) = q^*_{vi\lambda}. \tag{4.49}$$

Consequently, $\boldsymbol{p}(\hat{K}_{i\lambda})$ is designed as $\boldsymbol{p}(\hat{K}_{i\lambda}) = p(\hat{K}_{i\lambda})\boldsymbol{d}_{i\lambda}$.

Note that existence of a solution with $\underline{p} \leq p(\hat{K}_{i\lambda}) \leq \overline{p}$ to (4.49) is guaranteed by (4.48). The following theorem shows that, combined with the Interpolation Approach, the Majorization Condition 7 and the Pinpoints Condition 9 imply the Monotonicity and Gradient Condition 8.

Theorem 4.9. Assume that $\boldsymbol{w}(\hat{K})$ and $\boldsymbol{p}(\hat{K})$ are designed for a set of $L + 1$ pinpoints $\{\hat{K}_i\}$, $i = 0, \ldots, L$, with $\hat{K}_0 = \underline{K} < \hat{K}_1, < \ldots < \hat{K}_L = \overline{K}$. Let Conditions 7 and 9 be met for the pinpoints. After completing the function using the Interpolation Approach, $\boldsymbol{w}(\hat{K})$, $\boldsymbol{p}(\hat{K})$, and $q_v^*(\hat{K})$ functions satisfy both the Majorization Condition 7 and the Monotonicity and Gradient Condition 8 for $\underline{K} \leq \hat{K} \leq \overline{K}$.

The proof of Theorem 4.9 is given in Appendix C.5.

Let us consider the case when all virtual packets are encoded using random block codes with the same input distribution but with their rate parameters satisfying $r_{v1} > r_{v2} > \ldots > r_{vV}$. The Majorization Condition enables the system to shift observation weights, as $\hat{K}$ increases, either toward the low rate virtual packets (by using the simplified condition given in (4.45)) or toward the high rate virtual packets (by using (4.44)). Such flexibility can help to move the system equilibrium closer to its ideal value. Nevertheless, according to our observations, for most of the cases of interest, performance gain obtained by varying the observation vector in $\hat{K}$ is often minor compared with a carefully optimized system design either using a single virtual packet or using multiple virtual packets but with a constant observation vector.

5

Summary

As opposed to their relative abundance in wireline communication, transmission power and channel bandwidth are the most precious resources to wireless systems. Due to this reason, architectural problems faced by wireless part of the networks are fundamentally different from those of the wireline networks. While both "efficiency" and "modularity" are critical to the operation of a wireless network, classical information theory and network theory each only emphasizes one side of the core concerns.

This monograph addressed the architectural inefficiency problem at the bottom two layers of a wireless network. Due to the "single-hop cellular structure" widely seen in wireless systems, we considered a multiple access communication environment. Technical results of the monograph are centered around a proposal to mitigate the corresponding architectural inefficiency by enhancing the interface between the physical and the data-link layers. In the enhanced physical-link layer interface, a link layer user can be equipped with multiple transmission options corresponding to different communication settings such as different power, rate, and antenna beam combinations. Navigation through these transmission options gives data link layer users the capability of exploiting advanced wireless properties such as power and rate adjustments in communication adaptation. This is particularly important to the efficiency of distributed networking systems when communication optimization cannot be done fully at the physical layer.

5.1 Key Technical Results

At the physical layer, we considered multiple access communication over a discrete-time memoryless channel, and focused on channel coding within one time slot. Assume that each transmitter is equipped with an ensemble of channel codes, and can *arbitrarily* choose a code from the ensemble to encode its message. Coding decisions are not shared among the transmitters or with the receiver. With channel and code ensemble information, the receiver either decodes the messages of interest or reports collision. The receiver should choose an "operation region" in the space of the coding vectors. The receiver is expected to decode the messages reliably if the vector of coding decisions of the transmitters happens to locate inside the operation region, and to report collision reliably if the coding decision vector happens to locate outside the operation region.

With a key extension on the definition of "communication error" from its classical meaning of correct message decoding to the revised meaning of providing the expected outcome (which may include correct message decoding and collision report), we were able to define the asymptotic achievability of an operation region by taking the codeword length (or time slot length) to infinity. The maximum achievable region was defined as the distributed capacity of the discrete-time memoryless channel and was shown to coincide with the classical Shannon channel capacity in a sense explained in the monograph. We obtained distributed capacities of the channel under various communication models including joint message decoding, single user message decoding, communication with the existence of an interfering user, and communication over a compound channel.

When the codeword length is finite, we derived achievable performance bounds by considering the tradeoff between the choice of the operation region (including the extended consideration of operation region and operation margin) and the generalized error performance of the system. The performance bounds are valid for any codeword length. They become tight when the codeword length is large in value.

At the data link layer, we considered the problem of distributed medium access control with each link layer user having multiple transmission options. To simplify the problem, we assumed that the multiple access network should consist of an unknown number of homogeneous users each having a saturated message queue. In each time slot, a user should randomly determine its transmission activity according to an associated probability vector whose entries correspond to the available transmission and idling options. The receiver is assumed to make a judgement on the success/failure status of one or multiple presumed virtual packet transmissions. Success probabilities of the virtual packets are estimated and fed back to the users. We developed a distributed medium access control framework that can adapt the transmission probability vectors of the users to a unique equilibrium that is not far from optimal with respect to a chosen symmetric network utility. We showed that the framework can incorporate a general link layer channel model and can be used to optimize a general symmetric network utility. Simulation results are provided to demonstrate the optimality and convergence performance of the proposed distributed medium access control algorithm.

A key step that supported the correct functioning of the proposed medium access control algorithm is the development of a proper contention measure of the multiple access channel. The contention measure, being common to all users, enabled the users to estimate the user number in the system and to adjust their transmission schemes to a common target.

5.2 Research Timeline

The first investigation of the extended channel coding theory was due to Luo and Emphremides [57][58]. The original motivation was to develop channel coding theorems to model bursty message arrivals (which eventually was translated into an arbitrary choice of communication rate) and to also incorporate collision detection at the receiver. The basic idea of extending the communication error definition from correct message decoding to providing the expected communication outcome was proposed. Coding theorems also covered the cases of joint message decoding as well as single user message decoding. Unfortunately, [57][58] tried to characterize fundamental performance limitation of the system using a rate region and also to give a transmitter the freedom of generating each single codeword using a specific input distribution. These efforts unnecessarily complicated the theorems and their corresponding proofs.

In [85], Wang and Luo considered the case of finite codeword length, and derived a set of achievable bounds for the tradeoff between the choice of the operation region and the error performance of the systems. Error performance measure was originally chosen in [85] as the maximum probability of all types of error events. In [86], the results were extended to distributed communication over a compound channel. The idea of using an interfering user (or a virtual user) to model the impact of the compound channel was proposed. These results were jointly presented later in [87]. While derivations of the error probability bounds given in [85][86][87] are more or less standard, revising the exponential bound used in [28] turned out to be challenging. The corresponding approach present in [87] was inspired by the key reference of [64], which is still listed as an "unpublished" paper.

In [55], Luo presented a set of relatively clean results for the achievable regions and the achievable error performance bounds. Achievable region was first defined in the space of the coding vectors. Recognizing the existence of multiple types of communication error events, a generalized error performance measure was introduced to support the assignment of different weights to different error types. For the purpose of complexity reduction, a two step decoding approach was proposed to first detect the coding vector and then to decode the messages of interest.

Finally, in this monograph, we defined the achievable region as a property of the communication channel, without depending it on the choice of the code ensembles of the transmitters. Such a revision enabled the converse proof of the maximum achievable region, which led to the definition and the characterization of the distributed channel capacity. From the detailed derivations of the error performance bounds given in [55], it can be seen that there are numerous ways to obtain different achievable error performance bounds of the system. In this monograph, we improved the bounds given in [55] in a way that we believe is the most relevant to the potential applications of these results.

Compared with the physical layer research, where analytical approaches were extended from the existing tools developed in the classical channel coding literature, research on the data link layer problems faced a different set of challenges due to the lack of similar prior investigation in the literature. In [80], Tang, Zhao, and Luo formulated the distributed medium access control problem from a game theoretic perspective and analyzed the condition for the existence of a unique Nash equilibrium. In [79], we formulated the problem using a stochastic approximation model. A virtual packet was introduced to establish a channel contention measure that is common to all users. This approach guaranteed the equality of the transmission probabilities of all users at any equilibrium, and consequently simplified the theoretical analysis of the system. However, due to the difficult of proving a key monotonicity property of the theoretical channel contention measure, we were only able to develop the distributed medium access control algorithm for a class of multi-packet reception channels with the assumption that each user should only have a single transmission option. The framework was later extended to a system with a general link layer channel model and a general symmetric network utility. Finally, with the relaxation of a key monotoniticy property, and the change from a direct-design approach to a search-assisted approach, we were able to extend the proposed medium access control framework to distributed link layer networking with each user being equipped with multiple transmission options. A majorization condition was introduced to enable the use of multiple virtual packets in the system model.

Acknowledgements

This work was supported by the National Science Foundation under Grants CCF-1420608 and CNS-1618960. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendices

A

Proofs of Theorems in Section 2

A.1 Proof of Theorem 2.2

The basic idea of the achievability proof follows Shannon's typical sequence arguments, and the converse proof is based on Fano's inequality. Because both approaches are more or less standard, we will go through the proof briefly with explanations only on details specific to the distributed communication model.

A.1.1 The Achievability Part

Consider an arbitrary random code ensemble $\mathcal{G}$ with a finite cardinality $|\mathcal{G}| = M$. Given codeword length N, code ensemble corresponding to the codeword length is denoted by $\mathcal{G}^{(N)}$. We will first show the existence of a decoding algorithm to achieve $\lim_{N\to\infty} P_e^{(N)}(g) = 0$ for all $g \in \mathcal{G}$.

Let $\epsilon > 0$ be a small constant. We define the set $A_\epsilon^{(N)}(g)$ of jointly typical channel input and output sequences $(X^{(N)}, Y^{(N)})$ for code index g as follows

$$A_\epsilon^{(N)}(g) = \Big\{(X^{(N)}, Y^{(N)}) \in \mathcal{X}^{(N)} \times \mathcal{Y}^{(N)}\Big|$$
$$\left|-\frac{1}{N}\log p_g(X^{(N)}) - H_g(X)\right| < \epsilon,$$

$$\left|-\frac{1}{N}\log p_g(Y^{(N)}) - H_g(Y)\right| < \epsilon,$$
$$\left|-\frac{1}{N}\log p_g(X^{(N)}, Y^{(N)}) - H_g(X,Y)\right| < \epsilon,$$
$$\left.\left|-\frac{1}{N}\log p_{Y^{(N)}}(Y^{(N)}) - H_g(Y)\right| < \epsilon\right\}, \tag{A.1}$$

where $H_g()$ and $p_g()$ denote respectively the entropy function and the probability function with respect to joint distribution $P_{XY} = P_{Y|X}P_{gX}$, i.e., with respect to input distribution P_{gX}, while $p_{Y^{(N)}}()$ denotes the probability function with respect to the empirical distribution obtained from sequence $Y^{(N)}$.

Note that the definition of typical sequence set given in (A.1) is stronger than the classical Shannon's definition [71], but it is weaker than the definition of strong typical sequence set given in [8] and [21]. The reason we added the condition involving the empirical distributions in (A.1) is that, it enables us to bridge probability bounds of a channel input and output sequence pair that appears in multiple typical sequence sets corresponding to different code indices.

Let $X_g^{(N)}(w)$ denote the randomly generated codeword corresponding to message w and code index g. Upon observing channel output sequence $Y^{(N)}$, the receiver searches for all message and code index pairs $(\hat{w}, \hat{g})$ such that $\hat{g} \in C_d$ and $(X_{\hat{g}}^{(N)}(\hat{w}), Y^{(N)}) \in A_\epsilon^{(N)}(\hat{g})$. If one and only one message and code index pair is found, the receiver outputs $(\hat{w}, \hat{g})$ as its decoding outcome. Otherwise, the receiver reports collision.

Denote the actual message and code index pair by (w, g). It is easy to see that

$$\lim_{N\to\infty} Pr\{(X_g^{(N)}(w), Y^{(N)}) \in A_\epsilon^{(N)}(g)\} = 1. \tag{A.2}$$

We define an event "Error$_1$" as follows,

$$\text{Error}_1 : \exists(\tilde{w}, \tilde{g}) \text{ with } \tilde{g} \in C_d, (\tilde{w}, \tilde{g}) \neq (w, g),$$
$$\text{and } (X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}) \in A_\epsilon^{(N)}(\tilde{g}). \tag{A.3}$$

Depending on the value of g, there are two types of communication errors. First, if $g \in C_d$, according to (A.2), the receiver will find (w, g) as one of the decoding candidates. The receiver should output (w, g) unless event "Error_1" happens, in which case the receiver will be confused and will report collision. Second, if $g \notin C_d$, because we do not regard correct decoding as an error event, communication error in this case refers to the event that the receiver outputs an erroneous decoding estimate $(\tilde{w}, \tilde{g}) \neq (w, g)$. A necessary condition for the error event is again the event of "Error_1".

Next, we will show that $\lim_{N\to\infty} Pr\{\text{Error}_1\} = 0$, irrespective of whether $g \in C_d$ or not. Let us assume that event "Error_1" does happen. Because $\tilde{g} \in C_d$, according to the definition of C_d, we have

$$r_{\tilde{g}} < I_{\tilde{g}}(X; Y). \tag{A.4}$$

Also because $(\tilde{w}, \tilde{g}) \neq (w, g)$, $X_{\tilde{g}}^{(N)}(\tilde{w})$ and the channel output sequence are generated independently. Therefore, according to the standard typical sequence argument [71][20], for each message and code index pair $(\tilde{w}, \tilde{g})$ and for large enough N, the probability that the receiver will find $(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}) \in A_{\epsilon}^{(N)}(\tilde{g})$ should satisfy the following upper bound

$$\begin{aligned}
&\frac{1}{N} \log[Pr\{(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}) \in A_{\epsilon}^{(N)}(\tilde{g})\}] \\
&\leq (H_{\tilde{g}}(X, Y) + \epsilon) - (H_{\tilde{g}}(X) - \epsilon) - (H_g(Y) - \epsilon) \\
&\leq (H_{\tilde{g}}(X, Y) + \epsilon) - (H_{\tilde{g}}(X) - \epsilon) - (H_{\tilde{g}}(Y) - 3\epsilon) \\
&= -I_{\tilde{g}}(X; Y) + 5\epsilon,
\end{aligned} \tag{A.5}$$

where the last term in the first inequality is due to the fact that the channel sequence is generated using code index g, and the probability bound is translated in the second inequality to one corresponding to code index $\tilde{g}$ because both entropy bounds are associated with the same output sequence $Y^{(N)}$ with the same empirical distribution.

Therefore, according to the union bound, for large enough N, probability of event "Error_1" can be upper bounded by

$$\begin{aligned}
Pr\{\text{Error}_1\} &\leq \sum_{\tilde{g} \in C_d} \sum_{\tilde{w}} Pr\{(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}) \in A_{\epsilon}^{(N)}(\tilde{g})\} \\
&\leq \sum_{\tilde{g} \in C_d} \exp\left[N\left(r_{\tilde{g}} - I_{\tilde{g}}(X; Y) + 5\epsilon\right)\right].
\end{aligned} \tag{A.6}$$

This implies that we can find a small enough ϵ to ensure

$$\lim_{N\to\infty} Pr\{\text{Error}_1\} = 0 \tag{A.7}$$

Achievability of C_d then follows.

A.1.2 The Converse Part

Consider an operation region R that is asymptotically achievable. Let $g \in R$ be an arbitrary point in R. We will show that $g \in C_d^c$ must be true.

Let (w, g) be the actual message vector and code index vector pair. We assume that g is known to the receiver. We will also skip g in the subscription to simplify the notations. Because $g \in R$, according to the definition of achievable region, the receiver should output $\hat{w} = w$ with an asymptotic probability of one. Or in other words, $\lim_{N\to\infty} P_e^{(N)} = 0$.

Let $\epsilon > 0$ be an arbitrary small constant. According to Fano's inequality [23][20], for large enough N, we have

$$\begin{aligned}
&r_g \leq \frac{1}{N}H(w) + \epsilon \\
&= \frac{1}{N}H\left(w|Y^{(N)}\right) + \frac{1}{N}I\left(w;Y^{(N)}\right) + \epsilon \\
&< \frac{1}{N} + \frac{1}{N}P_e^{(N)}\log(|w|) + \frac{1}{N}I\left(w;Y^{(N)}\right) + \epsilon \\
&\leq \frac{1}{N}I\left(w;Y^{(N)}\right) + 2\epsilon \\
&\leq \frac{1}{N}I\left(X^{(N)};Y^{(N)}\right) + 2\epsilon \\
&= I\left(X;Y\right) + 2\epsilon,
\end{aligned} \tag{A.8}$$

where the inequality in the second line from the last is obtained due to the data processing inequality [20], and equality in the last line is due to the fact that the channel is memoryless and codeword symbols are generated independently. By taking N to infinity and taking ϵ to 0, (A.8) implies that $r_g \leq I\left(X;Y\right)$. Therefore we must have $g \in C_d^c$. Because $g \in R$ is chosen arbitrarily, $R \subseteq C_d^c$ must be true.

A.2 Proof of Theorem 2.6

This proof is purposely organized in the same structure as the proof of Theorem 2.2, so that readers can use the proof of Theorem 2.2 as a reference to help navigating through the more notation-complicated derivations presented in this proof.

A.2.1 The Achievability Part

Consider an arbitrary random code ensemble $\boldsymbol{\mathcal{G}}$ with a finite cardinality $|\mathcal{G}_k| = M$, $\forall k \in \{1, \dots, K\}$. Given codeword length N, code ensemble corresponding to the codeword length is denoted by $\boldsymbol{\mathcal{G}}^{(N)}$. We will first show the existence of a decoding algorithm to achieve $\lim_{N\to\infty} P_e^{(N)}(\boldsymbol{g}) = 0$ for all $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$.

Let $\epsilon > 0$ be a small constant. We define the set $A_\epsilon^{(N)}(\boldsymbol{g})$ of typical sequences $(\boldsymbol{X}^{(N)}, Y^{(N)})$ for code index $\boldsymbol{g}$ as follows

$$
\begin{aligned}
&A_\epsilon^{(N)}(\boldsymbol{g}) = \Big\{(\boldsymbol{X}^{(N)}, Y^{(N)}) \in \boldsymbol{\mathcal{X}}^{(N)} \times \mathcal{Y}^{(N)} \Big| \forall S \subseteq \{1, \dots, K\}, \\
&\left| -\frac{1}{N} \log p_{\boldsymbol{g}}(\boldsymbol{X}_S^{(N)}) - H_{\boldsymbol{g}}(\boldsymbol{X}_S) \right| < \epsilon, \\
&\left| -\frac{1}{N} \log p_{\boldsymbol{g}}(\boldsymbol{X}_S^{(N)}, Y^{(N)}) - H_{\boldsymbol{g}}(\boldsymbol{X}_S, Y) \right| < \epsilon, \\
&\left| -\frac{1}{N} \log p_{(\boldsymbol{X}_S^{(N)}, Y^{(N)})}(\boldsymbol{X}_S^{(N)}, Y^{(N)}) - H_{\boldsymbol{g}}(\boldsymbol{X}_S, Y) \right| < \epsilon \Big\}, \qquad \text{(A.9)}
\end{aligned}
$$

where $H_{\boldsymbol{g}}()$ and $p_{\boldsymbol{g}}()$ denote respectively the entropy function and the probability function with respect to joint distribution $P_{\boldsymbol{X}Y}$ under input distribution $P_{\boldsymbol{g}\boldsymbol{X}}$, i.e., $P_{\boldsymbol{X}Y} = P_{Y|\boldsymbol{X}} \prod_{k=1}^{K} P_{g_k X_k}$, while $p_{(\boldsymbol{X}_S^{(N)}, Y^{(N)})}()$ denotes the probability function with respect to the empirical distribution obtained from vector sequence pair $(\boldsymbol{X}_S^{(N)}, Y^{(N)})$.

