Distributed Capacity of A Multiple Access Channel

Yanru Tang, Faeze Heydaryan, and Jie Luo
ECE Dept., Colorado State Univ., Fort Collins, CO 80523
Email: {yrtang, faezeh66, rockey}@colostate.edu

***Abstract*—In distributed communication, each transmitter prepares an ensemble of channel codes corresponding to different values of communication parameters. To encode a message, a transmitter chooses a channel code individually without sharing the coding choice with other transmitters or with the receiver. Upon receiving the block of channel output symbols, the receiver either decodes the messages of interest if a pre-determined reliability requirement can be met, or reports collision otherwise. Revised from the existing distributed channel coding theorems, distributed capacity of a discrete-time memoryless multiple access channel is defined and derived under the assumption that codeword length can be taken to infinity. An improved achievable error performance bound is presented for the case when codeword length is finite.**[1]

I. Introduction

Classical channel coding assumes that users in a communication party should jointly optimize their channel codes, and transmit encoded messages to the receiver over a long time duration. Overhead of achieving the required user coordination is often ignored based on the fundamental assumption that coordinated message transmission should dominate the communication process. However, this assumption is increasingly challenged by the dynamic packet-based communication activities in data networks. In a wireless network, not only messages can be short and bursty, coordinating a large number of users can also be expensive or infeasible in terms of overhead. A significant proportion of messages in existing wireless networks such as Wi-Fi systems are transmitted using distributed protocols where users make their communication decisions individually. Featured by opportunistic channel access and occasional packet collision, the distributed communication model does not fall into the classical channel coding framework. Its fundamental limits therefore cannot be understood without extending the classical coding tools.

Distributed channel coding theory, presented in [3][7][2], assumes that each transmitter should be equipped with an ensemble of channel codes as opposed to one code. Code ensembles are shared off-line with the receiver, e.g., by specifying codebook generation algorithms in the physical layer protocol. Different codes can correspond to different communication parameter settings such as different rate and power combinations. During online communication, possibly depending on a link layer decision, each transmitter individually chooses a code to encode a message. Without knowing the coding choices of the users, a receiver either decodes the messages of interest if a pre-determined decoding reliability requirement can be met, or reports collision otherwise. An achievable region is defined in [3][2] as the set of code index vectors that support asymptotic reliable message recovery. Error performance bounds in the case of finite codeword length were obtained in [7][2].

In this paper, we present two extensions to the distributed channel coding theorems given in [3][7][2]. First, in [3][2], achievable regions were defined not only as a function of the communication channel, but also as a function of the code ensembles selected by the users. We revise the definition to one that only depends on the communication channel. Such a revision enabled the definition of the distributed channel capacity, which is supported by the existing achievability proof and a new but rather straightforward converse proof. Second, error probability in a communication system is often dominated by a small number of error events. In a distributed communication system, different error events may or may not correspond to different code index vectors of the users. In [2, Theorem 3], the obtained achievable error performance bound contains a term that equals the probability of the worst case error event multiplies the number of code index vectors outside the operation region. If the latter parameter takes a large value, the corresponding error performance bound can be very loose. We revise the derivation to obtain a performance bound that essentially replaces the particular term with a summation of error probabilities each corresponding to one code index vector. The new error performance bound is tighter because it is unlikely to scale in the number of code index vectors.

We want to point out that, when the multiple access channel is user symmetric, including the case of a user symmetric compound channel with possibly user asymmetric realizations, the problem formulation and the results in this paper can be related to the unsourced random access model proposed in [4] by assuming that each active user should randomly generate a temporary identity and include it as part of the message. If temporary identities of different users are different, such an operation ensures that different users should access different sections of the same codebook, and this is equivalent to the case of active users using different codecooks.

Proofs of the Theorems in this paper can be found in [6].

