Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Simultaneous Nonunique Decoding Is Rate-Optimal

Bernd Bandemer
University of California, San Diego
La Jolla, CA 92093, USA
Email: bandemer@ucsd.edu

Abbas El Gamal
Stanford University
Stanford, CA 94305, USA
Email: abbas@ee.stanford.edu

Young-Han Kim
University of California, San Diego
La Jolla, CA 92093, USA
Email: yhk@ucsd.edu

***Abstract*—It is shown that simultaneous nonunique decoding is rate-optimal for the general K-sender, L-receiver discrete memoryless interference channel when encoding is restricted to randomly generated codebooks, superposition coding, and time sharing. This result implies that the Han–Kobayashi inner bound for the two-user-pair interference channel cannot be improved simply by using a better decoder such as the maximum likelihood decoder. It also generalizes and extends previous results by Baccelli, El Gamal, and Tse on Gaussian interference channels with point-to-point Gaussian random codebooks and shows that the Cover–van der Meulen inner bound with no common auxiliary random variable on the capacity region of the broadcast channel can be improved to include the superposition coding inner bound simply by using simultaneous nonunique decoding. The key to proving the main result 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.**

I. Introduction

Consider the discrete memoryless interference channel (DM-IC) with K sender–receiver (user) pairs $p(y_1, \dots, y_K|x_1, \dots, x_K)$. What is the set of simultaneously achievable rate tuples? What coding scheme achieves this *capacity region*? Answering these questions involves joint optimization of the encoding and decoding functions, which has proved elusive even for the case of $K = 2$.

In this paper, we take a simpler modular approach to these questions. We restrict the encoding functions to the class of randomly generated codebooks with superposition coding and time sharing. This class includes, for example, the random codebook ensemble used in the Han–Kobayashi coding scheme [1]. We investigate the optimal rate region achievable by this class of random code ensembles. Our main result is to show that *simultaneous nonunique decoding* (SND) [2]–[4], in which each receiver attempts to recover the unique codeword from its intended sender along with codewords from interfering senders, achieves the optimal rate region.

This result has several implications.

- It establishes indirectly the rate region achieved by using the given random code ensemble and the optimal decoding rule for each code realization that minimizes the probability of decoding error, namely, maximum likelihood decoding (MLD).
- It shows that treating interference as noise at the decoder performs worse than SND in general.
- It shows that the Han–Kobayashi inner bound [1], [2], [4, Theorem 6.4], which was established using a typicality-based simultaneous decoding rule, cannot be improved by using MLD.
- It generalizes the result for K-user-pair Gaussian interference channels with point-to-point Gaussian random codes in [5] to arbitrary (not necessarily Gaussian) random codes with time sharing and superposition coding.
- It shows that the interference decoding rate region for the three-user-pair deterministic interference channel in [6] is the optimal rate region achievable by point-to-point random codes and time sharing.

We illustrate our result and its implications via the following two simple examples.

A. Interference Channels with Two User Pairs

Consider the two-user-pair discrete memoryless interference channel (2-DM-IC) $p(y_1, y_2|x_1, x_2)$ depicted in Figure 1.

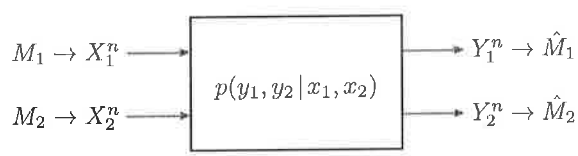


Figure 1. Two-user-pair discrete memoryless interference channel.

Given a product input pmf $p(x_1)\,p(x_2)$, consider a random code ensemble that consists of randomly generated codewords $x_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, and $x_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, each drawn according to $\prod_{i=1}^n p_{X_1}(x_{1i})$ and $\prod_{i=1}^n p_{X_2}(x_{2i})$, respectively. What is the set of achievable rate pairs (R_1, R_2) under this class of randomly generated point-to-point codes? Instead of analyzing the performance of MLD directly, we instead consider the rate region achievable by a *suboptimal* (in the sense of error probability) decoder and show that this rate region is optimal. Note that such an indirect approach has been used to prove coding theorems for discrete memoryless point-to-point channels with randomly generated codes, where joint typicality decoding achieves the same rate as MLD.