As explained in the proof of Theorem 2.2, the definition of typical sequence set given in (A.9) is stronger than the classical Shannon's definition [71], but it is weaker than the definition of strong typical sequence set given in [8] and [21]. The reason we added the condition involving the empirical distributions in (A.9) is that, it enables us to bridge probability bounds of a partial channel input and output sequence pair that appears in multiple typical sequence sets corresponding to different code index vectors.

Let $\boldsymbol{X}_{\boldsymbol{g}}^{(N)}(\boldsymbol{w})$ denote the randomly generated codeword vector sequence corresponding to message vector $\boldsymbol{w}$ and code index vector $\boldsymbol{g}$. Upon observing channel output sequence $Y^{(N)}$, the receiver searches for all message vector and code index vector pairs $(\hat{\boldsymbol{w}}, \hat{\boldsymbol{g}})$ such that $\hat{\boldsymbol{g}} \in \boldsymbol{C}_{d1}$ and $(\boldsymbol{X}_{\hat{\boldsymbol{g}}}^{(N)}(\hat{\boldsymbol{w}}), Y^{(N)}) \in A_{\epsilon}^{(N)}(\hat{\boldsymbol{g}})$. If only one message vector and code index vector pair is found, or, if multiple vector pairs are found but all vector pairs correspond to the same message and code index pair $(\hat{w}_1, \hat{g}_1)$ for user 1, then the receiver outputs $(\hat{w}_1, \hat{g}_1)$ as its decoding outcome. Otherwise, the receiver reports collision for user 1.

Denote the actual message vector and code index vector pair by $(\boldsymbol{w}, \boldsymbol{g})$. It is easy to see that

$$\lim_{N\to\infty} Pr\{(\boldsymbol{X}_{\boldsymbol{g}}^{(N)}(\boldsymbol{w}), Y^{(N)}) \in A_{\epsilon}^{(N)}(\boldsymbol{g})\} = 1. \tag{A.10}$$

We define an event "Error$_1$" as follows,

$$\begin{aligned} \text{Error}_1 : \exists(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}}) \text{ with } \tilde{\boldsymbol{g}} \in \boldsymbol{C}_{d1}, (\tilde{w}_1, \tilde{g}_1) \neq (w_1, g_1), \\ \text{and } (\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{(N)}(\tilde{\boldsymbol{w}}), Y^{(N)}) \in A_{\epsilon}^{(N)}(\tilde{\boldsymbol{g}}). \end{aligned} \tag{A.11}$$

Depending on the value of $\boldsymbol{g}$, there are two types of communication errors. First, if $\boldsymbol{g} \in \boldsymbol{C}_{d1}$, according to (A.10), the receiver will find (w_1, g_1) as one of the decoding candidates. The receiver should output (w_1, g_1) unless event "Error$_1$" happens, in which case the receiver will be confused and will report collision. Second, if $\boldsymbol{g} \notin \boldsymbol{C}_{d1}$, because we do not regard correct decoding for user 1 as an error event, communication error in this case refers to the event that the receiver outputs an erroneous decoding estimate $(\tilde{w}_1, \tilde{g}_1) \neq (w_1, g_1)$. A necessary condition for the error event is again the event of "Error$_1$".

Next, we show that $\lim_{N\to\infty} Pr\{\text{Error}_1\} = 0$, irrespective of whether $\boldsymbol{g} \in \boldsymbol{C}_{d1}$ or not. To do that, we first need to define another event "Error$_{1S}$", for an arbitrary user subset $S \subseteq \{1, \dots, K\}$ with $1 \in S$, as follows

$$\begin{aligned} \text{Error}_{1S} : \exists(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}}) \text{ with } \tilde{\boldsymbol{g}} \in \boldsymbol{C}_{d1}, (\tilde{\boldsymbol{w}}_{\bar{S}}, \tilde{\boldsymbol{g}}_{\bar{S}}) = (\boldsymbol{w}_{\bar{S}}, \boldsymbol{g}_{\bar{S}}), \\ \text{and with } (\tilde{w}_k, \tilde{g}_k) \neq (w_k, g_k), \forall k \in S, \end{aligned}$$

$$\text{such that } (\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}}(\tilde{\boldsymbol{w}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}}). \tag{A.12}$$

Event "Error$_{1S}$" represents the situation that the receiver finds a decoding candidate $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}})$. $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}})$ is identical to the actual vector pair $(\boldsymbol{w}, \boldsymbol{g})$ on the part corresponding to users not in S, while $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}})$ differs from $(\boldsymbol{w}, \boldsymbol{g})$ on the part corresponding to users in S. We will show that $\lim_{N\to\infty} Pr\{\text{Error}_{1S}\} = 0$, $\forall S$, which implies $\lim_{N\to\infty} Pr\{\text{Error}_1\} = 0$.

Let us assume that event "Error$_{1S}$" does happen for a given $S \subseteq \{1, \dots, K\}$ with $1 \in S$. Because $\tilde{\boldsymbol{g}} \in \boldsymbol{C}_{d1}$, according to the definition of $\boldsymbol{C}_{d1}$, we can find a user subset $\tilde{S} \subseteq S$ with $1 \in \tilde{S}$, such that

$$\sum_{k\in\tilde{S}} r_{\tilde{g}_k} < I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}; Y|\boldsymbol{X}_{\bar{S}}). \tag{A.13}$$

Consider part of the vectors corresponding to user subset $\tilde{S} \cup \bar{S}$. We say that a partial codeword vector sequence $\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}})$ is jointly typical with $Y^{(N)}$ with respect to code index vector $\tilde{\boldsymbol{g}}$, denoted by $(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})$, if there exists a codeword vector sequence $\boldsymbol{X}^{(N)}$ with $\boldsymbol{X}^{(N)}_{\tilde{S}\cup\bar{S}} = \boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}})$, such that $(\boldsymbol{X}^{(N)}, Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})$.

Because $(\tilde{w}_k, \tilde{g}_k) \neq (w_k, g_k)$ for all $k \in \tilde{S} \subseteq S$, $\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}})$ and the channel output sequence are generated independently. Therefore, according to the standard typical sequence argument [71][20], for each partial message vector and code index vector pair $(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}, \tilde{\boldsymbol{g}})$ and for large enough N, the probability of finding $(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})$ should satisfy the following upper bound

$$\begin{aligned}
&\frac{1}{N}\log[Pr\{(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})\}] \\
&\leq (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}, \boldsymbol{X}_{\bar{S}}, Y) + \epsilon) - (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}) - \epsilon) - (H_{\boldsymbol{g}}(\boldsymbol{X}_{\bar{S}}, Y) - \epsilon) \\
&\leq (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}, \boldsymbol{X}_{\bar{S}}, Y) + \epsilon) - (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}) - \epsilon) - (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\bar{S}}, Y) - 3\epsilon) \\
&= -I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}; \boldsymbol{X}_{\bar{S}}, Y) + 5\epsilon \\
&= -I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}; Y|\boldsymbol{X}_{\bar{S}}) + 5\epsilon,
\end{aligned} \tag{A.14}$$

where the last term in the first inequality is due to the fact that the corresponding channel input and output sequences are generated using code index vector $\boldsymbol{g}$, and the probability bound is translated in the second inequality to one corresponding to code index vector $\tilde{\boldsymbol{g}}$ because both entropy bounds are associated with the same partial vector sequence $(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_{\bar{S}}}(\boldsymbol{w}_{\bar{S}}), Y^{(N)})$ with the same empirical distribution.

By assumption, we have $(\tilde{\boldsymbol{w}}_{\bar{S}}, \tilde{\boldsymbol{g}}_{\bar{S}}) = (\boldsymbol{w}_{\bar{S}}, \boldsymbol{g}_{\bar{S}})$. Therefore, according to the union bound, for large enough N, $Pr\{\text{Error}_{1S}\}$ can be upper bounded by

$$Pr\{\text{Error}_{1S}\} \le \sum_{\tilde{\boldsymbol{g}}_S, \tilde{\boldsymbol{g}} \in \boldsymbol{C}_{d1}} \sum_{\tilde{\boldsymbol{w}}_{\tilde{S}}} Pr\{(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})\}$$

$$\le \sum_{\tilde{\boldsymbol{g}}_S, \tilde{\boldsymbol{g}} \in \boldsymbol{C}_{d1}} \exp\left[N\left(\sum_{k\in\tilde{S}} r_{\tilde{g}_k} - I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}}; \boldsymbol{X}_{\bar{S}}, Y) + 5\epsilon\right)\right]. \quad \text{(A.15)}$$

This implies that we can find a small enough ϵ to ensure

$$\lim_{N\to\infty} Pr\{\text{Error}_{1S}\} = 0. \quad \text{(A.16)}$$

Consequently, $\lim_{N\to\infty} Pr\{\text{Error}_1\} = 0$. Achievability of $\boldsymbol{C}_{d1}$ for user 1 then follows.

A.2.2 The Converse Part

Consider an operation region $\boldsymbol{R}_1$ that is asymptotically achievable for user 1. Let $\boldsymbol{g} \in \boldsymbol{R}_1$ be an arbitrary code index vector in $\boldsymbol{R}_1$. We will show that $\boldsymbol{g} \in \boldsymbol{C}^c_{d1}$ must be true.

Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code index vector pair. We assume $\boldsymbol{g}$ is known to the receiver. We will skip $\boldsymbol{g}$ in the subscription to simplify the notations. Because $\boldsymbol{g} \in \boldsymbol{R}_1$, according to the definition of achievable region, the receiver should output $\hat{w}_1 = w_1$ with an asymptotic probability of one.

Let $S \subseteq \{1, \dots, K\}$ be an arbitrary user subset with $1 \in S$. Assume that codewords of users in $\bar{S}$ are known at the receiver. Because the message of user 1 is correctly decoded with an asymptotic probability of one, there must exist a user subset $\tilde{S} \subseteq S$ with $1 \in \tilde{S}$ such that, with an asymptotic probability of one, the receiver can jointly decode the messages of users in $\tilde{S}$ by regarding the input symbols from users in $S \setminus \tilde{S}$ as interference. Denote the probability that the receiver is not able to recover the message of all users in $\tilde{S}$ as $P^{(N)}_e(\tilde{S})$, we have $\lim_{N\to\infty} P^{(N)}_e(\tilde{S}) = 0$.

Let $\epsilon > 0$ be an arbitrary small constant. According to Fano's inequality, for large enough N, we have

$$\begin{aligned}
&\sum_{k\in\tilde{S}} r_{g_k} \leq \frac{1}{N}H(\boldsymbol{w}_{\tilde{S}}) + \epsilon = \frac{1}{N}H\left(\boldsymbol{w}_{\tilde{S}}|\boldsymbol{X}_{\bar{S}}^{(N)}(\boldsymbol{w}_{\bar{S}})\right) + \epsilon \\
&= \frac{1}{N}H\left(\boldsymbol{w}_{\tilde{S}}|\boldsymbol{X}_{\bar{S}}^{(N)}(\boldsymbol{w}_{\bar{S}}), Y^{(N)}\right) + \frac{1}{N}I\left(\boldsymbol{w}_{\tilde{S}}; Y^{(N)}|\boldsymbol{X}_{\bar{S}}^{(N)}(\boldsymbol{w}_{\bar{S}})\right) + \epsilon \\
&< \frac{1}{N} + \frac{1}{N}P_e^{(N)}(\tilde{S})\log(|\boldsymbol{w}_{\tilde{S}}|) + \frac{1}{N}I\left(\boldsymbol{w}_{\tilde{S}}; Y^{(N)}|\boldsymbol{X}_{\bar{S}}^{(N)}(\boldsymbol{w}_{\bar{S}})\right) + \epsilon \\
&\leq \frac{1}{N}I\left(\boldsymbol{w}_{\tilde{S}}; Y^{(N)}|\boldsymbol{X}_{\bar{S}}^{(N)}(\boldsymbol{w}_{\bar{S}})\right) + 2\epsilon \\
&\leq \frac{1}{N}I\left(\boldsymbol{X}_{\tilde{S}}^{(N)}; Y^{(N)}|\boldsymbol{X}_{\bar{S}}^{(N)}(\boldsymbol{w}_{\bar{S}})\right) + 2\epsilon \\
&= I\left(\boldsymbol{X}_{\tilde{S}}; Y|\boldsymbol{X}_{\bar{S}}\right) + 2\epsilon,
\end{aligned} \tag{A.17}$$

where inequality in the second line from the last is obtained due to the data processing inequality [20], and equality in the last line is due to the fact that the channel is memoryless and codeword symbols are generated independently. By taking N to infinity and taking ϵ to 0, (A.17) implies that $\sum_{k\in\tilde{S}} r_{g_k} \leq I\left(\boldsymbol{X}_{\tilde{S}}; Y|\boldsymbol{X}_{\bar{S}}\right)$. Because this holds for every user subset $S \subseteq \{1,\ldots,K\}$ with $1 \in S$, we must have $\boldsymbol{g} \in \boldsymbol{C}_{d1}^c$. Because $\boldsymbol{g} \in \boldsymbol{R}_1$ is chosen arbitrarily, $\boldsymbol{R}_1 \subseteq \boldsymbol{C}_{d1}^c$ therefore must be true.

A.3 Proof of Corollary 2.7

Note that, if a region is achievable under the alternative error probability definition (2.15), it is also achievable under error probability definition (2.13). Therefore, converse part of the corollary is implied by the converse part of Theorem 2.6.

Let $\boldsymbol{\mathcal{G}}$ be the code ensemble vector of the users. Given $\boldsymbol{\mathcal{G}}$, we first find a small positive constant $\sigma > 0$, such that for all $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$, either $\boldsymbol{g} \notin \boldsymbol{C}_{d1}$ or $\boldsymbol{g} \in \boldsymbol{C}_{d1}^{(\sigma)}$, where $\boldsymbol{C}_{d1}^{(\sigma)}$ is defined as follows

$$\boldsymbol{C}_{d1}^{(\sigma)} = \left\{\boldsymbol{g} \left| \begin{array}{l} \boldsymbol{g} = (\boldsymbol{r}_{\boldsymbol{g}}, \boldsymbol{P}_{\boldsymbol{g}\boldsymbol{X}}), \forall S \subseteq \{1,\ldots,K\}, 1 \in S, \exists \tilde{S} \subseteq S, 1 \in \tilde{S}, \\ \text{such that, } \sum_{k\in\tilde{S}} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}}; Y|\boldsymbol{X}_{\bar{S}}) - \sigma \end{array} \right.\right\}. \tag{A.18}$$

Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code index vector pair. In the proof of Theorem 2.6, we already showed that, the probability for the receiver to output an erroneous message for user 1 is asymptotically zero. To prove the corollary, we need to show that, if $\boldsymbol{g} \notin \boldsymbol{C}_{d1}$, then the probability for the receiver to output the correct message for user 1, as opposed to reporting collision, is also asymptotically zero. Therefore, in the rest of the proof, we assume $\boldsymbol{g} \notin \boldsymbol{C}_{d1}$.

Let $\epsilon > 0$ be a small positive constant. Define the set $A_\epsilon^{(N)}(\boldsymbol{g})$ of typical sequences $(\boldsymbol{X}^{(N)}, Y^{(N)})$ for code index $\boldsymbol{g}$ as in (A.9). Let $S \subseteq \{1, \ldots, K\}$ be an arbitrary user subset with $1 \in S$. We define event "Error$_{2S}$" as follows

$$\begin{aligned}\text{Error}_{2S} : \exists(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}}) &\text{ with } \tilde{\boldsymbol{g}} \in \boldsymbol{C}_{d1}^{(\sigma)},\\ &(\tilde{\boldsymbol{w}}_{\bar{S}\cup\{1\}}, \tilde{\boldsymbol{g}}_{\bar{S}\cup\{1\}}) = (\boldsymbol{w}_{\bar{S}\cup\{1\}}, \boldsymbol{g}_{\bar{S}\cup\{1\}}),\\ &\text{and with } (\tilde{w}_k, \tilde{g}_k) \neq (w_k, g_k), \forall k \in S \setminus \{1\},\\ &\text{such that } (\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{(N)}(\tilde{\boldsymbol{w}}), Y^{(N)}) \in A_\epsilon^{(N)}(\tilde{\boldsymbol{g}}).\end{aligned} \tag{A.19}$$

Event "Error$_{2S}$" represents the situation that the receiver finds a decoding candidate $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}})$. $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}})$ is identical to the actual vector pair $(\boldsymbol{w}, \boldsymbol{g})$ on the part corresponding to users in $\bar{S} \cup \{1\}$, while $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}})$ differs from $(\boldsymbol{w}, \boldsymbol{g})$ on the part corresponding to users not in $\bar{S} \cup \{1\}$. We will show that $\lim_{N\to\infty} Pr\{\text{Error}_{2S}\} = 0$ for all S.

Suppose that event "Error$_{2S}$" does happen for a given S with $1 \in S$. To obtain $\lim_{N\to\infty} Pr\{\text{Error}_{2S}\} = 0$, our key objective is to show that there exists a user subset $\tilde{S} \subseteq S$ with $1 \in \tilde{S}$ and $\tilde{S} \setminus \{1\} \neq \phi$, such that

$$\sum_{k\in\tilde{S}\setminus\{1\}} r_{\tilde{g}_k} < I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}; Y | \boldsymbol{X}_{\bar{S}\cup\{1\}}) - \sigma + 4\epsilon. \tag{A.20}$$

Consider part of the vectors corresponding to user subset $\tilde{S} \cup \bar{S}$. As in Section A.2, we say that a partial codeword vector sequence $\boldsymbol{X}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}^{(N)}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}})$ is jointly typical with $Y^{(N)}$ with respect to code index vector $\tilde{\boldsymbol{g}}$, denoted by $(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}^{(N)}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A_\epsilon^{(N)}(\tilde{\boldsymbol{g}})$, if there exists a codeword vector sequence $\boldsymbol{X}^{(N)}$ with $\boldsymbol{X}_{\tilde{S}\cup\bar{S}}^{(N)} = \boldsymbol{X}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}^{(N)}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}})$, such that $(\boldsymbol{X}^{(N)}, Y^{(N)}) \in A_\epsilon^{(N)}(\tilde{\boldsymbol{g}})$.

