# Evaluation of an Achievable Rate Region for the Broadcast Channel

BRUCE E. HAJEK AND MICHAEL B. PURSLEY, SENIOR MEMBER, IEEE

*Abstract*—**The problem of transmission of separate messages to each of two receivers over a general binary-input broadcast channel is investigated. A new approach to a class of information-theoretic problems is developed and applied to obtain bounds on the cardinalities of auxiliary random variables. These bounds permit the calculation of two different regions of achievable rate pairs which are derived from the Cover–van der Meulen region $\mathfrak{R}$ of achievable rate triples. Numerical evaluation of these regions of rate pairs for two examples demonstrates that the region $\mathfrak{R}$ can be enlarged. This enlargement is accomplished by making $\mathfrak{R}$ internally consistent, as the true capacity region must be. The results display complex interactions between common and separate information in broadcast problems.**

## I. INTRODUCTION

NEARLY four years ago T. M. Cover and E. C. van der Meulen independently established an achievable rate region $\mathfrak{R}$ for the general discrete memoryless broadcast channel (see [3] and [12]). The general broadcast situation, which van der Meulen [12], [13] refers to as situation $(K, \text{III})$, involves the transmission of separate messages at rates $R_1$ and $R_2$ to each of two receivers and transmission of a common message at rate $R_0$ to both receivers. The general broadcast channel problem is to find a computable characterization of the capacity region $\mathfrak{R}^*$, which is the set of all rate triples $(R_1, R_2, R_0)$ such that the messages can be transmitted reliably over the broadcast channel.

There were two key problems left open by Cover and van der Meulen. First, they did not show that $\mathfrak{R}$ is the true capacity region (i.e., no converse was given). Second, they did not demonstrate that $\mathfrak{R}$ is computable; that is, they did not establish bounds on the cardinalities of certain auxiliary random variables.

Manuscript received November 11, 1976; revised June 18, 1978. This work was supported in part by the National Science Foundation under Grants ENG75-20864 and ENG75-22621 and in part by the Joint Services Electronics Program under Contract DAAB-07-72-C-0259. Portions of this paper were presented at the 1977 IEEE International Symposium on Information Theory, Cornell University, Ithaca, NY, October 1977, and at the 1977 International Conference on Communications, Chicago, IL, June 1977.

B. E. Hajek was with the Coordinated Science Laboratory, University of Illinois. He is now with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.

M. B. Pursley is with the Coordinated Science Laboratory and the Department of Electrical Engineering, University of Illinois, Urbana, IL 61801.

Even before the papers of Cover and van der Meulen appeared, both of these problems had already been solved by Gallager [5] for an important special case known as the *degraded* broadcast channel (see also the work of Wyner [14] and Ahlswede and Körner [1]). However, both the determination of $\mathfrak{R}^*$ and the question of the computability of $\mathfrak{R}$ remain unsolved for the general case.

The present paper makes a contribution to both of these problems. We are primarily interested in the broadcast situation in which there is no common message. In this situation, termed situation $(K, \text{I})$ by van der Meulen [12], [13], two separate message components are to be transmitted, and each component is of interest to only one receiver. This is an important problem in its own right, and we feel that its solution will be a major step toward the solution of the general broadcast problem. A rate pair $(R_1, R_2)$ is defined to be achievable for this situation if $(R_1, R_2, 0)$ is an achievable rate triple.

The most obvious region of rate pairs to consider is the set $\mathfrak{R}_0$ of all $(R_1, R_2)$ such that $(R_1, R_2, 0) \in \mathfrak{R}$. This is the region presented by van der Meulen [12, eq. (30)] for situation $(K, \text{I})$. One contribution of the present paper is to show that $\mathfrak{R}_0$ (and hence $\mathfrak{R}$) can be enlarged by employing a more general method of extracting an achievable set of rate pairs from the Cover–van der Meulen region. A larger region is the set $\hat{\mathfrak{R}}_0$ of all rate pairs of the form $(R_1 + S_1, R_2 + S_2)$, where $S_i \geq 0$ $(i = 1, 2)$ and $(R_1, R_2, S_1 + S_2) \in \mathfrak{R}$. While it is clear from the definition that the region $\hat{\mathfrak{R}}_0$ is achievable,[1] it is not clear that there is a broadcast channel for which $\hat{\mathfrak{R}}_0$ is strictly larger than $\mathfrak{R}_0$. That this is indeed the case will follow from our results on the computation of $\mathfrak{R}_0$. The *main result* of this paper is the proof that for binary-input broadcast channels, the regions $\mathfrak{R}_0$ and $\hat{\mathfrak{R}}_0$ are computable. This amounts to obtaining bounds on the range of certain auxiliary random variables. For situation $(K, \text{I})$ such bounds have previously been obtained only for the degraded broadcast channel [5]. Bounds have also been obtained for the transmission of degraded messages (situation $(K, \text{II})$ of van der Meulen [12], [13]) by Körner and Marton [10].

[1] Achievability of the region $\hat{\mathfrak{R}}_0$ follows from the observation that the common message capability of a code can be shared between the two separate message components (provided there is no requirement for confidentiality of the messages).

0018-9448/79/0100-0036$00.75 

As an application we compute $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$ for two examples, and we find that there are pairs in $\hat{\mathcal{R}}_0$ which are not in $\mathcal{R}_0$. A *side result* of this fact is that $\mathcal{R}$ can be enlarged. Indeed, if we define $\hat{\mathcal{R}}$ to be the convex hull of the union of $\mathcal{R}$ with the set of all $(R_1, R_2, 0)$ for which $(R_1, R_2) \in \hat{\mathcal{R}}_0$, then $\hat{\mathcal{R}}$ will be strictly larger than $\mathcal{R}$ whenever $\hat{\mathcal{R}}_0$ is strictly larger than $\mathcal{R}_0$ (since the latter condition implies that the intersection of $\hat{\mathcal{R}}$ with the $R_0 = 0$ plane is strictly larger than the intersection of $\mathcal{R}$ with the $R_0 = 0$ plane).

In the present paper we do not prove any new random coding theorems. Instead, we show by actual calculation of $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$ that there is something to be gained from the more general method of extracting rate pairs from $\mathcal{R}$. The result that $\hat{\mathcal{R}}_0$ is strictly larger than $\mathcal{R}_0$ for some channels is new. This fact was previously unknown primarily because both regions were never before calculated for the same channel. The obstacle was the lack of bounds on the cardinalities of auxiliary random variables needed to compute $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$. An additional result obtained in the paper is a new alternative characterization of the region $\hat{\mathcal{R}}_0$ for the general broadcast channel.