II. Multiple Access with Single User Decoding

Consider a multiple access system with K transmitters (users) and one receiver. Time is slotted with each time slot equaling the length of N channel symbols, and this is

[1]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.

also the length of a codeword. Throughout the paper, we only consider channel coding within one time slot. We use bold font variable to represent a vector whose entries are the corresponding variables of all users. The discrete-time memoryless channel is modeled by a conditional distribution $P_{Y|\boldsymbol{X}}$, where $\boldsymbol{X} = [X_1, \ldots, X_K] \in \boldsymbol{\mathcal{X}}$ is the channel input symbol vector with $\boldsymbol{\mathcal{X}}$ being the vector of finite input alphabets, and $Y \in \mathcal{Y}$ is the channel output symbol with $\mathcal{Y}$ being the finite output alphabets. We assume that channel input alphabet $\mathcal{X}_k$ should be known at user k, for $k = 1, \ldots, K$, and conditional distribution $P_{Y|\boldsymbol{X}}$ should be known at the receiver.

Each transmitter, say user k, is equipped with an ensemble of M channel codes, denoted by $\mathcal{G}_k^{(N)} = \{g_{k1}, \ldots, g_{kM}\}$. Let $\boldsymbol{\mathcal{G}}^{(N)}$ denote the vector of code ensembles of all users. Let $\boldsymbol{g} = [g_1, \ldots, g_K]$ be a code index vector. We say $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}^{(N)}$ if $g_k \in \mathcal{G}_k^{(N)}$ for all $1 \leq k \leq K$. For each user k, each index $g_k \in \mathcal{G}_k^{(N)}$ represents a random block code described as follows [5][3]. Let $\mathcal{L}_{g_k} = \left\{\mathcal{C}_{g_k\theta_k} : \theta_k \in \Theta_k^{(N)}\right\}$ be a library of codebooks, indexed by a set $\Theta_k^{(N)}$. Each codebook contains $\lfloor e^{Nr_{g_k}} \rfloor$ codewords of length N, where r_{g_k} is a pre-determined parameter termed the "communication rate" of code g_k. Let $[\mathcal{C}_{g_k\theta_k}(w_k)]_j$ denote the jth symbol of the codeword corresponding to message w_k in codebook $\mathcal{C}_{g_k\theta_k}$. At the beginning of each time slot, a codebook index θ_k is generated randomly according to a distribution $\gamma_k^{(N)}$. The distribution $\gamma_k^{(N)}$ and the codebooks $\mathcal{C}_{g_k\theta_k}$, $\forall g_k \in \mathcal{G}_k^{(N)}$, are chosen such that random variables $X_{g_k w_k j} : \theta_k \to [\mathcal{C}_{g_k\theta_k}(w_k)]_j$, $\forall j, w$ and $\forall g_k$, are i.i.d. according to a pre-determined input distribution $P_{g_k X_k}$. Assume that code library $\mathcal{L}_{g_k}$ and the value of θ_k are both known at the receiver. That is, the receiver knows the randomly generated codebook of g_k, and this is true for all codes and for all users. Note that this can be achieved by sharing the random codebook generation algorithms with the receiver. In the above description, a random block code g_k is characterized by its communication rate r_{g_k} and its input distribution $P_{g_k X_k}$. With an abuse of the notation, we regard $g_k = (r_{g_k}, P_{g_k X_k})$ as a variable representing a rate and distribution pair of user k, which is not a function of the codeword length N. Similarly, we regard $\boldsymbol{g} = (\boldsymbol{r_g}, \boldsymbol{P_{gX}})$ as a vector variable representing the rate and distribution pairs of all users. We will use "code space" to refer to the space of $\boldsymbol{g}$, which is also the space of rate vector and distribution vector pairs. We use $\boldsymbol{\mathcal{G}}$, i.e., without superscription (N), to represent a code ensemble in the code space where each $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$ represents a point in the code space.

At the beginning of each time slot, we assume that each user, say user k, arbitrarily chooses a code $g_k \in \mathcal{G}_k^{(N)}$, maps a message w_k to a codeword $X_{g_k}^{(N)}(w_k)$, and then sends the codeword through the channel. Here "arbitrary" refers to the assumption that the coding choice is not controlled by, and even its statistical information may not be known to, the physical layer transmitter. Assume that $(\boldsymbol{w}, \boldsymbol{g})$ is the actual message vector and code index vector chosen by the transmitters. Let $\boldsymbol{X}_{\boldsymbol{g}}^{(N)}(\boldsymbol{w})$ be the vector of codewords. We assume that neither $\boldsymbol{g}$ nor $\boldsymbol{w}$ is known at the receiver.