Note that, for each partial message vector and code index vector pair $(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}, \tilde{\boldsymbol{g}})$ and for large enough N, the probability that the receiver will find $(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})$ should satisfy the following upper bound

$$\begin{aligned}
&\frac{1}{N}\log[Pr\{(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})\}] \\
&\leq (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}, \boldsymbol{X}_{\bar{S}\cup\{1\}}, Y) + \epsilon) - (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}) - \epsilon) \\
&\quad -(H_{\boldsymbol{g}}(\boldsymbol{X}_{\bar{S}\cup\{1\}}, Y) - \epsilon) \\
&\leq (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}, \boldsymbol{X}_{\bar{S}\cup\{1\}}, Y) + \epsilon) - (H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}) - \epsilon) \\
&\quad -(H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\bar{S}\cup\{1\}}, Y) - 3\epsilon) \\
&= -I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}; \boldsymbol{X}_{\bar{S}\cup\{1\}}, Y) + 5\epsilon. \qquad (A.21)
\end{aligned}$$

According to the union bound, if (A.20) is true, for large enough N, $Pr\{\text{Error}_{2S}\}$ should be upper bounded by

$$\begin{aligned}
&Pr\{\text{Error}_{2S}\} \\
&\leq \sum_{\tilde{\boldsymbol{g}}_{S\setminus\{1\}}, \tilde{\boldsymbol{g}}\in \boldsymbol{C}^{(\sigma)}_{d1}} \sum_{\tilde{\boldsymbol{w}}_{\tilde{S}\setminus\{1\}}} Pr\{(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{\tilde{S}\cup\bar{S}}}(\tilde{\boldsymbol{w}}_{\tilde{S}\cup\bar{S}}), Y^{(N)}) \in A^{(N)}_{\epsilon}(\tilde{\boldsymbol{g}})\} \\
&\leq \sum_{\tilde{\boldsymbol{g}}_{S\setminus\{1\}}, \tilde{\boldsymbol{g}}\in \boldsymbol{C}^{(\sigma)}_{d1}} \exp\left[N\left(\sum_{k\in\tilde{S}\setminus\{1\}} r_{\tilde{g}_k} - I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}; \boldsymbol{X}_{\bar{S}\cup\{1\}}, Y)\right.\right. \\
&\quad \left.\left. +5\epsilon\right)\right] \\
&< M^{|S|-1}\exp[N(-\sigma + 9\epsilon)]. \qquad (A.22)
\end{aligned}$$

Given σ, (A.22) implies that we can find a small enough ϵ to ensure $\lim_{N\to\infty} Pr\{\text{Error}_{2S}\} = 0$.

Next, we prove that (A.20) indeed holds. We first show that, there exists a user subset S_1 with $S_1 \cap S = \phi$, such that for all user subsets $S_2 \subseteq \bar{S} \setminus S_1$, we have

$$\sum_{k\in S_2} r_{g_k} + r_{g_1} \geq I_{\boldsymbol{g}}(\boldsymbol{X}_{S_2\cup\{1\}}; Y|\boldsymbol{X}_{S_1}). \qquad (A.23)$$

Because $\boldsymbol{g} \notin \boldsymbol{C}_{d1}$, according to (2.16), there exists a user subset S_3 with $1 \in S_3$, such that the following bound holds for all user subsets $\tilde{S}_3 \subseteq S_3$ with $1 \in \tilde{S}_3$,

$$\sum_{k\in\tilde{S}_3} r_{g_k} \geq I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}_3}; Y|\boldsymbol{X}_{\bar{S}_3}). \qquad (A.24)$$

In fact, (A.23) holds if we choose $S_1 = \bar{S}_3 \cap \bar{S}$. To see this, note that with $S_1 = \bar{S}_3 \cap \bar{S}$, for any user subset $S_2 \subseteq \bar{S} \setminus S_1$, we must have $S_2 \cap \bar{S}_3 = \phi$ and $1 \notin S_2$. In other words, $\{1\} \cup S_2 \subseteq S_3$. Consequently, according to (A.24), we have

$$\begin{aligned}
&\sum_{k \in S_2} r_{g_k} + r_{g_1} \geq I_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}; Y | \boldsymbol{X}_{\bar{S}_3}) \\
&= H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}) - H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}} | Y, \boldsymbol{X}_{\bar{S}_3}) \\
&\geq H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}) - H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}} | Y, \boldsymbol{X}_{S_1}) \\
&= I_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}; Y | \boldsymbol{X}_{S_1}),
\end{aligned} \tag{A.25}$$

where inequality in the third line is due to the fact that $S_1 \subseteq \bar{S}_3$.

Now, let S_1 be the user subset that satisfies (A.23). If the receiver outputs another message vector and code index vector pair $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{g}})$ with $\tilde{\boldsymbol{g}} \in \boldsymbol{C}_{d1}^{(\sigma)}$, $(\tilde{\boldsymbol{w}}_{\bar{S} \cup \{1\}}, \tilde{\boldsymbol{g}}_{\bar{S} \cup \{1\}}) = (\boldsymbol{w}_{\bar{S} \cup \{1\}}, \boldsymbol{g}_{\bar{S} \cup \{1\}})$, and $(\tilde{w}_k, \tilde{g}_k) \neq (w_k, g_k)$ for all $k \in S \setminus \{1\}$, it implies that partial codeword vector sequence and channel output sequence pair $\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_{\bar{S} \cup \{1\}}}(\boldsymbol{w}_{\bar{S} \cup \{1\}}), Y^{(N)}\right)$ are jointly typical with respect to both $A_\epsilon^{(N)}(\boldsymbol{g})$ and $A_\epsilon^{(N)}(\tilde{\boldsymbol{g}})$. Therefore, for all user subsets $S_2 \subseteq \bar{S} \setminus S_1$, we can translate the bounds of (A.23) to the following one associated with $\tilde{\boldsymbol{g}}$.

$$\begin{aligned}
&\sum_{k \in S_2} r_{\tilde{g}_k} + r_{\tilde{g}_1} \geq I_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}; Y | \boldsymbol{X}_{S_1}) \\
&= H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}) - H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}} | Y, \boldsymbol{X}_{S_1}) \\
&= H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}) - H_{\boldsymbol{g}}(\boldsymbol{X}_{S_2 \cup \{1\}}, \boldsymbol{X}_{S_1}, Y) + H_{\boldsymbol{g}}(\boldsymbol{X}_{S_1}, Y) \\
&\geq H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2 \cup \{1\}}) - [H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2 \cup \{1\}}, \boldsymbol{X}_{S_1}, Y) + 2\epsilon] \\
&\quad + [H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_1}, Y) - 2\epsilon] \\
&\geq H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2 \cup \{1\}}) - [H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2 \cup \{1\}} | Y, \boldsymbol{X}_{S_1}) + 4\epsilon] \\
&= I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2 \cup \{1\}}; Y | \boldsymbol{X}_{S_1}) - 4\epsilon.
\end{aligned} \tag{A.26}$$

By assumption, $\tilde{\boldsymbol{g}} \in C_{d1}^{(\sigma)}$, and therefore we can find a user subset $\tilde{S} \cup S_2$ with $\tilde{S} \subseteq S$, $\{1\} \in \tilde{S}$, and $S_2 \subseteq \bar{S} \setminus S_1$, such that

$$\sum_{k \in S_2 \cup \tilde{S}} r_{\tilde{g}_k} < I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2 \cup \tilde{S}}; Y | \boldsymbol{X}_{S_1}) - \sigma. \tag{A.27}$$

Subtracting (A.26) from (A.27), we get

$$
\begin{aligned}
\sum_{k\in\tilde{S}\setminus\{1\}} r_{\tilde{g}_k} &< I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2\cup\tilde{S}};Y|\boldsymbol{X}_{S_1}) - \sigma - I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{S_2\cup\{1\}};Y|\boldsymbol{X}_{S_1}) + 4\epsilon \\
&= I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}};Y|\boldsymbol{X}_{S_2\cup S_1\cup\{1\}}) - \sigma + 4\epsilon \\
&= H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}) - H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}|Y,\boldsymbol{X}_{S_2\cup S_1\cup\{1\}}) - \sigma + 4\epsilon \\
&\le H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}) - H_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}}|Y,\boldsymbol{X}_{\bar{S}\cup\{1\}}) - \sigma + 4\epsilon \\
&= I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}};Y|\boldsymbol{X}_{\bar{S}\cup\{1\}}) - \sigma + 4\epsilon.
\end{aligned} \tag{A.28}
$$

Note that (A.28) implies $\tilde{S}\setminus\{1\} \ne \phi$, because otherwise $\sum_{k\in\tilde{S}\setminus\{1\}} r_{\tilde{\boldsymbol{g}}k} = I_{\tilde{\boldsymbol{g}}}(\boldsymbol{X}_{\tilde{S}\setminus\{1\}};Y|\boldsymbol{X}_{\bar{S}\cup\{1\}}) = 0$, with a small enough ϵ, (A.28) leads to an invalid inequality of $0 < 0 - \sigma + 4\epsilon < 0$.

With the proof of (A.20), conclusion of the corollary then follows.

A.4 Proof of Corollary 2.8

From the definition, it is easy to see that $\boldsymbol{C}_{dS_0} \subseteq \bigcap_{k\in S_0} \boldsymbol{C}_{dk}$. Next, we will show that $\boldsymbol{C}_{dS_0} \supseteq \bigcap_{k\in S_0} \boldsymbol{C}_{dk}$. In other words, if $\boldsymbol{g}$ is a code index vector satisfying $\boldsymbol{g} \in \boldsymbol{C}_{dk}$ for all $k \in S_0$, we will show that $\boldsymbol{g} \in \boldsymbol{C}_{dS_0}$ must be true.

For any user subset S with $S\cap S_0 \ne \phi$, let $\tilde{k} \in S\cap S_0$ be an arbitrary user in $S \cap S_0$. Because $\boldsymbol{g} \in \boldsymbol{C}_{d\tilde{k}}$, according to (2.16), we can find a subset $\tilde{S}_1 \subseteq S$ with $\tilde{k} \in \tilde{S}_1 \cap S_0$, such that

$$
\sum_{k\in\tilde{S}_1} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}_1};Y|\boldsymbol{X}_{\bar{S}}). \tag{A.29}
$$

Now consider user subset $S \setminus \tilde{S}_1$. If $(S \setminus \tilde{S}_1) \cap S_0 \ne \phi$, for the same reason explained above, we can find another user subset $\tilde{S}_2 \subseteq S \setminus \tilde{S}_1$ with $\tilde{S}_2 \cap S_0 \ne \phi$ such that

$$
\sum_{k\in\tilde{S}_2} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}_2};Y|\boldsymbol{X}_{\overline{S\setminus\tilde{S}_1}}). \tag{A.30}
$$

Apply this procedure recursively. That is, if for an integer $j > 0$ such that $(S \setminus \tilde{S}_1 \setminus \ldots \setminus \tilde{S}_{j-1}) \cap S_0 \neq \phi$, we can find user subset $\tilde{S}_j \subseteq S \setminus \tilde{S}_1 \setminus \ldots \setminus \tilde{S}_{j-1}$ with $\tilde{S}_j \cap S_0 \neq \phi$, such that

$$\sum_{k\in\tilde{S}_j} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}_j}; Y|\boldsymbol{X}_{\overline{S\setminus\tilde{S}_1\setminus\ldots\setminus\tilde{S}_{j-1}}}). \tag{A.31}$$

Let the procedure be carried out till we find an integer $J > 0$ with $(S \setminus \tilde{S}_1 \setminus \ldots \setminus \tilde{S}_J) \cap S_0 = \phi$. Define $\tilde{S} = \bigcup_{j=1}^{J} \tilde{S}_j$. Because (A.31) holds for all $j \leq J$, and user subsets $\tilde{S}_j$ for $j = 1, \ldots, J$ are mutually exclusive, we must have

$$\sum_{k\in\tilde{S}} r_{g_k} < \sum_{j=1}^{J} I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}_j}; Y|\boldsymbol{X}_{\overline{S\setminus\tilde{S}_1\setminus\ldots\setminus\tilde{S}_{j-1}}}) = I_{\boldsymbol{g}}(\boldsymbol{X}_{\tilde{S}}; Y|\boldsymbol{X}_{\bar{S}}). \tag{A.32}$$

This implies that user subset $\tilde{S}$ with $S \cap S_0 \subseteq \tilde{S} \subseteq S$ can be found to satisfy (A.32) for any user subset S with $S \cap S_0 \neq \phi$. Consequently, $\boldsymbol{g} \in \boldsymbol{C}_{dS_0}$ must be true.

B

Proofs of Theorems in Section 3

B.1 Proof of Theorem 3.1

Given the sequence of channel output symbols $Y^{(N)}$, for any channel input sequence $X^{(N)}$ and code index g, we define the weighted likelihood of the channel input sequence $L_g\left(X^{(N)}, Y^{(N)}\right)$ as follows.

$$L_g\left(X^{(N)}, Y^{(N)}\right) = P(Y^{(N)}|X^{(N)})e^{-N\alpha_g}. \tag{B.1}$$

Also define a constraint set $\mathcal{R}$ of message and code index pairs, where each code index should belong to the operation region and the weighted likelihood of the corresponding codeword should stay above a predetermined threshold.

$$\mathcal{R} = \left\{(w, g)|g \in R \text{ and } \forall \tilde{g} \notin R, L_g\left(X_g^{(N)}(w), Y^{(N)}\right) > e^{-N\tau_{[g,\tilde{g}]}(Y^{(N)})}\right\} \tag{B.2}$$

where $\tau_{[g,\tilde{g}]}(Y^{(N)})$ is a threshold function whose value will be specified later.

We assume that the following decoding algorithm is used at the receiver. Given $Y^{(N)}$, the receiver calculates the constraint set $\mathcal{R}$. If $\mathcal{R}$ is empty, the receiver reports collision. Otherwise, the receiver outputs the message and code index pair $(\hat{w}, \hat{g}) \in \mathcal{R}$ with the maximum weighted likelihood value. In other words,

$$(\hat{w}, \hat{g}) = \operatorname*{argmax}_{(w,g)\in\mathcal{R}} L_g\left(X_g^{(N)}(w), Y^{(N)}\right). \tag{B.3}$$

We define the probabilities of the following three types of error events. First, under the assumption that (w, g) with $g \in R$ is the actual message and code index pair, we define $P_{m[g,\tilde{g}]}$ as the probability that the weighted likelihood of (w, g) is no larger than that of another message and code index pair $(\tilde{w}, \tilde{g})$ corresponding to code index $\tilde{g} \in R$.

$$\begin{aligned} P_{m[g,\tilde{g}]} = Pr\Big\{ & \exists\tilde{w}, (\tilde{w}, \tilde{g}) \neq (w, g), \text{ such that} \\ & L_g\left(X_g^{(N)}(w), Y^{(N)}\right) \leq L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}\right)\Big\}, \text{ for } g, \tilde{g} \in R. \end{aligned} \tag{B.4}$$

Second, again under the assumption that (w, g) with $g \in R$ is the actual message and code index pair, we define $P_{t[g,\tilde{g}]}$ as the probability that the weighted likelihood of (w, g) is no larger than the threshold associated with code index $\tilde{g} \notin R$.

$$P_{t[g,\tilde{g}]} = Pr\left\{L_g\left(X_g^{(N)}(w), Y^{(N)}\right) \leq e^{-N\tau_{[g,\tilde{g}]}(Y^{(N)})}\right\}, \text{ for } g \in R, \tilde{g} \notin R \tag{B.5}$$

Third, under the assumption that $(\tilde{w}, \tilde{g})$ with $\tilde{g} \notin R$ is the actual message and code index pair, we define $P_{i[\tilde{g},g]}$ as the probability that there exists at least one message and code index pair (w, g) with $g \in R$, such that the weighted likelihood of (w, g) is larger than the threshold associated with code index $\tilde{g}$.

$$\begin{aligned} P_{i[\tilde{g},g]} = Pr\left\{\exists w \text{ with } L_g\left(X_g^{(N)}(w), Y^{(N)}\right) > e^{-N\tau_{[g,\tilde{g}]}(Y^{(N)})}\right\}, \\ \text{for } \tilde{g} \notin R, g \in R. \end{aligned} \tag{B.6}$$

With the above probability definitions, generalized error performance of the system can be upper bounded by

$$\text{GEP} \leq \sum_{g\in R}\left[\sum_{\tilde{g}\in R} P_{m[g,\tilde{g}]}e^{-N\alpha_g} + \sum_{\tilde{g}\notin R}\left(P_{t[g,\tilde{g}]}e^{-N\alpha_g} + P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}}\right)\right]. \tag{B.7}$$

Next, we will derive upper-bounds for each of the three terms on the right hand side of (B.7).

Step I: Upper-bounding $P_{m[g,\tilde{g}]}e^{-N\alpha_g}$.

Under the assumption that (w, g) with $g \in R$ is the actual message and code index pair, and with $\tilde{g} \in R$, we write $P_{m[g,\tilde{g}]}e^{-N\alpha_g}$ as follows

$$P_{m[g,\tilde{g}]}e^{-N\alpha_g} = E_{X^{(N)}}\left[\sum_{Y^{(N)}} L_g\left(X_g^{(N)}(w), Y^{(N)}\right)\phi_{m[g,\tilde{g}]}\right], \tag{B.8}$$

where $E_{X^{(N)}}[.]$ denotes the expectation operation over the random codebook generation, and $\phi_{m[g,\tilde{g}]}$ is an indicator function with $\phi_{m[g,\tilde{g}]} = 1$ if there exists a $\tilde{w}$ with $(\tilde{w}, \tilde{g}) \neq (w, g)$ such that $L_g\left(X_g^{(N)}(w), Y^{(N)}\right) \leq L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}\right)$, and $\phi_{m[g,\tilde{g}]} = 0$ if such a message $\tilde{w}$ cannot be found.

Let $\rho > 0$ and $s \geq 0$ be two arbitrary constants. As shown in [64], we can upper bound $\phi_{m[g,\tilde{g}]}$ by

$$\phi_{m[g,\tilde{g}]} \leq \left(\frac{\sum_{\tilde{w}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}\right)\right]^{\frac{s}{\rho}}}{\left[L_g\left(X_g^{(N)}(w), Y^{(N)}\right)\right]^{\frac{s}{\rho}}}\right)^{\rho}. \tag{B.9}$$

Substitute (B.9) into (B.8) yields

$$\begin{aligned}
P_{m[g,\tilde{g}]}e^{-N\alpha_g} &\leq E_{X^{(N)}}\left[\sum_{Y^{(N)}}\left[L_g\left(X_g^{(N)}(w), Y^{(N)}\right)\right]^{1-s}\right.\\
&\quad \left.\times\left(\sum_{\tilde{w}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}\right)\right]^{\frac{s}{\rho}}\right)^{\rho}\right]\\
&= \sum_{Y^{(N)}} E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)}(w), Y^{(N)}\right)\right]^{1-s}\right]\\
&\quad \times E_{X^{(N)}}\left[\left(\sum_{\tilde{w}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}\right)\right]^{\frac{s}{\rho}}\right)^{\rho}\right],
\end{aligned} \tag{B.10}$$

where the last equality is due to the assumption that $(\tilde{w}, \tilde{g}) \neq (w, g)$, and therefore codewords $X_{\tilde{g}}^{(N)}(\tilde{w})$ and $X_g^{(N)}(w)$ are generated independently.

Now assume that $0 < \rho \leq 1$, we can further bound $P_{m[g,\tilde{g}]}e^{-N\alpha_g}$ by

$$
\begin{aligned}
P_{m[g,\tilde{g}]}e^{-N\alpha_g} &\le \sum_{Y^{(N)}} E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)}(w), Y^{(N)}\right)\right]^{1-s}\right] \\
&\quad \times \left(\sum_{\tilde{w}} E_{X^{(N)}}\left[\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}), Y^{(N)}\right)\right]^{\frac{s}{\rho}}\right]\right)^{\rho} \\
&= e^{N\rho r_{\tilde{g}}} \sum_{Y^{(N)}} E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)}, Y^{(N)}\right)\right]^{1-s}\right] \\
&\quad \times \left(E_{X^{(N)}}\left[\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}, Y^{(N)}\right)\right]^{\frac{s}{\rho}}\right]\right)^{\rho},
\end{aligned} \tag{B.11}
$$

where in the last equality we removed $\tilde{w}$ in the notation because the corresponding terms are not functions of the messages after taking the expectation operations.