A few months after the submission of our original manuscript, Gelfand [6] obtained the capacity region for the Blackwell channel [12]. One of the referees has stated that this result and other more recent results of Gelfand and Pinsker [7] and of Marton [11] show that at least for deterministic broadcast channels the region $\hat{\mathcal{R}}_0$ is not the capacity region. However, $\hat{\mathcal{R}}_0$ is the capacity region for some class of broadcast channels, and this class contains all of the degraded broadcast channels. Furthermore, $\hat{\mathcal{R}}_0$ is an inner bound to the capacity region for any discrete memoryless broadcast channel. Hence evaluation of $\hat{\mathcal{R}}_0$ is of considerable importance for at least two classes of channels: broadcast channels for which $\hat{\mathcal{R}}_0$ is the capacity region and broadcast channels for which the capacity region (or a tighter bound to the capacity region) either is unknown or is not computable.

In addition to the specific results described above, another contribution of the paper is our new approach to the broadcast channel problem that is employed in Section II. This approach is based on a representation theory for the auxiliary random variables that arise in the information-theoretic descriptions of $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$. The basic idea is that the original auxiliary random variables are represented by simpler random variables in a way that preserves certain key properties of the joint distribution of the auxiliary and broadcast channel random variables (e.g., see Theorem 1). The applications of the representation theory that arise in the paper (the proofs of Theorems 1–3) require somewhat less than the full generality of the theory as presented in our recent report [8]. Hence only a simplified version is given here, and this version is developed as needed in the proofs of Theorems 1–3 rather than as a separate topic as in [8]. We feel that the representation theory is useful for a much broader range of information-theoretic problems than the broadcast channel problem considered in this paper.

### A. The General Broadcast Problem: Notation and Preliminaries

A two-receiver discrete memoryless broadcast channel $\mathcal{K} = (\mathcal{X}, p_1, p_2, \mathcal{Y}_1, \mathcal{Y}_2)$ consists of a finite "input" alphabet $\mathcal{X}$, finite "output" alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and transition probability functions $p_1(i|k)$ and $p_2(j|k)$ defined for $k \in \mathcal{X}$, $i \in \mathcal{Y}_1$, and $j \in \mathcal{Y}_2$. Let $\mathcal{X}^n$, $\mathcal{Y}_1^n$, and $\mathcal{Y}_2^n$ denote the sets of $n$-sequences with elements from $\mathcal{X}$, $\mathcal{Y}_1$, and $\mathcal{Y}_2$, respectively, and let $p_1^n$ and $p_2^n$ denote the $n$th-order memoryless extensions of $p_1$ and $p_2$, respectively. An $(n, M_1, M_2, M_0, \epsilon)$ code for the channel consists of $M = M_1 M_2 M_0$ codewords $x_{ijk} \in \mathcal{X}^n$, $M_1 M_0$ disjoint subsets $\mathcal{A}_{i,k} \subset \mathcal{Y}_1^n$, and $M_2 M_0$ disjoint subsets $\mathcal{B}_{j,k} \subset \mathcal{Y}_2^n$, $1 \leqslant i \leqslant M_1$, $1 \leqslant j \leqslant M_2$, $1 \leqslant k \leqslant M_0$, such that

$$M^{-1} \sum_{i,j,k} \left( \sum_{y \in \mathcal{A}_{i,k}^c} p_1^n(y|x_{ijk}) + \sum_{y \in \mathcal{B}_{j,k}^c} p_2^n(y|x_{ijk}) \right) \leqslant \epsilon.$$

A triple $(R_1, R_2, R_0)$ is achievable if for each $\epsilon > 0$ there exists an $(n, M_1, M_2, M_0, \epsilon)$ code with $n^{-1} \log M_i \geqslant R_i - \epsilon$, $i = 0, 1, 2$. Throughout this paper logarithms will be to the base two.

The broadcast channel, as just defined, models a single sender broadcasting information to two receivers. A rate triple $(R_1, R_2, R_0)$ is achievable if the sender can transmit separate messages reliably to the first receiver at rate $R_1$ and to the second receiver at rate $R_2$ and can simultaneously send a common message sequence to both receivers at rate $R_0$.

Auxiliary random variables are introduced through the notion of a test channel. A random vector $(\boldsymbol{U}, X)$ is defined to be a *test channel* for the broadcast channel $\mathcal{K} = (\mathcal{X}, p_1, p_2, \mathcal{Y}_1, \mathcal{Y}_2)$ if the components of the random vector $\boldsymbol{U}$ are mutually independent and if the random variable $X$ has range $\mathcal{X}$. The components $U_i$ of $\boldsymbol{U}$ are the test channel inputs. The range of $U_i$ is denoted by $\mathcal{U}_i$ and is called the alphabet for the $i$th input to the test channel. Given a test channel $(\boldsymbol{U}, X)$, we say that the random vector $\boldsymbol{Y} = (Y_2, Y_2)$ is an output of the broadcast channel $\mathcal{K}$ if, for each $i = 1, 2$,

$$P[Y_i = j | X = k] = p_i(j|k)$$

for $k \in \mathcal{X}$, $j \in \mathcal{Y}_i$, and if $(\boldsymbol{U}, X, \boldsymbol{Y})$ is a Markov chain (i.e., $\boldsymbol{U}$ and $\boldsymbol{Y}$ are conditionally independent given $X$). The vector $(\boldsymbol{U}, X, \boldsymbol{Y})$ may be considered to be a cascade of the test channel $(\boldsymbol{U}, X)$ with the broadcast channel $\mathcal{K}$. Given a set $\mathcal{D}$ of test channels, we denote by $\mathcal{D}\mathcal{K}$ the set of all $(\boldsymbol{U}, X, \boldsymbol{Y})$ such that $(\boldsymbol{U}, X) \in \mathcal{D}$ and $\boldsymbol{Y}$ is an output of $\mathcal{K}$ corresponding to input $X$.

Let $\mathcal{P}$ (respectively, $\hat{\mathcal{P}}$) be the set of all test channels $(\boldsymbol{U}, X)$, with $\boldsymbol{U} = (U_1, U_2)$ (respectively, $\boldsymbol{U} = (U_1, U_2, U_0)$), such that the test channel input $\boldsymbol{U}$ has finite range. The Cover–van der Meulen region of achievable rate triples is

the convex closure of all rate triples $(R_1, R_2, R_0)$ for which there exists some $(\boldsymbol{U}, X, \boldsymbol{Y}) \in \hat{\mathcal{P}}\mathcal{K}$ such that

$$R_1 \leqslant I(U_1; Y_1|U_0),$$
$$R_2 \leqslant I(U_2; Y_2|U_0),$$
$$R_0 \leqslant \min\{I(U_0; Y_1|U_1), I(U_0; Y_2|U_2)\},$$
$$R_1 + R_0 \leqslant I(U_0, U_1; Y_1),$$

and

$$R_2 + R_0 \leqslant I(U_0, U_2; Y_2).$$

Throughout the paper $\mathbb{R}$ denotes the real line, and $\mathbb{R}_+$ denotes the nonnegative real line. If $X$ and $Y$ are random variables, $X \sim Y$ means that $X$ and $Y$ have the same distribution.

