Optimal Achievable Rates for Interference Networks With Random Codes

Bernd Bandemer, *Member, IEEE*, Abbas El Gamal, *Fellow, IEEE*, and Young-Han Kim, *Fellow, IEEE*

***Abstract*— The optimal rate region for interference networks is characterized when encoding is restricted to random code ensembles with superposition coding and time sharing. A simple *simultaneous nonunique decoding* rule, under which each receiver decodes for the intended message as well as the interfering messages, is shown to achieve this optimal rate region regardless of the relative strengths of signal, interference, and noise. This result implies that the Han–Kobayashi bound, the best known inner bound on the capacity region of the two-user pair interference channel, cannot be improved merely by using the optimal maximum likelihood decoder.**

***Index Terms*— Han–Kobayashi bound, interference network, maximum likelihood decoding, network information theory, random code ensemble, superposition coding, simultaneous decoding.**

I. Introduction

CONSIDER a communication scenario in which multiple senders communicate independent messages to multiple receivers over a network with interference. What is the set of simultaneously achievable rate tuples for reliable communication? What coding scheme achieves this *capacity region*? Answering these questions involves joint optimization of the encoding and decoding functions, which has remained elusive even for the case of two sender–receiver pairs.

With a complete theory in terra incognita, we take in this paper a simpler modular approach to these questions. Instead of searching for the optimal encoding functions, suppose rather that the encoding functions are restricted to realizations of a given random code ensemble of a certain structure. What is the set of simultaneously achievable rate tuples so that the probability of decoding error, when *averaged over the random code ensemble*, can be made arbitrarily small? To be specific, we focus on random code ensembles with superposition coding and time sharing of independent and identically distributed (i.i.d.) codewords. This class of random code ensembles includes those used in the celebrated Han–Kobayashi coding scheme [9].

We characterize the set $\mathscr{R}^*$ of rate tuples achievable by the random code ensemble for an interference network as the intersection of rate regions for its component multiple access channels in which each receiver recovers its intended messages as well as appropriately chosen unintended messages. More specifically, the rate region $\mathscr{R}^*$ for the interference network with senders $[1:K] = \{1, 2, \ldots, K\}$, each communicating an independent message, and receivers $[1:L]$, each required to recover a subset $\mathcal{D}_1, \ldots, \mathcal{D}_L \subseteq [1:K]$ of messages, is

$$\mathscr{R}^* = \bigcap_{l\in[1:L]} \bigcup_{\mathcal{S}\supseteq\mathcal{D}_l} \mathscr{R}_{\mathrm{MAC}(\mathcal{S},\,l)}. \tag{1}$$

Here $\mathscr{R}_{\mathrm{MAC}(\mathcal{S},\,l)}$ denotes the set of rate tuples achievable by the random code ensemble for the *multiple access channel* with senders $\mathcal{S}$ and receiver l when the codewords from the other senders $[1:K]\setminus\mathcal{S}$ are treated as random noise.

A direct approach to proving this result would be to analyze the average performance of the optimal decoding rule for each realization of the random code ensemble that minimizes the probability of decoding error, namely, maximum likelihood decoding (MLD). This analysis, however, is unnecessarily cumbersome. We instead take an indirect yet conducive approach that is common in information theory. First, we show that any rate tuple inside $\mathscr{R}^*$ is achieved by using the typicality-based *simultaneous nonunique decoding* (SND) rule [6], [7], [13], in which each receiver attempts to recover the codewords from its intended senders and (potentially nonuniquely) the codewords from interfering senders. Second, we show that if the average probability of error of MLD for the random code ensemble is asymptotically zero, then its rate tuple must lie in $\mathscr{R}^*$. The key to proving the second step is to show that after a maximal set of messages has been recovered, the remaining signal at each receiver is distributed essentially independently and identically. The two-step approach taken here is reminiscent of the random coding proof for the capacity of the point-to-point channel [16], wherein a suboptimal (in the sense of the probability of error) decoding rule based on the notion of joint typicality can achieve the same rate as MLD when used for random code ensembles.

Our result has several implications.

- It shows that incorporating the structure of interference into decoding, when properly done as in MLD and SND, always achieves higher or equal rates compared

Manuscript received November 27, 2012; revised March 29, 2015; accepted July 29, 2015. Date of publication September 4, 2015; date of current version November 18, 2015. This work was supported in part by the Korea Communications Commission within the Research and Development under Program KCA-2012-11-921-04-001 through the Electronics and Telecommunications Research Institute and in part by the National Science Foundation under Grant CCF-1320895. This paper was presented at the 2012 Allerton Conference on Communication, Control and Computing.

B. Bandemer was with the Information Theory and Applications Center, University of California at San Diego, La Jolla CA 92093 USA. He is now with Rasa Networks, San Jose, CA 95131 USA (e-mail: bandemer@alumni.stanford.edu).

A. El Gamal is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: abbas@ee.stanford.edu).

Y.-H. Kim is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: yhk@ucsd.edu).

Communicated by Y. Steinberg, Associate Editor at Large.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2015.2476792

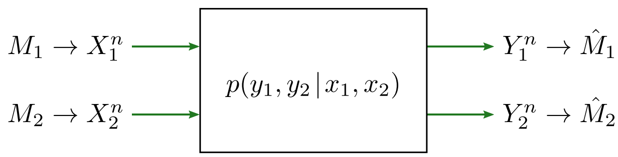


Fig. 1. Two-user-pair discrete memoryless interference channel.

to treating interference as random noise; thus, the traditional wisdom of distinguishing between decoding for the interference at high signal-to-noise ratio and ignoring the interference at low signal-to-noise ratio does not provide any improvement on achievable rates.

- It shows that the Han–Kobayashi inner bound [9], [6], [7, Th. 6.4], which was established using the random code ensemble and a typicality-based simultaneous decoding rule, cannot be improved by using a more powerful decoding rule such as MLD.
- It generalizes the result by Motahari and Khandani [12], and Baccelli *et al.* [2] on the optimal rate region for K-user-pair Gaussian interference channels with point-to-point Gaussian random code ensembles to arbitrary (not necessarily Gaussian) random code ensembles with time sharing and superposition coding.
- It shows that the interference decoding rate region for the three-user-pair deterministic interference channel in [3] is the optimal rate region achievable by point-to-point random code ensembles and time sharing.

As the simplest example of a general interference network, consider the two-user-pair discrete memoryless interference channel (2-DM-IC) $p(y_1, y_2|x_1, x_2)$ with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, depicted in Fig. 1. Here sender $j = 1, 2$ wishes to communicate a message to its respective receiver via n transmissions over the shared interference channel. Each message M_j, $j = 1, 2$, is separately encoded into a codeword $X_j^n = (X_{j1}, X_{j2}, \ldots, X_{jn})$ and transmitted over the channel. Upon receiving the sequence Y_j^n, receiver $j = 1, 2$ finds an estimate $\hat{M}_j$ of the message M_j.

We now consider the standard random coding analysis for inner bounds on the set of achievable rate pairs (the capacity region) of the 2-DM-IC. Given a product input pmf $p(x_1)\,p(x_2)$, suppose that the codewords $x_j^n(m_j)$, $m_j \in [1 : 2^{nR_j}] = \{1, 2, \ldots, 2^{nR_j}\}$, for $j = 1, 2$ are generated randomly, each drawn according to $\prod_{i=1}^n p_{X_j}(x_{ji})$.

We recall the rate regions achieved by employing the following simple suboptimal decoding rules, described for receiver 1 (see [7, Sec. 6.2]).

- *Treating Interference as Noise (IAN):* Receiver 1 finds the unique message $\hat{m}_1$ such that $(x_1^n(\hat{m}_1), y_1^n)$ is jointly typical. (See the end of this section for the definition of joint typicality.) It can be shown that the average probability of decoding error for receiver 1 tends to zero as $n \to \infty$ if

$$R_1 < I(X_1; Y_1). \tag{2}$$

The corresponding rate region (IAN region) is depicted in Fig. 2(a).

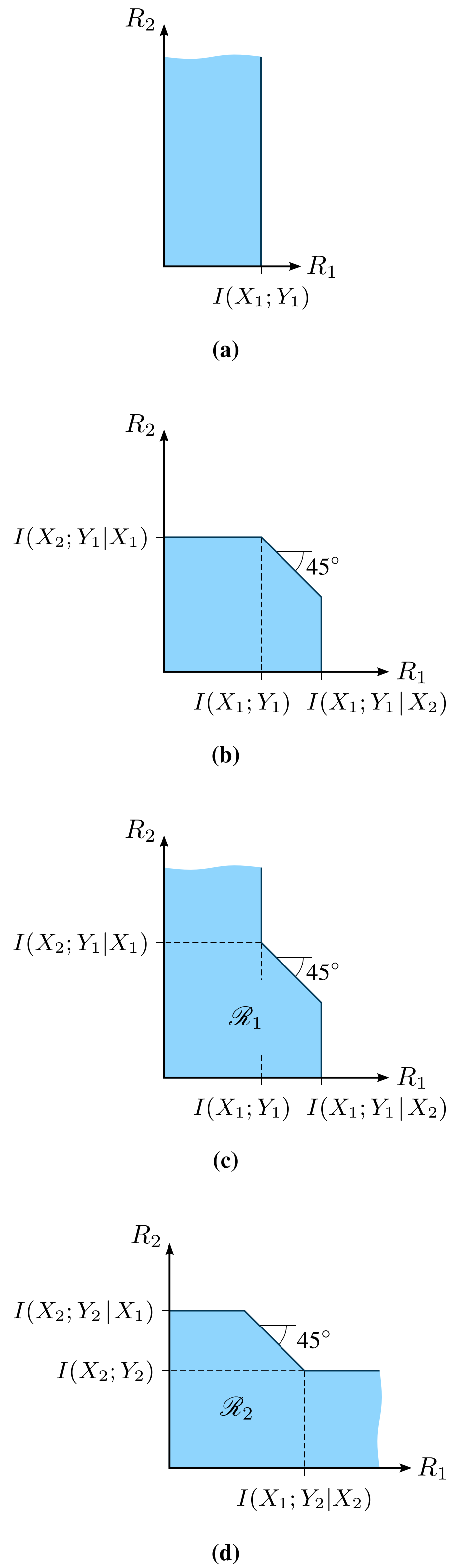


Fig. 2. Achievable rate regions for the 2-DM-IC: (a) treating interference as noise, (b) using simultaneous decoding, (c) using simultaneous nonunique decoding ($\mathscr{R}_1$); note that $\mathscr{R}_1$ is the union of the regions in (a) and (b); and (d) using simultaneous nonunique decoding at receiver 2 ($\mathscr{R}_2$).

- *Simultaneous Decoding (SD):* Receiver 1 finds the unique message pair $(\hat{m}_1, \hat{m}_2)$ such that $(x_1^n(\hat{m}_1), x_2^n(\hat{m}_2), y_1^n)$ is jointly typical. The average probability of decoding error for receiver 1 tends to zero as $n \to \infty$ if

$$R_1 < I(X_1; Y_1 \mid X_2), \tag{3a}$$
$$R_2 < I(X_2; Y_1 \mid X_1), \tag{3b}$$
$$R_1 + R_2 < I(X_1, X_2; Y_1). \tag{3c}$$

The corresponding rate region (SD region) is depicted in Fig. 2(b).

Now, consider simultaneous nonunique decoding (SND) in which receiver 1 finds the unique $\hat{m}_1$ such that $(x_1^n(\hat{m}_1), x_2^n(m_2), y_1^n)$ is jointly typical *for some* m_2. Clearly, any rate pair in the SD rate region (3a–3c) is achievable via SND. Less obviously, any rate pair in the IAN region (2) is also achievable via SND as we show in the achievability proof of Theorem 1 in Section II. Hence, SND can achieve any rate pair in the union of the IAN and SD regions, that is, the rate region $\mathscr{R}_1$ as depicted in Fig. 2(c). Similarly, the average probability of decoding error for receiver 2 using SND tends to zero as $n \to \infty$ if the rate pair (R_1, R_2) is in $\mathscr{R}_2$, which is defined analogously by exchanging the roles of the two users (see Fig. 2(d)). Combining the decoding requirements for both receivers yields the rate region $\mathscr{R}_1 \cap \mathscr{R}_2$.

This rate region $\mathscr{R}_1 \cap \mathscr{R}_2$ turns out to be optimal for the given random code ensemble. As shown in the converse proof of Theorem 1, if the probability of error for MLD averaged over the random code ensemble tends to zero as $n \to \infty$, then the rate pair (R_1, R_2) must reside inside the closure of $\mathscr{R}_1 \cap \mathscr{R}_2$. Thus, SND achieves the same rate region as MLD (for random code ensembles of the given structure).

The rest of the paper is organized as follows. For simplicity of presentation, in Section II we formally define the problem for the two-user-pair interference channel and establish our main result for the random code ensemble with time sharing and no superposition coding. In Section III, we extend our result to a multiple-sender multiple-receiver discrete memoryless interference network in which each sender has a single message and wishes to communicate it to a subset of the receivers. This extension includes superposition coding with an arbitrary number of layers. In Section IV, we specialize the result to the Han–Kobayashi coding scheme for the two-user-pair interference channel. Most technical proofs are deferred to the Appendices.