Let $X_n^{(N)}$ and $Y_n^{(N)}$ denote the nth symbols of $X^{(N)}$ and $Y^{(N)}$ respectively. By following a derivation similar to the one presented in [28, Section II], we have

$$
\begin{aligned}
&E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)}, Y^{(N)}\right)\right]^{1-s}\right] \\
&= \sum_{X^{(N)}} \left[P(Y^{(N)}|X^{(N)})e^{-N\alpha_g}\right]^{1-s} P_g(X^{(N)}) \\
&= \sum_{X^{(N)}} \prod_{n=1}^{N} \left[P(Y_n^{(N)}|X_n^{(N)})e^{-\alpha_g}\right]^{1-s} P_g(X_n^{(N)}) \\
&= \prod_{n=1}^{N} \sum_{X} \left[P(Y_n^{(N)}|X)e^{-\alpha_g}\right]^{1-s} P_g(X).
\end{aligned} \tag{B.12}
$$

Meanwhile, we also have,

$$
\begin{aligned}
&\left(E_{X^{(N)}}\left[\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}, Y^{(N)}\right)\right]^{\frac{s}{\rho}}\right]\right)^{\rho} \\
&= \left(\prod_{n=1}^{N} \sum_{X} \left[P(Y_n^{(N)}|X)e^{-\alpha_{\tilde{g}}}\right]^{\frac{s}{\rho}} P_{\tilde{g}}(X)\right)^{\rho} \\
&= \prod_{n=1}^{N} \left(\sum_{X} \left[P(Y_n^{(N)}|X)e^{-\alpha_{\tilde{g}}}\right]^{\frac{s}{\rho}} P_{\tilde{g}}(X)\right)^{\rho}.
\end{aligned} \tag{B.13}
$$

Therefore, (B.11) implies that

$$
\begin{aligned}
P_{m[g,\tilde{g}]}e^{-N\alpha_g} &\leq e^{N\rho r_{\tilde{g}}} \sum_{Y^{(N)}} \prod_{n=1}^{N} \left(\sum_X \left[P(Y_n^{(N)}|X) e^{-\alpha_g} \right]^{1-s} P_g(X) \right) \\
&\quad \times \left(\sum_X \left[P(Y_n^{(N)}|X) e^{-\alpha_{\tilde{g}}} \right]^{\frac{s}{\rho}} P_{\tilde{g}}(X) \right)^{\rho} \\
&= e^{N\rho r_{\tilde{g}}} \left\{ \sum_Y \left(\sum_X \left[P(Y|X) e^{-\alpha_g} \right]^{1-s} P_g(X) \right) \right. \\
&\quad \left. \times \left(\sum_X \left[P(Y|X) e^{-\alpha_{\tilde{g}}} \right]^{\frac{s}{\rho}} P_{\tilde{g}}(X) \right)^{\rho} \right\}^N \\
&= \exp\left(-N \left[-\rho r_{\tilde{g}} - \log \sum_Y \left(\sum_X \left[P(Y|X) e^{-\alpha_g} \right]^{1-s} P_g(X) \right) \right. \right. \\
&\quad \left. \left. \times \left(\sum_X \left[P(Y|X) e^{-\alpha_{\tilde{g}}} \right]^{\frac{s}{\rho}} P_{\tilde{g}}(X) \right)^{\rho} \right] \right). \qquad \text{(B.14)}
\end{aligned}
$$

Because (B.14) holds for all $0 < \rho \leq 1$ and $s \geq 0$, and becomes trivial for $s > 1$, we conclude that

$$
P_{m[g,\tilde{g}]}e^{-N\alpha_g} \leq \exp(-NE_m(g,\tilde{g})), \qquad \text{(B.15)}
$$

where $E_m(g,\tilde{g})$ is given in (3.5).

Step II: Upper-bounding $P_{t[g,\tilde{g}]}e^{-N\alpha_g}$.

Under the assumption that (w, g) with $g \in R$ is the actual message and code index pair, we write $P_{t[g,\tilde{g}]}e^{-N\alpha_g}$ as follows

$$
P_{t[g,\tilde{g}]}e^{-N\alpha_g} = E_{X^{(N)}} \left[\sum_{Y^{(N)}} L_g\left(X_g^{(N)}(w), Y^{(N)} \right) \phi_{t[g,\tilde{g}]} \right], \qquad \text{(B.16)}
$$

where $\phi_{t[g,\tilde{g}]}$ is an indicator function with

$$
\phi_{t[g,\tilde{g}]} = \begin{cases} 1 & \text{if } L_g\left(X_g^{(N)}(w), Y^{(N)} \right) \leq e^{-N\tau_{[g,\tilde{g}]}(Y^{(N)})} \\ 0 & \text{otherwise} \end{cases}. \qquad \text{(B.17)}
$$

Let $s_1 \geq 0$ be an arbitrary constant. We can upper bound $\phi_{t[g,\tilde{g}]}$ by

$$
\phi_{t[g,\tilde{g}]} \leq \frac{e^{-Ns_1\tau_{[g,\tilde{g}]}(Y^{(N)})}}{\left[L_g\left(X_g^{(N)}(w), Y^{(N)} \right) \right]^{s_1}}. \qquad \text{(B.18)}
$$

Substitute (B.18) into (B.16), we get

$$
\begin{aligned}
&P_{t[g,\tilde{g}]}e^{-N\alpha_g} \\
&\leq E_{X^{(N)}}\left[\sum_{Y^{(N)}}\left[L_g\left(X_g^{(N)}(w),Y^{(N)}\right)\right]^{1-s_1}e^{-Ns_1\tau_{[g,\tilde{g}]}(Y^{(N)})}\right] \\
&= \sum_{Y^{(N)}}E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{1-s_1}\right]e^{-Ns_1\tau_{[g,\tilde{g}]}(Y^{(N)})}. \quad \text{(B.19)}
\end{aligned}
$$

We again removed w in the notation because the result after taking the expectation operation is not a function of the message.

Step III: Upper-bounding $P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}}$.

Under the assumption that $(\tilde{w},\tilde{g})$ with $\tilde{g}\notin R$ is the actual message and code index pair, with $g\in R$, we write $P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}}$ as follows

$$
P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}} = E_{X^{(N)}}\left[\sum_{Y^{(N)}}\sum_{\tilde{g}\notin R}L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}),Y^{(N)}\right)\phi_{i[g,\tilde{g}]}\right], \quad \text{(B.20)}
$$

where $\phi_{i[g,\tilde{g}]}$ is an indicator function with

$$
\phi_{i[g,\tilde{g}]} = \begin{cases} 1 & \text{if } \exists w, L_g\left(X_g^{(N)}(w),Y^{(N)}\right) > e^{-N\tau_{[g,\tilde{g}]}(Y^{(N)})} \\ 0 & \text{otherwise} \end{cases}. \quad \text{(B.21)}
$$

Let $s_2\geq 0$ and $\tilde{\rho}>0$ be two arbitrary constants. We can upper bound $\phi_{i[g,\tilde{g}]}$ by

$$
\phi_{i[g,\tilde{g}]} \leq \left(\frac{\sum_w\left[L_g\left(X_g^{(N)}(w),Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}}{e^{-N\frac{s_2}{\tilde{\rho}}\tau_{[g,\tilde{g}]}(Y^{(N)})}}\right)^{\tilde{\rho}}. \quad \text{(B.22)}
$$

Substitute (B.22) into (B.20) to obtain

$$
\begin{aligned}
\sum_{\tilde{g}\notin R}P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}} &\leq E_{X^{(N)}}\Bigg[\sum_{Y^{(N)}}\sum_{\tilde{g}\notin R}L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}),Y^{(N)}\right) \\
&\quad\times\left(\sum_w\left[L_g\left(X_g^{(N)}(w),Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}\right)^{\tilde{\rho}}e^{Ns_2\tau_{[g,\tilde{g}]}(Y^{(N)})}\Bigg] \\
&= \sum_{Y^{(N)}}E_{X^{(N)}}\left[\sum_{\tilde{g}\notin R}L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}(\tilde{w}),Y^{(N)}\right)\right]
\end{aligned}
$$

$$\times E_{X^{(N)}}\left[\left(\sum_{w}\left[L_g\left(X_g^{(N)}(w),Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}\right)^{\tilde{\rho}}\right]e^{Ns_2\tau_{[g,\tilde{g}]}(Y^{(N)})}, \tag{B.23}$$

where the last equality is due to the assumption that $(\tilde{w},\tilde{g})\neq(w,g)$, and therefore codewords $X_{\tilde{g}}^{(N)}(\tilde{w})$ and $X_g^{(N)}(w)$ are generated independently.

Now assume that $0<\tilde{\rho}\leq 1$, (B.23) further implies that

$$\sum_{\tilde{g}\notin R}P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}}\leq\sum_{Y^{(N)}}E_{X^{(N)}}\left[\sum_{\tilde{g}\notin R}L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)},Y^{(N)}\right)\right]$$
$$\times e^{N\tilde{\rho}r_g}\left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}\right]\right)^{\tilde{\rho}}e^{Ns_2\tau_{[g,\tilde{g}]}(Y^{(N)})}. \tag{B.24}$$

Step IV: Choosing $\tau_{[g,\tilde{g}]}(Y^{(N)})$.

Given $g\in R$, $\tilde{g}\notin R$, $Y^{(N)}$, and auxiliary variables $s_1\geq 0$, $s_2\geq 0$, $0<\tilde{\rho}\leq 1$, we choose $\tau_{[g,\tilde{g}]}(Y^{(N)})$ to match the following pairs of terms on the right hand sides of (B.19) and (B.24),

$$E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{1-s_1}\right]e^{-Ns_1\tau_{[g,\tilde{g}]}(Y^{(N)})}$$
$$=e^{Ns_2\tau_{[g,\tilde{g}]}(Y^{(N)})}E_{X^{(N)}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)},Y^{(N)}\right)\right]$$
$$\times e^{N\tilde{\rho}r_g}\left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}\right]\right)^{\tilde{\rho}}. \tag{B.25}$$

(B.25) implies that

$$e^{-N\tau_{[g,\tilde{g}]}(Y^{(N)})}=\left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{1-s_1}\right]\right)^{-\frac{1}{s_1+s_2}}$$
$$\times\left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}\right]\right)^{\frac{\tilde{\rho}}{s_1+s_2}}$$
$$\times\left(E_{X^{(N)}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)},Y^{(N)}\right)\right]\right)^{\frac{1}{s_1+s_2}}e^{N\frac{\tilde{\rho}}{s_1+s_2}r_g}. \tag{B.26}$$

Substitute (B.26) into (B.19), we get

$$P_{t[g,\tilde{g}]}e^{-N\alpha_g} \leq \sum_{Y^{(N)}} \left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{1-s_1}\right]\right)^{\frac{s_2}{s_1+s_2}}$$
$$\times\left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}\right]\right)^{\frac{s_1\tilde{\rho}}{s_1+s_2}}$$
$$\times\left(E_{X^{(N)}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)},Y^{(N)}\right)\right]\right)^{\frac{s_1}{s_1+s_2}} e^{N\frac{s_1\tilde{\rho}}{s_1+s_2}r_g}. \tag{B.27}$$

Let $s_2 < \tilde{\rho}$ and $s_1 = 1 - \frac{s_2}{\tilde{\rho}}$[1]. Do a variable change with $\rho = \frac{\tilde{\rho}(\tilde{\rho}-s_2)}{\tilde{\rho}-(1-\tilde{\rho})s_2}$ and $s = 1-\frac{\tilde{\rho}-s_2}{\tilde{\rho}-(1-\tilde{\rho})s_2}$. We have $1-s_1 = \frac{s_2}{\tilde{\rho}} = \frac{s}{s+\rho}$, $\frac{s_2}{s_1+s_2} = s$, $\frac{s_1\tilde{\rho}}{s_1+s_2} = \rho$, $\frac{s_1}{s_1+s_2} = 1-s$. Inequality (B.27) then becomes

$$P_{t[g,\tilde{g}]}e^{-N\alpha_g} \leq \sum_{Y^{(N)}} \left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{\frac{s}{s+\rho}}\right]\right)^{s+\rho}$$
$$\times\left(E_{X^{(N)}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)},Y^{(N)}\right)\right]\right)^{1-s} e^{N\rho r_g}. \tag{B.28}$$

Let $X_n^{(N)}$ and $Y_n^{(N)}$ denote the nth symbols of $X^{(N)}$ and $Y^{(N)}$ respectively. We have

$$\left(E_{X^{(N)}}\left[\left[L_g\left(X_g^{(N)},Y^{(N)}\right)\right]^{\frac{s}{s+\rho}}\right]\right)^{s+\rho}$$
$$= \left(\sum_{X^{(N)}}\left[P(Y^{(N)}|X^{(N)})e^{-N\alpha_g}\right]^{\frac{s}{s+\rho}} P_g(X^{(N)})\right)^{s+\rho}$$
$$= \left(\sum_{X^{(N)}}\prod_{n=1}^{N}\left[P(Y_n^{(N)}|X_n^{(N)})e^{-\alpha_g}\right]^{\frac{s}{s+\rho}} P_g(X_n^{(N)})\right)^{s+\rho}$$
$$= \left(\prod_{n=1}^{N}\sum_{X}\left[P(Y_n^{(N)}|X)e^{-\alpha_g}\right]^{\frac{s}{s+\rho}} P_g(X)\right)^{s+\rho}$$
$$= \prod_{n=1}^{N}\left(\sum_{X}\left[P(Y_n^{(N)}|X)e^{-\alpha_g}\right]^{\frac{s}{s+\rho}} P_g(X)\right)^{s+\rho}. \tag{B.29}$$

Similarly,

[1] This implies that $s_1 + s_2 > 0$ in (B.26) and (B.27).

$$\begin{aligned}
&\left(E_{X^{(N)}}\left[L_{\tilde{g}}\left(X_{\tilde{g}}^{(N)}, Y^{(N)}\right)\right]\right)^{1-s} \\
&= \left(\prod_{n=1}^{N}\sum_{X} P(Y_n^{(N)}|X)e^{-\alpha_{\tilde{g}}}P_{\tilde{g}}(X)\right)^{1-s} \\
&= \prod_{n=1}^{N}\left(\sum_{X} P(Y_n^{(N)}|X)e^{-\alpha_{\tilde{g}}}P_{\tilde{g}}(X)\right)^{1-s}. \qquad \text{(B.30)}
\end{aligned}$$

Consequently, (B.28) implies that

$$\begin{aligned}
P_{t[g,\tilde{g}]}e^{-N\alpha_g} &\leq \sum_{Y^{(N)}}\prod_{n=1}^{N}\left(\sum_{X}\left[P(Y_n^{(N)}|X)e^{-\alpha_g}\right]^{\frac{s}{s+\rho}}P_g(X)\right)^{s+\rho} \\
&\times\left(\sum_{X}P(Y_n^{(N)}|X)e^{-\alpha_{\tilde{g}}}P_{\tilde{g}}(X)\right)^{1-s} e^{N\rho r_g} \\
&= \left\{\sum_{Y}\left(\sum_{X}\left[P(Y|X)e^{-\alpha_g}\right]^{\frac{s}{s+\rho}}P_g(X)\right)^{s+\rho}\right. \\
&\left.\times\left(\sum_{X}P(Y|X)e^{-\alpha_{\tilde{g}}}P_{\tilde{g}}(X)\right)^{1-s}\right\}^{N} e^{N\rho r_g} \\
&= \exp\left(-N\left[-\rho r_g - \log\left(\sum_{Y}\left(\sum_{X}\left[P(Y|X)e^{-\alpha_g}\right]^{\frac{s}{s+\rho}}P_g(X)\right)^{s+\rho}\right.\right.\right. \\
&\left.\left.\left.\times\left(\sum_{X}P(Y|X)e^{-\alpha_{\tilde{g}}}P_{\tilde{g}}(X)\right)^{1-s}\right)\right]\right). \qquad \text{(B.31)}
\end{aligned}$$

Similarly, by substituting (B.26) into (B.24), and with the same derivations, we also get the same bound for $P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}}$.

Because s and ρ can take any value with the constraints of $0 < \rho \leq 1$ and $0 \leq s \leq 1-\rho$, the bounds further lead to

$$\begin{aligned}
P_{t[g,\tilde{g}]}e^{-N\alpha_g} &\leq \exp(-NE_i(g,\tilde{g})) \\
P_{i[\tilde{g},g]}e^{-N\alpha_{\tilde{g}}} &\leq \exp(-NE_i(g,\tilde{g})), \qquad \text{(B.32)}
\end{aligned}$$

where $E_i(g,\tilde{g})$ is given in (3.5).

Finally, substituting (B.15) and (B.32) into (B.7) yields the conclusion of the theorem.

B.2 Proof of Theorem 3.3

Given the sequence of channel output symbols $Y^{(N)}$, for any channel input vector sequence $\boldsymbol{X}_D^{(N)}$ and code index vector $\boldsymbol{g}$, we define the weighted likelihood of the channel input sequence, denoted by $L_{\boldsymbol{g}}\left(\boldsymbol{X}_D^{(N)}, Y^{(N)}\right)$, as follows.

$$L_{\boldsymbol{g}}\left(\boldsymbol{X}_D^{(N)}, Y^{(N)}\right) = P(Y^{(N)}|\boldsymbol{X}_D^{(N)}, \boldsymbol{g}_{\bar{D}})e^{-N\alpha_{\boldsymbol{g}}}. \tag{B.33}$$

For every user subset $S \subseteq D$, we define a constraint set $\boldsymbol{\mathcal{R}}_S$ of message vector and code index vector pairs. Each code index vector in the constraint set should belong to the operation region and weighted likelihood of the corresponding codeword vector should stay above a pre-determined threshold.

$$\begin{aligned}\boldsymbol{\mathcal{R}}_S = \Big\{ & (\boldsymbol{w}_D, \boldsymbol{g}) \,| \boldsymbol{g} \in \boldsymbol{R}_D \text{ and } \forall \tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D \text{ with } \boldsymbol{g}_S = \tilde{\boldsymbol{g}}_S, \\ & L_{\boldsymbol{g}}\left(\boldsymbol{X}_D^{(N)}, Y^{(N)}\right) > e^{-N\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}(\boldsymbol{X}_S^{(N)}, Y^{(N)})} \Big\},\end{aligned} \tag{B.34}$$

where $\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}(\boldsymbol{X}_S^{(N)}, Y^{(N)})$ is a threshold function whose value depends on $\boldsymbol{X}_S^{(N)}$ and $Y^{(N)}$ and will be specified later. We further define constraint set $\boldsymbol{\mathcal{R}}_P$ as the intersection of $\boldsymbol{\mathcal{R}}_S$ for all S.

$$\boldsymbol{\mathcal{R}}_P = \bigcap_{S \subseteq D} \boldsymbol{\mathcal{R}}_S. \tag{B.35}$$

Assume the following decoding algorithm at the receiver. Given $Y^{(N)}$, the receiver first calculates constraint sets $\boldsymbol{\mathcal{R}}_S$ for all $S \subseteq D$ to obtain constraint set $\boldsymbol{\mathcal{R}}_P$. The receiver reports collision for all regular users if $\boldsymbol{\mathcal{R}}_P$ is empty. Otherwise, the receiver outputs $(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}) \in \boldsymbol{\mathcal{R}}_P$ with the maximum weighted likelihood value. In other words,

$$(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}) = \underset{(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}) \in \boldsymbol{\mathcal{R}}_P}{\text{argmax}}\, L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right). \tag{B.36}$$

Next, we need to define the notation $(\boldsymbol{w}_D, \boldsymbol{g}) \stackrel{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ that will be extensively used to simplify the expressions in the proof.

$$(\boldsymbol{w}_D, \boldsymbol{g}) \stackrel{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}}) : (\boldsymbol{w}_S, \boldsymbol{g}_S) = (\tilde{\boldsymbol{w}}_S, \tilde{\boldsymbol{g}}_S),$$
$$(w_k, g_k) \neq (\tilde{w}_k, \tilde{g}_k), \forall k \in D \setminus S. \tag{B.37}$$

In other words, $(\boldsymbol{w}_D, \boldsymbol{g}) \stackrel{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ means that the two message vector and code index vector pairs are equal for regular users in S and are different for regular users not in S. The term does not imply any assumption on the code indices of the interfering users.