### B. The Broadcast Situation with No Common Message

The projection of $\mathcal{R}$ on the $R_0 = 0$ plane yields the two-dimensional region $\mathcal{R}_0$. An alternative description of $\mathcal{R}_0$ given by van der Meulen [12] is $\mathcal{R}_0 = \overline{\text{co}}\, C(\mathcal{P})$, where "co" denotes convex hull, "$\overline{\text{co}}$" denotes the closure of the convex hull, and

$$C(\mathcal{D}) = \{(R_1, R_2) | R_1 \leqslant I(U_1; Y_1), R_2 \leqslant I(U_2; Y_2), (\boldsymbol{U}, X, \boldsymbol{Y}) \in \mathcal{D}\mathcal{K}\}$$

for any set $\mathcal{D}$ of two-input test channels for the broadcast channel $\mathcal{K}$.

The region $\hat{\mathcal{R}}_0$, which we have defined by

$$\hat{\mathcal{R}}_0 = \{(R_1 + S_1, R_2 + S_2) | S_1 \geqslant 0, S_2 \geqslant 0, (R_1, R_2, S_1 + S_2) \in \mathcal{R}\},$$

has the following very useful alternative characterization which is proved in Appendix A.

*Proposition 1:* If $\mathcal{C}$ is the set of all rate pairs $(T_1, T_2)$ such that

$$T_1 \leqslant I(U_1, U_0; Y_1), \tag{1}$$
$$T_2 \leqslant I(U_2, U_0; Y_2), \tag{2}$$

and

$$T_1 + T_2 \leqslant I(U_1; Y_1|U_0) + I(U_2; Y_2|U_0) + \min\{I(U_0; Y_1), I(U_0; Y_2)\}, \tag{3}$$

then $\hat{\mathcal{R}}_0 = \overline{\text{co}}\, \mathcal{C}$.

For the general two-receiver binary-input memoryless broadcast channel, we will show (Theorem 3) that $\hat{\mathcal{R}}_0$ can be calculated by considering only those test channels $((U_1, U_2, U_0), X)$ for which $U_0$ is binary, $U_1$ and $U_2$ are ternary, and $X$ is a (deterministic) function of $(U_1, U_2, U_0)$. The proof of this employs a similar result for $\mathcal{R}_0$ (Theorem 2); namely, $\mathcal{R}_0 = \mathcal{R}_b \triangleq \overline{\text{co}}\, C(\mathcal{P}_b)$, where $\mathcal{P}_b$ is the collection of test channels $((U_1, U_2), X) \in \mathcal{P}$ such that $U_1$ and $U_2$ are binary and either $X \equiv U_1 \wedge U_2$ or $X \equiv U_1 \vee U_2$ (where "$\wedge$" denotes minimum and "$\vee$" denotes maximum). In fact, the proof of the result for $\hat{\mathcal{R}}_0$ is essentially an extension of the proof of the result for $\mathcal{R}_0$. Indeed, there are only four basic information quantities involved in the above characterization of $\hat{\mathcal{R}}_0$: $I(U_0; Y_i)$ and $I(U_i; Y_i|U_0)$, for $i = 1$ and 2. Roughly speaking, after selecting $U_0$ to make the quantities $I(U_0; Y_i)$ large, the remaining problem of maximizing $I(U_i; Y_i|U_0)$ is equivalent to the maximization of the quantities $I(U_i; Y_i)$, which we also encounter in computing $\mathcal{R}_0$.

## II. Cardinality Bounds for $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$

In this section we will consider the regions $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$ for a general two-receiver binary-input memoryless broadcast channel $\mathcal{K} = (\{0,1\}, p_1 p_2, \mathcal{Y}_1, \mathcal{Y}_2)$. We begin with an alternative characterization of $\mathcal{R}_0 = \overline{\text{co}}\, C(\mathcal{P})$. For $\boldsymbol{\lambda} \in \mathbb{R}_+^2$ define

$$c(\boldsymbol{\lambda}) = \sup\{\Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}, X) | (\boldsymbol{U}, X) \in \mathcal{P}\}$$

where

$$\Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}, X) \triangleq \lambda_1 I(U_1; Y_1) + \lambda_2 I(U_2; Y_2) \tag{4}$$

where $\boldsymbol{Y}$ is any output of $\mathcal{K}$ corresponding to the test channel input $(\boldsymbol{U}, X)$. Then $\mathcal{R}_0$ is the collection of pairs $(R_1, R_2) \in \mathbb{R}_+^2$ dominated by the family of lines $\{\lambda_1 x_1 + \lambda_2 x_2 = c(\boldsymbol{\lambda}) | \boldsymbol{\lambda} \in \mathbb{R}_+^2\}$. To prove that $\mathcal{R}_0 = \mathcal{R}_b$, we need only show that given $\boldsymbol{\lambda} \in \mathbb{R}_+^2$ and $(\boldsymbol{U}, X) \in \mathcal{P}$, there exists a $(\boldsymbol{U}^*, X^*) \in \mathcal{P}_b$ such that $\Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}, X) \leqslant \Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}^*, X^*)$. Our approach will be to first "reduce" the cardinality of the range of $U_2$. That is, we will produce $(\boldsymbol{U}^*, X^*) \in \mathcal{P}$ such that $U_1$ and $U_1^*$ have the same distribution, $U_2^*$ is binary, $X^*$ is a function of $\boldsymbol{U}^*$, and $\Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}, X) \leqslant \Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}^*, X^*)$. A second application of this procedure will then prove that $U_1^*$ can also be chosen to be binary. Finally, we show that $(\boldsymbol{U}^*, X^*)$ can be chosen so that either $X^* \equiv U_1^* \vee U_2^*$ or $X^* \equiv U_1^* \wedge U_2^*$ (i.e., $(\boldsymbol{U}^*, X^*) \in \mathcal{P}_b$), which establishes the result $\mathcal{R}_0 = \mathcal{R}_b$. Let $\boldsymbol{\lambda} \in \mathbb{R}_+^2$ and $(\boldsymbol{U}, X) \in \mathcal{P}$ be fixed throughout this section.