First consider the rate regions achievable by the following simple suboptimal decoding rules and their achievable rate regions, described for receiver 1 (cf. [4]).

This research was supported in part by the Korea Communications Commission under the R&D program KCA-2012-11-921-04-001 (ETRI).

- *Treating interference as noise (IAN).* Receiver 1 finds the unique $\hat{m}_1$ such that $(x_1^n(\hat{m}_1), y_1^n)$ is jointly typical. The average probability of decoding error for receiver 1 tends to zero as $n \to \infty$ if

$$R_1 < I(X_1; Y_1). \tag{1}$$

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

- *Simultaneous (unique) 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

$$R_1 < I(X_1; Y_1 \,|\, X_2), \tag{2a}$$
$$R_2 < I(X_2; Y_1 \,|\, X_1), \tag{2b}$$
$$R_1 + R_2 < I(X_1, X_2; Y_1). \tag{2c}$$

The corresponding SD region is depicted in Figure 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 (2) is achievable via SND. Less obviously, any rate pair in the IAN region (1) is also achievable via SND. 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 Figure 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. Combining the decoding requirements for both receivers yields the rate region $\mathscr{R}_1 \cap \mathscr{R}_2$. In the converse proof of Theorem 1 in Section II, we show that the sufficient condition $(R_1, R_2) \in \mathscr{R}_1 \cap \mathscr{R}_2$ is in fact necessary, that is, no decoder (even the maximum likelihood decoder) can achieve a rate pair outside the closure of $\mathscr{R}_1 \cap \mathscr{R}_2$.

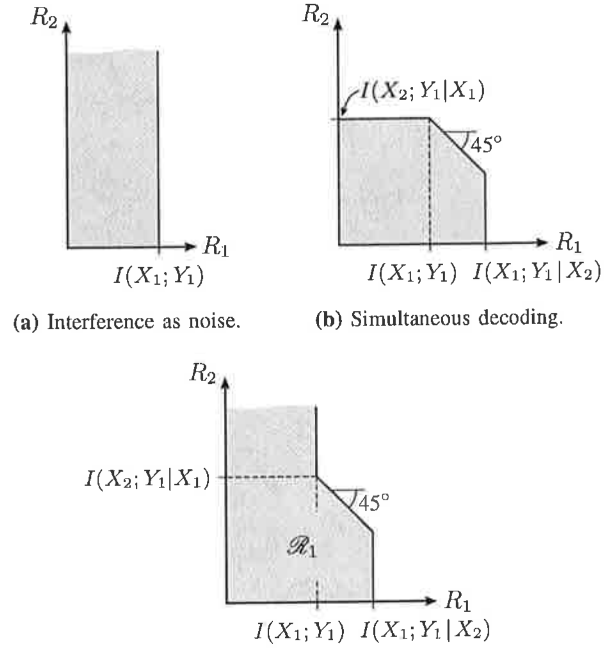
B. Broadcast Channels with Two Receivers

The previous example assumes randomly generated point-to-point codebooks. To illustrate our result for superposition coding, consider a special case of the Cover–van der Meulen coding scheme [7], [8], [4, Equation (8.8)] for the two-receiver discrete memoryless broadcast channel $p(y_1, y_2|x)$ with private messages depicted in Figure 3. In this scheme, we fix a pmf $p(u_1)p(u_2)$ and a function $x(u_1, u_2)$ and randomly generate sequences $u_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$ and $u_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, each according to $\prod_{i=1}^n p_{U_1}(u_{1i})$ and $\prod_{i=1}^n p_{U_2}(u_{2i})$, respectively. To communicate the message pair (m_1, m_2), the sender transmits $x_i = x(u_{1i}(m_1), u_{2i}(m_2))$ for $i \in [1:n]$.