We now define the probabilities of the following three types of error events.

First, assume that $(\boldsymbol{w}_D, \boldsymbol{g})$ with $\boldsymbol{g} \in \boldsymbol{R}_D$ is the actual message vector and code index vector pair. For any user subset $S \subset D$, we define $P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ as the probability of the error event that the weighted likelihood of $(\boldsymbol{w}_D, \boldsymbol{g})$ is no larger than that of another message vector and code index vector pair $(\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ with $(\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D, \boldsymbol{g})$ and $\tilde{\boldsymbol{g}} \in \boldsymbol{R}_D$.

$$P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} = Pr\Big\{\exists \tilde{\boldsymbol{w}}_D, (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D, \boldsymbol{g}), \text{ such that}$$
$$L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right) \leq L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)\Big\}$$
$$\text{for } \boldsymbol{g}, \tilde{\boldsymbol{g}} \in \boldsymbol{R}_D \text{ with } \boldsymbol{g}_S = \tilde{\boldsymbol{g}}_S. \tag{B.38}$$

Second, again assume that $(\boldsymbol{w}_D, \boldsymbol{g})$ with $\boldsymbol{g} \in \boldsymbol{R}_D$ is the actual message vector and code index vector pair. For any user subset $S \subseteq D$, we define $P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ as the probability of the error event that the weighted likelihood of $(\boldsymbol{w}_D, \boldsymbol{g})$ is no larger than the threshold associated with code index $\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D$ with $\boldsymbol{g}_S = \tilde{\boldsymbol{g}}_S$ and user subset S.

$$P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} = Pr\left\{L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right) \leq e^{-N\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}(\boldsymbol{X}_S^{(N)}, Y^{(N)})}\right\},$$
$$\text{for } \boldsymbol{g} \in \boldsymbol{R}_D, \tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D \text{ with } \boldsymbol{g}_S = \tilde{\boldsymbol{g}}_S. \tag{B.39}$$

Third, assume that $(\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ with $\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D$ is the actual message vector and code index vector pair. For any user subset $S \subseteq D$, we define $P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]}$ as the probability of the error event that there exists at least one message vector and code index vector pair $(\boldsymbol{w}_D, \boldsymbol{g})$ with $(\boldsymbol{w}_D, \boldsymbol{g}) \stackrel{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ and $\boldsymbol{g} \in \boldsymbol{R}_D$, such that the weighted likelihood of

$(\boldsymbol{w}_D, \boldsymbol{g})$ is larger than the threshold associated with code index $\tilde{\boldsymbol{g}}$ and user subset S.

$$P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]} = Pr\Big\{\exists \tilde{\boldsymbol{w}}_D, (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D, \boldsymbol{g}), \text{ such that } L_{\boldsymbol{g}}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_D}(\boldsymbol{w}_D), Y^{(N)}\right) > e^{-N\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}(\boldsymbol{X}^{(N)}_S, Y^{(N)})}\Big\}$$
$$\text{for } \tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D, \boldsymbol{g} \in \boldsymbol{R}_D \text{ with } \boldsymbol{g}_S = \tilde{\boldsymbol{g}}_S. \quad \text{(B.40)}$$

Note that, when $\tilde{\boldsymbol{g}} \in \widehat{\boldsymbol{R}}_D$, a decoding output with $(\boldsymbol{w}_D, \boldsymbol{g}) \stackrel{D}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ is not considered as an error. While if $\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D \cup \widehat{\boldsymbol{R}}_D$, all decoding outputs are considered as error events. Consequently, with the above probability definitions, generalized error performance of the system can be upper bounded by

$$\begin{aligned} \text{GEP}_D \le \sum_{\boldsymbol{g}\in\boldsymbol{R}_D} \Bigg\{ \sum_{S\subset D} \Bigg[\sum_{\tilde{\boldsymbol{g}}\in\boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S=\boldsymbol{g}_S} P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}} \\ + \sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S=\boldsymbol{g}_S} \left(P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}} + P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]} e^{-N\alpha_{\tilde{\boldsymbol{g}}}} \right) \Bigg] \\ + \sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}_D\cup\widehat{\boldsymbol{R}}_D, \tilde{\boldsymbol{g}}_D=\boldsymbol{g}_D} \left(P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},D]} e^{-N\alpha_{\boldsymbol{g}}} + P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},D]} e^{-N\alpha_{\tilde{\boldsymbol{g}}}} \right) \Bigg\} \end{aligned} \quad \text{(B.41)}$$

Next, we will derive upper-bounds for each of the terms on the right-hand side of (B.41).

Step I: Upper-bounding $P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}}$, with $S \subset D$.

Under the assumption that $(\boldsymbol{w}_D, \boldsymbol{g})$ with $\boldsymbol{g} \in \boldsymbol{R}_D$ is the actual message vector and code index vector pair, and with $\tilde{\boldsymbol{g}} \in \boldsymbol{R}_D$, we write $P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}}$ as follows

$$P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}} = E_{\boldsymbol{X}^{(N)}_D} \left[\sum_{Y^{(N)}} L_{\boldsymbol{g}} \left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_D}(\boldsymbol{w}_D), Y^{(N)} \right) \phi_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} \right], \quad \text{(B.42)}$$

where $E_{\boldsymbol{X}^{(N)}_D}[.]$ denotes the expectation operation over the random codebook generation, and $\phi_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ is an indicator function with $\phi_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} = 1$ if there exists a $\tilde{\boldsymbol{w}}_D$ with $(\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D, \boldsymbol{g})$ and $\tilde{\boldsymbol{g}} \in \boldsymbol{R}_D$ such that

$L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right) \leq L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)$, and $\phi_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} = 0$ if such a message vector $\tilde{\boldsymbol{w}}_D$ cannot be found.

Let $\rho > 0$ and $s \geq 0$ be two arbitrary constants. As shown in [64], we can upper bound $\phi_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ by

$$\phi_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} \leq \left(\frac{\sum_{\tilde{\boldsymbol{w}}_D, (\tilde{\boldsymbol{w}}_D,\tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D,\boldsymbol{g})} \left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)\right]^{\frac{s}{\rho}}}{\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{\frac{s}{\rho}}} \right)^{\rho} \tag{B.43}$$

Substitute (B.43) into (B.42) yields

$$\begin{aligned} P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}} &\leq E_{\boldsymbol{X}_D^{(N)}} \left[\sum_{Y^{(N)}} \left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{1-s} \right. \\ &\quad \left. \times \left(\sum_{\tilde{\boldsymbol{w}}_D, (\tilde{\boldsymbol{w}}_D,\tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D,\boldsymbol{g})} \left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)\right]^{\frac{s}{\rho}} \right)^{\rho} \right] \\ &= \sum_{Y^{(N)}} E_{\boldsymbol{X}_S^{(N)}} \left[E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{1-s} \right] \right. \\ &\quad \left. \times E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left(\sum_{\tilde{\boldsymbol{w}}_D, (\tilde{\boldsymbol{w}}_D,\tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D,\boldsymbol{g})} \left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)\right]^{\frac{s}{\rho}} \right)^{\rho} \right] \right], \end{aligned} \tag{B.44}$$

where the last equality is due to the assumption that $(\tilde{w}_k, \tilde{g}_k) \neq (w_k, g_k)$, $\forall k \in D \setminus S$, and therefore codewords $\boldsymbol{X}_{\tilde{\boldsymbol{g}}_{D\setminus S}}^{(N)}(\tilde{\boldsymbol{w}}_{D\setminus S})$ and $\boldsymbol{X}_{\boldsymbol{g}_{D\setminus S}}^{(N)}(\boldsymbol{w}_{D\setminus S})$ are generated independently.

Now assume that $0 < \rho \leq 1$, we can further bound $P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}}$ by

$$\begin{aligned} &P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}} \\ &\leq \sum_{Y^{(N)}} E_{\boldsymbol{X}_S^{(N)}} \left[E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{1-s} \right] \right. \\ &\quad \left. \times \left(\sum_{\tilde{\boldsymbol{w}}_D, (\tilde{\boldsymbol{w}}_D,\tilde{\boldsymbol{g}}) \stackrel{S}{=} (\boldsymbol{w}_D,\boldsymbol{g})} E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)\right]^{\frac{s}{\rho}} \right] \right)^{\rho} \right] \end{aligned}$$

$$= e^{N\rho\sum_{k\in D\setminus S} r_{\tilde{g}_k}} \sum_{Y^{(N)}} E_{\boldsymbol{X}_S^{(N)}} \left[E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{1-s}\right] \right.$$
$$\left. \times \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s}{\rho}}\right]\right)^{\rho} \right], \tag{B.45}$$

where in the last equality we removed the message variables in the notation because the corresponding terms are not functions of the messages after taking the expectation operations.

Let $\boldsymbol{X}_{\boldsymbol{g}_D n}^{(N)}$, $X_{kn}^{(N)}$ and $Y_n^{(N)}$ denote the nth symbol vector of $\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}$, the nth symbol of $X_k^{(N)}$, and the nth symbol of $Y^{(N)}$, respectively. By following a derivation similar to the one presented in [28, Section II], we have

$$E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{1-s}\right]$$
$$= \sum_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[P(Y^{(N)}|\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, \boldsymbol{X}_{D\setminus S}^{(N)}, \boldsymbol{g}_{\bar{D}}) e^{-N\alpha_{\boldsymbol{g}}}\right]^{1-s} \prod_{k\in D\setminus S} P_{g_k}(X_k^{(N)})$$
$$= \sum_{\boldsymbol{X}_{D\setminus S}^{(N)}} \prod_{n=1}^{N} \left[P(Y_n^{(N)}|\boldsymbol{X}_{\boldsymbol{g}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S n}^{(N)}, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}}\right]^{1-s} \prod_{k\in D\setminus S} P_{g_k}(X_{kn}^{(N)})$$
$$= \prod_{n=1}^{N} \sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y_n^{(N)}|\boldsymbol{X}_{\boldsymbol{g}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}}\right]^{1-s} \prod_{k\in D\setminus S} P_{g_k}(X_k). \tag{B.46}$$

Meanwhile, we also have

$$\left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s}{\rho}}\right]\right)^{\rho}$$
$$= \left(\prod_{n=1}^{N} \sum_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[P(Y_n^{(N)}|\boldsymbol{X}_{\tilde{\boldsymbol{g}}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}^{(N)}, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}}\right]^{\frac{s}{\rho}} \right.$$
$$\left. \times \prod_{k\in D\setminus S} P_{\tilde{g}_k}(X_k^{(N)}) \right)^{\rho}$$
$$= \prod_{n=1}^{N} \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y_n^{(N)}|\boldsymbol{X}_{\tilde{\boldsymbol{g}}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}}\right]^{\frac{s}{\rho}} \prod_{k\in D\setminus S} P_{\tilde{g}_k}(X_k) \right)^{\rho}$$

(B.47)

Because $(\tilde{\boldsymbol{w}}_S, \tilde{\boldsymbol{g}}_S) = (\boldsymbol{w}_S, \boldsymbol{g}_S)$, (B.45) implies that

$$\begin{aligned}
P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}} &\le e^{N\rho\sum_{k\in D\setminus S} r_{\tilde{g}_k}} \sum_{Y^{(N)}} \sum_{\boldsymbol{X}_S^{(N)}} \prod_{n=1}^{N} \prod_{k\in S} P_{g_k}(X_{kn}^{(N)}) \\
&\times \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y_n^{(N)} | \boldsymbol{X}_{\boldsymbol{g}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{1-s} \prod_{k\in D\setminus S} P_{g_k}(X_k) \right) \\
&\times \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y_n^{(N)} | \boldsymbol{X}_{\tilde{\boldsymbol{g}}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \right]^{\frac{s}{\rho}} \prod_{k\in D\setminus S} P_{\tilde{g}_k}(X_k) \right)^{\rho} \\
&= e^{N\rho\sum_{k\in D\setminus S} r_{\tilde{g}_k}} \left\{ \sum_{Y} \sum_{\boldsymbol{X}_S} \prod_{k\in S} P_{g_k}(X_k) \right. \\
&\times \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y | \boldsymbol{X}_D, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{1-s} \prod_{k\in D\setminus S} P_{g_k}(X_k) \right) \\
&\times \left. \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y | \boldsymbol{X}_D, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \right]^{\frac{s}{\rho}} \prod_{k\in D\setminus S} P_{\tilde{g}_k}(X_k) \right)^{\rho} \right\}^{N}.
\end{aligned} \tag{B.48}$$

Because (B.48) holds for all $0 < \rho \le 1$ and $s \ge 0$, and becomes trivial for $s > 1$, we conclude that

$$P_{m[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}} \le \exp(-N E_{mD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)), \tag{B.49}$$

where $E_{mD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)$ is given in (3.13).

Step II: Upper-bounding $P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}}$, with $S \subseteq D$.

Under the assumption that $(\boldsymbol{w}_D, \boldsymbol{g})$ with $\boldsymbol{g} \in \boldsymbol{R}_D$ is the actual message vector and code index vector pair, we write $P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}}$ as follows

$$P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}} = E_{\boldsymbol{X}_D^{(N)}} \left[\sum_{Y^{(N)}} L_{\boldsymbol{g}} \left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)} \right) \phi_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} \right], \tag{B.50}$$

where $\phi_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ is an indicator function with

$$\phi_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} = \begin{cases} 1 & \text{if } L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right) \le e^{-N\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}(\boldsymbol{w}_S), Y^{(N)}\right)} \\ 0 & \text{otherwise} \end{cases}. \tag{B.51}$$

Let $s_1 \ge 0$ be an arbitrary constant. We can upper bound $\phi_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ by

$$\phi_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} \le \frac{e^{-Ns_1\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}(\boldsymbol{w}_S), Y^{(N)}\right)}}{\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{s_1}}. \tag{B.52}$$

Substitute (B.52) into (B.50), we get

$$\begin{aligned} P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}} &\le E_{\boldsymbol{X}_D^{(N)}}\left[\sum_{Y^{(N)}}\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{1-s_1}\right. \\ &\quad \left.\times e^{-Ns_1\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}(\boldsymbol{w}_S), Y^{(N)}\right)}\right] \\ &= \sum_{Y^{(N)}} E_{\boldsymbol{X}_S^{(N)}}\left[E_{\boldsymbol{X}_{D\setminus S}^{(N)}}\left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{1-s_1}\right]\right. \\ &\quad \left.\times e^{-Ns_1\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, Y^{(N)}\right)}\right]. \end{aligned} \tag{B.53}$$

We again removed the message variable in the notation because, after taking the expectation operation, the result is not a function of the message.

Step III: Upper-bounding $P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]}e^{-N\alpha_{\tilde{\boldsymbol{g}}}}$, with $S \subseteq D$.

Under the assumption that $(\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ with $\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D$ is the actual message vector and code index vector pair, and with $\boldsymbol{g} \in \boldsymbol{R}_D$, we write $P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]}e^{-N\alpha_{\tilde{\boldsymbol{g}}}}$ as follows

$$P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]}e^{-N\alpha_{\tilde{\boldsymbol{g}}}} = E_{\boldsymbol{X}_D^{(N)}}\left[\sum_{Y^{(N)}} L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)\phi_{i[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\right],$$

(B.54)

where $\phi_{i[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ is an indicator function with $\phi_{i[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} = 1$ if there exists a $\boldsymbol{w}_D$, with $(\boldsymbol{w}_D, \boldsymbol{g}) \overset{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})$ and $\boldsymbol{g} \in \boldsymbol{R}_D$, to satisfy inequality $L_{\boldsymbol{g}}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_D}(\boldsymbol{w}_D), Y^{(N)}\right) > e^{-N\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_S}(\boldsymbol{w}_S), Y^{(N)}\right)}$, and $\phi_{i[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} = 0$ otherwise.

Let $s_2 \geq 0$ and $\tilde{\rho} > 0$ be two arbitrary constants. We can upper bound $\phi_{i[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}$ by

$$\phi_{i[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} \leq \left(\frac{\sum_{\boldsymbol{w}_D, (\boldsymbol{w}_D, \boldsymbol{g}) \overset{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})} \left[L_{\boldsymbol{g}}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_D}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}}{e^{-N\frac{s_2}{\tilde{\rho}}\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_S}(\boldsymbol{w}_S), Y^{(N)}\right)}} \right)^{\tilde{\rho}}. \tag{B.55}$$

Substitute (B.55) into (B.54) to obtain

$$\begin{aligned}
P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]} e^{-N\alpha_{\tilde{\boldsymbol{g}}}} &\leq E_{\boldsymbol{X}^{(N)}_D} \Bigg[\sum_{Y^{(N)}} \\
& L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_D}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right) e^{N s_2 \tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_S}(\boldsymbol{w}_S), Y^{(N)}\right)} \\
& \times \left(\sum_{\boldsymbol{w}_D, (\boldsymbol{w}_D, \boldsymbol{g}) \overset{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})} \left[L_{\boldsymbol{g}}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_D}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}} \right)^{\tilde{\rho}} \Bigg] \\
&= \sum_{Y^{(N)}} E_{\boldsymbol{X}^{(N)}_S} \Bigg[e^{N s_2 \tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_S}(\boldsymbol{w}_S), Y^{(N)}\right)} \\
& \times E_{\boldsymbol{X}^{(N)}_{D\setminus S}} \left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_D}(\tilde{\boldsymbol{w}}_D), Y^{(N)}\right)\right] \\
& \times E_{\boldsymbol{X}^{(N)}_{D\setminus S}} \left[\left(\sum_{\boldsymbol{w}_D, (\boldsymbol{w}_D, \boldsymbol{g}) \overset{S}{=} (\tilde{\boldsymbol{w}}_D, \tilde{\boldsymbol{g}})} \left[L_{\boldsymbol{g}}\left(\boldsymbol{X}^{(N)}_{\boldsymbol{g}_D}(\boldsymbol{w}_D), Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}} \right)^{\tilde{\rho}} \right] \Bigg]
\end{aligned} \tag{B.56}$$

where the last equality is due to the assumption that $(\tilde{w}_k, \tilde{g}_k) \neq (w_k, g_k)$, $\forall k \in D \setminus S$, and therefore codewords $\boldsymbol{X}^{(N)}_{\tilde{\boldsymbol{g}}_{D\setminus S}}(\tilde{\boldsymbol{w}}_{D\setminus S})$ and $\boldsymbol{X}^{(N)}_{\boldsymbol{g}_{D\setminus S}}(\boldsymbol{w}_{D\setminus S})$ are generated independently.