Let $\mathcal{U}_1$ and $\mathcal{U}_2$ denote the ranges[2] of $U_1$ and $U_2$, and suppose $\|\mathcal{U}_1\| = n$ and $\|\mathcal{U}_2\| = m$. By "relabeling" we can suppose that $\mathcal{U}_1 = \{1, \cdots, n\}$, $\mathcal{U}_2 = \{1, \cdots, m\}$, $P[X = 1 | U_1 = i]$ is nonincreasing in $i$, and $P[X = 1 | U_2 = j]$ is nondecreasing in $j$.

Before exhibiting the random variables $(\boldsymbol{U}^*, X^*)$, we will construct an intermediate pair of random variables $(\tilde{\boldsymbol{U}}, \tilde{X})$ contained in a subset $\tilde{\mathcal{P}}_n$ of $\mathcal{P}$. Let $\mathcal{T}_n$ be the collection of all $\boldsymbol{t} \in \mathbb{R}^n$ such that $1 \geqslant t_1 \geqslant \cdots \geqslant t_n \geqslant 0$, and let $\mathcal{S}_n$ be the collection of all $n$-dimensional probability vectors. For convenience let $s_0 = t_{n+1} = 0$ and $t_0 = 1$ in this section. Define $\tilde{\mathcal{P}}_n$ as the set of all $(\tilde{\boldsymbol{U}}, \tilde{X}) \in \mathcal{P}$ for which there exists a $(\boldsymbol{s}, \boldsymbol{t}) \in \mathcal{S}_n \times \mathcal{T}_n$ such that

$$P[\tilde{U}_1 = i] = s_i, \qquad P[\tilde{U}_2 = j] = t_j - t_{j+1}, \qquad \tilde{X} = \chi_{\{\tilde{U}_1 \leqslant \tilde{U}_2\}}, \tag{5}$$

[2]Although these sets are assumed to be finite in this paper, we show in [8] that such a restriction is not necessary and when removed gives a more general result. The finiteness of $\mathcal{U}_1$ and $\mathcal{U}_2$ is assumed here to simplify the presentation of the basic results.

where $1 \leqslant i \leqslant n$, $0 \leqslant j \leqslant m$, and $\chi$ denotes the indicator function (i.e., $\tilde{X}=1$ if $\tilde{U}_1 \leqslant \tilde{U}_2$ and $\tilde{X}=0$ otherwise). It follows that

$$P[\tilde{X}=1|\tilde{U}_1=i]=t_i, \qquad r_j \triangleq P[\tilde{X}=1|\tilde{U}_2=j]=\sum_{k=0}^{j} s_k, \tag{6}$$

and that

$$P[\tilde{X}=1]=\sum_{i=1}^{n} s_i t_i=\sum_{j=0}^{n}(t_j-t_{j+1})r_j. \tag{7}$$

Whenever random variables $(\tilde{U},\tilde{X})$ are related to vectors $(s,t)$ as in (5), we shall write $(s,t)\leftrightarrow(\tilde{U},\tilde{X})$. Now if we let

$$s_i=P[U_1=i], \qquad t_i=P[X=1|U_1=i], \qquad 1 \leqslant i \leqslant n, \tag{8}$$

then we obtain a particular $(\tilde{U},\tilde{X})\in\tilde{\mathfrak{P}}_n$ such that $(s,t)\leftrightarrow(\tilde{U},\tilde{X})$. We summarize the properties of $(\tilde{U},\tilde{X})$ in the following theorem.

*Theorem 1:* If $\boldsymbol{Y}$ (respectively, $\tilde{\boldsymbol{Y}}$) is an output of $\mathcal{K}$ corresponding to input $(\boldsymbol{U},X)$ (respectively, $(\tilde{\boldsymbol{U}},\tilde{X})$), then

a) $(U_1,X)\sim(\tilde{U}_1,\tilde{X})$,
b) $I(U_i;Y_i)\leqslant I(\tilde{U}_i;\tilde{Y}_i), \qquad i=1,2.$

*Remark:* Clearly b) implies that $\Phi_\lambda(\boldsymbol{U},X)\leqslant\Phi_\lambda(\tilde{\boldsymbol{U}},X)$.

*Proof:* Assertion a) is obvious from (5), (6), and (8). Assertion b) for $i=1$ follows from a), which (along with the definition of channel outputs) implies that $(U_1,X,Y_1)\sim(\tilde{U}_1,\tilde{X},\tilde{Y}_1)$. Similarly $Y_2\sim\tilde{Y}_2$ so that $H(Y_2)=H(\tilde{Y}_2)$. Hence to prove b) for $i=2$, it suffices to show that

$$H(Y_2|U_2)\geqslant H(\tilde{Y}_2|\tilde{U}_2). \tag{9}$$

Since $(\boldsymbol{U},X,\boldsymbol{Y})$ is a Markov chain for which the conditional discrete density function for $Y_i$ given $X$ is $p_i$, the conditional entropy $H(Y_i|U_i)$ can be written as an expectation of a function of $P[X=1|U_i]$. First, note that for $j\in\mathcal{Y}_i$,

$$P[Y_i=j|U_i]=\sum_{k=0}^{1} p_i(j|k)P[X=k|U_i]$$
$$=L_{i,j}(P[X=1|U_i])$$

where $L_{i,j}:[0,1]\to[0,1]$ is defined by

$$L_{i,j}(\alpha)=\alpha p_i(j|1)+(1-\alpha)p_i(j|0)$$

for $j\in\mathcal{Y}_i$, $\alpha\in[0,1]$, and $i=1,2$. Hence

$$H(Y_i|U_i)=E\left\{\sum_j \varphi[L_{i,j}(P[X=1|U_i])]\right\}$$
$$=E\{\Lambda_i(P[X=1|U_i])\} \tag{10}$$

where $\varphi(p)=-p\log p$ and

$$\Lambda_i(\alpha)=\sum_j \varphi[L_{i,j}(\alpha)]. \tag{11}$$

Since $L_{i,j}$ is affine and $\varphi$ is concave, $\varphi\circ L_{i,j}$ is also concave. Therefore, $\Lambda_i$ is the sum of (finitely many) concave functions, and so it is concave.

Notice that $\Lambda_i$ is specified completely by the broadcast channel transition probability $p_i$ and that (10) is established using only the facts that $(\boldsymbol{U},X,\boldsymbol{Y})$ is a Markov chain and that the conditional distribution of $Y_i$ given $X$ is determined by $p_i$; the joint distribution of $Y_1$ and $Y_2$ is immaterial.