Throughout we closely follow the notation in [7]. In particular, for $X \sim p(x)$ and $\varepsilon \in (0, 1)$, we define the set of ε-typical n-sequences x^n (or the typical set in short) [15] as

$$\mathcal{T}_\varepsilon^{(n)}(X) = \big\{x^n : |\#\{i : x_i = x\}/n - p(x)| \le \varepsilon p(x) \text{ for all } x \in \mathcal{X}\big\}.$$

For a tuple of random variables $(X_1, \ldots, X_k)$, the joint typical set $\mathcal{T}_\varepsilon^{(n)}(X_1, \ldots, X_k)$ is defined as the typical set $\mathcal{T}_\varepsilon^{(n)}((X_1, \ldots, X_k))$ for a single random variable $(X_1, \ldots, X_k)$. The joint typical set $\mathcal{T}_\varepsilon^{(n)}(X_\mathcal{S})$ for a subtuple $X_\mathcal{S} = (X_k : k \in \mathcal{S})$ is defined similarly for each $\mathcal{S} \subseteq [1 : k]$. We use $\delta(\varepsilon) > 0$ to denote a generic function of $\varepsilon > 0$ that tends to zero as $\varepsilon \to 0$. Similarly, we use $\varepsilon_n \ge 0$ to denote a generic function of n that tends to zero as $n \to \infty$.

II. DM-IC With Two User Pairs

Consider the two-user-pair discrete memoryless interference channel (2-DM-IC) $p(y_1, y_2 \mid x_1, x_2)$ introduced in the previous section (see Fig. 1). A $(2^{nR_1}, 2^{nR_2}, n)$ code $\mathcal{C}_n$ for the 2-DM-IC consists of

- two message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$,
- two encoders, where encoder 1 assigns a codeword $x_1^n(m_1)$ to each message $m_1 \in [1 : 2^{nR_1}]$ and encoder 2 assigns a codeword $x_2^n(m_2)$ to each message $m_2 \in [1 : 2^{nR_2}]$, and
- two decoders, where decoder 1 assigns an estimate $\hat{m}_1$ or an error message e to each received sequence y_1^n and decoder 2 assigns an estimate $\hat{m}_2$ or an error message e to each received sequence y_2^n.

We assume that the message pair (M_1, M_2) is uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The average probability of error for the code $\mathcal{C}_n$ is defined as

$$P_e^{(n)}(\mathcal{C}_n) = \mathsf{P}\big\{(\hat{M}_1, \hat{M}_2) \ne (M_1, M_2)\big\}.$$

A rate pair (R_1, R_2) is said to be *achievable* for the 2-DM-IC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes $\mathcal{C}_n$ such that $\lim_{n\to\infty} P_e^{(n)}(\mathcal{C}_n) = 0$. The *capacity region* $\mathscr{C}$ of the 2-DM-IC is the closure of the set of achievable rate pairs (R_1, R_2).

We now limit our attention to a randomly generated code ensemble with a special structure. Let $p = p(q, x_1, x_2) = p(q)\, p(x_1|q)\, p(x_2|q)$ be a given pmf on $\mathcal{Q} \times \mathcal{X}_1 \times \mathcal{X}_2$, where $\mathcal{Q}$ is a finite alphabet. Suppose that the codewords $X_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, and $X_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, that constitute the codebook, are generated randomly as follows:

- Let $Q^n \sim \prod_{i=1}^n p_Q(q_i)$.
- Let $X_1^n(m_1) \sim \prod_{i=1}^n p_{X_1|Q}(x_{1i}|q_i)$, $m_1 \in [1 : 2^{nR_1}]$, conditionally independent given Q^n.
- Let $X_2^n(m_2) \sim \prod_{i=1}^n p_{X_2|Q}(x_{2i}|q_i)$, $m_2 \in [1 : 2^{nR_2}]$, conditionally independent given Q^n.

Each instance $\{(x_1^n(m_1), x_2^n(m_2)) : (m_1, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]\}$ of such generated codebooks, along with the corresponding optimal decoders, constitutes a $(2^{nR_1}, 2^{nR_2}, n)$ code. We refer to the random code ensemble generated in this manner as the $(2^{nR_1}, 2^{nR_2}, n; p)$ *random code ensemble*.

Definition 1 (Random Coding Optimal Rate Region): Given a pmf $p = p(q)\, p(x_1|q)\, p(x_2|q)$, the *optimal rate region* $\mathscr{R}^*(p)$ *achievable by the* p*-distributed random code ensemble* is the closure of the set of rate pairs (R_1, R_2) such that the sequence of $(2^{nR_1}, 2^{nR_2}, n; p)$ random code ensembles $\mathcal{C}_n$ satisfies

$$\lim_{n\to\infty} \mathsf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)] = 0,$$

where the expectation is with respect to the random code ensemble $\mathcal{C}_n$.

To characterize the random coding optimal rate region, we define $\mathscr{R}_1(p)$ to be the set of rate pairs (R_1, R_2) such that

$$R_1 \le I(X_1; Y_1 \mid Q) \tag{4a}$$

or

$$R_2 \le I(X_2; Y_1 \mid X_1, Q), \tag{4b}$$
$$R_1 + R_2 \le I(X_1, X_2; Y_1 \mid Q). \tag{4c}$$

Similarly, define $\mathscr{R}_2(p)$ by making the index substitution $1 \leftrightarrow 2$. We are now ready to state the main result of the section.

Theorem 1: Given a pmf $p = p(q)\,p(x_1|q)\,p(x_2|q)$, the optimal rate region of the DM-IC $p(y_1, y_2|x_1, x_2)$ achievable by the p-distributed random code ensemble is

$$\mathscr{R}^*(p) = \mathscr{R}_1(p) \cap \mathscr{R}_2(p).$$

Before we prove the theorem, we point out a few important properties of the random coding optimal rate region.

Remark 1 (MAC Form): Let $\mathscr{R}_{1,\mathrm{IAN}}(p)$ be the set of rate pairs (R_1, R_2) such that

$$R_1 \le I(X_1; Y_1 \mid Q),$$

that is, the achievable rate (region) for the point-to-point channel $p(y_1|x_1)$ by treating the interfering signal X_2 as noise. Let $\mathscr{R}_{1,\mathrm{SD}}(p)$ be the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X_1; Y_1 \mid X_2, Q), \\ R_2 &\le I(X_2; Y_1 \mid X_1, Q), \\ R_1 + R_2 &\le I(X_1, X_2; Y_1 \mid Q), \end{aligned}$$

that is, the achievable rate region for the multiple access channel $p(y_1|x_1, x_2)$ by decoding for both messages M_1 and M_2 simultaneously. Then, we can express $\mathscr{R}_1(p)$ as

$$\mathscr{R}_1(p) = \mathscr{R}_{1,\mathrm{IAN}}(p) \cup \mathscr{R}_{1,\mathrm{SD}}(p),$$

which is referred to as the *MAC form* of $\mathscr{R}_1(p)$, since it is the union of the achievable rate regions of 1-sender and 2-sender multiple access channels. The region $\mathscr{R}_2(p)$ can be expressed similarly as the union of the *interference-as-noise region* $\mathscr{R}_{2,\mathrm{IAN}}(p)$ and the *simultaneous-decoding region* $\mathscr{R}_{2,\mathrm{SD}}(p)$. Hence the optimal rate region $\mathscr{R}^*(p)$ can be expressed as

$$\begin{aligned} \mathscr{R}^*(p) = \big(\mathscr{R}_{1,\mathrm{IAN}}(p) \cap \mathscr{R}_{2,\mathrm{IAN}}(p)\big) \cup \big(\mathscr{R}_{1,\mathrm{IAN}}(p) \cap \mathscr{R}_{2,\mathrm{SD}}(p)\big) \\ \cup \big(\mathscr{R}_{1,\mathrm{SD}}(p) \cap \mathscr{R}_{2,\mathrm{IAN}}(p)\big) \cup \big(\mathscr{R}_{1,\mathrm{SD}}(p) \cap \mathscr{R}_{2,\mathrm{SD}}(p)\big), \end{aligned} \tag{5}$$

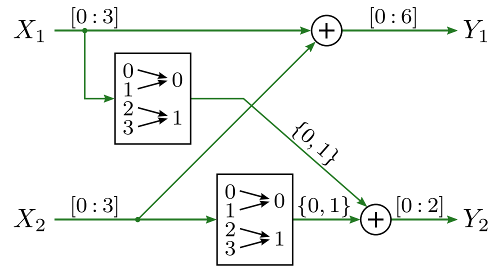
which is achieved by taking the union over all possible combinations of treating interference as noise and simultaneous decoding at the two receivers.

Remark 2 (Min Form): The region $\mathscr{R}_1(p)$ in (4a–4c) can be equivalently characterized as the set of rate pairs (R_1, R_2) such that

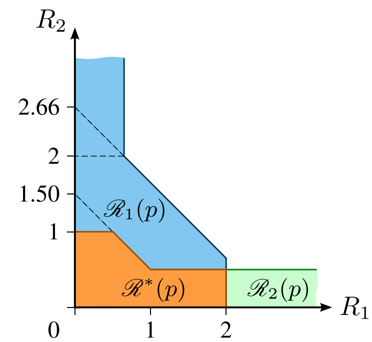
$$R_1 \le I(X_1; Y_1 \mid X_2, Q), \tag{6a}$$

$$R_1 + \min\{R_2,\ I(X_2; Y_1 \mid X_1, Q)\} \le I(X_1, X_2; Y_1 \mid Q). \tag{6b}$$

The minimum term in (6b) can be interpreted as the effective rate of the interfering signal X_2 at the receiver Y_1, which is a monotone increasing function of R_2 and saturates at the maximum possible rate for distinguishing X_2 codewords; see [3]. When R_2 is small, all X_2 codewords are distinguishable and the effective rate equals the actual code rate. In comparison, when R_2 is large, the codewords are not distinguishable and the effective rate equals $I(X_2; Y_1 \mid X_1, Q)$, which is the maximum achievable rate for the channel from X_2 to Y_1.



(a)



(b)

Fig. 3. An example for nonconvex $\mathscr{R}^*(p)$. (a) Channel block diagram. (b) Regions $\mathscr{R}_1(p)$, $\mathscr{R}_2(p)$, and $\mathscr{R}^*(p)$ for $Q = \emptyset$ and $X_1, X_2 \sim \mathrm{Unif}[0:3]$.

Remark 3 (Nonconvexity): The random coding optimal rate region $\mathscr{R}^*(p)$ is not convex in general. This is exemplified by the deterministic 2-DM-IC in Fig. 3.

A direct approach to proving Theorem 1 would be to analyze the performance of maximum likelihood decoding:

$$\hat{m}_1 = \arg\max_{m_1} \frac{1}{2^{nR_2}} \sum_{m_2} \prod_{i=1}^{n} p_{Y_1|X_1,X_2}(y_{1i} \mid x_{1i}(m_1), x_{2i}(m_2)),$$

$$\hat{m}_2 = \arg\max_{m_2} \frac{1}{2^{nR_1}} \sum_{m_1} \prod_{i=1}^{n} p_{Y_2|X_1,X_2}(y_{2i} \mid x_{1i}(m_1), x_{2i}(m_2))$$

for the p-distributed random code. Instead of performing this analysis, which is quite complicated (if possible), we establish the achievability of $\mathscr{R}^*(p)$ by the suboptimal simultaneous nonunique decoding rule, which uses the notion of joint typicality. We then show that if the average probability of error of the $(2^{nR_1}, 2^{nR_2}, n; p)$ random code ensemble tends to zero as $n \to \infty$, then the rate pair (R_1, R_2) must lie in $\mathscr{R}^*(p)$.

A. Proof of Achievability

Each receiver uses *simultaneous nonunique decoding*. Receiver 1 declares that $\hat{m}_1$ is sent if it is the unique message among $[1 : 2^{nR_1}]$ such that

$$\big(q^n, x_1^n(\hat{m}_1), x_2^n(m_2), y_1^n\big) \in \mathcal{T}_\varepsilon^{(n)} \quad \text{for some } m_2 \in [1 : 2^{nR_2}].$$

If there is no such message or more than one, it declares an error. Similarly, receiver 2 finds the unique message $\hat{m}_2 \in [1 : 2^{nR_2}]$ such that

$$\big(q^n, x_1^n(m_1), x_2^n(\hat{m}_2), y_2^n\big) \in \mathcal{T}_\varepsilon^{(n)} \quad \text{for some } m_1 \in [1 : 2^{nR_1}].$$

To analyze the probability of decoding error averaged over the random codebook ensemble, assume without loss of generality that $(M_1, M_2) = (1, 1)$ is sent. Receiver 1 makes an error only if one or both of the following events occur:

$$\begin{aligned}\mathcal{E}_1 &= \big\{(Q^n, X_1^n(1), X_2^n(1), Y_1^n) \notin \mathcal{T}_\varepsilon^{(n)}\big\},\\ \mathcal{E}_2 &= \big\{(Q^n, X_1^n(m_1), X_2^n(m_2), Y_1^n) \in \mathcal{T}_\varepsilon^{(n)} \\ &\qquad \text{for some } m_1 \neq 1 \text{ and some } m_2\big\}.\end{aligned}$$

By the law of large numbers, $\mathsf{P}(\mathcal{E}_1)$ tends to zero as $n \to \infty$.