Now assume that $0 < \tilde{\rho} \le 1$, (B.56) further implies that

$$P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]} e^{-N\alpha_{\tilde{\boldsymbol{g}}}} \le \sum_{Y^{(N)}} E_{\boldsymbol{X}_S^{(N)}} \left[e^{Ns_2 \tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, Y^{(N)}\right)} \right.$$
$$\times E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)}\right)\right] e^{N\tilde{\rho}\sum_{k\in D\setminus S} r_{g_k}}$$
$$\left. \times \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}} \right] \right)^{\tilde{\rho}} \right]. \tag{B.57}$$

Step IV: Choosing $\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, Y^{(N)}\right)$, for $S \subseteq D$.

Given $\boldsymbol{g} \in \boldsymbol{R}_D$, S, $Y^{(N)}$, and auxiliary variables $s_1 \ge 0$, $s_2 \ge 0$, $0 < \tilde{\rho} \le 1$, we choose $\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, Y^{(N)}\right)$ to match the following pairs of terms on the right hand sides of (B.53) and (B.57),

$$E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{1-s_1} \right] e^{-Ns_1 \tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, Y^{(N)}\right)}$$
$$= e^{Ns_2 \tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, Y^{(N)}\right)} E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)}\right)\right]$$
$$\times e^{N\tilde{\rho}\sum_{k\in D\setminus S} r_{g_k}} \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}} \right] \right)^{\tilde{\rho}} \Bigg]. \tag{B.58}$$

(B.58) implies that

$$e^{-N\tau_{[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}\left(\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, Y^{(N)}\right)} =$$
$$\left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{1-s_1} \right] \right)^{-\frac{1}{s_1+s_2}}$$
$$\times \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}} \right] \right)^{\frac{\tilde{\rho}}{s_1+s_2}} e^{N\frac{\tilde{\rho}}{s_1+s_2}\sum_{k\in D\setminus S} r_{g_k}}$$
$$\times \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)}\right)\right] \right)^{\frac{1}{s_1+s_2}}. \tag{B.59}$$

Substitute (B.59) into (B.53), we get

$$P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}} \leq \sum_{Y^{(N)}} E_{\boldsymbol{X}_S^{(N)}}\Bigg[\left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}}\left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{1-s_1}\right]\right)^{\frac{s_2}{s_1+s_2}} \times \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}}\left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s_2}{\tilde{\rho}}}\right]\right)^{\frac{s_1\tilde{\rho}}{s_1+s_2}} e^{N\frac{s_1\tilde{\rho}}{s_1+s_2}\sum_{k\in D\setminus S} r_{g_k}} \times \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}}\left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)}\right)\right]\right)^{\frac{s_1}{s_1+s_2}}\Bigg]. \tag{B.60}$$

Let $s_2 < \tilde{\rho}$ and $s_1 = 1 - \frac{s_2}{\tilde{\rho}}$[2]. Do a variable change with $\rho = \frac{\tilde{\rho}(\tilde{\rho}-s_2)}{\tilde{\rho}-(1-\tilde{\rho})s_2}$ and $s = 1 - \frac{\tilde{\rho}-s_2}{\tilde{\rho}-(1-\tilde{\rho})s_2}$, which implies that $1 - s_1 = \frac{s_2}{\tilde{\rho}} = \frac{s}{s+\rho}$, $\frac{s_2}{s_1+s_2} = s$, $\frac{s_1\tilde{\rho}}{s_1+s_2} = \rho$, $\frac{s_1}{s_1+s_2} = 1 - s$. Inequality (B.60) then becomes

$$P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]}e^{-N\alpha_{\boldsymbol{g}}} \leq e^{N\rho\sum_{k\in D\setminus S} r_{g_k}} \sum_{Y^{(N)}} E_{\boldsymbol{X}_S^{(N)}}\Bigg[\left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}}\left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s}{s+\rho}}\right]\right)^{s+\rho} \times \left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}}\left[L_{\tilde{\boldsymbol{g}}}\left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)}\right)\right]\right)^{1-s}\Bigg]. \tag{B.61}$$

Let $\boldsymbol{X}_{\boldsymbol{g}_D n}^{(N)}$, $X_{kn}^{(N)}$ and $Y_n^{(N)}$ denote the nth symbol vector of $\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}$, the nth symbol of $X_k^{(N)}$, and the nth symbol of $Y^{(N)}$, respectively. We have

$$\left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}}\left[\left[L_{\boldsymbol{g}}\left(\boldsymbol{X}_{\boldsymbol{g}_D}^{(N)}, Y^{(N)}\right)\right]^{\frac{s}{s+\rho}}\right]\right)^{s+\rho} = \left(\sum_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[P(Y^{(N)}|\boldsymbol{X}_{\boldsymbol{g}_S}^{(N)}, \boldsymbol{X}_{D\setminus S}^{(N)}, \boldsymbol{g}_{\bar{D}})e^{-N\alpha_{\boldsymbol{g}}}\right]^{\frac{s}{s+\rho}} \times \prod_{k\in D\setminus S} P_{g_k}(X_k^{(N)})\right)^{s+\rho}$$

[2] This implies that $s_1 + s_2 > 0$ in (B.59) and (B.60).

$$= \left(\sum_{\boldsymbol{X}_{D\setminus S}^{(N)}} \prod_{n=1}^{N} \left[P(Y_n^{(N)} | \boldsymbol{X}_{\boldsymbol{g}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S n}^{(N)}, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{\frac{s}{s+\rho}} \right.$$

$$\left. \times \prod_{k \in D\setminus S} P_{g_k}(X_{kn}^{(N)}) \right)^{s+\rho}$$

$$= \prod_{n=1}^{N} \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y_n^{(N)} | \boldsymbol{X}_{\boldsymbol{g}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{\frac{s}{s+\rho}} \right.$$

$$\left. \times \prod_{k \in D\setminus S} P_{g_k}(X_k) \right)^{s+\rho}. \tag{B.62}$$

Similarly,

$$\left(E_{\boldsymbol{X}_{D\setminus S}^{(N)}} \left[\sum_{\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S = \boldsymbol{g}_S} L_{\tilde{\boldsymbol{g}}} \left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}_D}^{(N)}, Y^{(N)} \right) \right] \right)^{1-s}$$

$$= \prod_{n=1}^{N} \left(\sum_{\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S = \boldsymbol{g}_S} \sum_{\boldsymbol{X}_{D\setminus S}} P(Y_n^{(N)} | \boldsymbol{X}_{\tilde{\boldsymbol{g}}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \right.$$

$$\left. \times \prod_{k \in D\setminus S} P_{\tilde{g}_k}(X_k) \right)^{1-s}. \tag{B.63}$$

Consequently, (B.61) implies that

$$P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}} \le e^{N\rho \sum_{k\in D\setminus S} r_{g_k}} \sum_{Y^{(N)}} \sum_{\boldsymbol{X}_S^{(N)}} \prod_{n=1}^{N} \prod_{k\in S} P_{g_k}(X_{kn}^{(N)})$$

$$\left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y_n^{(N)} | \boldsymbol{X}_{\boldsymbol{g}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{\frac{s}{s+\rho}} \prod_{k \in D\setminus S} P_{g_k}(X_k) \right)^{s+\rho}$$

$$\times \left(\sum_{\boldsymbol{X}_{D\setminus S}} P(Y_n^{(N)} | \boldsymbol{X}_{\tilde{\boldsymbol{g}}_S n}^{(N)}, \boldsymbol{X}_{D\setminus S}, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \prod_{k \in D\setminus S} P_{\tilde{g}_k}(X_k) \right)^{1-s}$$

$$= e^{N\rho \sum_{k\in D\setminus S} r_{g_k}} \left\{ \sum_{Y} \sum_{\boldsymbol{X}_S} \prod_{k\in S} P_{g_k}(X_k) \right.$$

$$\times \left(\sum_{\boldsymbol{X}_{D\setminus S}} \left[P(Y|\boldsymbol{X}_D, \boldsymbol{g}_{\bar{D}}) e^{-\alpha_{\boldsymbol{g}}} \right]^{\frac{s}{s+\rho}} \prod_{k\in D\setminus S} P_{g_k}(X_k) \right)^{s+\rho}$$
$$\left.\times \left(\sum_{\boldsymbol{X}_{D\setminus S}} P(Y|\boldsymbol{X}_D, \tilde{\boldsymbol{g}}_{\bar{D}}) e^{-\alpha_{\tilde{\boldsymbol{g}}}} \prod_{k\in D\setminus S} P_{\tilde{g}_k}(X_k) \right)^{1-s} \right\}^N . \quad \text{(B.64)}$$

Similarly, by substituting (B.59) into (B.57), and with the same derivations, we also get the same bound for $P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]} e^{-N\alpha_{\tilde{\boldsymbol{g}}}}$.

Because s and ρ can take any value with the constraints of $0 < \rho \leq 1$ and $0 \leq s \leq 1-\rho$, the bounds further lead to

$$\begin{aligned} P_{t[\boldsymbol{g},\tilde{\boldsymbol{g}},S]} e^{-N\alpha_{\boldsymbol{g}}} &\leq \exp(-NE_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S)) \\ P_{i[\tilde{\boldsymbol{g}},\boldsymbol{g},S]} e^{-N\alpha_{\tilde{\boldsymbol{g}}}} &\leq \exp(-NE_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S)), \end{aligned} \quad \text{(B.65)}$$

where $\exp(-NE_{iD}(\boldsymbol{g},\tilde{\boldsymbol{g}},S))$ is given in (3.13).

Finally, substituting (B.49) and (B.65) into (B.41) yields the conclusion of the theorem.

B.3 Proof of Theorem 3.4

We assume the following decoding algorithm. Given the partition σ, let the receiver be equipped with $(D, \boldsymbol{R}_D)$ decoders corresponding to all $D \subseteq \{1, \ldots, K\}$ with $1 \in D$. After receiving the channel output symbol sequence $Y^{(N)}$, the receiver first carries out all the $(D, \boldsymbol{R}_D)$-decoding operations. If at least one $(D, \boldsymbol{R}_D)$ decoder outputs an estimated message for user 1 and decoding outputs of all $(D, \boldsymbol{R}_D)$ decoders agree on the message and code index estimate for user 1, then the receiver should output the corresponding estimate of user 1. Otherwise, the receiver reports collision for user 1.

Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code vector pair. We will show in the following that, irrespective of the value of $\boldsymbol{g}$, the receiver should give an acceptable output if all the $(D, \boldsymbol{R}_D)$ decoding operations are carried out correctly.

First, if $\boldsymbol{g} \in \boldsymbol{R}_1$, because the operation regions corresponding to all $D \subseteq \{1, \ldots, K\}$ with $1 \in D$ form a partition of $\boldsymbol{R}_1$, there must exist a user subset $\hat{D}$ with $1 \in \hat{D}$ such that $\boldsymbol{g} \in \boldsymbol{R}_{\hat{D}}$. Therefore, if the corresponding $(\hat{D}, \boldsymbol{R}_{\hat{D}})$ decoder operates correctly, it should output the correct message vector and code index vector estimate $(\boldsymbol{w}_{\hat{D}}, \boldsymbol{g}_{\hat{D}})$ for all regular users in $\hat{D}$, and this includes the correct message and code index estimate (w_1, g_1) for user 1. For any other user subset $D \neq \hat{D}$, if the corresponding $(D, \boldsymbol{R}_D)$ decoder operates correctly, it should either output the correct message and code index estimate for user 1, or output collision report for user 1. Therefore, under the assumption that all $(D, \boldsymbol{R}_D)$ decoders operate correctly, there is at least one correct message and code index output, and all decoding outputs agree with each other on user 1. Therefore, the receiver should output the correct message and code index for user 1.

Second, if $\boldsymbol{g} \in \widehat{\boldsymbol{R}}_1$, under the assumption that all $(D, \boldsymbol{R}_D)$ decoders operate correctly, the receiver should either output the correct message and code index for user 1, or report collision for user 1.

Third, if $\boldsymbol{g} \notin \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1$, we must have $\boldsymbol{g} \notin \boldsymbol{R}_D \cup \widehat{\boldsymbol{R}}_D$ for all $D \subseteq \{1, \ldots, K\}$ with $1 \in D$. Therefore, under the assumption that all $(D, \boldsymbol{R}_D)$ decoders operate correctly, all $(D, \boldsymbol{R}_D)$ decoders should output collision, and consequently the receiver should output collision for user 1.

Let $P_e(\boldsymbol{g})$ be the conditional error probability of the system. Let us use $P_{eD}(\boldsymbol{g})$ to denote the conditional error probability of the $(D, \boldsymbol{R}_D)$ decoder with user subset D. According to the above analysis and the union bound, we must have

$$P_e(\boldsymbol{g}) \leq \sum_{D, D \subseteq \{1,\ldots,K\}, 1 \in D} P_{eD}(\boldsymbol{g}). \tag{B.66}$$

Consequently, with the set of weight parameters $\{\alpha_{\boldsymbol{g}}\}$, we have

$$\begin{aligned} \text{GEP} &= \sum_{\boldsymbol{g}} P_e(\boldsymbol{g}) e^{-N\alpha_{\boldsymbol{g}}} \\ &\leq \sum_{D, D \subseteq \{1,\ldots,K\}, 1 \in D} \sum_{\boldsymbol{g}} P_{eD}(\boldsymbol{g}) e^{-N\alpha_{\boldsymbol{g}}} \\ &= \sum_{D, D \subseteq \{1,\ldots,K\}, 1 \in D} \text{GEP}_D. \end{aligned} \tag{B.67}$$

Conclusion of the theorem follows because (B.67) holds for all partition σ.

B.4 Proof of Theorem 3.5

Given channel output sequence $Y^{(N)}$, we define the weighted likelihood of code index vector $\boldsymbol{g}$ by $L_{\boldsymbol{g}}(Y^{(N)})$, as follows.

$$L_{\boldsymbol{g}}(Y^{(N)}) = P(Y^{(N)}|\boldsymbol{g})e^{-N\alpha_{\boldsymbol{g}}}. \tag{B.68}$$

Assume that the receiver finds $\hat{\boldsymbol{g}} = \text{argmax}_{\boldsymbol{g}} L_{\boldsymbol{g}}(Y^{(N)})$ with the maximum weighted likelihood. If $\hat{\boldsymbol{g}} \in \boldsymbol{R}$, the receiver outputs "true"; if $\hat{\boldsymbol{g}} \in \widehat{\boldsymbol{R}}$, the receiver outputs "neutral"; while if $\hat{\boldsymbol{g}} \notin \boldsymbol{R} \cup \widehat{\boldsymbol{R}}$ the receiver reports "false".

Let $\boldsymbol{g}$ be the actual code index vector. Define $P_{[\boldsymbol{g},\tilde{\boldsymbol{g}}]}$ as the probability that weighted likelihood of the actual code index vector is smaller than that of another code index vector $\tilde{\boldsymbol{g}}$.

$$P_{[\boldsymbol{g},\tilde{\boldsymbol{g}}]} = Pr\left\{L_{\tilde{\boldsymbol{g}}}(Y^{(N)}) > L_{\boldsymbol{g}}(Y^{(N)})\right\}. \tag{B.69}$$

General error performance of the system can therefore be upper bounded by

$$\text{GEP} \leq \sum_{\boldsymbol{g}\in\boldsymbol{R}} \sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}\cup\widehat{\boldsymbol{R}}} \left(P_{[\boldsymbol{g},\tilde{\boldsymbol{g}}]}e^{-N\alpha_{\boldsymbol{g}}} + P_{[\tilde{\boldsymbol{g}},\boldsymbol{g}]}e^{-N\alpha_{\tilde{\boldsymbol{g}}}}\right). \tag{B.70}$$

For any $s \geq 0$, we can upper bound $P_{[\boldsymbol{g},\tilde{\boldsymbol{g}}]}e^{-N\alpha_{\boldsymbol{g}}}$ by

$$\begin{aligned}
P_{[\boldsymbol{g},\tilde{\boldsymbol{g}}]}e^{-N\alpha_{\boldsymbol{g}}} &\leq \sum_{Y^{(N)}} L_{\boldsymbol{g}}(Y^{(N)}) \left(\frac{L_{\tilde{\boldsymbol{g}}}(Y^{(N)})}{L_{\boldsymbol{g}}(Y^{(N)})}\right)^s \\
&= \sum_{Y^{(N)}} [L_{\boldsymbol{g}}(Y^{(N)})]^{(1-s)}[L_{\tilde{\boldsymbol{g}}}(Y^{(N)})]^s \\
&= \left(\sum_{Y}[P(Y|\boldsymbol{g})e^{-\alpha_{\boldsymbol{g}}}]^{(1-s)}[P(Y|\tilde{\boldsymbol{g}})e^{-\alpha_{\tilde{\boldsymbol{g}}}}]^s\right)^N.
\end{aligned} \tag{B.71}$$

Substitute (B.71) into (B.70) gives the bound in (3.24).

C

Proofs of Theorems in Section 4

C.1 Proof of Theorem 4.3

The partial derivative of $q_v(p, K)$ with respect to p is given by

$$\begin{aligned}\frac{\partial q_v(p,K)}{\partial p} &= \sum_{j=0}^{K} \binom{K}{j} j p^{j-1}(1-p)^{K-j} C_{vj} \\ &\quad - \sum_{j=0}^{K} \binom{K}{j} p^j (K-j)(1-p)^{K-j-1} C_{vj} \\ &= -\sum_{j=0}^{K-1} K \binom{K-1}{j} p^j (1-p)^{K-1-j}(C_{vj} - C_{v(j+1)}) \\ &\leq 0, \end{aligned} \tag{C.1}$$

where the last inequality is due to the assumption that $C_{vj} \geq C_{v(j+1)}$ for all $j \geq 0$. Note that (C.1) holds with strict inequality if $K > J_{\epsilon_v}$ and $p(1-p) \neq 0$.

C.2 Proof of Theorem 4.4

Let us first consider the situation when $\frac{x^*}{N+b} \leq p_{\max}$.