If we define $a_j=P[U_2\geqslant j]$ and $b_j=P[X=1|U_2=j]$, then by (10),

$$H(Y_2|U_2)=\sum_{j=1}^{m}(a_j-a_{j+1})\Lambda_2(b_j). \tag{12}$$

Similarly (see (5) and (6)),

$$H(\tilde{Y}_2|\tilde{U}_2)=E\{\Lambda_2(P[\tilde{X}=1|\tilde{U}_2])\}$$
$$=\sum_{j=0}^{m}(t_j-t_{j+1})\Lambda_2(r_j). \tag{13}$$

Since our goal is to prove (9), we wish to compare (12) and (13).

Let $\gamma_1\geqslant\gamma_2\geqslant\cdots\geqslant\gamma_N$ be an ordering of $\{a_2,\cdots,a_m\}\cup\{t_1,\cdots,t_n\}$ (hence $N=m+n-1$), and let $\gamma_0=a_1=t_0=1$, $\gamma_{N+1}=a_{m+1}=t_{n+1}=0$. Recall that $1=a_1\geqslant\cdots\geqslant a_{m+1}=0$ and $1=t_0\geqslant\cdots\geqslant t_{n+1}=0$. Define $\rho(j)$ and $\tilde{\rho}(j)$ by

$$\rho(j)=k, \qquad \text{if } a_k\geqslant\gamma_j>a_{k+1}$$

and

$$\tilde{\rho}(j)=k, \qquad \text{if } t_k\geqslant\gamma_j>t_{k+1},$$

and let $b'_j=b_{\rho(j)}$ and $r'_j=r_{\tilde{\rho}(j)}$. With $\delta_j\triangleq\gamma_j-\gamma_{j+1}$, (12) and (13) become

$$H(Y_2|U_2)=\sum_{j=0}^{N}\delta_j\Lambda_2(b'_j) \tag{14}$$

and

$$H(\tilde{Y}_2|\tilde{U}_2)=\sum_{j=0}^{N}\delta_j\Lambda_2(r'_j), \tag{15}$$

respectively. Since both $b_k$ and $r_k$ are nondecreasing in $k$ and since $\rho(j)$ and $\tilde{\rho}(j)$ are both nondecreasing in $j$, it follows that both $b'_j$ and $r'_j$ are nondecreasing sequences. This, the concavity of $\Lambda_2$, and the next two lemmas will establish (9), completing the proof of Theorem 1.

*Lemma 1:* Let $x_0,x_1,\cdots,x_N$ and $y_0,y_1,\cdots,y_N$ be nondecreasing sequences of real numbers. Let $\xi_0,\xi_1,\cdots,\xi_N$ be a sequence of real numbers such that for each $k$ in the range $0\leqslant k\leqslant N$,

$$\sum_{j=k}^{N}\xi_j x_j\geqslant\sum_{j=k}^{N}\xi_j y_j \tag{16}$$

with equality for $k=0$. Then for any concave function $\Lambda$,

$$\sum_{j=0}^{N}\xi_j\Lambda(x_j)\leqslant\sum_{j=0}^{N}\xi_j\Lambda(y_j). \tag{17}$$

Lemma 1, which is proved in Appendix B, is a generalization of a result obtained in 1929 by Hardy, Littlewood, and Pólya (see [9, p. 89]). The generalization is due to Fuchs [4], and the proof that we give in Appendix B is essentially his.

*Lemma 2:* For $0 \leqslant k \leqslant N$,

$$\sum_{j=k}^{N} \delta_j b'_j \leqslant \sum_{j=k}^{N} \delta_j r'_j$$

with equality for $k=0$.

*Proof of Lemma 2:* Construct random variables $\Theta$ and $\tilde{\Theta}$ on $\{0,1,\cdots,N\}$ such that $(\Theta, U_2, X)$ and $(\tilde{\Theta}, U_2, \tilde{X})$ are Markov chains with

$$P[\Theta=j|U_2=k]=\delta_j(a_k-a_{k+1})^{-1}$$

if $k=\rho(j)$ and

$$P[\tilde{\Theta}=j|\tilde{U}_2=k]=\delta_j(t_k-t_{k+1})^{-1}$$

if $k=\tilde{\rho}(j)$. These probabilities are zero otherwise. Since $U_2$ is independent of $U_1$, $\Theta$ may also be chosen to be independent of $U_1$. Similarly, let $\tilde{\Theta}$ be independent of $\tilde{U}_1$. Note that $U_2=\rho(\Theta)$, that

$$P[\Theta=j]=P[\Theta=j|U_2=\rho(j)]P[U_2=\rho(j)]=\delta_j,$$

and that

$$\begin{aligned} b'_j &= P[X=1|U_2=\rho(j)] \\ &= P[X=1|U_2=\rho(j),\Theta=j]=P[X=1|\Theta=j]. \end{aligned}$$

Similarly, $\tilde{U}_2=\tilde{\rho}(\tilde{\Theta})$, $P[\tilde{\Theta}=j]=\delta_j$, and $r'_j=P[\tilde{X}=1|\tilde{\Theta}=j]$. Notice that $(\tilde{\Theta},\tilde{U}_2)$ is also independent of $\tilde{U}_1$, since $(\tilde{\Theta},\tilde{U}_2)=(\tilde{\Theta},\tilde{\rho}(\tilde{\Theta}))$. We then obtain, for $0\leqslant k\leqslant N$,

$$\begin{aligned} \sum_{j=k}^{N}\delta_j b'_j &= P[\Theta\geqslant k, X=1] \\ &= \sum_{i=1}^{n} P[\Theta\geqslant k, X=1|U_1=i]s_i \\ &\leqslant \sum_{i=1}^{n} \min\{P[\Theta\geqslant k|U_1=i], P[X=1|U_1=i]\}s_i \\ &= \sum_{i=1}^{n} \min\{P[\Theta\geqslant k], P[\tilde{X}=1|\tilde{U}_1=i]\}s_i \\ &= \sum_{i=1}^{n} \min\{P[\tilde{\Theta}\geqslant k], P[\tilde{U}_2\geqslant i]\}s_i \\ &= \sum_{i=1}^{n} \min\{P[\tilde{\Theta}\geqslant k], P[\tilde{\rho}(\tilde{\Theta})\geqslant i]\}s_i \\ &= \sum_{i=1}^{n} P[\tilde{\Theta}\geqslant k, \tilde{\rho}(\tilde{\Theta})\geqslant i]s_i \\ &= \sum_{i=1}^{n} P[\tilde{\Theta}\geqslant k, \tilde{U}_2\geqslant i|\tilde{U}_1=i]s_i \\ &= \sum_{i=1}^{n} P[\tilde{\Theta}\geqslant k, \tilde{X}=1|\tilde{U}_1=i]s_i \\ &= P[\tilde{\Theta}\geqslant k, \tilde{X}=1]=\sum_{j=k}^{N}\delta_j r'_j. \end{aligned} \tag{18}$$

Note that for $k=0$ equality holds in (18). Hence the proof of Lemma 2 is complete.