We bound $\mathsf{P}(\mathcal{E}_2)$ in two ways, which leads to the MAC form of $\mathscr{R}_1(p)$ in Remark 1. First, since the joint typicality of the quadruple $(Q^n, X_1^n(m_1), X_2^n(m_2), Y_1^n)$ for each m_2 implies the joint typicality of the triple $(Q^n, X_1^n(m_1), Y_1^n)$, we have

$$\mathcal{E}_2 \subseteq \big\{(Q^n, X_1^n(m_1), Y_1^n) \in \mathcal{T}_\varepsilon^{(n)} \text{ for some } m_1 \neq 1\big\} = \mathcal{E}_2'.$$

Then, by the packing lemma in [7, Sec. 3.2], $\mathsf{P}(\mathcal{E}_2')$ tends to zero as $n \to \infty$ if

$$R_1 < I(X_1; Y_1 \mid Q) - \delta(\varepsilon). \tag{7}$$

The second way to bound $\mathsf{P}(\mathcal{E}_2)$ is to partition $\mathcal{E}_2$ into the two events

$$\begin{aligned}\mathcal{E}_{21} &= \big\{(Q^n, X_1^n(m_1), X_2^n(1), Y_1^n) \in \mathcal{T}_\varepsilon^{(n)} \\ &\qquad \text{for some } m_1 \neq 1\big\},\\ \mathcal{E}_{22} &= \big\{(Q^n, X_1^n(m_1), X_2^n(m_2), Y_1^n) \in \mathcal{T}_\varepsilon^{(n)} \\ &\qquad \text{for some } m_1 \neq 1,\, m_2 \neq 1\big\}.\end{aligned}$$

Again by the packing lemma, $\mathsf{P}(\mathcal{E}_{21})$ and $\mathsf{P}(\mathcal{E}_{22})$ tend to zero as $n \to \infty$ if

$$R_1 < I(X_1; Y_1 \mid X_2, Q) - \delta(\varepsilon), \tag{8a}$$
$$R_1 + R_2 < I(X_1, X_2; Y_1 \mid Q) - \delta(\varepsilon). \tag{8b}$$

Thus we have shown that the average probability of decoding error at receiver 1 tends to zero as $n \to \infty$ if at least one of (7) and (8a, 8b) holds. Similarly, we can show that the average probability of decoding error at receiver 2 tends to zero as $n \to \infty$ if $R_2 < I(X_2; Y_2 \mid Q) - \delta(\varepsilon)$, or $R_2 < I(X_2; Y_2 \mid X_1, Q) - \delta(\varepsilon)$ and $R_1 + R_2 < I(X_1, X_2; Y_2 \mid Q) - \delta(\varepsilon)$. Since $\varepsilon > 0$ is arbitrary and $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$, this completes the proof of achievability for any rate pair (R_1, R_2) in the interior of $\mathscr{R}_1(p) \cap \mathscr{R}_2(p)$. □

Remark 4 (Comparison to Maximum Likelihood Decoding): It is instructive to consider the following progression of decoding rules for receiver 1.

1) Maximum likelihood decoding:

$$\begin{aligned}\hat{m}_1 &= \arg\max_{m_1} p(y_1^n \mid m_1)\\ &= \arg\max_{m_1} \frac{1}{2^{nR_2}} \sum_{m_2} p(y_1^n \mid m_1, m_2)\\ &= \arg\max_{m_1} \frac{1}{2^{nR_2}} \sum_{m_2} \prod_{i=1}^n p_{Y_1|X_1,X_2} \\ &\qquad\qquad (y_{1i} \mid x_{1i}(m_1), x_{2i}(m_2)),\end{aligned} \tag{9}$$

which is the optimal decoding rule.

2) Simultaneous maximum likelihood decoding:

$$\hat{m}_1 = \arg\max_{m_1} \max_{m_2} p(y_1^n \mid m_1, m_2),$$

which is equivalent to performing optimal decoding of the message pair (M_1, M_2) and then taking the first coordinate. Note the maximum over m_2 instead of the average as in (9).

3) Typicality score decoding:

$$\hat{m}_1 = \arg\min_{m_1} \min_{m_2} \varepsilon^\star(y_1^n, m_1, m_2),$$

where $\varepsilon^\star(y_1^n, m_1, m_2)$ is defined as the smallest ε such that

$$(q^n, x_1^n(m_1), x_2^n(m_2), y_1^n) \in \mathcal{T}_\varepsilon^{(n)}.$$

Here the notion of joint typicality plays the role of likelihood in previous decoding rules and $\varepsilon^\star$ captures the penalty for being atypical.

4) Simultaneous nonunique decoding: Find the unique $\hat{m}_1$ such that

$$(q^n, x_1^n(\hat{m}_1), x_2^n(m_2), y_1^n) \in \mathcal{T}_\varepsilon^{(n)} \quad \text{for some } m_2.$$

This is equivalent to performing typicality score decoding with predetermined threshold ε for $\varepsilon^\star(y_1^n, m_1, m_2)$; thus first forming a list of all (m_1, m_2) for which $\varepsilon^\star(y_1^n, m_1, m_2) \leq \varepsilon$, and then taking the first coordinate of the members of the list (if it is unique).

Starting from the optimal maximum likelihood decoding rule, each subsequent rule modifies its predecessor by "relaxing" one step. Nonetheless, these relaxation steps do not result in any significant loss in performance, as is evident in the rate-optimality of the simultaneous nonunique decoding rule.

Remark 5: As observed in [4] (see also (5) in Remark 1 above), each rate point in $\mathscr{R}^*(p)$ can alternatively be achieved by having each receiver specifically decode for either the desired message alone or both the desired and interfering messages.

B. Proof of the Converse

Fix a pmf $p = p(q)\,p(x_1|q)\,p(x_2|q)$ and let (R_1, R_2) be a rate pair achievable by the p-distributed random code ensemble. We prove that this implies that $(R_1, R_2) \in \mathscr{R}_1(p) \cap \mathscr{R}_2(p)$ as claimed. Here, we show the details for the inclusion $(R_1, R_2) \in \mathscr{R}_1(p)$; the proof for $(R_1, R_2) \in \mathscr{R}_2(p)$ follows similarly. With slight abuse of notation, let $\mathcal{C}_n$ denote the random codebook (and the time sharing sequence), namely $(Q^n, X_1^n(1), \ldots, X_1^n(2^{nR_1}), X_2^n(1), \ldots, X_2^n(2^{nR_2}))$.

First consider a fixed codebook $\mathcal{C}_n = c$. By Fano's inequality,

$$H(M_1 \mid Y_1^n, \mathcal{C}_n = c) \leq 1 + nR_1 P_e^{(n)}(c).$$

Taking the expectation over the random codebook $\mathcal{C}_n$, it follows that

$$H(M_1 \mid Y_1^n, \mathcal{C}_n) \leq 1 + nR_1\, \mathsf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)] \leq n\varepsilon_n, \tag{10}$$

where $\varepsilon_n \to 0$ as $n \to \infty$ since $\mathsf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)] \to 0$.

We prove the conditions in the min form (6a, 6b). To see that the first inequality is true, note that

$$\begin{aligned} n(R_1 - \varepsilon_n) &= H(M_1 \mid \mathcal{C}_n) - n\varepsilon_n \\ &\overset{(a)}{\le} I(M_1; Y_1^n \mid \mathcal{C}_n) \\ &\le I(X_1^n; Y_1^n \mid \mathcal{C}_n) \\ &\le I(X_1^n; Y_1^n, X_2^n \mid \mathcal{C}_n) \\ &= I(X_1^n; Y_1^n \mid X_2^n, \mathcal{C}_n) \\ &= H(Y_1^n \mid X_2^n, \mathcal{C}_n) - H(Y_1^n \mid X_1^n, X_2^n, \mathcal{C}_n) \\ &\overset{(b)}{\le} H(Y_1^n \mid X_2^n, Q^n) - H(Y_1^n \mid X_1^n, X_2^n, Q^n) \\ &\overset{(c)}{=} nI(X_1; Y_1 \mid X_2, Q), \end{aligned}$$

where (a) follows by (the averaged version of) Fano's inequality in (10), (b) follows by omitting some conditioning and using the memoryless property of the channel, and (c) follows since the tuple $(Q_i, X_{1i}, X_{2i}, Y_i)$ is i.i.d. for all i. Note that unlike conventional converse proofs where nothing can be assumed about the codebook structure, here we can take advantage of the properties of a given codebook generation procedure.

To prove the second inequality in (6), we need the following lemma, which is proved in Appendix A.

Lemma 1:

$$\lim_{n\to\infty} \frac{1}{n} H(Y_1^n \mid X_1^n, \mathcal{C}_n) = H(Y_1 \mid X_1, X_2, Q) + \min\{R_2, I(X_2; Y_1 \mid X_1, Q)\}.$$

The lemma states that depending on R_2, $(1/n)H(Y_1^n \mid X_1^n, \mathcal{C}_n)$ either tends to $H(Y_1 \mid X_1, Q)$, that is, the remaining received sequence after recovering the desired codeword looks like i.i.d. noise, or to $R_2 + H(Y_1 \mid X_1, X_2, Q)$, that is, the receiver can distinguish the interfering codeword from the noise.

Equipped with this lemma, we have

$$\begin{aligned} n(R_1 - \varepsilon_n) &\overset{(a)}{\le} I(X_1^n; Y_1^n \mid \mathcal{C}_n) \\ &= H(Y_1^n \mid \mathcal{C}_n) - H(Y_1^n \mid X_1^n, \mathcal{C}_n) \\ &\le H(Y_1^n \mid Q^n) - H(Y_1^n \mid X_1^n, \mathcal{C}_n) \\ &\overset{(b)}{\le} nH(Y_1 \mid Q) - nH(Y_1 \mid X_1, X_2, Q) \\ &\quad - \min\{nR_2, nI(X_2; Y_1 \mid X_1, Q)\} + n\varepsilon_n \\ &= nI(X_1, X_2; Y_1 \mid Q) \\ &\quad + \min\{nR_2, nI(X_2; Y_1 \mid X_1, Q)\} + n\varepsilon_n. \end{aligned}$$

Here, (a) follows by Fano's inequality and (b) follows by Lemma 1 with some ε_n that tends to zero as $n \to \infty$. The conditions for $\mathscr{R}_2(p)$ can be proved similarly. This completes the proof of the converse.

III. DM-IN With K Senders and L Receivers

We generalize the previous result to the K-sender, L-receiver discrete memoryless interference network ((K, L)-DM-IN) with input alphabets $\mathcal{X}_1, \dots, \mathcal{X}_K$, output alphabets $\mathcal{Y}_1, \dots, \mathcal{Y}_L$, and pmfs $p(y_1, \dots, y_L \mid x_1, \dots, x_K)$. In this network, each sender $k \in [1:K]$ communicates an independent message M_k at rate R_k and each receiver $l \in [1:L]$ wishes to recover the messages sent by a subset $\mathcal{D}_l \subseteq [1:K]$ of senders (also referred to as a demand set). The channel is depicted in Fig. 4.

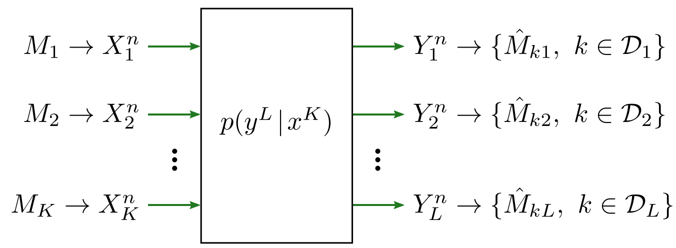


Fig. 4. Discrete memoryless interference network with K senders and L receivers.

More formally, a $(2^{nR_1}, \dots, 2^{nR_K}, n)$ code $\mathcal{C}_n$ for the (K, L)-DM-IN consists of

- K message sets $[1:2^{nR_1}]$, …, $[1:2^{nR_K}]$,
- K encoders, where encoder $k \in [1:K]$ assigns a codeword $x_k^n(m_k)$ to each message $m_k \in [1:2^{nR_k}]$,
- L decoders, where decoder $l \in [1:L]$ assigns estimates $\hat{m}_{kl}$, $k \in \mathcal{D}_l$, or an error message e to each received sequence y_l^n.

We assume that the message tuple $(M_1, \dots, M_K)$ is uniformly distributed over $[1:2^{nR_1}] \times \cdots \times [1:2^{nR_K}]$. The average probability of error for the code $\mathcal{C}_n$ is defined as

$$P_e^{(n)}(\mathcal{C}_n) = \mathsf{P}\big\{\hat{M}_{kl} \ne M_k \text{ for some } l \in [1:L],\ k \in \mathcal{D}_l\big\}.$$

A rate tuple $(R_1, \dots, R_K)$ is said to be *achievable* for the DM-IN if there exists a sequence of $(2^{nR_1}, \dots, 2^{nR_K}, n)$ codes $\mathcal{C}_n$ such that $\lim_{n\to\infty} P_e^{(n)}(\mathcal{C}_n) = 0$. The *capacity region* $\mathscr{C}$ of the (K, L)-DM-IN is the closure of the set of achievable rate tuples $(R_1, \dots, R_K)$.

As in Section II, we limit our attention to a randomly generated code ensemble with a special structure. Let $p = p(q)\, p(x_1|q) \cdots p(x_K|q)$ be a given pmf on $\mathcal{Q} \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_K$, where $\mathcal{Q}$ is a finite alphabet. Suppose that codewords $X_k^n(m_k)$, $m_k \in [1:2^{nR_k}]$, $k \in [1:K]$, are generated randomly as follows.

- Let $Q^n \sim \prod_{i=1}^n p_Q(q_i)$.
- For each $k \in [1:K]$ and $m_k \in [1:2^{nR_k}]$, let $X_k^n(m_k) \sim \prod_{i=1}^n p_{X_k|Q}(x_{ki}|q_i)$, conditionally independent given Q^n.