If each receiver decodes for its intended codeword while treating the other codeword as noise, we obtain the rate region consisting of all rate pairs (R_1, R_2) such that

$$R_1 < I(U_1; Y_1), \tag{3a}$$
$$R_2 < I(U_2; Y_2). \tag{3b}$$

However, it follows from Theorem 1 in Section II that if each receiver uses SND instead, we obtain the rate region



(a) Interference as noise. **(b)** Simultaneous decoding.

(c) Simultaneous nonunique decoding.

Figure 2. Achievable rate regions for receiver 1 in the 2-DM-IC: (a) treating interference as noise, (b) using simultaneous decoding, and (c) using simultaneous nonunique decoding ($\mathscr{R}_1$). Note that $\mathscr{R}_1$ is the union of the regions in (a) and (b).

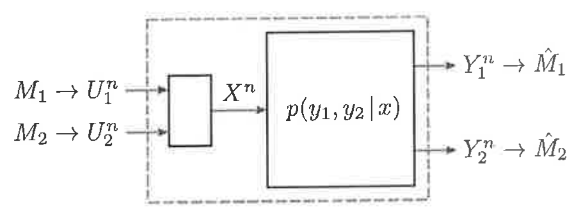


Figure 3. Broadcast channel with Cover–van der Meulen coding.

$\mathscr{R} = \mathscr{R}_1 \cap \mathscr{R}_2$, where $\mathscr{R}_1$ consists of all rate pairs (R_1, R_2) such that

$$R_1 < I(U_1; Y_1),$$

or

$$R_1 < I(U_1; Y_1 \,|\, U_2),$$
$$R_1 + R_2 < I(U_1, U_2; Y_1).$$

The set $\mathscr{R}_2$ is similarly defined by exchanging the subscripts 1 and 2.

The region $\mathscr{R}$ is strictly larger than (3). In particular, it includes as a special case the superposition inner bound [4, Theorem 5.1] consisting of all rate pairs (R_1, R_2) such that

$$R_1 < I(X; Y_1 \,|\, U_2), \tag{4a}$$
$$R_2 < I(U_2; Y_2), \tag{4b}$$
$$R_1 + R_2 < I(X; Y_1). \tag{4c}$$

In contrast, it was shown in [9] that the rate region in (3) does *not* in general include the superposition coding inner bound (4). Hence this non-inclusion is due to the use of a suboptimal decoding rule rather than superposition codebook generation.

The rest of the paper is organized as follows. For simplicity of presentation, in Section II we establish our result for the two-user-pair interference channel with only randomly generated codebooks and time sharing. The key part is the converse proof in which we show that after the desired message and potentially the undesired message have been decoded, the remaining signal is essentially independent and identically distributed (i.i.d.) in time. In Section III, we extend our result to a multiple-sender multiple-receiver discrete memoryless interference channel in which each sender has a single message and wishes to send it to a subset of the receivers. In Section IV, we specialize this extension to the Han–Kobayashi coding scheme for the two-user-pair interference channel. Details of the proof for the results in Sections III and IV are omitted due to space limitation. We use the notation in [4] throughout.

II. DM-IC WITH TWO USER PAIRS

Consider the two-user-pair discrete memoryless interference channel (2-DM-IC) $p(y_1, y_2 \,|\, x_1, x_2)$ with input alphabets $\mathcal{X}_1$, $\mathcal{X}_2$ and output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, as depicted in Figure 1. Define a $(2^{nR_1}, 2^{nR_2}, n)$ code $\mathcal{C}_n$, the probability of decoding error $P_e^{(n)}(\mathcal{C}_n)$ of a code, achievability of a rate pair (R_1, R_2), and the capacity region $\mathscr{C}$ of the 2-DM-IC in the standard way, see [4, Chapter 6].

We now limit our attention to a randomly generated code ensemble with a special structure. Let $p = p(q)\, p(x_1|q)\, p(x_2|q)$ 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 $\{x_1^n(m_1) : m_1 \in [1{:}2^{nR_1}]\}$, $\{x_2^n(m_2) : m_2 \in [1:2^{nR_2}]\}$ 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*.