We now return to the proof of the main result of this section, as outlined in the first paragraph. Suppose $(\boldsymbol{s}',\boldsymbol{t}') \in \mathcal{S}_n \times \mathcal{T}_n$ and $(\boldsymbol{s}',\boldsymbol{t}')\leftrightarrow(\boldsymbol{U}',X')\in\tilde{\mathcal{P}}_n$. Recall that this means (see (5))

$$P[U'_1=i]=s_i, \qquad P[U'_2=j]=t_j-t_{j+1}, \qquad X'=\chi_{\{U'_1\leqslant U'_2\}}.$$

We may write

$$\Phi_\lambda(\boldsymbol{U}',X')=\Phi_\lambda^{(1)}(\boldsymbol{U}',X')-\Phi_\lambda^{(2)}(\boldsymbol{U}',X')$$

where we define

$$\Phi_\lambda^{(1)}(\boldsymbol{U}',X')=\lambda_1 H(Y'_1)+\lambda_2 H(Y'_2)$$

and

$$\Phi_\lambda^{(2)}(\boldsymbol{U}',X')=\lambda_1 H(Y'_1|U'_1)+\lambda_2 H(Y'_2|U'_2).$$

As before, $\boldsymbol{Y}'$ is any output of the channel corresponding to input $(\boldsymbol{U}',X')$. Since $(\boldsymbol{U}',X')$ is completely determined by $(\boldsymbol{s}',\boldsymbol{t})$, we may consider $\Phi_\lambda(\boldsymbol{U}',X')$ and $\Phi_\lambda^{(i)}(\boldsymbol{U}',X')$, $i=1,2$, to be functions of $(\boldsymbol{s}',\boldsymbol{t}')$. For example, if $s'_0=t'_{n+1}=0$ and $t'_0=1$, then (10) may be applied to yield

$$\Phi_\lambda^{(2)}(\boldsymbol{s}',\boldsymbol{t}')=\lambda_1\sum_{i=1}^{n} s'_i\Lambda_1(t'_i)+\lambda_2\sum_{j=0}^{n}(t'_j-t'_{j+1})\Lambda_2\left(\sum_{i=0}^{j} s'_i\right). \tag{19}$$

Recall that $\boldsymbol{s}$ and $\boldsymbol{t}$ were given in (8) and that $(\tilde{\boldsymbol{U}},\tilde{X})\in\tilde{\mathcal{P}}_n$ was constructed according to $(\boldsymbol{s},\boldsymbol{t})\leftrightarrow(\tilde{\boldsymbol{U}},\tilde{X})$. Let $\mathcal{D}\subset\mathcal{T}_n$ consist of those $\boldsymbol{t}'\in\mathcal{T}_n$ such that

$$\sum_{i=1}^{n} s_i t'_i = P[\tilde{X}=1]. \tag{20}$$

Then by Theorem 1 and the fact that $\boldsymbol{t}\in\mathcal{D}$,

$$\Phi_\lambda(\boldsymbol{U},X)\leqslant\Phi_\lambda(\boldsymbol{s},\boldsymbol{t})\leqslant \sup\{\Phi_\lambda(\boldsymbol{s},\boldsymbol{t}')|\boldsymbol{t}'\in\mathcal{D}\}. \tag{21}$$

If $(\boldsymbol{U}',X')\leftrightarrow(\boldsymbol{s},\boldsymbol{t}')$ and $\boldsymbol{t}'\in\mathcal{D}$, then (20) implies that $P[\tilde{X}=1]=P[X'=1]$. This, in turn, implies that $\Phi_\lambda^{(1)}(\boldsymbol{s},\cdot)$ is constant on $\mathcal{D}$, because the entropies of the channel outputs depend only on the distribution of the input $X'$. On the other hand, we observe from (19) that $\Phi_\lambda^{(2)}(\boldsymbol{s},\cdot)$ is concave on $\mathcal{T}_n$, and hence on $\mathcal{D}$, since the sum of linear and concave functions is concave. Since $\mathcal{D}$ is a compact convex subset of $\mathbb{R}^n_+$, the concave (continuous) function $\Phi_\lambda^{(2)}(\boldsymbol{s},\cdot)$ is minimized over $\mathcal{D}$ at an extreme point of $\mathcal{D}$, let us say $\boldsymbol{t}^*$. Then if $(\boldsymbol{U}^*,X^*)\in\tilde{\mathcal{P}}_n$ is constructed according to $(\boldsymbol{s},\boldsymbol{t}^*)\leftrightarrow(\boldsymbol{U}^*,X^*)$, it follows that

$$\Phi_\lambda(\boldsymbol{U},X)\leqslant\sup\{\Phi_\lambda(\boldsymbol{s},\boldsymbol{t}')|\boldsymbol{t}'\in\mathcal{D}\}=\Phi_\lambda(\boldsymbol{U}^*,X^*).$$

Since $\boldsymbol{t}^*=(t_1,t_2,\cdots,t_n)$ is an extreme point of $\mathcal{D}$, we claim that for some integers $j$ and $k$ for which $0\leqslant j<k\leqslant n$,

$$1=t_0^*=\cdots=t_j^*>t_{j+1}^*=\cdots=t_k^*\geqslant t_{k+1}^*=\cdots=t_{n+1}^*=0 \tag{22}$$

where we have (by convention) let $t_0^*=1$ and $t_{n+1}^*=0$. To see that this claim is true, suppose that $\boldsymbol{t}^*$ is an extreme point of $\mathcal{D}$ but that $\boldsymbol{t}^*$ does not have the form indicated in (22). Then there must exist integers $a$, $b$, and $c$ satisfying

$0 \leqslant a < b < c \leqslant n$ such that

$$1 \geqslant t_a^* > t_{a+1}^* = t_b^* > t_{b+1}^* = t_c^* > t_{c+1}^* \geqslant 0.$$