According to the definition of $q_v^*(\hat{p})$ in (4.16), we have

$$\frac{dq_v^*(\hat{p})}{d\hat{p}} = \frac{q_N(\hat{p}) - q_{N+1}(\hat{p})}{p_N - p_{N+1}} + \frac{\hat{p} - p_{N+1}}{p_N - p_{N+1}} \frac{dq_N(\hat{p})}{d\hat{p}} + \frac{p_N - \hat{p}}{p_N - p_{N+1}} \frac{dq_{N+1}(\hat{p})}{d\hat{p}}. \tag{C.2}$$

Write $\hat{K} = N + 1 - \lambda$ with $\lambda \in (0, 1]$. We have

$$\hat{p} - p_{N+1} = \frac{x^*}{\hat{K} + b} - \frac{x^*}{N + 1 + b} = \frac{\lambda}{N + 1 + b}\hat{p}, \tag{C.3}$$

and

$$p_N - \hat{p} = \frac{x^*}{N + b} - \frac{x^*}{\hat{K} + b} = \frac{1 - \lambda}{N + b}\hat{p}. \tag{C.4}$$

Meanwhile, because function $q_{N+1}(\hat{p})$ can be decomposed as

$$\begin{aligned} q_{N+1}(\hat{p}) &= \sum_{j=0}^{N+1} \binom{N+1}{j} \hat{p}^j (1 - \hat{p})^{N+1-j} C_{vj} \\ &= \hat{p} \sum_{j=0}^{N} \binom{N}{j} \hat{p}^j (1 - \hat{p})^{N-j} C_{v(j+1)} \\ &\quad + (1 - \hat{p}) \sum_{j=0}^{N} \binom{N}{j} \hat{p}^j (1 - \hat{p})^{N-j} C_{vj}, \end{aligned} \tag{C.5}$$

we have

$$q_N - q_{N+1} = \hat{p} \sum_{j=0}^{N} \binom{N}{j} \hat{p}^j (1 - \hat{p})^{N-j} (C_{vj} - C_{v(j+1)}). \tag{C.6}$$

Furthermore, the derivatives of $q_N(\hat{p})$ and $q_{N+1}(\hat{p})$ are given by

$$\frac{dq_N(\hat{p})}{d\hat{p}} = -\sum_{j=0}^{N} (N - j) \binom{N}{j} \hat{p}^j (1 - \hat{p})^{N-j-1} (C_{vj} - C_{v(j+1)}), \tag{C.7}$$

and

$$\frac{dq_{N+1}(\hat{p})}{d\hat{p}} = -\sum_{j=0}^{N} (N + 1) \binom{N}{j} \hat{p}^j (1 - \hat{p})^{N-j} (C_{vj} - C_{v(j+1)}). \tag{C.8}$$

Substitute the above results into (C.2), we get

$$(p_N - p_{N+1}) \frac{dq_v^*(\hat{p})}{\hat{p}} = \hat{p} \sum_{j=0}^{N} \binom{N}{j} \hat{p}^j (1 - \hat{p})^{N-j} (C_{vj} - C_{v(j+1)})$$

$$
\begin{aligned}
&-\frac{\lambda}{N+1+b}\hat{p}\sum_{j=0}^{N}\binom{N}{j}\hat{p}^{j}(1-\hat{p})^{N-j-1}(N-j)(C_{vj}-C_{v(j+1)}) \\
&-\frac{1-\lambda}{N+b}\hat{p}\sum_{j=0}^{N}(N+1)\binom{N}{j}\hat{p}^{j}(1-\hat{p})^{N-j}(C_{vj}-C_{v(j+1)}) \\
&=\hat{p}\sum_{j=0}^{N}\binom{N}{j}\hat{p}^{j}(1-\hat{p})^{N-j-1}(C_{vj}-C_{v(j+1)}) \\
&\quad\times\left(1-\hat{p}-\frac{\lambda(N-j)}{N+1+b}-\frac{(1-\lambda)(1-\hat{p})(N+1)}{N+b}\right) \\
&=\hat{p}\sum_{j=0}^{N}\binom{N}{j}\hat{p}^{j}(1-\hat{p})^{N-j-1}(C_{vj}-C_{v(j+1)}) \\
&\quad\times\left(\frac{\lambda((1-\hat{p})(N+1+b)-N+j)}{N+1+b}+\frac{(1-\lambda)(1-\hat{p})(b-1)}{N+b}\right).
\end{aligned}
\tag{C.9}
$$

Note that, for all $j\geq 0$, we have

$$
\begin{aligned}
&\frac{\lambda((1-\hat{p})(N+1+b)-N+j)}{N+1+b}\geq\frac{\lambda((1-p_N)(N+1+b)-N+j)}{N+1+b} \\
&\geq\frac{\lambda(b-x^*+j)}{N+1+b}.
\end{aligned}
\tag{C.10}
$$

Therefore, $\frac{dq_v^*(\hat{p})}{d\hat{p}}\geq 0$ if $b\geq 1$ and the following inequality is satisfied.

$$
\sum_{j=0}^{N}\binom{N}{j}\hat{p}^{j}(1-\hat{p})^{N-j-1}(C_{vj}-C_{v(j+1)})(b-x^*+j)\geq 0. \tag{C.11}
$$

It is easy to see that (C.11) holds if $b\geq x^*-\gamma_{\epsilon_v}$, with γ_{ϵ_v} being defined in (4.18).

Furthermore, if we have both $b>1$ and $b>x^*-J_{\epsilon_v}$ holding with strict inequality, and $C_{vj}>C_{v(j+1)}$ for at least one $j\leq N$, then $\frac{dq_v^*(\hat{p})}{d\hat{p}}>0$ should also hold with strict inequality for $\hat{p}\in(0,p_{\max})$.

Next, consider the case when $\frac{x^*}{N+b}\geq p_{\max}$. It is easy to see that $\frac{dq_v^*(\hat{p})}{d\hat{p}}=0$ if $\frac{x^*}{\hat{K}+b}\geq p_{\max}$. If $\frac{x^*}{\hat{K}+b}<p_{\max}$ but $\frac{x^*}{N+b}\geq p_{\max}$ on the other hand, we can write $\hat{K}=N+1-\lambda$ with $0<\lambda\leq N+1+b-\frac{x^*}{p_{\max}}$.

Consequently, (C.2) and (C.3) still hold. But (C.4) should be changed to

$$p_N - \hat{p} = p_{\max} - \frac{x^*}{\hat{K} + b} \leq \frac{1-\lambda}{N+b}\hat{p}. \tag{C.12}$$

As a result, (C.9) becomes

$$(p_N - p_{N+1})\frac{dq_v^*(\hat{p}}{\hat{p}} \geq \hat{p}\sum_{j=0}^{N}\binom{N}{j}\hat{p}^j(1-\hat{p})^{N-j-1}(C_{vj} - C_{v(j+1)})$$
$$\times\left(\frac{\lambda((1-\hat{p})(N+1+b) - N + j)}{N+1+b} + \frac{(1-\lambda)(1-\hat{p})(b-1)}{N+b}\right). \tag{C.13}$$

By following the rest of the derivations, it can be seen that conclusion of the theorem still holds.

C.3 Proof of Theorem 4.5

First, because $b > \max\{1, x^* - J_{\epsilon_v}\}$ holds with strict inequality, the theoretical channel contention measure $q_v^*(\hat{p})$ is strictly increasing in $\hat{p}$ for $\hat{p} \in (0, p_{\max})$. Given user number K, $q_v(\hat{p}, K)$ is non-increasing in $\hat{p}$. Therefore, if $K \geq J_{\epsilon_v}$, then $\hat{p} = p^* = \frac{x^*}{K+b}$ is the only solution to $q_v(\hat{p}, K) = q_v^*(\hat{p})$. When $K < J_{\epsilon_v}$ on the other hand, we have $q_v(\hat{p}, K) > q_v^*(\hat{p})$ for all $\hat{p} \in [0, p_{\max})$. This implies that $\boldsymbol{p}^* = \min\{p_{\max}, \frac{x^*}{K+b}\}\mathbf{1}$ is the only equilibrium of the system.

Second, we show that there exists a constant $\epsilon > 0$, such that $\frac{dq_v^*(\hat{p})}{d\hat{p}} \geq \epsilon > 0$ for all $\hat{p} < p_{\max}$. Note that $\hat{p} < p_{\max}$ implies $\hat{K} > J_{\epsilon_v}$. From (C.9) and (C.10), we get

$$\frac{dq_v^*(\hat{p})}{\hat{p}} \geq \frac{\hat{p}}{p_N - p_{N+1}}\binom{N}{J_{\epsilon_v}}\hat{p}^{J_{\epsilon_v}}(1-\hat{p})^{N-J_{\epsilon_v}-1}(C_{vJ_{\epsilon_v}} - C_{v(J_{\epsilon_v}+1)})$$
$$\times\left(\frac{\lambda(b - x^* + J_{\epsilon_v})}{N+1+b} + \frac{(1-\lambda)(1-\hat{p})(b-1)}{N+b}\right). \tag{C.14}$$

Because the right hand side of (C.14) has a positive limit when $\hat{p} \to 0$, we can find two small positive constants $\epsilon_0, \epsilon_1 > 0$, such that $\frac{dq_v^*(\hat{p})}{\hat{p}} \geq \epsilon_0$ for all $\hat{p} \leq \epsilon_1$. On the other hand, when $\epsilon_1 \leq \hat{p} < p_{\max}$, because $b > \max\{1, x^* - \gamma_{\epsilon_v}\}$ holds with strict inequality, we can find a small positive constant $\epsilon_2 > 0$, such that the right hand side of (C.14) is no less than ϵ_2. Therefore, by choosing $\epsilon = \min\{\epsilon_0, \epsilon_2\}$, we have

$$\frac{dq_v^*(\hat{p})}{d\hat{p}} \geq \epsilon > 0, \qquad \text{for all } \hat{p} < p_{\max}. \tag{C.15}$$

Third, let $q_v^{*-1}(.)$ be the inverse function of $q_v^*(p)$. For any given transmission probability vector $\boldsymbol{p}$, transmission probability target $\hat{p}$ is obtained by

$$\hat{p} = q_v^{*-1}(q_v) = q_v^{*-1}(q_v(\boldsymbol{p}, K)). \tag{C.16}$$

Because $\frac{dq_v^*(\hat{p})}{d\hat{p}} \geq \epsilon > 0$, we can find a constant $K_{l1} > 0$ such that

$$|\hat{p}_1 - \hat{p}_2| \leq K_{l1}|q_{v1} - q_{v2}|, \tag{C.17}$$

for all $\hat{p}_1 = q_v^{*-1}(q_{v1})$ and $\hat{p}_2 = q_v^{*-1}(q_{v2})$. In the meantime, since $q_v = q_v(\boldsymbol{p}, K)$ is Lipschitz continuous in $\boldsymbol{p}$ for any given K, there must exist a constant $K_{l2} > 0$ to satisfy

$$|q_{v1} - q_{v2}| \leq K_{l2}\|\boldsymbol{p}_1 - \boldsymbol{p}_2\|, \tag{C.18}$$

for all $q_{v1} = q_v(\boldsymbol{p}_1, K)$ and $q_{v2} = q_v(\boldsymbol{p}_2, K)$. Consequently, by combining (C.17) and (C.18), we have

$$|\hat{p}_1 - \hat{p}_2| \leq K_{l1}K_{l2}\|\boldsymbol{p}_1 - \boldsymbol{p}_2\|, \tag{C.19}$$

for all $\hat{p}_1 = q_v^{*-1}(q_v(\boldsymbol{p}_1, K))$ and $\hat{p}_2 = q_v^{*-1}(q_v(\boldsymbol{p}_2, K))$. This implies that the probability target function given in (C.16) satisfies the Lipschitz Condition 2.

Finally, when the system is noisy, the receiver can choose to measure q_v over an extended number of time slots, namely increasing the value of Q introduced in Step 2 of the proposed MAC algorithm. If users maintain their transmission probabilities during the Q times slots, it is often the case that the potential measurement bias in the system can be reduced arbitrarily close to zero with a large enough Q. Therefore, the Mean and Bias Condition 1 is also satisfied.

Consequently, convergence of the distributed probability adaptation is supported by Theorems 4.1 and 4.2.

C.4 Proof of Theorem 4.8

We first show that the associated ODE of the system should have a unique equilibrium at $\boldsymbol{P}^* = \mathbf{1} \otimes \boldsymbol{p}(K)$. According to Condition 8 and Theorem 4.4, $q_v^*(\hat{K})$ should be strictly decreasing in $\hat{K}$ for $\hat{K} \geq J_{\epsilon_v}(\boldsymbol{w}(\underline{K}), \boldsymbol{d}(\underline{K}))$. If (4.42) has two solutions at $\hat{K}_1$ and $\hat{K}_2$ with $\hat{K}_1 < \hat{K}_2$, it implies that

$$[\boldsymbol{w}(\hat{K}_1) - \boldsymbol{w}(\hat{K}_2)]^T \boldsymbol{q}_v = q_v^*(\hat{K}_1) - q_v^*(\hat{K}_2). \tag{C.20}$$

Because $\boldsymbol{q}_v$ satisfies $q_{v1} \leq q_{v2} \leq \ldots \leq q_{vV}$, (C.20) implies that there must exists a $j \leq V$ with

$$\sum_{i=j}^{V} w_i(\hat{K}_1) - \sum_{i=j}^{V} w_i(\hat{K}_2) \geq q_v^*(\hat{K}_1) - q_v^*(\hat{K}_2), \tag{C.21}$$

which contradicts with the Majorization Condition (4.44). Therefore, (4.42) can have at most one solution. This implies that any equilibrium of the ODE must take the form of $\boldsymbol{P}^* = \mathbf{1} \otimes \boldsymbol{p}(\hat{K})$ for some $\hat{K} \geq J_{\epsilon_v}(\boldsymbol{w}(\underline{K}), \boldsymbol{d}(\underline{K}))$.

Assume that the actual user number satisfies $\underline{K} < K < \overline{K}$. With all the users setting their transmission probability vectors at $\boldsymbol{p}(\hat{K})$, due to Items 3 and 4 in Condition 8, if $K > \hat{K}$ and $\hat{K}$ is an integer, we must have

$$\boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), K) < \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), \lfloor \hat{K} \rfloor) = q_v^*(\hat{K}). \tag{C.22}$$

When $K > \hat{K}$ and $\hat{K}$ is not an integer, we have

$$\begin{aligned} \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), K) &< \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), \lfloor \hat{K} \rfloor) \\ \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), K) &\leq \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), \lfloor \hat{K} \rfloor + 1), \end{aligned} \tag{C.23}$$

which imply that

$$\boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), K) < q_v^*(\hat{K}). \tag{C.24}$$

On the other hand, if $K < \hat{K}$ and $\hat{K}$ is an integer, we must have

$$\boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), K) > \boldsymbol{w}(\hat{K})^T \boldsymbol{q}_v(\boldsymbol{p}(\hat{K}), \hat{K}) = q_v^*(\hat{K}). \tag{C.25}$$

When $K < \hat{K}$ and $\hat{K}$ is not an integer, we have

$$\boldsymbol{w}(\hat{K})^T\boldsymbol{q}_v(\boldsymbol{p}(\hat{K}),K) > \boldsymbol{w}(\hat{K})^T\boldsymbol{q}_v(\boldsymbol{p}(\hat{K}),\lfloor\hat{K}\rfloor+1)$$
$$\boldsymbol{w}(\hat{K})^T\boldsymbol{q}_v(\boldsymbol{p}(\hat{K}),K) \geq \boldsymbol{w}(\hat{K})^T\boldsymbol{q}_v(\boldsymbol{p}(\hat{K}),\lfloor\hat{K}\rfloor), \qquad \text{(C.26)}$$

which also imply that

$$\boldsymbol{w}(\hat{K})^T\boldsymbol{q}_v(\boldsymbol{p}(\hat{K}),K) > q_v^*(\hat{K}). \qquad \text{(C.27)}$$

Consequently, (4.42) must have a unique solution at $\hat{K} = K$. When $K \leq \underline{K}$ or $K \geq \overline{K}$ on the other hand, uniqueness of the solution to (4.42) can be seen by following the proof of Theorem 4.5.

Next, according to Condition 8, for $\underline{K} \leq \hat{K} \leq \overline{K}$, $\boldsymbol{p}(\hat{K})$ is Lipschitz continuous in $\hat{K}$, and $q_v^*(\hat{K})$ satisfies (4.33). Combined with the Head and Tail Condition 6 and the fact that $p(\hat{K})$ is designed for $\hat{K} \leq \underline{K}$ and $\hat{K} \geq \overline{K}$ according to the guideline given in Section 4.2, we conclude that $\hat{\boldsymbol{p}}(\boldsymbol{q}_v)$ is Lipschitz continuous in $\boldsymbol{q}_v$. Because $\boldsymbol{q}_v(\boldsymbol{P},K)$ is also Lipschitz continuous in $\boldsymbol{P}$, $\hat{\boldsymbol{p}}$ must be Lipschitz continuous in $\boldsymbol{P}$, which satisfies Condition 2.

Finally, the Mean and Bias Condition is also satisfied because, by assumption, one can increase Q in Step 2 of the proposed MAC algorithm to reduce the potential measurement bias in $\boldsymbol{q}_v$ arbitrarily close to zero.

C.5 Proof of Theorem 4.9

Because Item 2 in Condition 7 and Items 2, 3, and 4 in Condition 8 hold by assumption, we only need to prove Item 1 in Condition 8. That is, with the Interpolation Approach, $\boldsymbol{p}(\hat{K})$ should be Lipschitz continuous in $\hat{K}$. To avoid unnecessary notation complication, we use $\frac{dp(\hat{K})}{d\hat{K}}$ to represent the derivative of $p(\hat{K})$ if $p(\hat{K})$ is differentiable at $\hat{K}$. If $p(\hat{K})$ is only continuous but not differentiable at $\hat{K}$, then $\frac{dp(\hat{K})}{d\hat{K}}$ represents one or an arbitrary subderivative of $p(\hat{K})$. If $p(\hat{K})$ is not continuous at $\hat{K}$, then $\frac{dp(\hat{K})}{d\hat{K}}$ should take the values of $\pm\infty$. Note that the adoption such a notation does not imply a continuity assumption on $p(\hat{K})$.

Let $i \in \{1,\ldots,L\}$ and $0 \leq \lambda < 1$ be chosen arbitrarily. Let $\hat{K} = \hat{K}_{i\lambda}$. To simplify the discussion, we assume that the neighboring two pinpoints satisfy $\hat{K}_{i+1} = \hat{K}_i + 1$, i.e., they are one integer apart from each other[1]. Consequently, by writing $\hat{K} = (1-\lambda)\hat{K}_i + \lambda\hat{K}_{i+1}$ as a function of λ, we have $\frac{dp(\hat{K})}{d\hat{K}} = \frac{dp(\lambda)}{d\lambda}$.

To bound $\frac{dp(\lambda)}{d\lambda}$, we consider two different expressions of $q_v^*(\hat{K})$. The first expression is given by

$$q_v^*(\lambda) = (1-\lambda)q_v^*(\hat{K}_i) + \lambda q_v^*(\hat{K}_{i+1}). \tag{C.28}$$

Taking derivative with respect to λ, we get $\frac{dq_v^*(\lambda)}{d\lambda} = q_v^*(\hat{K}_{i+1}) - q_v^*(\hat{K}_i)$. Because both $q_v^*(\hat{K}_{i+1})$ and $q_v^*(\hat{K}_i)$ are bounded, we can find a constant $0 < \overline{\Delta}_1$, such that

$$\left|\frac{dq_v^*(\lambda)}{d\lambda}\right| \leq \overline{\Delta}_1. \tag{C.29}$$

Next, consider the second expression of $q_v^*(\hat{K})$ given below.

$$q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda}) = \boldsymbol{w}_{i\lambda}^T[(1-\lambda)\boldsymbol{q}_v(p_{i\lambda}\boldsymbol{d}_{i\lambda}, \hat{K}_i) + \lambda\boldsymbol{q}_v(p_{i\lambda}\boldsymbol{d}_{i\lambda}, \hat{K}_{i+1})]. \tag{C.30}$$

Taking derivative with respect to λ, we get

$$\begin{aligned} \frac{dq_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{d\lambda} &= \frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial\lambda} \\ &+ \left[\frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial\boldsymbol{w}_{i\lambda}}\right]^T \frac{d\boldsymbol{w}_{i\lambda}}{d\lambda} + \left[\frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial\boldsymbol{d}_{i\lambda}}\right]^T \frac{d\boldsymbol{d}_{i\lambda}}{d\lambda} \\ &+ \frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial p_{i\lambda}}\frac{dp_{i\lambda}}{d\lambda}. \end{aligned} \tag{C.31}$$

Let us consider each of the terms on the right hand side of (C.31). First, we have

$$\frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial\lambda} = \boldsymbol{w}_{i\lambda}^T[\boldsymbol{q}_v(p_{i\lambda}\boldsymbol{d}_{i\lambda}, \hat{K}_{i+1}) - \boldsymbol{q}_v(p_{i\lambda}\boldsymbol{d}_{i\lambda}, \hat{K}_i)]. \tag{C.32}$$

Because all terms on the right hand side of (C.32) are bounded, there exists a constant $\overline{\Delta}_2 > 0$ such that

$$\left|\frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial\lambda}\right| \leq \overline{\Delta}_2. \tag{C.33}$$

[1]The proof can be easily extended to the case when this assumption does not hold.