Each instance of codebooks generated in this manner, along with the corresponding optimal decoders, constitutes a $(2^{nR_1}, \dots, 2^{nR_K}, n)$ code. We refer to the random code ensemble thus generated as the $(2^{nR_1}, \dots, 2^{nR_K}, n; p)$ *random code ensemble*.

Definition 2 (Random Coding Optimal Rate Region): Given a pmf $p = p(q)\, p(x_1|q) \cdots p(x_K|q)$, the *optimal rate region* $\mathscr{R}^*(p)$ *achievable by the* p*-distributed random code ensemble* is the closure of the set of rate tuples $(R_1, \dots, R_K)$ such that the sequence of the $(2^{nR_1}, \dots, 2^{nR_K}, n; p)$ random code ensembles $\mathcal{C}_n$ satisfies

$$\lim_{n\to\infty} \mathsf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)] = 0,$$

where the expectation is with respect to the random code ensemble $\mathcal{C}_n$.

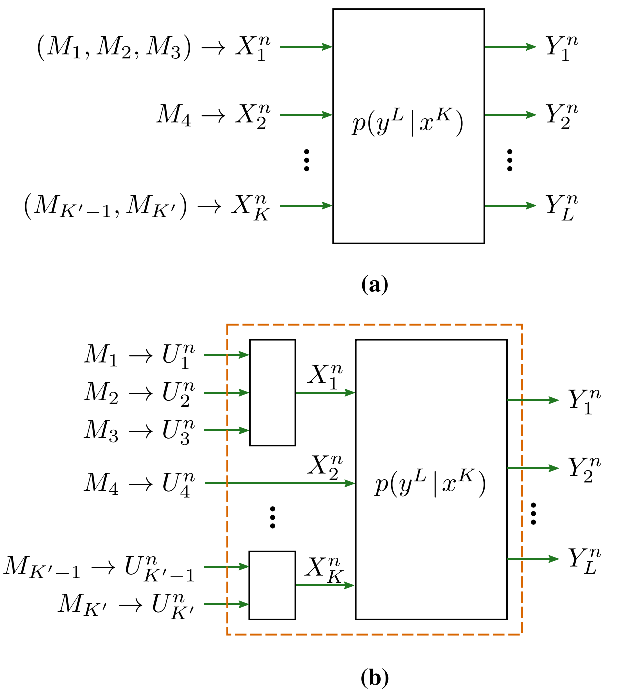


Fig. 5. The class of (K, L)-DM-INs includes superposition coding with an arbitrary number of layers. (a) Multiple messages per sender via superposition coding. (b) Equivalent channel with a single message per sender.

We note that the p-distributed random code ensemble for the (K, L)-DM-IN captures superposition coding with an arbitrary number of layers. Suppose that there are K senders, some of which need to communicate multiple messages (see Fig. 5(a)). In superposition coding, each message at a sender is encoded into a codeword $U_{k'}^n$ and the sender combines (superimposes) all such codewords. By merging the combining functions at the sender with the physical channel $p(y^L|x^K)$, we obtain a (K', L)-DM-IN $p(y^L|u^{K'})$ with "virtual" inputs $U_{k'}$, $k' \in [1 : K']$, as illustrated in Fig. 5(b).

Define the rate region $\mathscr{R}_1(p)$ as

$$\mathscr{R}_1(p) = \bigcup_{\substack{\mathcal{S} \subseteq [1:K],\\ \mathcal{D}_1 \subseteq \mathcal{S}}} \mathscr{R}_{\mathrm{MAC}(\mathcal{S})}(p), \tag{11}$$

where $\mathscr{R}_{\mathrm{MAC}(\mathcal{S})}(p)$ is the achievable rate region for the multiple access channel from the set of senders $\mathcal{S}$ to receiver 1, i.e., the set of rate tuples $(R_1, \ldots, R_K)$ such that

$$R_{\mathcal{T}} = \sum_{j \in \mathcal{T}} R_j \le I(X_{\mathcal{T}}; Y_1 \mid X_{\mathcal{S} \setminus \mathcal{T}}, Q) \quad \text{for all } \mathcal{T} \subseteq \mathcal{S}.$$

Note that the set $\mathscr{R}_{\mathrm{MAC}(\mathcal{S})}(p)$ corresponds to the rate region achievable by decoding for the messages from the senders $\mathcal{S}$, which contains all desired messages and possibly some interfering messages. Also note that $\mathscr{R}_{\mathrm{MAC}(\mathcal{S})}(p)$ contains upper bounds only on the rates R_k, $k \in \mathcal{S}$, of the active senders $\mathcal{S}$ in the MAC. The signals from the inactive senders in $\mathcal{S}^{\mathrm{c}}$ are treated as noise and the corresponding rates R_k for $k \in \mathcal{S}^{\mathrm{c}}$ are unconstrained. Consequently, $\mathscr{R}_1(p)$ is unbounded in the coordinates R_k for $k \in [1 : K] \setminus \mathcal{D}_1$.

The region $\mathscr{R}_1(p)$ in (11) can equivalently be written as the set of rate tuples $(R_1, \ldots, R_K)$ such that for all $\mathcal{U} \subseteq [1 : K] \setminus \mathcal{D}_1$ and for all $\mathcal{D}$ with $\emptyset \subset \mathcal{D} \subseteq \mathcal{D}_1$,

$$\begin{aligned} R_{\mathcal{D}} + \min_{\mathcal{U}' \subseteq \mathcal{U}} \big(R_{\mathcal{U}'} + I(X_{\mathcal{U} \setminus \mathcal{U}'}; Y_1 \mid X_{\mathcal{D}}, X_{\mathcal{U}'}, X_{[1:K] \setminus \mathcal{D} \setminus \mathcal{U}}, Q) \big) \\ \le I(X_{\mathcal{D}}, X_{\mathcal{U}}; Y_1 \mid X_{[1:K] \setminus \mathcal{D} \setminus \mathcal{U}}, Q). \end{aligned} \tag{12}$$

As in the case of the 2-DM-IC, each argument of each term in the minimum represents a different mode of signal saturation. The equivalence between the MAC form (11) and the min form (12) can be proved by identifying the largest set of decodable interfering messages as in [12]. For completeness, we provide a proof in Appendix B.

Remark 6: The MAC and min forms of $\mathscr{R}_1(p)$ are duals to each other in the following sense. The condition for $(R_1, \ldots, R_K) \in \mathscr{R}_1(p)$ in the MAC form (11) can be expressed as

$$\begin{aligned} \exists \mathcal{S} \subseteq [1 : K], \quad \mathcal{D}_1 \subseteq \mathcal{S} : \\ \forall \mathcal{T} \subseteq \mathcal{S} : \\ R_{\mathcal{T}} \le I(X_{\mathcal{T}}; Y_1 \mid X_{\mathcal{S} \setminus \mathcal{T}}, Q). \end{aligned} \tag{13}$$

The conditions in the min form (12) can be rewritten[1] as

$$\begin{aligned} \forall \mathcal{V} \subseteq [1 : K], \quad \mathcal{V} \cap \mathcal{D}_1 \ne \emptyset : \\ \exists \mathcal{V}' \subseteq \mathcal{V}, \quad \mathcal{V}' \cap \mathcal{D}_1 = \mathcal{V} \cap \mathcal{D}_1 : \\ R_{\mathcal{V}'} \le I(X_{\mathcal{V}'}; Y_1 \mid X_{[1:K] \setminus \mathcal{V}}, Q). \end{aligned} \tag{14}$$

Both conditions involve a set of messages from the senders $\mathcal{S}$ (or $\mathcal{V}$) and its subset $\mathcal{T}$ (or $\mathcal{V}'$), and impose a mutual information upper bound on the sum rate over the subset. The key difference is the order of the quantifiers $\forall$ and $\exists$.

Analogous to $\mathscr{R}_1(p)$, define the regions $\mathscr{R}_2(p), \ldots, \mathscr{R}_L(p)$ for receivers $2, \ldots, L$ by making appropriate index substitutions. We are now ready to state the main result for the (K, L)-DM-IN.

Theorem 2: Given a pmf $p = p(q)\, p(x_1|q) \cdots p(x_K|q)$, the optimal rate region of the (K, L)-DM-IN $p(y^L|x^K)$ with demand sets $\mathcal{D}_1, \ldots, \mathcal{D}_L$ achievable by the p-distributed random code ensemble is

$$\mathscr{R}^*(p) = \bigcap_{l \in [1:L]} \mathscr{R}_l(p).$$

Note that, as for its 2-DM-IC counterpart, this region is not convex in general.

Example 1: Consider the K-user-pair Gaussian interference network

$$Y_l = \sum_{k=1}^{K} g_{kl} X_k + Z_l, \quad l \in [1 : K],$$

where $Z_l \sim \mathrm{N}(0, 1)$ and g_{kl} are channel gains from sender k to receiver l. Assume the Gaussian random code ensemble with $X_k \sim \mathrm{N}(0, 1)$, $k \in [1 : K]$. The optimal rate region achievable

[1] To see this, first note that the minimum terms on the left hand side of (12) represent a set of conditions of which at least one has to be true, then use the identity

$$\begin{aligned} & I(X_{\mathcal{D}}, X_{\mathcal{U}}; Y_1 \mid X_{[1:K] \setminus \mathcal{D} \setminus \mathcal{U}}, Q) \\ & \quad - I(X_{\mathcal{U} \setminus \mathcal{U}'}; Y_1 \mid X_{\mathcal{D}}, X_{\mathcal{U}'}, X_{[1:K] \setminus \mathcal{D} \setminus \mathcal{U}}, Q) \\ & \quad = I(X_{\mathcal{D}}, X_{\mathcal{U}'}; Y_1 \mid X_{[1:K] \setminus \mathcal{D} \setminus \mathcal{U}}, Q), \end{aligned}$$

and finally, let $\mathcal{V} = \mathcal{U} \cup \mathcal{D}$ and $\mathcal{V}' = \mathcal{U}' \cup \mathcal{D}$.

by this random code ensemble was established in [2] and [12], and can be recovered from Theorem 2 by letting $K = L$, $\mathcal{D}_k = \{k\}$ for $k \in [1:K]$, and applying the discretization procedure in [7, Sec. 3.4]. Theorem 2 generalizes this result in several directions, since (a) it applies to non-Gaussian networks, (b) it applies to non-Gaussian random code ensembles (which is crucial to analyze the performance under a fixed constellation), and (c) it includes coded time sharing and superposition coding.

Example 2: Consider the deterministic interference channel with three sender–receiver pairs (3-DIC) [3], where

$$\begin{aligned} Y_1 &= f_1(g_{11}(X_1), h_1(g_{21}(X_2), g_{31}(X_3))),\\ Y_2 &= f_2(g_{22}(X_2), h_2(g_{32}(X_3), g_{12}(X_1))),\\ Y_3 &= f_3(g_{33}(X_3), h_3(g_{13}(X_1), g_{23}(X_2))) \end{aligned}$$

for some loss functions g_{kl} and combining functions h_k and f_k, $k, l \in \{1, 2, 3\}$. The combining functions are supposed to be injective in each argument. This setting is of interest since it contains as special cases the El Gamal–Costa two-user-pair interference channel [8], for which the Han–Kobayashi coding scheme achieves the capacity region, and the Avestimehr–Diggavi–Tse q-ary expansion deterministic (QED) interference channel [1], which approximates Gaussian interference networks in the high-power regime. The 3-DIC is an instance of a (K, L)-DM-IN with $L = K = 3$ and $\mathcal{D}_k = \{k\}$ for $k \in [1:K]$. The interference decoding inner bound on the 3-DIC capacity region in [3] coincides with the region in Theorem 2 in its min form. Beyond the results in [3], Theorem 2 establishes that the interference decoding inner bound is in fact optimal given the codebook structure. Note that for the 3-DIC channel, we can identify each minimum term with a specific signal in the channel block diagram for which the term counts the number of distinguishable sequences.

Proof of Theorem 2: We focus only on receiver 1 for which M_k, $k \in \mathcal{D}_1$, are the desired messages and M_k, $k \in \mathcal{D}_1^c = [1:K] \setminus \mathcal{D}_1$, are interfering messages. Achievability is proved using simultaneous nonunique decoding. Receiver 1 declares that $\hat{m}_{\mathcal{D}_1}$ is sent if it is the unique message tuple such that

$$\big(q^n, x^n_{\mathcal{D}_1}(\hat{m}_{\mathcal{D}_1}), x^n_{\mathcal{D}_1^c}(m_{\mathcal{D}_1^c}), y_1^n\big) \in \mathcal{T}_\varepsilon^{(n)} \quad \text{for some } m_{\mathcal{D}_1^c},$$

where $x^n_{\mathcal{D}_1}(\hat{m}_{\mathcal{D}_1})$ is the tuple of $x^n_k(\hat{m}_k)$, $k \in \mathcal{D}_1$, and similarly, $x^n_{\mathcal{D}_1^c}(m_{\mathcal{D}_1^c})$ is the tuple of $x^n_k(m_k)$, $k \in \mathcal{D}_1^c$. The analysis follows similar steps as in Subsection II-A.

To prove the converse, fix a pmf p and let $(R_1, \ldots, R_K)$ be a rate tuple that is achievable by the p-distributed random code ensemble. We need the following generalization of Lemma 1, which is proved in Appendix C.