Choose $\epsilon_1$ and $\epsilon_2$ such that $0 < \epsilon_1 + \epsilon_2 < t_b^* - t_{b+1}^*$, $0 \leqslant \epsilon_1 < t_a^* - t_{a+1}^*$, $0 \leqslant \epsilon_2 < t_c^* - t_{c+1}^*$, and

$$\epsilon_1 \sum_{i=a+1}^{b} s_i = \epsilon_2 \sum_{i=b+1}^{c} s_i.$$

Define $t_i^{(1)} = t_i^{(2)} = t_i^*$, for $0 \leqslant i \leqslant a$ and $c+1 \leqslant i \leqslant n+1$. For $a+1 \leqslant i \leqslant b$ let $t_i^{(1)} = t_i^* + \epsilon_1$ and $t_i^{(2)} = t_i^* - \epsilon_1$, and for $b+1 \leqslant i \leqslant c$ let $t_i^{(1)} = t_i^* - \epsilon_2$ and $t_i^{(2)} = t_i^* + \epsilon_2$. From the definition of $\mathcal{D}$ (see (20)) it is clear that $\boldsymbol{t}^{(m)} = (t_1^{(m)}, t_2^{(m)}, \cdots, t_n^{(m)}) \in \mathcal{D}$, for $m = 1, 2$. It is easy to see that $\boldsymbol{t}^* = \frac{1}{2}(\boldsymbol{t}^{(1)} + \boldsymbol{t}^{(2)})$. Since this contradicts the fact that $\boldsymbol{t}^*$ is an extreme point of $\mathcal{D}$, the claim must be true. Since $(\boldsymbol{s}, \boldsymbol{t}^*) \leftrightarrow (\boldsymbol{U}^*, X^*)$, it follows from the form of $\boldsymbol{t}^*$ given in (22) that $P[U_2^* = i] = 0$ unless $i = j$ or $i = k$; that is, $U_2^*$ is essentially a binary random variable.

Summarizing, we have started with $\boldsymbol{\lambda} \in \mathbb{R}_+^2$ and $(\boldsymbol{U}, X) \in \mathcal{P}$ and then produced $(\boldsymbol{U}^*, X^*)$, where $U_1^* \sim U_1$, $U_2^* \in \{j, k\}$ with probability one, $X^* = \chi_{\{U_1^* \leqslant U_2^*\}}$, and $\Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}, X) \leqslant \Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}^*, X^*)$. By repeating this procedure we can similarly show that $U_1^*$ can be taken to be distributed on some set $\{j', k'\}$. The special form of the function $X^* = \chi_{\{U_1^* \leqslant U_2^*\}}$ then insures that, after "relabeling," $U_1^*$ and $U_2^*$ can be supposed to have range $\{0, 1\}$ and $X^*$ can be taken to be either $U_1^* \wedge U_2^*$ or $U_1^* \vee U_2^*$. We have thus proven the following theorem.

*Theorem 2:* For the general two-receiver binary-input memoryless broadcast channel, $\mathcal{R}_0 = \mathcal{R}_b$.

*Remark:* Since the $(\boldsymbol{U}^*, X^*)$ constructed from $(\boldsymbol{U}, X)$ in the proof of Theorem 2 satisfied $P[X^* = 1] = P[X = 1]$, we also have the following generalization:

$$\begin{aligned} c(\boldsymbol{\lambda}, p) &\triangleq \sup \left\{ \Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}, X) \mid (\boldsymbol{U}, X) \in \tilde{\mathcal{P}},\ P[X = 1] = p \right\} \\ &= \max \left\{ \Phi_{\boldsymbol{\lambda}}(\boldsymbol{U}, X) \mid (\boldsymbol{U}, X) \in \mathcal{P}_b,\ P[X = 1] = p \right\}. \end{aligned} \tag{23}$$

This will be used in the proof of Theorem 3. We will also use the fact that $c(\boldsymbol{\lambda}, p)$ is continuous in $p$, which is easily proved from (23).

In Appendix C we prove the following result for the broadcast channel $\mathcal{K}$.

*Theorem 3:* For the general two-receiver binary-input memoryless broadcast channel, the achievable rate region $\hat{\mathcal{R}}_0$ can be obtained by considering only test channels $((U_1, U_2, U_0), X)$ in $\hat{\mathcal{P}}_b$ (i.e., $U_0$ is binary, $U_1$ and $U_2$ are ternary, and $X$ is a deterministic function of $(U_1, U_2, U_0)$).

## III. Examples—$\hat{\mathcal{R}}_0$ Compared with $\mathcal{R}_0$

In the previous section we obtained bounds on the cardinalities of the ranges of the auxiliary random variables employed in the characterization of $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$. Both regions were derived from the region of achievable rate triples $\mathcal{R}$ previously given by Cover and van der Meulen. However, to the best of our knowledge, both $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$ have not been calculated for the same channel. By calculating $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$ for two simple examples, we will show that $\hat{\mathcal{R}}_0$ is sometimes larger than $\mathcal{R}_0$. This fact, as revealed by out second example, is the result of complex interplay between "separate" and "common" mutual information quantities. As noted in Section I, the fact that $\hat{\mathcal{R}}_0$ is larger than $\mathcal{R}_0$ implies that $\hat{\mathcal{R}}$ is larger than $\mathcal{R}$.

### A. A Binary Symmetric Broadcast Channel

The binary symmetric broadcast channel (BSBC) is a simple example of a degraded broadcast channel, the capacity of which is already known [13]. We consider it here to clarify the relation of $\mathcal{R}_0$ to $\hat{\mathcal{R}}_0$. $\hat{\mathcal{R}}_0$ is the capacity region of rate pairs for the BSBC and, indeed, any degraded channel. This is easily seen by comparing our expression for $\mathcal{C}$ with the capacity region for degraded channels as given in [5] or [13]. Thus the capacity region of degraded channels may be extracted from the Cover–van der Meulen region of rate triples. This was originally proved by van der Meulen ([12, p. 187]) in the context of his situation $(K, \text{II})$.

Now consider the particular BSBC with crossover probabilities zero and 0.25. That is, consider $\mathcal{K} = (\{0, 1\}, p_1, p_2, \{0, 1\})$, where $p_1(0|0) = p_1(1|1) = 1$ and $p_2(0|0) = p_2(1|1) = 0.75$. We shall compute $\mathcal{R}_0$ for this channel. Consider $((U_1, U_2), X) \in \mathcal{P}_b$ such that $P[U_i = 1] = \alpha_i$, $P[U_i = 0] = 1 - \alpha_i$, and $X \equiv U_1 \wedge U_2$. If $(Y_1, Y_2)$ is an output of $\mathcal{K}$ corresponding to $(\boldsymbol{U}, X)$, then $I[U_i, Y_i] = t_i(\alpha_1, \alpha_2)$, where

$$t_1(\alpha_1, \alpha_2) \triangleq h(\alpha_1 \alpha_2) - \alpha_1 h(\alpha_2)$$

and

$$t_2(\alpha_1, \alpha_2) \triangleq h(\alpha_1 \alpha_2 * 0.25) - \alpha_2 h(\alpha_1 * 0.25) - (1 - \alpha_2) h(0.25)$$

where $p * q = p(1-q) + (1-p)q$. Because of the symmetry between $X = 0$ and $X = 1$ for the BSBC, there is no need to consider $X \equiv U_1 \vee U_2$ so that, by Theorem 2,

$$\mathcal{R}_0 = \text{co} \left\{ (R_1, R_2) \mid R_i \leqslant t_i(\alpha_1, \alpha_2), \text{ for some } 0 \leqslant \alpha_1, \alpha_2 \leqslant 1 \right\}.$$

It is easy to show numerically that $\mathcal{R}_0$ is simply the region one obtains by time sharing. On the other hand, the set of all achievable rate pairs, which consists of those $(R_1, R_2)$ for which $R_1 \leqslant 1 - h^{-1}(h(R_2) * 0.25)$ (see [13]), is strictly larger than the time-sharing region.