Second, because

$$\left\| \frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial \boldsymbol{w}_{i\lambda}} \right\| \leq \|\boldsymbol{q}_v(p_{i\lambda}\boldsymbol{d}_{i\lambda}, \hat{K}_i)\| \leq V, \tag{C.34}$$

there exists a constant $\overline{\Delta}_3 > 0$, such that

$$\left| \left[\frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial \boldsymbol{w}_{i\lambda}} \right]^T \frac{d\boldsymbol{w}_{i\lambda}}{d\lambda} \right| \leq \overline{\Delta}_3. \tag{C.35}$$

Third, by following a derivation similar to (C.1), we can see that there exists a constant $\overline{\Delta}_4 > 0$, such that

$$\left| \left[\frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial \boldsymbol{d}_{i\lambda}} \right]^T \frac{d\boldsymbol{d}_{i\lambda}}{d\lambda} \right| \leq \overline{\Delta}_4. \tag{C.36}$$

Fourth, according to Item 2 and 3 in Condition 9, we have $\hat{K}_{i\lambda} > \hat{K}_i \geq J_{\epsilon_v}(\boldsymbol{w}_{i\lambda}, \boldsymbol{d}_{i\lambda})$, and $\underline{p} \leq p_{i\lambda} \leq \overline{p}$. From the derivation of (C.1), we can see that there exists a constant $\underline{\Delta}_1 > 0$, such that

$$\left| \frac{\partial q_v^*(\lambda, \boldsymbol{w}_{i\lambda}, p_{i\lambda}\boldsymbol{d}_{i\lambda})}{\partial p_{i\lambda}} \right| \geq \underline{\Delta}_1. \tag{C.37}$$

Because (C.28) and (C.30) must equal each other, combining (C.33), (C.35), (C.36) and (C.37), we can see that there exists a constant $K_g > 0$, such that $\left\| \frac{d\boldsymbol{p}(\hat{K})}{d\hat{K}} \right\| \leq K_g$. With the extended definition of $\frac{d\boldsymbol{p}(\hat{K})}{d\hat{K}}$, as explained at the beginning of the proof, $\left\| \frac{d\boldsymbol{p}(\hat{K})}{d\hat{K}} \right\| \leq K_g$ means that $\boldsymbol{p}(\hat{K})$ is Lipschitz continuous in $\hat{K}$.

References

[1] Ahlswede, R. 1971. “Multi-way Communication Channels”. In: *IEEE ISIT*. Tsahkadsor, Armenia, USSR. 23–52.

[2] Ahlswede, R., N. Cai, S. Li, and R. Yeung. 2000. “Network Information Flow”. *IEEE Transactions on Information Theory*. 46(July): 1204–1216.

[3] Ahmadi, S. 2009. “An Overview of Next-generation Mobile WiMAX Technology”. *IEEE Communications Magazine*. 47(June): 84–98.

[4] Anantharam, V. 1991. “The Stability Region of The Finite-user Slotted ALOHA Protocol”. *IEEE Transactions on Information Theory*. 37(May): 535–540.

[5] Anantharam, V. and S. Verdu. 1996. “Bits Through Queues”. *IEEE Transactions on Information Theory*. 42(Jan.): 4–18.

[6] Baccelli, F. and B. Blaszczyszyn. 2009. “Stochastic Geometry and Wireless Networks”. *Foundations and Trends in Networking*. 3: 249–449.

[7] Beharghaven, V., A. Demers, S. Shenker, and L. Zhang. 1994. “MACAW: A Media Access Protocol for Wireless LAN’s”. In: *ACM SIGCOMM*. London, UK.

[8] Berger, T. 1977. *Multiterminal Source Coding*. G. Longo, Ed. New York: Springer-Verlag.

[9] Bergmans, P. 1973. “Random Coding Theorem for Broadcast Channels with Degraded Components”. *IEEE Transactions on Information Theory*. 19(Mar.): 197–207.

[10] Bergmans, P. 1974. “A Simple Converse for Broadcast Channels with Additive White Gaussian Noise”. *IEEE Transactions on Information Theory*. 20(Mar.): 279–280.

[11] Bertsekas, D. and R. Gallager. 1992. *Data Networks*. 2nd. Prentice Hall.

[12] Bianchi, G. 2000. “Performance Analysis of the IEEE 802.11 Distributed Coordination Function”. *IEEE Journal on Selected Areas in Communications*. 18(Mar.): 535–547.

[13] Blokh, E. and V. Zyablov. 1982 (In Russian). *Linear Concatenated Codes*. Moscow: Nauka.

[14] Borkar, V. and S. Meyn. 2000. “The O.D.E Method for Convergence of Stochastic Approximation and Reinforcement Learning”. *SIAM Journal on Control and Optimization*. 38(Jan.): 447–469.

[15] Byers, J., M. Luby, and A. Rege. 1998. “A Digital Fountain Approach to Reliable Distribution of Bulk Data”. In: *ACM SIGCOMM*. Vancouver, Canada.

[16] Capetanakis, J. 1979. “Tree Algorithms for Packets Broadcast Channel”. *IEEE Transactions on Information Theory*. IT-25(Sept.): 505–515.

[17] Celik, G., G. Zussman, W. Khan, and E. Modiano. 2010. “MAC for Networks with Multipacket Reception Capability and Spatially Distributed Nodes”. *IEEE Transactions on Mobile Computing*. 9(Feb.): 226–240.

[18] Cidon, I., H. Kodesh, and M. Sidi. 1988. “Erasure, Capture and Random Power Level Section in Multi-Access Systems”. *IEEE Transactions on Communications*. 36(Mar.): 263–271.

[19] Cover, T. 1972. “Broadcast Channels”. *IEEE Transactions on Information Theory*. 18(Jan.): 2–14.

[20] Cover, T. and J. Thomas. 2005. *Elements of Information Theory*. 2nd. Wiley Interscience.

[21] Csiszar, I. and J. Korner. 1981. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press.

[22] Ephremides, A. and B. Hajek. 1998. “Information Theory and Communication Networks: An Unconsummated Union”. *IEEE Transactions on Information Theory*. 44(Oct.): 2416–2434.

[23] Fano, R. 1961. *Transmission of Information: A Statistical Theory of Communications*. Cambridge, Massachusetts: M.I.T. Press.

[24] Farkas, L. and T. Kói. 2015. “Random Access and Source-Channel Coding Error Exponents for Multiple Access Channels”. *IEEE Transactions on Information Theory*. 61(Apr.): 3029–3040.

[25] Feinstein, A. 1955. “Error Bounds in Noisy Channels Without Memory”. *IEEE Transactions on Information Theory*. 1(Sept.): 13–14.

[26] Forney, G. 1966. *Concatenated Codes*. Cambridge, MA: MIT Press.

[27] Fu, B., Y. Xiao, H. Deng, and H. Zeng. 2014. “A Survey of Cross-Layer Designs in Wireless Networks”. *IEEE Communications Surveys & Tutorials*. 16(First Quarter): 110–126.

[28] Gallager, R. 1965. “A Simple Derivation of The Coding Theorem and Some Applications”. *IEEE Transactions on Information Theory*. 11(Jan.): 3–18.

[29] Gallager, R. 1968. *Information Theory and Reliable Communication*. Wiley.

[30] Gallager, R. 1974. “Capacity and Coding for Degraded Broadcast Channels”. *Problemy Peredachi Informatsii*. 10(Mar.): 3–14.

[31] Gallager, R. 1976. “Basic Limits on Protocol Information in Data Communication Networks”. *IEEE Transactions on Information Theory*. 22(July): 385–398.

[32] Gallager, R. 1978. “Conflict Resolution in Random Access Broadcast Networks”. In: *AFOSR Workshop Communication Theory Applications*. Provincetown, MA. 74–76.

[33] Garces, R. and J. Garcia-Luna-Aceves. 2000. “Collision Avoidance and Resolution Multiple Access for Multichannel Wireless Networks”. In: *IEEE INFOCOM*. Tel Aviv, Israel. 595–602.

[34] Gau, R. 2011. "Tree/Stack Splitting with Remainder for Distributed Wireless Medium Acccess Control with Multipacket Reception". *IEEE Transactions on Wireless Communications*. 10(Nov.): 3909–3923.

[35] Ghez, S., S. Verdú, and S. Schwartz. 1988. "Stability Properties of Slotted ALOHA with Multipacket Reception Capability". *IEEE Transactions on Automatic Control*. 33(July): 640–649.

[36] Ghez, S., S. Verdú, and S. Schwartz. 1989. "Optimal Decentralized Control in The Random Access Multipacket Channel". *IEEE Transactions on Automatic Control*. 34(Nov.): 1153–1163.

[37] Goldsmith, A., M. Effros, R. Koetter, M. Medard, A. Ozdaglar, and L. Zheng. 2011. "Beyond Shannon: The Quest for Fundamental Performance Limits of Wireless Ad Hoc Networks". *IEEE Communications Magazine*. 49(May): 195–205.

[38] Gupta, P. and P. Kumar. 2000. "The Capacity of Wireless Networks". *IEEE Transactions on Information Theory*. 46(Mar.): 388–404.

[39] Gupta, P. and P. Kumar. 2003. "Towards an Information Theory of Large Networks: An Achievable Rate Region". *IEEE Transactions on Information Theory*. 49(Aug.): 1877–1894.

[40] Haenggi, M. 2012. *Stochastic Geometry for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press.

[41] Hajek, B. 1985. "Stochastic Approximation Methods for Decentralized Control of Multiaccess Communications". *IEEE Transactions on Information Theory*. IT-31(Mar.): 176–184.

[42] Hajek, B. and T. Loon. 1982. "Decentralized Dynamic Control of A Multiaccess Broadcast Channel". *IEEE Transactions on Automatic Control*. 27(June): 559–569.

[43] Ho, T., R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger. 2006. "A Random Linear Network Coding Approach to Multicast". *IEEE Transactions on Information Theory*. 52(Oct.): 4413–4430.

[44] Hui, J. 1984. "Multiple Accessing for The Collision Channel without Feedback". *IEEE Journal on Selected Areas in Communications*. SAC-2(July): 575–582.

[45] "IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications". 2007. *IEEE Std 802.11 - 2007*: 1–1076.

[46] Karlin, S. and H. Taylor. 1975. *A First Course in Stochastic Processes*. 2nd ed. San Diego, CA: Academic.

[47] Kelly, F. and I. MacPhee. 1987. "The Number of Packets Transmitted by Collision Detect Random Access Schemes". *Annals of Probability*. 15(Apr.): 1557–1568.

[48] Kiefer, J. and J. Wolfowitz. 1952. "Stochastic Estimation of The Maximum of A Regression Function". *The Annals of Mathematical Statistics*. 23(Mar.): 462–466.

[49] Koetter, R. and M. Médard. 2003. "An Algebraic Approach to Network Coding". *IEEE Transactions on Networking*. 11(May): 782–795.

[50] Kushner, H. and G. Yin. 1997. *Stochastic Approximation Algorithms and Applications*. New York: Springer Verlag.

[51] Lau, C. and C. Leung. 1992. "Capture Models for Model Packet Radio Networks". *IEEE Transactions on Communications*. 40(May): 917–925.

[52] Li, S., E. Yeung, and N. Cai. 2003. "Linear Network Coding". *IEEE Transactions on Information Theory*. IT-49(Feb.): 371–381.

[53] Liao, H. 1972. *Multiple Access Channels. Ph. D. Thesis*. University of Hawaii, Honolulu.

[54] Luby, M. 2002. "LT Codes". In: *IEEE FOCS*. Vancouver, Canada.

[55] Luo, J. 2015. "A Generalized Channel Coding Theory for Distributed Communication". *IEEE Transactions on Communications*. 63(Apr.): 1043–1056.

[56] Luo, J. and A. Ephremides. 2006. "On the Throughput, Capacity and Stability Regions of Random Multiple Access". *IEEE Transactions on Information Theory*. 52(June): 2593–2607.

[57] Luo, J. and A. Ephremides. 2010. "A Channel Coding Approach for Random Access Communication with Bursty Sources". In: *IEEE ISIT*. Austin, TX.
[58] Luo, J. and A. Ephremides. 2012. "A New Approach to Random Access: Reliable Communication and Reliable Collision Detection". *IEEE Transactions on Information Theory*. 58(Feb.): 989–1002.
[59] Luo, W. 1999. "Stability of N Interacting Queues in Random-access Systems". *IEEE Transactions on Information Theory*. 45(May): 1579–1587.
[60] Massey, J. and P. Mathys. 1985. "The Collision Channel without Feedback". *IEEE Transactions on Information Theory*. IT-31(Mar.): 192–204.
[61] Mathias, C. 2003. "Bluetooth is dead". *EE Times*. Oct. URL: http://www.eetimes.com/document.asp?doc_id=1147339.
[62] Metzner, J. 1976. "On Improving Utilization in ALOHA Networks". *IEEE Transactions on Communications*. COM-24(Apr.): 447–448.
[63] Minero, P., M. Franceschetti, and D. Tse. 2012. "Random Access: An Information-Theoretic Perspective". *IEEE Transactions on Information Theory*. 58(Feb.): 909–930.
[64] Montanari, A. and G. Forney. 2001, unpublished. "On Exponential Error Bounds for Random Codes on the DMC". URL: www.stanford.edu/montanar/PAPERS/FILEPAP/dmc.ps.
[65] Polyanskiy, Y., V. Poor, and S. Verdú. 2010. "Channel Coding Rate in The Finite Blocklength Regime". *IEEE Transactions on Information Theory*. 56(May): 2307–2359.
[66] Qin, X. and R. Berry. 2004. "Opportunistic Splitting Algorithms for Wireless Networks". In: *IEEE INFOCOM*. Hong Kong, China. 1662–1672.
[67] Rao, R. and A. Ephremides. 1988. "On The Stability of Interacting Queues in A Multiple-access System". *IEEE Transactions on Information Theory*. 34(Sept.): 918–930.

[68] Robbins, H. and S. Monro. 1951. “A Stochastic Approximation Method”. *The Annals of Mathematical Statistics*. 22(Mar.): 400–407.

[69] Russell, A. 2013. “OSI: The Internet That Wasn’t”. *IEEE Spectrum*. 50(July): 39–43.

[70] Shamai, S., I. Teletar, and S. Verdú. 2007. “Fountain Capacity”. *IEEE Transactions on Information Theory*. 53(Nov.): 4372–4376.

[71] Shannon, C. 1948. “A Mathematical Theory of Communication”. *Bell System Technical Journal*. 27(July): 379–423, 623–656.

[72] Shannon, C. 1949. “Communication in The Presence of Noise”. *Proceedings of Institute of Radio Engineers*. 37(Jan.): 10–21.

[73] Shannon, C., R. Gallager, and E. Berlekamp. 1967. “Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. I and II”. *Information and Control*. 10: 65–103, 522–552.

[74] Shokrollahi, A. 2006. “Raptor Codes”. *IEEE Transactions on Information Theory*. 52(June): 2551–2567.

[75] Sidi, M. and I. Cidon. 1985. “Splitting Protocols in Presence of Capture”. *IEEE Transactions on Information Theory*. 31(Mar.): 295–301.

[76] Srivastava, V. and M. Motani. 2005. “Cross-Layer Design: A Survey and The Road Ahead”. *IEEE Communications Magazine*. 43(Dec.): 112–119.

[77] Sun, W., O. Lee., Y. Shin, S. Kim, C. Yang, H. Kim, and S. Choi. 2014. “Wi-Fi Could be Much More”. *IEEE Communications Magazine*. 52(Nov.): 22–29.

[78] Szpankowski, W. 1994. “Stability Conditions for Some Distributed Systems: Buffered Random Access Systems”. *Advances in Applied Probability*. 26(Feb.): 498–515.

[79] Tang, Y., F. Heydaryan, and J. Luo. 2016. “On Utility Optimization in Distributed Multiple Access over a Multi-packet Reception Channel”. In: *IEEE ISIT*. Barcelona, Spain.

[80] Tang, Y., T. Zhao, and J. Luo. 2014. “Medium Access Control Game with An Enhanced Physical-Link Layer Interface”. In: *IEEE ISIT*. Honolulu, HI.

[81] Telatar, I. 1999. “Capacity of Multi-antenna Gaussian Channels”. *European Transactions on Telecommunications*. 10(Nov.): 585–595.

[82] Tsybakov, B. and V. Mikhailov. 1979. “Ergodicity of A Slotted ALOHA System”. *Problemy Peredachi Informatsii*. 15(Apr.): 73–87.

[83] Tsybakov, B. and V. Mikhailov. 1980. “Random Multiple Access of Packets: Part and Try Algorithm”. *Problems of Information Transmission*. 16(Oct.): 65–79.

[84] Viterbi, A. 1967. “Error Bounds for Convolutional Codes and An Asymptotically Optimum Decoding Algorithm”. *IEEE Transactions on Information Theory*. IT-13(Apr.): 260–269.

[85] Wang, Z. and J. Luo. 2010. “Achievable Error Exponent of Channel Coding in Random Access Communication”. In: *IEEE ISIT*. Austin, TX.

[86] Wang, Z. and J. Luo. 2011. “Coding Theorems for Random Access Communication over Compound Channel”. In: *IEEE ISIT*. St. Petersburg, Russia.

[87] Wang, Z. and J. Luo. 2012. “Error Performance of Channel Coding in Random Access Communication”. *IEEE Transactions on Information Theory*. 58(June): 3961–3974.

[88] Wang, Z. and J. Luo. 2013. “Fountain Communication using Concatenated Codes”. *IEEE Transactions on Communications*. 61(Feb.): 443–454.

[89] Weingarten, H., Y. Steinberg, and S. Shamai. 2006. “The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel”. *IEEE Transactions on Information Theory*. 52(Sept.): 3936–3964.

[90] Wyner, A. 1974. “Recent Results in The Shannon Theory”. *IEEE Transactions on Information Theory*. 20(Jan.): 2–10.

[91] Xue, F. and P. Kumar. 2005. “Network Coding Theory”. *Foundations and Trends in Communications and Information Theory*. 2: 241–381.

[92] Xue, F. and P. Kumar. 2006. "Scaling Laws for Ad Hoc Wireless Networks: An Information Theoretic Approach". *Foundations and Trends in Networking*. 1: 145–270.

[93] Yeh, E. 2012. "Fundamental Performance Limits in Cross-layer Wireless Optimization: Throughput, Delay, and Energy". *Foundations and Trends in Communications and Information Theory*. 9: 1–112.

[94] Yu, W., W. Rhee, S. Boyd, and J. Cioffi. 2004. "Iterative Water-filling for Gaussian Vector Multiple Access Channels". *IEEE Transactions on Information Theory*. 50(Jan.): 145–151.

[95] Yu, Y. and G. Giannakis. 2007. "High-throughput Random Acccess Using Successive Interference Cancellation in A Tree Algorithm". *IEEE Transactions on Information Theory*. 53(Dec.): 4628–4639.

[96] Zimmermann, H. 1980. "OSI Reference Model-The ISO Model of Architecture for Open Systems Interconnection". *IEEE Transactions on Communications*. 28(Apr.): 425–432.