Definition 1 (Random code 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* 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 codes $\mathcal{C}_n$ 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.

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

$$R_1 \le I(X_1; Y_1 \,|\, Q) \tag{5a}$$

or

$$R_2 \le I(X_2; Y_1 \,|\, X_1, Q), \tag{5b}$$
$$R_1 + R_2 \le I(X_1, X_2; Y_1 \,|\, Q). \tag{5c}$$

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 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 optimal rate region.

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

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

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 $\mathscr{R}_{1,\text{SD}}(p)$ be the set of rate pairs (R_1, R_2) such that

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

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

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

which is referred to as the *MAC form* of $\mathscr{R}_1(p)$, since it is the union of the capacity 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* $\mathscr{R}_{2,\text{IAN}}(p)$ and the *simultaneous-decoding region* $\mathscr{R}_{2,\text{SD}}(p)$. Hence the optimal rate region $\mathscr{R}^*(p)$ can be expressed as

$$\begin{aligned} &(\mathscr{R}_{1,\text{IAN}}(p) \cap \mathscr{R}_{2,\text{IAN}}(p)) \cup (\mathscr{R}_{1,\text{IAN}}(p) \cap \mathscr{R}_{2,\text{SD}}(p)) \\ &\cup (\mathscr{R}_{1,\text{SD}}(p) \cap \mathscr{R}_{2,\text{IAN}}(p)) \cup (\mathscr{R}_{1,\text{SD}}(p) \cap \mathscr{R}_{2,\text{SD}}(p)), \end{aligned} \tag{6}$$

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 (5) can be equivalently characterized as the set of rate pairs (R_1, R_2) such that

$$R_1 \le I(X_1; Y_1 \,|\, X_2, Q), \tag{7a}$$
$$R_1 + \min\{R_2, I(X_2; Y_1 \,|\, X_1, Q)\} \le I(X_1, X_2; Y_1 \,|\, Q). \tag{7b}$$

The $\min\{\cdot\}$ term represents the effective rate of the interfering signal X_2 at the receiver Y_1, which is a monotone increasing function of R_2 and reaches saturation at the maximum possible rate for distinguishing X_2 codewords; see [6]. 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.

Remark 3 (Nonconvexity). The optimal rate region $\mathscr{R}^*(p)$ is not convex in general.

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. This analysis, however, is unnecessarily cumbersome. We instead take an indirect yet conducive approach that is common in information theory. We first 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 tends to zero as $n \to \infty$, then (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 such that

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

If there is no such index or more than one, it declares an error. To analyze the probability of decoding error averaged over codebooks, 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:

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

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

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

Then, by the packing lemma in [4, Section 3.2], $\mathsf{P}(\mathcal{E}_2')$ tends to zero as $n \to \infty$ if

$$R_1 < I(X_1; Y_1 \mid Q) - \delta(\varepsilon). \tag{8}$$

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

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

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

$$R_1 < I(X_1; Y_1 \mid X_2, Q) - \delta(\varepsilon), \tag{9a}$$
$$R_1 + R_2 < I(X_1, X_2; Y_1 \mid Q) - \delta(\varepsilon). \tag{9b}$$

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 (8) or (9) 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. As observed in [10], each rate point in $\mathscr{R}^*(p)$ can alternatively be achieved by using simultaneous *unique* decoding rules at each receiver according to (6).

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. We prove that this implies $(R_1, R_2) \in \mathscr{R}_1(p) \cap \mathscr{R}_2(p)$ as claimed in Theorem 1. 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 codebooks for both transmitters, 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 codebooks $\mathcal{C}_n$, 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 (7). To see that the first inequality is true, note that

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

$$\overset{\text{(c)}}{\leq} \sum_{i=1}^{n} (H(Y_{1i} \,|\, X_{2i}, Q_i) - H(Y_{1i} \,|\, X_{1i}, X_{2i}, Q_i))$$
$$= \sum_{i=1}^{n} I(X_{1i}; Y_{1i} \,|\, X_{2i}, Q_i)$$
$$\overset{\text{(d)}}{=} nI(X_1; Y_1 \,|\, X_2, Q),$$

where (a) follows from (the averaged version of) Fano's inequality in (10), (b) increases the mutual information by introducing an additional term, and (c) follows by omitting some conditioning and by the memoryless property of the channel. In (d), we use the fact that the tuple $(Q_i, X_{1i}, X_{2i}, Y_i)$ is identically distributed 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 (7), we need the following lemma, which is proved in the Appendix.