We assume that the receiver is only interested in decoding the message of user 1, but can choose to decode the messages of some other users if necessary. Because users choose their codes arbitrarily, reliable message decoding is not always possible. Upon receiving the channel output symbol sequence $Y^{(N)} = [Y_1, Y_2, \ldots, Y_N]$, 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. As in [7][2], we assume that the receiver should choose an "operation region" $\boldsymbol{R}_1$ in the code space. Without knowing the actual message vector and code index vector pair $(\boldsymbol{w}, \boldsymbol{g})$, 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 $\boldsymbol{g} \notin \boldsymbol{R}_1$. 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.

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

Note that in the above error probability definition, for $\boldsymbol{g} \notin \boldsymbol{R}_1$, we regard both correct message decoding and collision report as acceptable channel outcomes. In other words, collision report is not strictly enforced for $\boldsymbol{g} \notin \boldsymbol{R}_1$. A more general error probability definition will be discussed in Section IV.

Definition 1: We say that an operation region $\boldsymbol{R}_1$ is asymptotically achievable for a multiple access channel $P_{Y|\boldsymbol{X}}$ for user 1, if for all finite M and all code ensemble vectors $\boldsymbol{\mathcal{G}}$ with each entry of code ensemble having a cardinality of M, decoding algorithms can be designed for the sequence of random code ensembles $\boldsymbol{\mathcal{G}}^{(N)}$ to achieve $\lim_{N\to\infty} P_e^{(N)}(\boldsymbol{g}) = 0, \forall \boldsymbol{g} \in \boldsymbol{\mathcal{G}}$.

Compared with the achievable region definition given in [2, Section III], the achievable region preseneted in Definition 1 is only a function of the region and the multiple access channel. It does not depend on the particular code ensembles $\boldsymbol{\mathcal{G}}$ chosen by the users. The following theorem is directly implied by the achievable region definition and the error probability definition given in (1).

Theorem 1: 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.

The following theorem characterizes the maximum achievable region of multiple access channel $P_{Y|\boldsymbol{X}}$ for user 1.

Theorem 2: 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.

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

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

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 $\boldsymbol{C}_{d1}$.

Theorem 2 can be extended from decoding for a single user to decoding for a user subset.

Definition 2: Let $S_0 \subseteq \{1, \ldots, K\}$ be a user subset. We say that an operation region $\boldsymbol{R}_{S_0}$ is asymptotically achievable for 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.

Corollary 1: For a discrete memoryless multiple access channel $P_{Y|\boldsymbol{X}}$ with finite input and output alphabets, let C_{dk} be the maximum achievable region for user k. The expression of C_{dk} can be obtained from (2) by replacing user index 1 with user index k. Let $S_0 \subseteq \{1, \ldots, K\}$ be a user subset. The maximum achievable region for user subset S_0 is given by

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

where ϕ is the empty set.

Note that, according to [2, Theorem 5], Theorem 2 and Corollary 1 still hold even if we strictly enforce collision report for $\boldsymbol{g} \notin \boldsymbol{R}_1$, i.e., if (1) is replaced by

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

With the support of Theorem 2 and Corollary 1, we define $\boldsymbol{C}_{d1}$ as the “distributed capacity” for user 1, and $\boldsymbol{C}_{dS_0}$ as the “distributed capacity” for user subset S_0, of multiple access channel $P_{Y|\boldsymbol{X}}$. Interestingly, the distributed capacity can indeed be regarded as an extension to the classical Shannon capacity in the following sense.