Lemma 2: If $\mathcal{D}_1 \subseteq \mathcal{S} \subseteq [1:K]$, then

$$\lim_{n\to\infty} \frac{1}{n} H(Y_1^n \,|\, X^n_{\mathcal{S}}, \mathcal{C}_n) = H(Y_1 \,|\, X_{[1:K]}, Q) + \min_{\mathcal{U} \subseteq \mathcal{S}^c} \big(R_{\mathcal{U}} + I(X_{(\mathcal{S}\cup\mathcal{U})^c}; Y_1 \,|\, X_{\mathcal{S}\cup\mathcal{U}}, Q)\big).$$

We now establish (12) as follows. Fix a subset of desired message indices, $\mathcal{D} \subseteq \mathcal{D}_1$, and a subset of interfering message indices, $\mathcal{U} \subseteq \mathcal{D}_1^c$. Then

$$\begin{aligned} &n(R_{\mathcal{D}} - \varepsilon_n)\\ &\overset{(a)}{\le} I(X^n_{\mathcal{D}}; Y_1^n \,|\, \mathcal{C}_n)\\ &\le I(X^n_{\mathcal{D}}; Y_1^n, X^n_{(\mathcal{D}\cup\mathcal{U})^c} \,|\, \mathcal{C}_n)\\ &\le I(X^n_{\mathcal{D}}; Y_1^n \,|\, X^n_{(\mathcal{D}\cup\mathcal{U})^c}, \mathcal{C}_n)\\ &= H(Y_1^n \,|\, X^n_{(\mathcal{D}\cup\mathcal{U})^c}, \mathcal{C}_n) - H(Y_1^n \,|\, X^n_{\mathcal{U}^c}, \mathcal{C}_n)\\ &\overset{(b)}{\le} nH(Y_1 \,|\, X_{(\mathcal{D}\cup\mathcal{U})^c}, Q) - nH(Y_1 \,|\, X_{[1:K]}, Q)\\ &\quad - n \cdot \min_{\mathcal{U}' \subseteq \mathcal{U}} \big(R_{\mathcal{U}'} + I(X_{(\mathcal{U}^c\cup\mathcal{U}')^c}; Y_1 \,|\, X_{\mathcal{U}^c\cup\mathcal{U}'}, Q)\big) + n\varepsilon_n\\ &= nI(X_{\mathcal{D}\cup\mathcal{U}}; Y_1^n \,|\, X_{(\mathcal{D}\cup\mathcal{U})^c}, Q)\\ &\quad - n \cdot \min_{\mathcal{U}' \subseteq \mathcal{U}} \big(R_{\mathcal{U}'} + I(X_{\mathcal{U}\setminus\mathcal{U}'}; Y_1 \,|\, X_{(\mathcal{U}\setminus\mathcal{U}')^c}, Q)\big) + n\varepsilon_n, \end{aligned}$$

where (a) follows by Fano's inequality and (b) follows by Lemma 2. This completes the proof of the converse. □

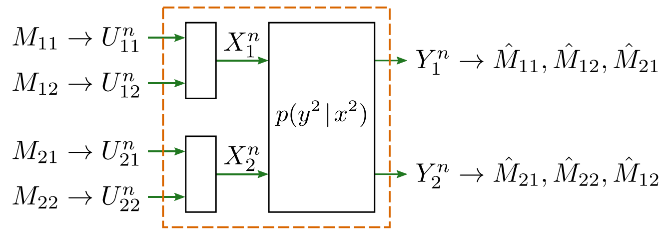


Fig. 6. Han–Kobayashi coding scheme.

IV. Application to the Han–Kobayashi Coding Scheme

We revisit the two-user-pair DM-IC in Fig. 1. The best known inner bound on the capacity region is achieved by the Han–Kobayashi coding scheme [9]. In this scheme, the message M_1 is split into common and private messages M_{12} and M_{11} at rates R_{12} and R_{11}, respectively, such that $R_1 = R_{12} + R_{11}$. Similarly M_2 is split into common and private messages M_{21} and M_{22} at rates R_{21} and R_{22} such that $R_2 = R_{22} + R_{21}$. More specifically, the scheme uses random codebook generation and coded time sharing as follows. Fix a pmf $p = p(q)\,p(u_{11}|q)\,p(u_{12}|q)\,p(u_{21}|q)\,p(u_{22}|q)\,p(x_1|u_{11}, u_{12}, q)\,p(x_2|u_{21}, u_{22}, q)$, where the latter two conditional pmfs represent deterministic mappings $x_1(u_{11}, u_{12})$ and $x_2(u_{21}, u_{22})$. Randomly generate a coded time sharing sequence $q^n \sim \prod_{i=1}^n p_Q(q_i)$. For each $k, k' \in \{1, 2\}$ and $m_{kk'} \in [1 : 2^{nR_{kk'}}]$, randomly and conditionally independently generate a sequence $u^n_{kk'}(m_{kk'})$ according to $\prod_{i=1}^n p_{U_{kk'}|Q}(u_{kk'i}|q_i)$. To communicate message pair (m_{11}, m_{12}), sender 1 transmits $x_{1i} = x_1(u_{11i}, u_{12i})$ for $i \in [1:n]$, and analogously for sender 2. Receiver $k = 1, 2$ recovers its intended message M_k and the common message from the other sender (although it is not required to). While this decoding scheme helps reduce the effect of interference, it results in additional constraints on the rates for common messages. The Han–Kobayashi coding scheme is illustrated in Fig. 6.

Let $\mathscr{R}_{\mathrm{HK},1}(p)$ be defined as the set of rate tuples $(R_{11}, R_{12}, R_{21}, R_{22})$ such that

$$R_{11} \le I(U_{11}; Y_1 \,|\, U_{12}, U_{21}, Q), \tag{15a}$$
$$R_{12} \le I(U_{12}; Y_1 \,|\, U_{11}, U_{21}, Q), \tag{15b}$$
$$R_{21} \le I(U_{21}; Y_1 \,|\, U_{11}, U_{12}, Q), \tag{15c}$$
$$R_{11} + R_{12} \le I(U_{11}, U_{12}; Y_1 \,|\, U_{21}, Q), \tag{15d}$$
$$R_{11} + R_{21} \le I(U_{11}, U_{21}; Y_1 \,|\, U_{12}, Q), \tag{15e}$$
$$R_{12} + R_{21} \le I(U_{12}, U_{21}; Y_1 \,|\, U_{11}, Q), \tag{15f}$$
$$R_{11} + R_{12} + R_{21} \le I(U_{11}, U_{12}, U_{21}; Y_1 \,|\, Q). \tag{15g}$$

Similarly, define $\mathscr{R}_{\mathrm{HK},2}(p)$ by making the sender/receiver index substitutions $1 \leftrightarrow 2$ in the definition of $\mathscr{R}_{\mathrm{HK},1}(p)$. As shown by Han and Kobayashi [9], the coding scheme achieves any rate pair (R_1, R_2) that is in the interior of

$$\begin{aligned}\mathscr{R}_{\mathrm{HK}} &= \mathrm{Proj}_{4\to 2}\left(\textstyle\bigcup_p \mathscr{R}_{\mathrm{HK},1}(p) \cap \mathscr{R}_{\mathrm{HK},2}(p)\right)\\ &= \textstyle\bigcup_p \mathrm{Proj}_{4\to 2}\Big(\mathscr{R}_{\mathrm{HK},1}(p) \cap \mathscr{R}_{\mathrm{HK},2}(p)\Big),\end{aligned} \tag{16}$$

where $\mathrm{Proj}_{4\to 2}$ is the projection that maps the 4-dimensional (convex) set of rate tuples $(R_{11}, R_{12}, R_{21}, R_{22})$ into a 2-dimensional rate region of rate pairs $(R_1, R_2) = (R_{11} + R_{12}, R_{21} + R_{22})$ and the unions are taken over all pmfs $p = p(q)\, p(u_{11}|q)\, p(u_{12}|q)\, p(u_{21}|q)\, p(u_{22}|q)\, p(x_1|u_{11}, u_{12}, q)\, p(x_2|u_{21}, u_{22}, q)$.

We are interested in finding the rate region that is achievable by the Han–Kobayashi encoding functions in conjunction with the *optimal* decoding functions. To this end, note that by combining the channel and the deterministic mappings as indicated by the dashed box in Fig. 6, the channel $(U_{11}, U_{12}, U_{21}, U_{22}) \to (Y_1, Y_2)$ is a $(4, 2)$-DM-IN. After removing the artificial requirement for each decoder to recover the interfering sender's common message, the message demands are $\mathcal{D}_1 = \{11, 12\}$ and $\mathcal{D}_2 = \{21, 22\}$. Moreover, the Han–Kobayashi encoding scheme is in fact the p-distributed random code ensemble applied to this network, as defined in Section III.

Definition 3: The *optimal rate region $\mathscr{R}_{opt}$ achievable by the Han–Kobayashi random code ensembles* is defined as

$$\mathscr{R}_{\mathrm{opt}} = \mathrm{Proj}_{4\to 2}\left(\bigcup_p \mathscr{R}^*(p)\right) = \bigcup_p \mathrm{Proj}_{4\to 2}\Big(\mathscr{R}^*(p)\Big),$$

where the union is over pmfs of the form $p = p(q)\, p(u_{11}|q)\, p(u_{12}|q)\, p(u_{21}|q)\, p(u_{22}|q)\, p(x_1|u_{11}, u_{12})\, p(x_2|u_{21}, u_{22})$ with the latter two factors representing deterministic mappings $x_1(u_{11}, u_{12})$ and $x_2(u_{21}, u_{22})$, and $\mathscr{R}^*(p)$ is the optimal rate region achievable by the $p(q)\, p(u_{11}|q)\; p(u_{12}|q)\, p(u_{21}|q)\, p(u_{22}|q)$-distributed random code ensemble for the $(4, 2)$-DM-IN $p(y_1, y_2|u_{11}, u_{12}, u_{21}, u_{22}) = p_{Y_1,Y_2|X_1,X_2}(y_1, y_2|x_1(u_{11}, u_{12}), x_2(u_{21}, u_{22}))$ (cf. Definition 2).

Then Theorem 2 implies the following.

Theorem 3: $\mathscr{R}_{\mathrm{opt}} = \mathscr{R}_{\mathrm{HK}}$.

Thus, the Han–Kobayashi inner bound is optimal when encoding is restricted to randomly generated codebooks, superposition coding, and coded time sharing. It cannot be enlarged by replacing the decoders used in the proof of (15a–15g) with optimal decoders.

Proof of Theorem 3: Applying Theorem 2 to the definition of $\mathscr{R}_{\mathrm{opt}}$ yields

$$\mathscr{R}_{\mathrm{opt}} = \mathrm{Proj}_{4\to 2}\left(\bigcup_p \mathscr{R}_1(p) \cap \mathscr{R}_2(p)\right),$$

where $\mathscr{R}_1(p)$ is the set of rate tuples $(R_{11}, R_{12}, R_{21}, R_{22})$ such that

$$R_{\mathcal{T}_1} \le I(U_{\mathcal{T}_1}; Y_1 \,|\, U_{\mathcal{S}_1\setminus\mathcal{T}_1}, Q) \quad \text{for all } \mathcal{T}_1 \subseteq \mathcal{S}_1 \tag{17}$$

for some $\mathcal{S}_1$ with $\{11, 12\} \subseteq \mathcal{S}_1 \subseteq \{11, 12, 21, 22\}$. Likewise, $\mathscr{R}_2(p)$ is the set of rate tuples that satisfy

$$R_{\mathcal{T}_2} \le I(U_{\mathcal{T}_2}; Y_2 \,|\, U_{\mathcal{S}_2\setminus\mathcal{T}_2}, Q) \quad \text{for all } \mathcal{T}_2 \subseteq \mathcal{S}_2 \tag{18}$$

for some $\mathcal{S}_2$ with $\{21, 22\} \subseteq \mathcal{S}_2 \subseteq \{11, 12, 21, 22\}$. Here, $\mathcal{S}_1$ and $\mathcal{S}_2$ contain the indices of the messages recovered by receivers 1 and 2, respectively.

In order to compare $\mathscr{R}_{\mathrm{opt}}$ to $\mathscr{R}_{\mathrm{HK}}$, recall (15) and (16) and the compact description of $\mathscr{R}_{\mathrm{HK}}$ in [6] as the set of all rate pairs (R_1, R_2) such that

$$R_1 \le I(U_{11}, U_{12}; Y_1 \,|\, U_{21}, Q), \tag{19a}$$
$$R_2 \le I(U_{21}, U_{22}; Y_2 \,|\, U_{12}, Q), \tag{19b}$$
$$\begin{aligned}R_1 + R_2 \le\; &I(U_{11}, U_{12}, U_{21}; Y_1 \,|\, Q)\\ &+ I(U_{22}; Y_2 \,|\, U_{12}, U_{21}, Q),\end{aligned} \tag{19c}$$
$$\begin{aligned}R_1 + R_2 \le\; &I(U_{12}, U_{21}, U_{22}; Y_2 \,|\, Q)\\ &+ I(U_{11}; Y_1 \,|\, U_{12}, U_{21}, Q),\end{aligned} \tag{19d}$$
$$\begin{aligned}R_1 + R_2 \le\; &I(U_{11}, U_{21}; Y_1 \,|\, U_{12}, Q)\\ &+ I(U_{12}, U_{22}; Y_2 \,|\, U_{21}, Q),\end{aligned} \tag{19e}$$
$$\begin{aligned}2R_1 + R_2 \le\; &I(U_{11}, U_{12}, U_{21}; Y_1 \,|\, Q)\\ &+ I(U_{11}; Y_1 \,|\, U_{12}, U_{21}, Q)\\ &+ I(U_{12}, U_{22}; Y_2 \,|\, U_{21}, Q),\end{aligned} \tag{19f}$$
$$\begin{aligned}R_1 + 2R_2 \le\; &I(U_{12}, U_{21}, U_{22}; Y_2 \,|\, Q)\\ &+ I(U_{22}; Y_2 \,|\, U_{12}, U_{21}, Q)\\ &+ I(U_{11}, U_{21}; Y_1 \,|\, U_{12}, Q)\end{aligned} \tag{19g}$$

for some pmf of the form $p = p(q)\, p(u_{11}|q)\, p(u_{12}|q)\, p(u_{21}|q)\, p(u_{22}|q)\, p(x_1|u_{11}, u_{12})\, p(x_2|u_{21}, u_{22})$, where the latter two factors represent deterministic mappings $x_1(u_{11}, u_{12})$ and $x_2(u_{21}, u_{22})$.