Lemma 1.

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

The statement of the lemma is twofold. If $R_2 \geq I(X_2; Y_1 \,|\, X_1, Q)$, then the right hand side evaluates to $H(Y_1 \,|\, X_1, Q)$. Thus, given the desired codeword, the remaining received sequence looks like i.i.d. noise. On the other hand, if $R_2 < I(X_2; Y_1 \,|\, X_1, Q)$, then the right hand side evaluates to the sum of the rate R_2 of the interfering message and the entropy rate of the i.i.d. channel noise. In other words, the interfering message M_2 is decodable.

Now, consider

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

Here, (a) follows by Fano's inequality, (b) single-letterizes the first term by making use of the codebook structure, and (c) holds for sufficiently large n by Lemma 1. The last line is equivalent to the second inequality in (7). This completes the proof of the converse.

III. DM-IC with K Senders and L Receivers

We generalize the previous result to the K-sender, L-receiver discrete memoryless interference channel ((K,L)-DM-IC) with input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_K$, output alphabets $\mathcal{Y}_1, \ldots, \mathcal{Y}_L$, and channel pmfs $p(y_1, \ldots, y_L \,|\, x_1, \ldots, x_K)$. In this channel, 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 of senders $\mathcal{D}_l \subseteq [1:K]$. The channel is depicted in Figure 4.

More formally, a $(2^{nR_1}, \ldots, 2^{nR_K}, n)$ code $\mathcal{C}_n$ for the (K,L)-DM-IC 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 $\hat{m}_k^{(l)}$, $k \in \mathcal{D}_l$, or an error message e to each received sequence y_l^n.

We assume that the message tuple $(M_1, \ldots, M_K)$ 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

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

A rate tuple $(R_1, \ldots, R_K)$ is said to be *achievable* for the (K,L)-DM-IC if there exists a sequence of $(2^{nR_1}, \ldots, 2^{nR_K}, n)$ 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-IC is the closure of the set of achievable rate tuples $(R_1, \ldots, R_K)$.

As in Section II, we limit our attention to a randomly generated code ensemble with a special structure. Let $p = p(q)\, p(x_1|q) \ldots p(x_K|q)$ 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 fashion, along with the corresponding optimal decoders, constitutes a $(2^{nR_1}, \ldots, 2^{nR_K}, n)$ code. We refer to the random code ensemble generated in this manner as the $(2^{nR_1}, \ldots, 2^{nR_K}, n; p)$ *random code*.

Definition 2 (Random code optimal rate region). Given a pmf $p = p(q)\, p(x_1|q) \ldots p(x_K|q)$, the *optimal rate region* $\mathscr{R}^*(p)$ *achievable by the p-distributed random code* is the closure of the set of rate tuples $(R_1, \ldots, R_K)$ such that the sequence of the $(2^{nR_1}, \ldots, 2^{nR_K}, n; p)$ random codes $\mathcal{C}_n$ 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.

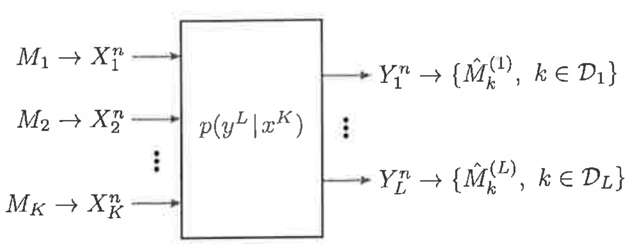


Figure 4. Discrete memoryless (K,L) interference channel.