Let $\boldsymbol{C}_d$ be the distributed capacity of the multiple access channel when the receiver is interested in decoding the messages of all users. According to Corollary 1, $\boldsymbol{C}_d$ is given by

$$\boldsymbol{C}_d = \left\{ \boldsymbol{g} \left| \boldsymbol{g} = (\boldsymbol{r}_{\boldsymbol{g}}, \boldsymbol{P}_{\boldsymbol{g}\boldsymbol{X}}), \forall S \subseteq \{1, \ldots, K\}, \sum_{k \in S} r_{g_k} < I_{\boldsymbol{g}}(\boldsymbol{X}_S; Y|\boldsymbol{X}_{\bar{S}}) \right. \right\}. \quad (2)$$

It is well known that Shannon capacity of the multiple access channel, denoted by $\boldsymbol{C}$, is given by

$$\boldsymbol{C} = \text{convex hull} \left(\left\{ \boldsymbol{r} \left| \exists \boldsymbol{P}_{\boldsymbol{X}}, \forall S \subseteq \{1, \ldots, K\}, \sum_{k \in S} r_k \leq I(\boldsymbol{X}_S; Y|\boldsymbol{X}_{\bar{S}}) \right. \right\} \right), \quad (3)$$

where $I(\boldsymbol{X}_S; Y|\boldsymbol{X}_{\bar{S}})$ is calculated with respect to joint distribution $P_{\boldsymbol{X}Y} = P_{Y|\boldsymbol{X}} \prod_{k=1}^{K} P_{X_k}$. From (2) and (3), we can see that the two capacity terms satisfy

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

Similar to classical channel coding theory, Theorem 2 and Corollary 1 hold even if input and output alphabets of the channel are continuous. One can also pose a constraint in the code space to limit the coding choices of the users, and to define the constrained distributed channel capacity.

Example 1: Consider a 2-user multiple access system over a discrete-time memoryless channel with additive Gaussian noise. The channel is modeled by $Y = X_1 + X_2 + V$, where V is the Gaussian noise with zero mean and variance N_0. Assume that each user can only choose random block codes with Gaussian input distribution of zero mean and variance P. With the input distributions being fixed, distributed capacity region of the multiple access channel can be represented by a region of the rate vectors $\boldsymbol{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$. According to Theorem 2, distributed capacity for user 1, i.e., when the receiver is only interested in decoding the message of user 1, of the multiple access channel is given by

$$\boldsymbol{C}_{d1} = \left\{ \boldsymbol{r} \left| \begin{array}{l} r_1 < \frac{1}{2} \log \left(1 + \frac{P}{N_0} \right), \\ \text{either } r_1 < \frac{1}{2} \log \left(1 + \frac{P}{P + N_0} \right) \\ \text{and/or } r_1 + r_2 < \frac{1}{2} \log \left(1 + \frac{2P}{N_0} \right) \end{array} \right. \right\}.$$

Similarly, distributed capacity for user 2 of the multiple access channel is given by

$$\boldsymbol{C}_{d2} = \left\{ \boldsymbol{r} \left| \begin{array}{l} r_2 < \frac{1}{2} \log \left(1 + \frac{P}{N_0} \right), \\ \text{either } r_2 < \frac{1}{2} \log \left(1 + \frac{P}{P + N_0} \right) \\ \text{and/or } r_1 + r_2 < \frac{1}{2} \log \left(1 + \frac{2P}{N_0} \right) \end{array} \right. \right\}.$$

The two regions are illustrated in Figure 1. It can be seen

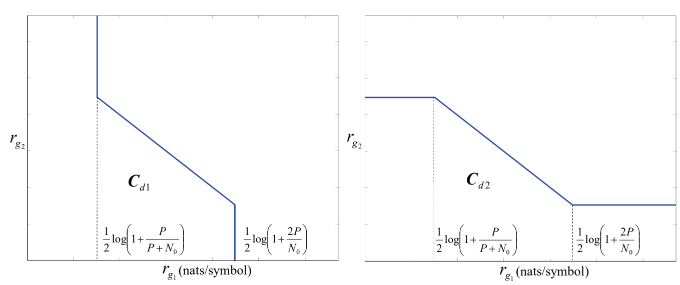

Fig. 1. Distributed capacities for each user of a 2-user Gaussian MAC.