It is easy to see that $\mathscr{R}_{\mathrm{HK}} \subseteq \mathscr{R}_{\mathrm{opt}}$. Choosing $\mathcal{S}_1 = \{11, 12, 21\}$ in (17), the resulting conditions coincide with the ones in (15), and the constituent sets satisfy the condition $\mathscr{R}_{\mathrm{HK},1}(p) \subseteq \mathscr{R}_1(p)$. Likewise, choosing $\mathcal{S}_2 = \{12, 21, 22\}$ in (18), $\mathscr{R}_{\mathrm{HK},2}(p) \subseteq \mathscr{R}_2(p)$, and the desired inclusion follows.

To show that $\mathscr{R}_{\mathrm{opt}} \subseteq \mathscr{R}_{\mathrm{HK}}$, note that conditions (17) and (18) must hold for some $\mathcal{S}_1 \supseteq \{11, 12\}$ and $\mathcal{S}_2 \supseteq \{21, 22\}$. For each of the 16 possible choices of $\mathcal{S}_1$ and $\mathcal{S}_2$, the resulting rate region is (directly or indirectly) included in $\mathscr{R}_{\mathrm{HK}}$ as follows (see Fig. 7).

- If $\mathcal{S}_1 = \{11, 12, 21\}$ and $\mathcal{S}_2 = \{21, 22, 12\}$, we obtain precisely $\mathscr{R}_{\mathrm{HK}}$ (depicted as a dashed box in the figure).

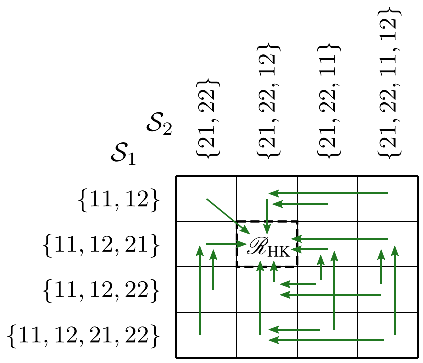


Fig. 7. Different cases of $\mathcal{S}_1$ and $\mathcal{S}_2$ for the region $\mathscr{R}_{\text{opt}}$ and the inclusion of the corresponding regions in $\mathscr{R}_{\text{HK}}$. An arrow from A to B means that the region achieved by case A is included in the region achieved by case B.

- If $\mathcal{S}_1 = \{11, 12, 21, 22\}$, both receivers decode for the messages with indices $\{21, 22\}$. This is equivalent to letting $U'_{21} = (U_{21}, U_{22})$, $U'_{22} = \emptyset$, and $\mathcal{S}'_1 = \{11, 12, 21\}$. A symmetric argument holds if $\mathcal{S}_2 = \{21, 22, 11, 12\}$.
- If $\mathcal{S}_1 = \{11, 12, 22\}$, then $\mathcal{S}_1$ can be replaced by $\{11, 12, 21\}$ by exchanging the roles of U_{21} and U_{22}. The exchange will not affect receiver 2, since the two auxiliary random variables play symmetric roles there. A symmetric argument holds if $\mathcal{S}_2 = \{21, 22, 11\}$.
- If $\mathcal{S}_1 = \{11, 12\}$ and $\mathcal{S}_2 = \{21, 22\}$, we apply Fourier–Motzkin elimination and arrive at
$$\begin{aligned} R_1 &\le I(X_1; Y_1 \mid Q), \\ R_2 &\le I(X_2; Y_2 \mid Q). \end{aligned}$$
This region is a subset of the one in (19a–19g) when the latter is specialized to $U_{12} = U_{21} = \emptyset$, $U_{11} = X_1$, and $U_{22} = X_2$.
- If $\mathcal{S}_1 = \{11, 12\}$ and $\mathcal{S}_2 = \{21, 22, 12\}$, Fourier–Motzkin elimination leads to
$$\begin{aligned} R_1 &\le I(X_1; Y_1 \mid Q), \\ R_1 &\le I(X_1; Y_1 \mid U_{12}, Q) + I(U_{12}; Y_2 \mid X_2, Q), \\ R_2 &\le I(X_2; Y_2 \mid U_{12}, Q), \\ R_1 + R_2 &\le I(X_1; Y_1 \mid U_{12}, Q) + I(U_{12}, X_2; Y_2 \mid Q). \end{aligned}$$
Again, this region is a subset of the one in (19a–19g), namely when the latter is specialized to $U_{21} = \emptyset$ and $U_{22} = X_2$. A symmetric argument holds if $\mathcal{S}_1 = \{11, 12, 21\}$ and $\mathcal{S}_2 = \{21, 22\}$.

This concludes the proof of Corollary 3. □

Remark 7: Chong, Motani, Garg, and El Gamal [6] proposed an alternative coding scheme with a different random codebook structure and showed that this scheme achieves $\mathscr{R}_{\text{HK}}$ in (16). More specifically, we fix a pmf $p = p(q)\,p(u_1, x_1|q)\,p(u_2, x_2|q)$. Randomly generate a coded time sharing sequence $q^n \sim \prod_{i=1}^n p_Q(q_i)$. For each $k \in \{1, 2\}$ and $m_{k1} \in [1 : 2^{nR_{k1}}]$, randomly and conditionally independently generate a sequence $u_k^n(m_{k1})$ according to $\prod_{i=1}^n p_{U_k|Q}(u_{ki}|q_i)$. For each $k \in \{1, 2\}$, $m_{k1} \in [1 : 2^{nR_{k1}}]$, and $m_{k2} \in [1 : 2^{nR_{k2}}]$, randomly and conditionally independently generate a sequence $x_k^n(m_{k1}, m_{k2})$ according to $\prod_{i=1}^n p_{X_k|U_k,Q}(x_{ki}|u_{ki}(m_{k1}, q_i))$. To communicate message pair (m_{11}, m_{12}), sender $k = 1, 2$ transmits $x_{ki}(m_{k1}, m_{k2})$ for $i \in [1 : n]$. A question arises whether using optimal decoders (and employing tighter performance analysis) would enlarge the achievable rate region of the coding scheme by Chong et al. The answer is negative, which can be shown by adapting the analysis technique in the proofs of Theorems 2 and 3 (see [18]).

V. Concluding Remarks

Taking a modular approach to the problem of finding the capacity region of the interference network, we have studied the performance of random code ensembles. This result provides a simple characterization of the rate region achievable by the optimal maximum likelihood decoding rule and invites more refined studies on the performance of random coding for interference networks, such as the error exponent analysis (see [10], [14]) and Verdú's finite-block performance bounds [17].

The optimal rate region can be achieved by simultaneous nonunique decoding, which fully incorporates the codebook structure of interfering signals. Although its performance can be achieved also by an appropriate combination of simultaneous decoding (SD) of strong interference and treating weak interference as noise (IAN) [2], [4], [12], simultaneous nonunique decoding provides a conceptual unification of SD and IAN, recovering all possible combinations of the two schemes at each receiver.

Finally, we remark that simultaneous nonunique decoding can be applied to encoding schemes beyond what is considered in this paper. For example, when combined with the restricted version of Marton's coding scheme [11] for the two-receiver broadcast channel $p(y_1, y_2|x)$ without the center codeword U_0 [7, Th. 8.3] under the random code ensemble $p = (p(u_1)p(u_2), x(u_1, u_2))$, simultaneous nonunique decoding can achieve any rate pair (R_1, R_2) such that
$$\begin{aligned} R_1 &< \tilde{R}_1, \\ R_2 &< \tilde{R}_2, \\ R_1 + R_2 &< \tilde{R}_1 + \tilde{R}_2 - I(U_1; U_2) \end{aligned}$$
for some $(\tilde{R}_1, \tilde{R}_2) \in \tilde{\mathscr{R}}_1 \cap \tilde{\mathscr{R}}_2$, where $\tilde{\mathscr{R}}_1$ consists of rate pairs such that
$$\tilde{R}_1 < I(U_1; Y_1)$$
or
$$\begin{aligned} \tilde{R}_1 &< I(U_1; Y_1, U_2), \\ \tilde{R}_1 + \tilde{R}_2 &< I(U_1, U_2; Y_1), \end{aligned}$$
and $\tilde{\mathscr{R}}_2$ is defined similarly by exchanging the subscripts 1 and 2. The optimality of this region, in any reasonable sense, remains open.

Appendix A
Proof of Lemma 1

Clearly, the right hand side of the equality is an upper bound to the left hand side, since
$$H(Y_1^n \mid X_1^n, \mathcal{C}_n) \le nH(Y_1 \mid X_1, Q),$$

and

$$\begin{aligned} H(Y_1^n \mid X_1^n, \mathcal{C}_n) &\le H(Y_1^n, M_2 \mid X_1^n, \mathcal{C}_n) \\ &= nR_2 + H(Y_1^n \mid X_1^n, X_2^n, \mathcal{C}_n) \\ &\le nR_2 + nH(Y_1 \mid X_1, X_2, Q), \end{aligned}$$

where we have used the codebook structure and the fact that the channel is memoryless.

To see that the right hand side is also a valid lower bound, note that

$$\begin{aligned} &H(Y_1^n \mid X_1^n, \mathcal{C}_n) \\ &= \underbrace{H(Y_1^n \mid X_1^n, \mathcal{C}_n, M_2)}_{\substack{=nH(Y_1|X_1,X_2) \\ =nH(Y_1|X_1,X_2,Q)}} + \underbrace{H(M_2)}_{=nR_2} - H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n). \end{aligned}$$

Next, we find an upper bound on $H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n)$ by showing that given X_1^n, $\mathcal{C}_n$, and Y_1^n, a relatively short list $\mathcal{L} \subseteq [1:2^{nR_2}]$ can be constructed that contains M_2 with high probability (the idea is similar to [7, proof of Lemma 22.1]). Without loss of generality, assume $M_2 = 1$. Fix an $\varepsilon > 0$ and define the random set

$$\mathcal{L} = \{m_2 : (Q^n, X_1^n, X_2^n(m_2), Y_1^n) \in \mathcal{T}_\varepsilon^{(n)}\}.$$

To analyze the cardinality $|\mathcal{L}|$, note that, for each $m_2 \neq 1$,

$$\begin{aligned} &\mathsf{P}\{(Q^n, X_1^n, X_2^n(m_2), Y_1^n) \in \mathcal{T}_\varepsilon^{(n)}\} \\ &= \sum_{q^n, x_1^n, x_2^n} \mathsf{P}\{Q^n = q^n, X_1^n = x_1^n, X_2^n(m_2) = x_2^n\} \\ &\quad \cdot \mathsf{P}\{(x_1^n, x_2^n, Y_1^n) \in \mathcal{T}_\varepsilon^{(n)}\} \\ &\overset{(a)}{\le} \sum_{q^n, x_1^n, x_2^n} \mathsf{P}\{Q^n = q^n, X_1^n = x_1^n, X_2^n(m_2) = x_2^n\} \\ &\quad \cdot 2^{-n(I(X_2;Y_1|X_1,Q)-\delta(\varepsilon))} \\ &= 2^{-n(I(X_2;Y_1|X_1,Q)-\delta(\varepsilon))}, \end{aligned}$$

where (a) follows by the joint typicality lemma. Thus, the cardinality $|\mathcal{L}|$ satisfies $|\mathcal{L}| \le 1 + B$, where B is a binomial random variable with $2^{nR_2} - 1$ trials and success probability at most $2^{-n(I(X_2;Y_1|X_1,Q)-\delta(\varepsilon))}$. The expected cardinality is therefore bounded as

$$\mathsf{E}(|\mathcal{L}|) \le 1 + 2^{n(R_2 - I(X_2;Y_1|X_1,Q)+\delta(\varepsilon))}. \tag{20}$$

Note that the true M_2 is contained in the list with high probability, i.e., $1 \in \mathcal{L}$, by the weak law of large numbers,

$$\mathsf{P}\{(Q^n, X_1^n, X_2^n(1), Y_1^n) \in \mathcal{T}_\varepsilon^{(n)}\} \to 1 \quad \text{as } n \to \infty.$$

Define the indicator random variable $E = \mathbb{I}(1 \in \mathcal{L})$, which therefore satisfies $\mathsf{P}\{E = 0\} \to 0$ as $n \to \infty$. Hence

$$\begin{aligned} &H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n) \\ &= H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n, E) + I(M_2; E \mid X_1^n, \mathcal{C}_n, Y_1^n) \\ &\le H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n, E) + 1 \\ &= 1 + \mathsf{P}\{E = 0\} \cdot H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n, E = 0) \\ &\quad + \mathsf{P}\{E = 1\} \cdot H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n, E = 1) \\ &\le 1 + nR_2\,\mathsf{P}\{E = 0\} + H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n, E = 1). \end{aligned}$$

For the last term, we argue that if M_2 is included in $\mathcal{L}$, then its conditional entropy cannot exceed $\log(|\mathcal{L}|)$:

$$\begin{aligned} &H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n, E = 1) \\ &\overset{(a)}{=} H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n, E = 1, \mathcal{L}, |\mathcal{L}|) \\ &\le H(M_2 \mid E = 1, \mathcal{L}, |\mathcal{L}|) \\ &= \sum_{l=0}^{2^{nR_2}} \mathsf{P}\{|\mathcal{L}| = l\} \cdot H(M_2 \mid E = 1, \mathcal{L}, |\mathcal{L}| = l) \\ &\le \sum_{l=0}^{2^{nR_2}} \mathsf{P}\{|\mathcal{L}| = l\} \cdot \log(l) \\ &= \mathsf{E}(\log(|\mathcal{L}|)) \\ &\overset{(b)}{\le} \log(\mathsf{E}(|\mathcal{L}|)) \\ &\overset{(c)}{\le} 1 + \max\{0, n(R_2 - I(X_2; Y_1|X_1, Q) + \delta(\varepsilon))\}, \end{aligned}$$

where (a) follows since the list $\mathcal{L}$ and its cardinality $|\mathcal{L}|$ are functions only of X_1^n, $\mathcal{C}_n$, and Y_1^n, (b) follows by Jensen's inequality, and (c) follows from (20) and the soft-max interpretation of the log-sum-exp function [5, p. 72].