Note that the setup discussed in Section II as well as the broadcast channel example in the introduction correspond to the special case of $K = L = 2$ and demand sets $\mathcal{D}_1 = \{1\}$ and $\mathcal{D}_2 = \{2\}$.

Define the rate region $\mathscr{R}_1(p)$ for receiver 1 in the MAC form as

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

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 $(R_1, \dots, R_K)$ such that

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

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

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

$$\begin{aligned} R_{\mathcal{D}} + \min_{\mathcal{U}'\subseteq\mathcal{U}} \{R_{\mathcal{U}'} + I(X_{\mathcal{U}\setminus\mathcal{U}'}; Y_1 \,|\, X_{\mathcal{D}}, X_{\mathcal{U}'}, X_{[1:K]\setminus\mathcal{D}\setminus\mathcal{U}}, Q)\} \\ \le I(X_{\mathcal{D}}, X_{\mathcal{U}}; Y_1 \,|\, X_{[1:K]\setminus\mathcal{D}\setminus\mathcal{U}}, Q). \quad (11) \end{aligned}$$

As in the case of 2-DM-IC, each argument of each term in the minimum represents a different mode of signal saturation.

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

Theorem 2. Given a pmf $p = p(q)\,p(x_1|q) \dots p(x_K|q)$, the optimal rate region of the (K, L)-DM-IC $p(y^L|x^K)$ with demand sets $\mathcal{D}_l$ achievable by the p-distributed random code is

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

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

In the following, we provide a sketch of the proof of Theorem 2. 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

$$\big(q^n, x_k^n(\hat{m}_k)\big|_{k\in\mathcal{D}_1}, x_k^n(m_k)\big|_{k\in[1:K]\setminus\mathcal{D}_1}, y_1^n\big) \in \mathcal{T}_\epsilon^{(n)}$$

for some $m_{[1:K]\setminus\mathcal{D}_1}$. The analysis proceeds analogously to Subsection II-A.

To prove the converse, fix a pmf p and let $(R_1, \dots, R_K)$ be a rate tuple that is achievable by the p-distributed random code. We focus only on receiver 1 for which M_k, $k \in \mathcal{D}_1$, are the desired messages and M_k, $k \in \mathcal{D}_1^c$, are undesired messages. We need the following generalization of Lemma 1, the proof of which is omitted.

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

$$\begin{aligned} &\lim_{n\to\infty} \tfrac{1}{n} H(Y_1^n \,|\, X_{\mathcal{S}}^n, \mathcal{C}_n) \\ &\quad = H(Y_1 \,|\, X_{[1:K]}, Q) \\ &\qquad + \min_{\mathcal{U}'\subseteq\mathcal{S}^c} \{R_{\mathcal{U}'} + I(X_{(\mathcal{S}\cup\mathcal{U}')^c}; Y_1 \,|\, X_{\mathcal{S}\cup\mathcal{U}'}, Q)\}. \end{aligned}$$

We now establish (11) as follows. Fix a set of desired message indices $\mathcal{D} \subseteq \mathcal{D}_1$ and a set of undesired message indices $\mathcal{U} \subseteq \mathcal{D}_1^c$. Then

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

where (a) follows from Fano's inequality, in (b), we have increased the mutual information by introducing an additional term, and (c) follows from Lemma 2 with $\mathcal{S} = \mathcal{U}^c$. The last line matches (11), and thus the proof is complete.

IV. Application to the Han–Kobayashi Scheme

Consider the two-user-pair DM-IC in Figure 1. The best known inner bound on the capacity region is achieved by the Han–Kobayashi coding scheme [1]. In this scheme, the message M_1 is split into common and private messages 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 with rates R_{21} and R_{22}, with $R_2 = R_{22} + R_{21}$. Receiver $k \in \{1, 2\}$ uniquely decodes 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.