that, closure of the intersection of $\boldsymbol{C}_{d1}$ and $\boldsymbol{C}_{d2}$, i.e., when the receiver is interested in decoding the messages of both users, coincides with the closure of the Shannon capacity region.

$$(\boldsymbol{C}_{d1} \cap \boldsymbol{C}_{d1})^c = \boldsymbol{C}^c = \left\{ \boldsymbol{r} \left| \begin{array}{l} r_1 < \frac{1}{2} \log \left(1 + \frac{P}{N_0} \right), r_2 < \frac{1}{2} \log \left(1 + \frac{P}{N_0} \right) \\ r_1 + r_2 < \frac{1}{2} \log \left(1 + \frac{2P}{N_0} \right) \end{array} \right. \right\}.$$

However, the same capacity region has different meanings under different communication models. In coordinated communication, Shannon capacity region suggests that users should jointly choose a rate vector within the capacity region to guarantee reliable message delivery. In distributed communication, on the other hand, users choose their rates individually. If the rate vector happens to locate inside the capacity region, the receiver can detect it and decode the messages reliably. If the rate vector happens to locate outside the capacity region, the receiver can reliably detect it and report collision.

III. Interfering User and Compound Channel

In this section, we extend the coding theorems presented in Section II to the case when the system has an "interfering user". As explained in [2], an interfering user can be a remote user whose codebook is unknown to the receiver, and hence its message is not decodable at the receiver. A "virtual" interfering user can also be used to model a compound channel whose realization affects the conditional channel distribution experienced by other users, but it is "virtual" in the sense of having no message to be decoded at the receiver [2].

Assume that, in addition to the K regular users indexed by $\{1, \dots, K\}$, there is an interfering user indexed as user 0. We assume that the interfering user is equipped with M communication options, denoted by $\mathcal{G}_0 = \{g_{01}, \dots, g_{0M}\}$. For convenience, we still call $\mathcal{G}_0$ a code ensemble and call $g_0 \in \mathcal{G}_0$ a code index. With the existence of the interfering user, the multiple access channel is now modeled by a conditional distribution $P_{Y|\boldsymbol{X}}(g_0)$, which is a function of the "coding" choice of the interfering user. Note that channel function $P_{Y|\boldsymbol{X}}(g_0)$ can be defined for a domain of g_0 that is beyond the ensemble $\mathcal{G}_0$. At the beginning of each time slot, assume that the interfering user should arbitrarily choose a "code" g_0, and this determines the multiple access channel $P_{Y|\boldsymbol{X}}(g_0)$ to be experienced by the regular users. The receiver knows the channel functions $P_{Y|\boldsymbol{X}}(g_0)$ for all $g_0 \in \mathcal{G}_0$, but does not know the value of g_0 chosen by the interfering user. Let vectors $\boldsymbol{g}$ and $\boldsymbol{\mathcal{G}}$ now contain the entry of the interfering user, while vectors $\boldsymbol{w}$ and $\boldsymbol{X}$ only contain the entries of the regular users.

As in Section II, we assume that the receiver is only interested in decoding the message of user 1. Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code index vector pair, unknown to the receiver. The receiver should choose an operation region $\boldsymbol{R}_1$ in the space of $\boldsymbol{g}$. 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 $\boldsymbol{g} \notin \boldsymbol{R}_1$.

Theorem 3: For a discrete-time memoryless multiple access channel $P_{Y|\boldsymbol{X}}(g_0)$ with finite input and output alphabets and with g_0 being the code index of an interfering user, conclusions of Theorems 1, 2, and Corollaries 1 still hold, if 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., excluding the interfering user.

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 in (1), then any subset of an achievable region should also be achievable. $\boldsymbol{C}_{d1}$ given in Theorem 2 is the maximum asymptotically achievable region for user 1, and $\boldsymbol{C}_{dS_0}$ given in Corollary 1 is the maximum asymptotically achievable region for user subset $S_0 \subseteq \{1, \dots, K\}$.

IV. Performance with A Finite Codeword Length

Following the system model introduced in Section III, in this section, we present the non-asymptotic analysis when the codeword length is finite and could be small in value. Throughout this section, codeword length N is assumed to be fixed at a constant.