Substituting back, we have

$$\begin{aligned} &H(M_2 \mid X_1^n, \mathcal{C}_n, Y_1^n) \\ &\quad \le 2 + nR_2\,\mathsf{P}\{E = 0\} \\ &\qquad + \max\{0, n(R_2 - I(X_2; Y_1|X_1, Q) + \delta(\varepsilon))\}, \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{n} H(Y_1^n \mid X_1^n, \mathcal{C}_n) \\ &\quad \ge H(Y_1|X_1, X_2, Q) + R_2 - \frac{2}{n} - R_2\,\mathsf{P}\{E = 0\} \\ &\qquad - \max\{0, R_2 - I(X_2; Y_1|X_1, Q) + \delta(\varepsilon)\} \\ &\quad \ge H(Y_1|X_1, X_2, Q) + \min\big\{R_2,\ I(X_2; Y_1|X_1, Q) - \delta(\varepsilon)\big\} \\ &\qquad - \frac{2}{n} - R_2\,\mathsf{P}\{E = 0\}. \end{aligned}$$

Taking the limit as $n \to \infty$, and noting that we are free to choose ε such that $\delta(\varepsilon)$ becomes arbitrarily small, the desired result follows.

Appendix B
Equivalence Between the Min and MAC Forms

Fix a distribution $p = p(q)\,p(x_1|q) \cdots p(x_K|q)$ and a rate tuple $(R_1, \ldots, R_K)$. We show that the conditions (13) and (14) are equivalent.

Proof That (13) Implies (14): We are given a set $\mathcal{S}$ with $\mathcal{D}_1 \subseteq \mathcal{S} \subseteq [1:K]$. Fix an arbitrary $\mathcal{V}$ with nonempty intersection $\mathcal{V} \cap \mathcal{D}_1$. Now consider $\mathcal{V}' = \mathcal{T} = \mathcal{S} \cap \mathcal{V}$. Note $\mathcal{V}' \cap \mathcal{D}_1 = \mathcal{V} \cap \mathcal{D}_1$ as required. Then,

$$\begin{aligned} R_{\mathcal{V}'} = R_{\mathcal{T}} &\overset{(a)}{\le} I(X_{\mathcal{T}}; Y_1 \mid X_{\mathcal{S}\setminus\mathcal{T}}, Q) \\ &\overset{(b)}{\le} I(X_{\mathcal{T}}; Y_1 \mid X_{\mathcal{S}\setminus\mathcal{V}}, X_{[1:K]\setminus\mathcal{S}\setminus\mathcal{V}}, Q) \\ &= I(X_{\mathcal{V}'}; Y_1 \mid X_{[1:K]\setminus\mathcal{V}}, Q), \end{aligned}$$

where (a) follows from (13), and (b) follows from the structure of p. □

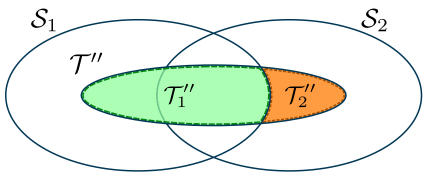


Fig. 8. Partitioning the set $\mathcal{T}'' \subseteq \mathcal{S}_1 \cup \mathcal{S}_2$.

Proof That (14) Implies (13): Denote a set $\mathcal{S} \subseteq [1:K]$ as *decodable* if

$$\forall \mathcal{T} \subseteq \mathcal{S} : R_{\mathcal{T}} \leq I(X_{\mathcal{T}}; Y_1 \,|\, X_{\mathcal{S}\setminus\mathcal{T}}, Q).$$

Then the following proposition holds, which is proved below.

Proposition 1: If $\mathcal{S}_1$ and $\mathcal{S}_2$ are decodable sets, then $\mathcal{S}_1 \cup \mathcal{S}_2$ is a decodable set.

To determine which messages are decodable, consider the optimization problem of maximizing $|\mathcal{S}|$ over decodable sets $\mathcal{S}$. From Proposition 1, a unique maximizer $\mathcal{S}^\star$ must exist, which is a superset of all decodable sets. Consider its complement $\overline{\mathcal{S}}^\star$. The intuitive reason for the messages indexed by $\overline{\mathcal{S}}^\star$ being *un*decodable is that the corresponding rates are too large. This notion is made precise in the following proposition, which is analogous to a property for the Gaussian case given in [2, Fact 1] and for which a proof is provided below.

Proposition 2: For all sets $\mathcal{U}$ with $\emptyset \subset \mathcal{U} \subseteq \overline{\mathcal{S}}^\star$, the rates satisfy

$$R_{\mathcal{U}} > I(X_{\mathcal{U}}; Y_1 \,|\, X_{\mathcal{S}^\star}, Q). \tag{21}$$

Assuming (13) is not true, there must be some desired message index that is not decodable, i.e., $\mathcal{D}_1 \nsubseteq \mathcal{S}^\star$, or equivalently, $\overline{\mathcal{S}}^\star \cap \mathcal{D}_1 \neq \emptyset$. Then we can choose $\mathcal{V} = \overline{\mathcal{S}}^\star$ in (14), yielding

$$\exists \mathcal{V}' \subseteq \overline{\mathcal{S}}^\star, \mathcal{V}' \cap \mathcal{D}_1 = \overline{\mathcal{S}}^\star \cap \mathcal{D}_1 : R_{\mathcal{V}'} \leq I(X_{\mathcal{V}'}; Y_1 \,|\, X_{\mathcal{S}^\star}, Q),$$

which contradicts (21). This proves that (14) implies (13). □

Proof of Proposition 1: Since $\mathcal{S}_1$ and $\mathcal{S}_2$ are decodable, we have

$$\begin{aligned} R_{\mathcal{T}} &\leq I(X_{\mathcal{T}}; Y_1 \,|\, X_{\mathcal{S}_1\setminus\mathcal{T}}, Q) \quad \text{for all } \mathcal{T} \subseteq \mathcal{S}_1, \\ R_{\mathcal{T}'} &\leq I(X_{\mathcal{T}'}; Y_1 \,|\, X_{\mathcal{S}_2\setminus\mathcal{T}'}, Q) \quad \text{for all } \mathcal{T}' \subseteq \mathcal{S}_2. \end{aligned}$$

and we need to show

$$R_{\mathcal{T}''} \leq I(X_{\mathcal{T}''}; Y_1 \,|\, X_{(\mathcal{S}_1\cup\mathcal{S}_2)\setminus\mathcal{T}''}, Q) \quad \text{for all } \mathcal{T}'' \subseteq \mathcal{S}_1 \cup \mathcal{S}_2.$$

Fix a subset $\mathcal{T}'' \subseteq \mathcal{S}_1 \cup \mathcal{S}_2$ and partition it as $\mathcal{T}'' = \mathcal{T}_1'' \cup \mathcal{T}_2''$ where $\mathcal{T}_1'' \subseteq \mathcal{S}_1$, $\mathcal{T}_2'' \subseteq \mathcal{S}_2$, $\mathcal{T}_1'' \cap \mathcal{T}_2'' = \emptyset$, and $\mathcal{T}_2'' \cap \mathcal{S}_1 = \emptyset$ (see Fig. 8).

Then

$$\begin{aligned} R_{\mathcal{T}''} &= R_{\mathcal{T}_1''} + R_{\mathcal{T}_2''} \\ &\leq I(X_{\mathcal{T}_1''}; Y_1 \,|\, X_{\mathcal{S}_1\setminus\mathcal{T}_1''}, Q) + I(X_{\mathcal{T}_2''}; Y_1 \,|\, X_{\mathcal{S}_2\setminus\mathcal{T}_2''}, Q) \\ &\leq I(X_{\mathcal{T}_1''}; Y_1 \,|\, X_{(\mathcal{S}_1\cup\mathcal{S}_2)\setminus\mathcal{T}''}, Q) \\ &\quad + I(X_{\mathcal{T}_2''}; Y_1 \,|\, X_{(\mathcal{S}_1\cup\mathcal{S}_2)\setminus\mathcal{T}''}, X_{\mathcal{T}_1''}, Q) \\ &= I(X_{\mathcal{T}_1''}, X_{\mathcal{T}_2''}; Y_1 \,|\, X_{(\mathcal{S}_1\cup\mathcal{S}_2)\setminus\mathcal{T}''}, Q), \end{aligned}$$

which concludes the proof. □

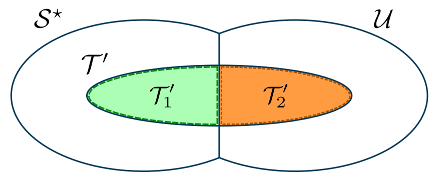


Fig. 9. Partitioning the set $\mathcal{T}' \subseteq \mathcal{S}^\star \cup \mathcal{U}$.

Proof of Proposition 2: Assume first that the proposition was not true. Then there must be a minimal $\mathcal{U}$ with $\emptyset \subset \mathcal{U} \subseteq \overline{\mathcal{S}}^\star$ such that

$$\begin{aligned} R_{\mathcal{U}} &\leq I(X_{\mathcal{U}}; Y_1 \,|\, X_{\mathcal{S}^\star}, Q), \\ R_{\mathcal{U}\setminus\mathcal{T}} &> I(X_{\mathcal{U}\setminus\mathcal{T}}; Y_1 \,|\, X_{\mathcal{S}^\star}, Q) \quad \text{for all } \mathcal{T} \text{ with } \emptyset \subset \mathcal{T} \subset \mathcal{U}. \end{aligned} \tag{22}$$

Now,

$$\begin{aligned} R_{\mathcal{T}} &= R_{\mathcal{U}} - R_{\mathcal{U}\setminus\mathcal{T}} \\ &\leq I(X_{\mathcal{U}}; Y_1 \,|\, X_{\mathcal{S}^\star}, Q) - I(X_{\mathcal{U}\setminus\mathcal{T}}; Y_1 \,|\, X_{\mathcal{S}^\star}, Q) \\ &= I(X_{\mathcal{T}}; Y_1 \,|\, X_{\mathcal{S}^\star}, X_{\mathcal{U}\setminus\mathcal{T}}, Q) \\ &\quad \text{for all } \mathcal{T} \text{ satisfying } \emptyset \subset \mathcal{T} \subset \mathcal{U}. \end{aligned}$$

Recalling (22), the last statement continues to hold for $\mathcal{T} = \mathcal{U}$. Thus,

$$R_{\mathcal{T}} \leq I(X_{\mathcal{T}}; Y_1 \,|\, X_{\mathcal{S}^\star}, X_{\mathcal{U}\setminus\mathcal{T}}, Q) \quad \text{for all } \mathcal{T} \subseteq \mathcal{U}. \tag{23}$$

We are going to show that $\mathcal{S}^\star \cup \mathcal{U}$ is decodable, which contradicts the definition of $\mathcal{S}^\star$ as the maximum decodable set since $\mathcal{U}$ is non-empty and does not intersect $\mathcal{S}^\star$. To this end, consider an arbitrary $\mathcal{T}' \subseteq \mathcal{S}^\star \cup \mathcal{U}$ and partition it as $\mathcal{T}' = \mathcal{T}_1' \cup \mathcal{T}_2'$ with $\mathcal{T}_1' \cap \mathcal{T}_2' = \emptyset$, $\mathcal{T}_1' \subseteq \mathcal{S}^\star$, and $\mathcal{T}_2' \subseteq \mathcal{U}$ (see Fig. 9).