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 can be limited to 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 $u_{kk'}^n(m_{kk'})$ according to $\prod_{i=1}^n p_{U_{kk'}|Q}(u_{kk'i}|q_i)$. To communicate the message pair (m_{11}, m_{12}), sender 1 transmits $x_{1i} = x_1(u_{11i}, u_{12i})$, and analogously for sender 2. The codebook structure is illustrated in Figure 5.

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

$$\begin{aligned}
R_{11} &\le I(U_{11}; Y_1 \mid U_{12}, U_{21}, Q),\\
R_{12} &\le I(U_{12}; Y_1 \mid U_{11}, U_{21}, Q),\\
R_{21} &\le I(U_{21}; Y_1 \mid U_{11}, U_{12}, Q),\\
R_{11} + R_{12} &\le I(U_{11}, U_{12}; Y_1 \mid U_{21}, Q),\\
R_{11} + R_{21} &\le I(U_{11}, U_{21}; Y_1 \mid U_{12}, Q),\\
R_{12} + R_{21} &\le I(U_{12}, U_{21}; Y_1 \mid U_{11}, Q),\\
R_{11} + R_{12} + R_{21} &\le I(U_{11}, U_{12}, U_{21}; Y_1 \mid Q).
\end{aligned}$$

Similarly, define $\mathscr{R}_{\text{HK},2}(p)$ by making the sender/receiver index substitutions $1 \leftrightarrow 2$ in the definition of $\mathscr{R}_{\text{HK},1}(p)$. The Han–Kobayashi inner bound can then be expressed as

$$\mathscr{R}_{\text{HK}} = \text{FM}\left(\bigcup_p \mathscr{R}_{\text{HK},1}(p) \cap \mathscr{R}_{\text{HK},2}(p)\right),$$

where FM 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 $(R_1, R_2) = (R_{11} + R_{12}, R_{21} + R_{22})$. This projection can be carried out by Fourier–Motzkin elimination. In [2], it is shown that the Han–Kobayashi inner bound can alternatively be achieved using only one auxiliary random variable per sender, superposition coding, and simultaneous nonunique decoding. The resulting equivalent characterization of $\mathscr{R}_{\text{HK}}$ is the set of all rate pairs (R_1, R_2) such that

$$\begin{aligned}
R_1 &\le I(X_1; Y_1 \mid U_{21}, Q), && \text{(12a)}\\
R_2 &\le I(X_2; Y_2 \mid U_{12}, Q), && \text{(12b)}\\
R_1 + R_2 &\le I(X_1, U_{21}; Y_1 \mid Q)\\
&\quad + I(X_2; Y_2 \mid U_{12}, U_{21}, Q), && \text{(12c)}\\
R_1 + R_2 &\le I(X_2, U_{12}; Y_2 \mid Q)\\
&\quad + I(X_1; Y_1 \mid U_{12}, U_{21}, Q), && \text{(12d)}\\
R_1 + R_2 &\le I(X_1, U_{21}; Y_1 \mid U_{12}, Q)\\
&\quad + I(X_2, U_{12}; Y_2 \mid U_{21}, Q), && \text{(12e)}\\
2R_1 + R_2 &\le I(X_1, U_{21}; Y_1 \mid Q)\\
&\quad + I(X_1; Y_1 \mid U_{12}, U_{21}, Q)\\
&\quad + I(X_2, U_{12}; Y_2 \mid U_{21}, Q), && \text{(12f)}\\
R_1 + 2R_2 &\le I(X_2, U_{12}; Y_2 \mid Q)\\
&\quad + I(X_2; Y_2 \mid U_{12}, U_{21}, Q)\\
&\quad + I(X_1, U_{21}; Y_1 \mid U_{12}, Q) && \text{(12g)}
\end{aligned}$$

for some pmf of the form $p(q)\, p(u_{12}, x_1|q)\, p(u_{21}, x_2|q)$.