As explained in [2], we will first need to consider an axillary decoder called the $(D, \boldsymbol{R}_D)$ decoder. Let $D \subseteq \{1, \dots, K\}$ be a subset of regular users with $1 \in D$. Assume that the receiver chooses an operation region $\boldsymbol{R}_D$ and an operation margin $\widehat{\boldsymbol{R}}_D$ both defined in the code space with $\boldsymbol{R}_D \cap \widehat{\boldsymbol{R}}_D = \phi$. A $(D, \boldsymbol{R}_D)$ decoder intends to decode the messages of all users in D by regarding signals from all other users as interference. Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code index vector pair. For $\boldsymbol{g} \in \boldsymbol{R}_D$, the decoder intends to decode the messages of users in D. For $\boldsymbol{g} \in \widehat{\boldsymbol{R}}_D$, the decoder intends to either decode the messages or to report collision for users in D. For $\boldsymbol{g} \notin \boldsymbol{R}_D \cup \widehat{\boldsymbol{R}}_D$, the decoder intends to enforce collision report for users in D. Let $(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D)$ be the estimated message vector and code index vector for users in D. Given $\boldsymbol{g}$, conditional error probability as a function of $\boldsymbol{g}$ is given by

$$P_e(\boldsymbol{g}) = \left\{ \begin{array}{l} \max\limits_{\boldsymbol{w}_D} Pr\{(\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D) \neq (\boldsymbol{w}_D, \boldsymbol{g}_D) | (\boldsymbol{w}_D, \boldsymbol{g})\}, \forall \boldsymbol{g} \in \boldsymbol{R}_D \\ \max\limits_{\boldsymbol{w}_D} 1 - Pr\left\{ \begin{array}{c} \text{"collision" or} \\ (\hat{\boldsymbol{w}}_D, \hat{\boldsymbol{g}}_D) = (\boldsymbol{w}_D, \boldsymbol{g}_D) \end{array} \middle| (\boldsymbol{w}_D, \boldsymbol{g}) \right\}, \\ \hfill \forall \boldsymbol{g} \in \widehat{\boldsymbol{R}}_D \\ \max\limits_{\boldsymbol{w}_D} 1 - Pr\left\{\text{"collision"} | (\boldsymbol{w}_D, \boldsymbol{g})\right\}, \forall \boldsymbol{g} \notin \boldsymbol{R}_D \cup \widehat{\boldsymbol{R}}_D \end{array} \right.$$

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

$$\left\{ \alpha_{\boldsymbol{g}} \middle| \alpha_{\boldsymbol{g}} \geq 0, \forall \boldsymbol{g} \in \boldsymbol{\mathcal{G}}, \sum_{\boldsymbol{g}} e^{-N\alpha_{\boldsymbol{g}}} = 1 \right\}. \tag{4}$$

We define the "generalized error performance" of the $(D, \boldsymbol{R}_D)$ decoder as

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

Let us use $P_{g_k}(X_k)$ to denote the probability of channel input symbol X_k under coding option g_k, and use $P(Y|\boldsymbol{X}_D, \boldsymbol{g}_{\bar{D}})$ to denote the conditional probability of channel output symbol Y given input symbol vector $\boldsymbol{X}_D$ for users in D, and code index vector $\boldsymbol{g}_{\bar{D}}$ for users not in D. The following theorem gives an achievable bound, improved from the corresponding bound presented in [2, Theorem 3], for the generalized error performance of the $(D, \boldsymbol{R}_D)$ decoder.