Then

$$\begin{aligned} R_{\mathcal{T}'} &= R_{\mathcal{T}_1'} + R_{\mathcal{T}_2'} \\ &\overset{(a)}{\leq} I(X_{\mathcal{T}_1'}; Y_1 \,|\, X_{\mathcal{S}^\star\setminus\mathcal{T}_1'}, Q) + I(X_{\mathcal{T}_2'}; Y_1 \,|\, X_{\mathcal{S}^\star}, X_{\mathcal{U}\setminus\mathcal{T}_2'}, Q) \\ &\overset{(b)}{\leq} I(X_{\mathcal{T}_1'}; Y_1 \,|\, X_{\mathcal{S}^\star\setminus\mathcal{T}_1'}, X_{\mathcal{U}\setminus\mathcal{T}_2'}, Q) \\ &\quad + I(X_{\mathcal{T}_2'}; Y_1 \,|\, X_{\mathcal{S}^\star\setminus\mathcal{T}_1'}, X_{\mathcal{U}\setminus\mathcal{T}_2'}, X_{\mathcal{T}_1'}, Q) \\ &= I(X_{\mathcal{T}_1'}, X_{\mathcal{T}_2'}; Y_1 \,|\, X_{(\mathcal{S}^\star\cup\mathcal{U})\setminus(\mathcal{T}_1'\cup\mathcal{T}_2')}, Q), \end{aligned}$$

where (a) follows from $\mathcal{S}^\star$ being decodable and (23), and in (b), we have augmented the first mutual information expression and rewritten the second one. This concludes the proof by contradiction. □

APPENDIX C
PROOF OF LEMMA 2

The proof proceeds along similar lines as the proof of Lemma 1. First, we show that the right hand side is a valid upper bound to the left hand side. For any $\mathcal{U} \subseteq \mathcal{S}^{\mathrm{c}}$,

$$\begin{aligned} H(Y_1^n \,|\, X_{\mathcal{S}}^n, \mathcal{C}_n) &\leq H(Y_1^n, M_{\mathcal{U}} \,|\, X_{\mathcal{S}}^n, \mathcal{C}_n) \\ &= nR_{\mathcal{U}} + H(Y_1^n \,|\, X_{\mathcal{S}}^n, X_{\mathcal{U}}^n, \mathcal{C}_n) \\ &\leq nR_{\mathcal{U}} + nH(Y_1 \,|\, X_{\mathcal{S}}, X_{\mathcal{U}}, Q) \\ &= nR_{\mathcal{U}} + nH(Y_1 \,|\, X_{[1:K]}, Q) \\ &\quad + I(X_{(\mathcal{S}\cup\mathcal{U})^{\mathrm{c}}}; Y_1 \,|\, X_{\mathcal{S}\cup\mathcal{U}}, Q), \end{aligned}$$

where we have used the codebook structure.

To see that the right hand side is a valid lower bound to the left hand side, note

$$H(Y_1^n \mid X_{\mathcal{S}}^n, \mathcal{C}_n) = nH(Y_1 | X_{[1:K]}, Q) + nR_{\mathcal{S}^c} - H(M_{\mathcal{S}^c} \mid X_{\mathcal{S}}^n, Y_1^n, \mathcal{C}_n).$$

Without loss of generality, assume $M_k = 1$, for $k \in \mathcal{S}^c$. Fix an $\varepsilon > 0$ and define the random set

$$\mathcal{L} = \{m_{\mathcal{S}^c} : (Q^n, X_i^n|_{i\in\mathcal{D}_1}, X_i^n(m_i)|_{i\in\mathcal{D}_1^c}, Y_1^n) \in \mathcal{T}_\varepsilon^{(n)} \text{ with } m_k = 1 \text{ for all } k \in \mathcal{D}_1^c \cap \mathcal{S}\}.$$

To analyze the cardinality $|\mathcal{L}|$, fix a $m_{\mathcal{S}^c}$ and consider the probability of $m_{\mathcal{S}^c} \in \mathcal{L}$. If $m_k \neq 1$ for all $k \in \mathcal{S}^c$, and $m_k = 1$ otherwise, then the joint typicality lemma implies

$$\mathsf{P}\{(Q^n, X_i^n|_{i\in\mathcal{D}_1}, X_i^n(m_i)|_{i\in\mathcal{D}_1^c}, Y_1^n) \in \mathcal{T}_\varepsilon^{(n)}\} \le 2^{-n(I(X_{\mathcal{S}^c}; Y_1 | X_{\mathcal{S}}, Q) - \delta(\varepsilon))},$$

and there are at most $2^{nR_{\mathcal{S}^c}}$ such $m_{\mathcal{S}^c}$. More generally, fix a subset $\mathcal{U} \subseteq \mathcal{S}^c$. If $m_k \neq 1$ for $k \in \mathcal{S}^c \setminus \mathcal{U}$, and $m_k = 1$ otherwise, then

$$\mathsf{P}\{(Q^n, X_i^n|_{i\in\mathcal{D}_1}, X_i^n(m_i)|_{i\in\mathcal{D}_1^c}, Y_1^n) \in \mathcal{T}_\varepsilon^{(n)}\} \le 2^{-n(I(X_{\mathcal{S}^c\setminus\mathcal{U}}; Y_1 | X_{\mathcal{S}}, X_{\mathcal{U}}, Q) - \delta(\varepsilon))},$$

and there are at most $2^{nR_{\mathcal{S}^c\setminus\mathcal{U}}}$ such $m_{\mathcal{S}^c}$. Thus,

$$\mathsf{E}(|\mathcal{L}|) \le \sum_{\mathcal{U}\subseteq\mathcal{S}^c} 2^{n(R_{\mathcal{S}^c\setminus\mathcal{U}} - I(X_{\mathcal{S}^c\setminus\mathcal{U}}; Y_1 | X_{\mathcal{S}}, X_{\mathcal{U}}, Q) + \delta(\varepsilon))}. \tag{24}$$

Define the indicator random variable $E = \mathbb{I}((1, 1, \ldots, 1) \in \mathcal{L})$, which satisfies $\mathsf{P}\{E = 0\} \to 0$ as $n \to \infty$ by the weak law of large numbers. Now

$$H(M_{\mathcal{S}^c} \mid X_{\mathcal{S}}^n, Y_1^n, \mathcal{C}_n) \le 1 + nR_{\mathcal{S}^c}\,\mathsf{P}\{E = 0\} + H(M_{\mathcal{S}^c} \mid X_{\mathcal{S}}^n, Y_1^n, \mathcal{C}_n, E = 1).$$

For the last term, we argue

$$\begin{aligned} &H(M_{\mathcal{S}^c} \mid X_{\mathcal{S}}^n, Y_1^n, \mathcal{C}_n, E = 1) \\ &\le \log(\mathsf{E}(|\mathcal{L}|)) \\ &\overset{(24)}{\le} \log\left(\sum_{\mathcal{U}\subseteq\mathcal{S}^c} 2^{n(R_{\mathcal{S}^c\setminus\mathcal{U}} - I(X_{\mathcal{S}^c\setminus\mathcal{U}}; Y_1 | X_{\mathcal{S}}, X_{\mathcal{U}}, Q) + \delta(\varepsilon))}\right) \\ &\le \max_{\mathcal{U}\subseteq\mathcal{S}^c} \big(n(R_{\mathcal{S}^c\setminus\mathcal{U}} - I(X_{\mathcal{S}^c\setminus\mathcal{U}}; Y_1 | X_{\mathcal{S}}, X_{\mathcal{U}}, Q) + \delta(\varepsilon))\big) + |\mathcal{S}^c|. \end{aligned}$$

Substituting back,

$$\begin{aligned} &H(M_{\mathcal{S}^c} \mid X_{\mathcal{S}}^n, Y_1^n, \mathcal{C}_n) \\ &\le 1 + |\mathcal{S}^c| + nR_{\mathcal{S}^c}\,\mathsf{P}\{E = 0\} \\ &\quad + \max_{\mathcal{U}\subseteq\mathcal{S}^c} \big(n(R_{\mathcal{S}^c\setminus\mathcal{U}} - I(X_{\mathcal{S}^c\setminus\mathcal{U}}; Y_1 | X_{\mathcal{S}}, X_{\mathcal{U}}, Q) + \delta(\varepsilon))\big), \end{aligned}$$

and finally,

$$\begin{aligned} &\frac{1}{n} H(Y_1^n \mid X_{\mathcal{S}}^n, \mathcal{C}_n) \\ &\ge H(Y_1 | X_{[1:K]}, Q) + R_{\mathcal{S}^c} - \frac{1 + |\mathcal{S}^c|}{n} - R_{\mathcal{S}^c}\,\mathsf{P}\{E = 0\} \\ &\quad - \max_{\mathcal{U}\subseteq\mathcal{S}^c} \big(R_{\mathcal{S}^c\setminus\mathcal{U}} - I(X_{\mathcal{S}^c\setminus\mathcal{U}}; Y_1 | X_{\mathcal{S}}, X_{\mathcal{U}}, Q) + \delta(\varepsilon)\big) \\ &= H(Y_1 | X_{[1:K]}, Q) - \frac{1 + |\mathcal{S}^c|}{n} - R_{\mathcal{S}^c}\,\mathsf{P}\{E = 0\} \\ &\quad + \min_{\mathcal{U}\subseteq\mathcal{S}^c} \big(R_{\mathcal{U}} + I(X_{\mathcal{S}^c\setminus\mathcal{U}}; Y_1 | X_{\mathcal{S}}, X_{\mathcal{U}}, Q) + \delta(\varepsilon)\big). \end{aligned}$$

Taking the limits $n \to \infty$ and $\varepsilon \to 0$ concludes the proof.

ACKNOWLEDGMENTS

The authors are grateful to Gerhard Kramer for an interesting conversation on optimal decoding rules that spurred interest in this research direction. They also would like to thank Jungwon Lee for enlightening discussions on minimum distance decoding for interference channels, which shaped the main ideas behind this paper.

REFERENCES

[1] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[2] F. Baccelli, A. El Gamal, and D. N. C. Tse, "Interference networks with point-to-point codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2582–2596, May 2011.

[3] B. Bandemer and A. El Gamal, "Interference decoding for deterministic channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2966–2975, May 2011.

[4] S. S. Bidokhti, V. M. Prabhakaran, and S. N. Diggavi, "Is non-unique decoding necessary?" in *Proc. IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, pp. 398–402, Jul. 2012.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[6] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han–Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, Jul. 2008.

[7] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[8] A. El Gamal and M. Costa, "The capacity region of a class of deterministic interference channels (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 343–346, Mar. 1982.

[9] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[10] E. A. Haroutunian, "Lower bound for the error probability of multiple-access channels," *Problemy Peredachi Informatsii*, vol. 11, no. 2, pp. 23–36, Jun. 1975.

[11] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[12] A. S. Motahari and A. K. Khandani, "To decode the interference or to consider it as noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1274–1283, Mar. 2011.

[13] C. Nair and A. El Gamal, "The capacity region of a class of three-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4479–4493, Oct. 2009.

[14] A. Nazari, A. Anastasopoulos, and S. S. Pradhan, "Error exponent for multiple-access channels: Lower bounds," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5095–5115, Sep. 2014.

[15] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[16] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[17] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, pp. 1–8, Oct. 2012.

[18] L. Wang, E. Şaşoğlu, B. Bandemer, and Y.-H. Kim, "A comparison of superposition coding schemes," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, pp. 2970–2974, Jul. 2013.

Bernd Bandemer (S'06–M'12) received the Dipl.-Ing. degree in Electrical and Computer Engineering in 2006 from Ilmenau University of Technology, Ilmenau, Germany and the Ph.D. degree in Electrical Engineering from Stanford University, Stanford, CA, in 2012. From 2012 to 2013, he was a postdoctoral researcher at the Information Theory and Applications Center at the University of California in San Diego. From 2013 to 2014, he was a research engineer at Bosch Research and Technology Center in Palo Alto, CA. In November 2014, he joined Rasa Networks in San Jose, CA, as Principal Data Scientist. His research interests include network information theory, wireless communications, machine learning and data mining.

Abbas El Gamal (S'71–M'73–SM'83–F'00) is the Hitachi America Professor in the School of Engineering and the Chair of the Department of Electrical Engineering at Stanford University. He received his B.Sc. Honors degree from Cairo University in 1972, and his M.S. in Statistics and Ph.D. in Electrical Engineering both from Stanford University in 1977 and 1978, respectively. From 1978 to 1980, he was an Assistant Professor of Electrical Engineering at USC. From 2004 to 2009, he was the Director of the Information Systems Laboratory at Stanford University. His research contributions have been in network information theory, FPGAs, and digital imaging devices and systems. He has authored or coauthored over 230 papers and holds 35 patents in these areas. He is coauthor of the book Network Information Theory (Cambridge Press 2011). He is a member of the US National Academy of Engineering and a Fellow of the IEEE. He received several honors and awards for his research contributions, including the the 2014 Viterbi Lecture, the 2013 Shannon Memorial Lecture, the 2012 Claude E. Shannon Award, the inaugural Padovani Lecture, and the 2004 INFOCOM Paper Award. He has been serving on the Board of Governors of the Information Theory Society since 2009 and is currently Junior Past President. He has played key roles in several semiconductor, EDA, and biotechnology startup companies.

Young-Han Kim (S'99–M'06–SM'12–F'15) received the B.S. degree with honors in electrical engineering from Seoul National University, Korea, in 1996 and the M.S. degrees in electrical engineering and statistics, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2001, 2006, and 2006, respectively. In July 2006, he joined the University of California, San Diego, where he is currently an Associate Professor of Electrical and Computer Engineering. His research interests are in statistical signal processing and information theory, with applications in communication, control, computation, networking, data compression, and learning. Dr. Kim is a recipient of the 2008 NSF Faculty Early Career Development (CAREER) Award, the 2009 US–Israel Binational Science Foundation Bergmann Memorial Award, the 2012 IEEE Information Theory Paper Award, and the 2015 James L. Massey Research & Teaching Award for Young Scholars. He served as an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and a Distinguished Lecturer for the IEEE Information Theory Society.