By combining the channel and the deterministic mappings as indicated by the dashed box in Figure 5, we note that $(U_{11}, U_{12}, U_{21}, U_{22}) \to (Y_1, Y_2)$ is a discrete memoryless $(4, 2)$ interference channel with message demand $\mathcal{D}_1 = \{11, 12\}$ and $\mathcal{D}_2 = \{21, 22\}$. Moreover, the Han-Kobayashi scheme uses the p-distributed random code, as defined in Section III. Thus Theorem 2 applies, and the optimal

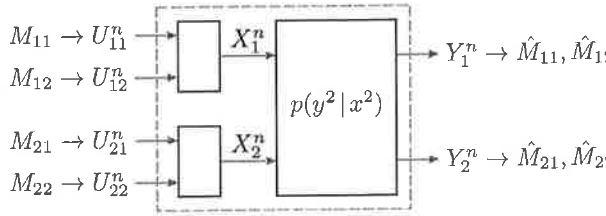


Figure 5. Interpretation of the Han–Kobayashi scheme as a $(4, 2)$-DM-IC.

rate region achievable under the Han–Kobayashi encoding functions is

$$\mathscr{R}_{\text{opt}} = \text{FM}\left(\bigcup_p \mathscr{R}_1(p) \cap \mathscr{R}_2(p)\right),$$

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

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

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 such tuples that satisfy

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

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 decoded by receivers 1 and 2, respectively.

The following can be proved by evaluating the region $\mathscr{R}_{\text{opt}}$ and comparing it to (12).

Corollary 1. The Han–Kobayashi inner bound is the optimal rate region for the 2-DM-IC, when encoding is restricted to randomly generated codebooks, superposition coding, and coded time sharing, i.e., $\mathscr{R}_{\text{HK}} = \mathscr{R}_{\text{opt}}$.

The corollary states that the Han–Kobayashi inner bound remains unchanged if the decoders are replaced by optimal ML decoders.

V. Conclusion

With the capacity region and the optimal coding scheme for the interference channel in terra incognito, we have taken a modular approach and studied the problem of optimal decoding for randomly generated codebooks, superposition coding, and time sharing. We showed that simultaneous nonunique decoding, by exploiting the codebook structure of interfering signals, achieves the performance of the optimal maximum likelihood decoding. As noted by Baccelli, El Gamal, and Tse [5] and Bidokhti, Prabhakaran, and Diggavi [10], the same performance can be achieved also by an appropriate combination of simultaneous decoding (SD) of strong interference and treating weak interference as noise (IAN), the latter of which has lower complexity. Nonetheless, simultaneous nonunique decoding provides a conceptual unification of SD and IAN, recovering all possible combinations of the two schemes at each receiver. Indeed, as with "the one ring to rule them all" [12], simultaneous nonunique decoding is the one rule that includes them all.

REFERENCES

[1] T. S. 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.

[2] 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.

[3] 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.

[4] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[5] 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.

[6] B. Bandemer and A. El Gamal, "Interference decoding for deterministic channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2966–2975, May 2011, arXiv:1001.4588.

[7] E. C. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 180–190, Mar. 1975.

[8] T. M. Cover, "An achievable rate region for the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 21, no. 4, pp. 399–404, Jul. 1975.

[9] A. A. Gohari, A. El Gamal, and V. Anantharam, "On Marton's inner bound for the general broadcast channel," in *Proceedings of ISIT 2010*, Austin, TX, Jun. 2010.

[10] S. S. Bidokhti, V. M. Prabhakaran, and S. N. Diggavi, "Is non-unique decoding necessary?" in *Proceedings of ISIT*, Boston, MA, Jul. 2012.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[12] J. R. R. Tolkien, *The Lord of the Rings*. Boston: Houghton Mifflin, 1954—1956, 3 volumes.

APPENDIX
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 the proof of Lemma 22.1 in [4]). 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 $m_2 \neq 1$,

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

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

$$\mathsf{E}|\mathcal{L}| \le 1 + 2^{n(R_2 - I(X_2;Y_1|X_1,Q) + \delta(\varepsilon))}. \tag{13}$$

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

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

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

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

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) stems from (13) and the soft-max interpretation of the log-sum-exp function [11, p. 72].

Substituting back, 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.