Theorem 4: Consider the distributed multiple access system described above. There exists a decoding algorithm such that GEP_D is upper bounded by

$$\begin{aligned}
&\mathrm{GEP}_D \leq \\
&\sum_{\boldsymbol{g}\in\boldsymbol{R}_D} \left\{ \sum_{S\subset D} \left[\sum_{\tilde{\boldsymbol{g}}\in\boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S=\boldsymbol{g}_S} \exp(-NE_{mD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)) \right.\right. \\
&\quad \left. +2 \sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S=\boldsymbol{g}_S} \exp(-NE_{iD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, S)) \right] \\
&\quad \left. +2 \sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}_D\cup\widehat{\boldsymbol{R}}_D, \tilde{\boldsymbol{g}}_D=\boldsymbol{g}_D} \exp(-NE_{iD}(\boldsymbol{g}, \tilde{\boldsymbol{g}}, D)) \right\}, \quad (5)
\end{aligned}$$

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

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

Compared with the bound presented in [2, Equation (7)], besides other minor improvements, the second and the third terms on the right hand side of (5) represent the key improvement that lead to a tighter bound. More specifically, error exponent represented in (5) and [2, Equation (7)] are identical. However, multiplication factor in front the second and the third terms on the right hand side of (5) equals 2, while the corresponding factor in [2, Equation (7)] equals $(1 + \sum_{\tilde{\boldsymbol{g}}\notin\boldsymbol{R}_D, \tilde{\boldsymbol{g}}_S=\boldsymbol{g}_S} 1)$. If the summation is dominated by a small number of terms, then error performance bound given in [2, Equation (7)] scales in the number of code index vectors of $\tilde{\boldsymbol{g}} \notin \boldsymbol{R}_D$, while the improved bound given in (5) does not have such a scaling problem.

Let us now consider the case when the receiver is only interested in decoding the message of user 1 but can choose to decode the messages of other users if necessary. Assume that the receiver should choose an operation region $\boldsymbol{R}_1$ and an operation margin $\widehat{\boldsymbol{R}}_1$ in the code space with $\boldsymbol{R}_1 \cap \widehat{\boldsymbol{R}}_1 = \phi$. Let $\boldsymbol{g}$ be the actual code index vector. The receiver intends to decode the message of user 1 for $\boldsymbol{g} \in \boldsymbol{R}_1$, to either decode the message of user 1 or to report collision for user 1 for $\boldsymbol{g} \in \widehat{\boldsymbol{R}}_1$, and to report collision for user 1 for $\boldsymbol{g} \notin \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1$.

Let $(\hat{w}_1, \hat{g}_1)$ be the message and code index estimate of user 1. Let $(\boldsymbol{w}, \boldsymbol{g})$ be the actual message vector and code index vector pair, conditional error probability of the system as a function of $\boldsymbol{g}$ is defined as

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

Let $\{\alpha_{\boldsymbol{g}}\}$ be a set of pre-determined weight parameters each being assigned to a code index vector $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$ and satisfying constraint (4). We define the "generalized error performance" of the system as

$$\mathrm{GEP} = \sum_{\boldsymbol{g}} P_e(\boldsymbol{g}) e^{-N\alpha_{\boldsymbol{g}}}. \quad (6)$$

According to [2, Theorem 4], an achievable bound on the generalized error performance of the system is given in the following theorem.

Theorem 5: 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

$$\begin{aligned}
&\boldsymbol{R}_1 = \bigcup_{D, D\subseteq\{1,\ldots,K\}, 1\in D} \boldsymbol{R}_D, \qquad \boldsymbol{R}_{D'} \cap \boldsymbol{R}_D = \phi, \\
&\forall D, D' \subseteq \{1, \ldots, K\}, D' \neq D, 1 \in D, D'.
\end{aligned}$$

There exists a decoding algorithm such that the generalized error performance defined in (6) is upper bounded by

$$\mathrm{GEP} \le \min_{\sigma} \sum_{D, D \subseteq \{1,\dots,K\}, 1 \in D} \mathrm{GEP}_D,$$

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 $\widehat{\boldsymbol{R}}_D = \boldsymbol{R}_1 \cup \widehat{\boldsymbol{R}}_1 \setminus \boldsymbol{R}_D$.

V. Error Exponents of Coordinated and Distributed Communications

Compared with a coordinated communication model where transmitters should always choose a coding scheme to support reliable decoding, a distributed communication system needs to prepare for the situation when transmitters choose coding options outside the operation region or even the capacity region. Receiver of a distributed communication system therefore has an extra responsibility to detect whether code index vector chosen by the transmitters is inside the operation region or not. While we have shown in Section III that the distributed capacity coincides with the Shannon capacity of the same channel, the achievable error performance bound of a distributed communication system is often inferior to that of the corresponding coordinated communication system. This suggests that the support of distributed communication does not come without a cost.

Example 2: 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 correspond to random block codes 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 the operation margin be an empty set, i.e., $\widehat{R}_1 = \phi$. Let $\{\alpha_g\}$ in (6) be chosen such that $e^{-N\alpha_1} = e^{-N\alpha_2} = 0.5$. In the case of a large codeword length, the scaling law of error probability bound (6) or (5) is determined by the following error exponent,

$$E_d = \min\{E_m(g=1, \tilde{g}=1), E_i(g=1, \tilde{g}=2)\}.$$

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

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

Note that $E_m(g=1, \tilde{g}=1)$ and $E_i(g=1, \tilde{g}=2)$ are computed using (6) with $D = \{1\}$ and $S = \phi$ as

$$\begin{aligned} E_m(g=1, \tilde{g}=1) &= \max_{0<\rho\le 1} -\rho r_1 + \max_{0\le s\le 1} -\log \sum_Y \\ &\times \left(\sum_X P(Y|X)^{1-s} P(X)\right) \left(\sum_X P(Y|X)^{\frac{s}{\rho}} P(X)\right)^{\rho} \\ &= \max_{0<\rho\le 1} -\rho r_1 - \log \sum_Y \left(\sum_X P(Y|X)^{\frac{1}{\rho+1}} P(X)\right)^{(\rho+1)}, \\ E_i(g=1, \tilde{g}=2) &= \max_{0<\rho\le 1} -\rho r + \max_{0\le s\le 1-\rho} \\ &-\log \sum_Y \left(\sum_X [P(Y|X)]^{\frac{s}{s+\rho}} P(X)\right)^{s+\rho} P(Y)^{1-s}. \end{aligned}$$

Figure 2 illustrates the two error exponents as functions of r_1. It can be seen that E_d achieved under the distributed communication model is only about half of the value of E_r. Note that the difference between E_d and E_r depends on the particular channel and the code ensemble at the transmitter.

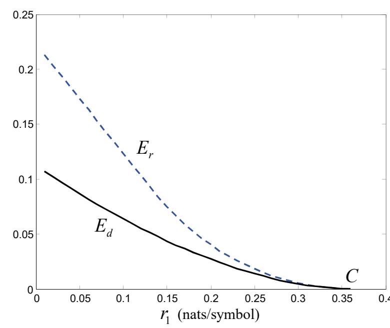

Fig. 2. Achievable error exponents of distributed and coordinated communications over a binary symmetric channel.

References

[1] R. Gallager, "A Simple Derivation of The Coding Theorem and Some Applications," *IEEE Trans. on Inform. Theory*, Vol. IT-11, pp. 3-18, Jan. 1965.

[2] J. Luo, "A Generalized Channel Coding Theory for Distributed Communication," *IEEE Trans. on Commun.*, Vol. 63, pp. 1043-1056, Apr. 2015.

[3] J. Luo and A. Ephremides, "A New Approach to Random Access: Reliable Communication and Reliable Collision Detection," *IEEE Trans. on Inform. Theory*, Vol. 58, pp. 379-423, Feb. 2012.

[4] Y. Polyanskiy, "A Perspective on Massive Random-Access," *IEEE ISIT*, Aachen, Germany, Jun. 2017.

[5] S. Shamai, I. Teletar, and S. Verdú, "Fountain capacity," *IEEE Trans. on Inform. Theory*, Vol. 53, pp. 4327–4376, Nov. 2007.

[6] Y. Tang, F. Heydaryan, and J. Luo, "Distributed Coding in A Multiple Access Environment," *Foundations and Trends in Networking*, Vol. 12, pp. 260-412, Now Publishers, 2018.

[7] Z. Wang and J. Luo, "Error Performance of Channel Coding in Random Access Communication," *IEEE Trans. on Inform. Theory*, Vol. 58, pp. 3961-3974, Jun. 2012.