### B. The Skewed Binary Broadcast Channel

We define the "skewed binary broadcast channel" to be $\mathcal{K} = (\{0, 1\}, p_1, p_2, \{0, 1\}, \{0, 1\})$, where $p_1(0|0) = p_2(1|1) = 1$ and $p_1(0|1) = p_2(1|0) = 0.5$. Using Theorem 2 and Theorem 3, we can evaluate $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$ numerically. Instead we will use a more revealing approach, which involves considering some of the quantities which were used to prove the results of Sections II.

It turns out that $\mathcal{R}_0$ is the time-sharing region for this channel. We will prove a stronger result first with a view

toward calculating $\hat{\mathcal{R}}_0$. We will begin by computing $c(\boldsymbol{\lambda},p)$ for $\boldsymbol{\lambda}\in\mathbb{R}_+^2$ and $0\leqslant p\leqslant 1$. (See the remark after Theorem 2 for definitions.) It was shown that $c(\boldsymbol{\lambda},p)$ can be computed using only test channels in $\mathcal{P}_b$. We will show that for the skewed binary broadcast channel, the set of test channels may be reduced even further—specifically, to the set of $(\boldsymbol{U},X)\in\mathcal{P}_b$ such that either $X=U_1$ or $X=U_2$. This follows from Proposition 2, which is proved in Appendix D.

*Proposition 2:* For all $\boldsymbol{\lambda}\in\mathbb{R}_+^2$ and $0\leqslant p\leqslant 1$,

$$c(\boldsymbol{\lambda},p)=R_{sb}^{p}\max\{\lambda_1,\lambda_2\} \tag{24}$$

where

$$R_{sb}^{p}\triangleq\max\{I(X;Y_1),I(X;Y_2)|P[X=1]=p\}. \tag{25}$$

(The variables $Y_i$ have conditional distribution $p_i$ given $X$.)

Recall that $\mathcal{R}_0$ may be described as the set of pairs $(R_1,R_2)\in\mathbb{R}_+^2$ such that, for all $\boldsymbol{\lambda}\in\mathbb{R}_+^2$,

$$\lambda_1R_1+\lambda_2R_2\leqslant c(\boldsymbol{\lambda})=\sup\{c(\boldsymbol{\lambda},p)|0\leqslant p\leqslant 1\}.$$

Hence the fact that $\mathcal{R}_0$ is the time-sharing region for the skewed binary broadcast channel follows from Proposition 2 with $\boldsymbol{\lambda}=(1,1)$.

We next address the problem of computing the region $\hat{\mathcal{R}}_0$. A key result is the following proposition, which is proved in Appendix E.

*Proposition 3:* For the skewed binary broadcast channel, $\hat{\mathcal{R}}_0$ may be obtained by using only those test channels $(\boldsymbol{U},X)\in\hat{\mathcal{P}}_b$ for which $U_1$, $U_2$, and $U_0$ are binary and

$$X=U_0U_1+(1-U_0)U_2. \tag{26}$$

The fact that it suffices to use test channels $(\boldsymbol{U},X)$ satisfying (26) when computing $\hat{\mathcal{R}}_0$ for the skewed binary broadcast channel greatly simplifies the computation. Indeed, this class of test channels may be parameterized by the three parameters $p=P[U_0=1]$, $\alpha=P[U_1=1]$, and $\beta=P[U_2=0]$. We have found that the line segment $\overline{\boldsymbol{P}_1,\boldsymbol{P}_2}$, where $\boldsymbol{P}_1=(0.2411\cdots,0.1205\cdots)$ and $\boldsymbol{P}_2=(0.1205\cdots,0.2411\cdots)$, is contained in the boundary of $\hat{\mathcal{R}}_0$. The fact that $\boldsymbol{P}_1\in\hat{\mathcal{R}}_0$ follows from using the test channel $(\boldsymbol{U},X)$ corresponding to $p=0.5$ and $\alpha=\beta=0.5-\sqrt{105}/30\cong 0.1584$ in (1)–(3). Then $\boldsymbol{P}_2\in\hat{\mathcal{R}}_0$ by symmetry. The fact that $\overline{\boldsymbol{P}_1\boldsymbol{P}_2}$ is contained in the boundary of $\hat{\mathcal{R}}_0$ will follow from the fact that

$$s^*\triangleq\max\{R_1+R_2|(R_1,R_2)\in\hat{\mathcal{R}}_0\}=0.3616\cdots. \tag{27}$$

That $s^*=0.3616\cdots$ can be shown as follows. Since $\boldsymbol{P}_1\in\hat{\mathcal{R}}_0$, it follows that $s^*\geqslant 0.3616\cdots$. From Proposition 1 (see (3) in particular) it follows that for any $\epsilon>0$ there is a $(\boldsymbol{U},X,\boldsymbol{Y})\in\hat{\mathcal{P}}\mathcal{X}$ such that

$$\begin{aligned}s^*-\epsilon&\leqslant I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)\\&\quad+\min\{I(U_0;Y_1),I(U_0;Y_2)\}\\&\leqslant I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)\\&\quad+\tfrac{1}{2}(I(U_0;Y_1)+I(U_0;Y_2)).\end{aligned} \tag{28}$$

Now using the fact that we need only consider test channels $(\boldsymbol{U},X)$ satisfying (26), we obtain from (28) that

$$s^*\leqslant\sup\{f(\alpha,\beta,p)|0\leqslant\alpha,\beta,p\leqslant 1\}$$

where $f(\alpha,\beta,p)$ is the quantity in (28) expressed in terms of the parameters $\alpha$, $\beta$, and $p$. The function $f(\alpha,\beta,p)$ is a smooth function so that it is straightforward (though tedious) to show that it attains its maximum value of $0.361643\cdots$ when $\alpha=\beta=0.5-\sqrt{105}/30$, $p=0.5$.

It is perhaps surprising that $\hat{\mathcal{R}}_0$ is substantially larger than the time-sharing region for this example, while $\mathcal{R}_0$ is the time-sharing region. In fact, given that $\mathcal{R}_0$ is the time-sharing region, it is easy to see that if any of the variables $U_0$, $U_1$, or $U_2$ is constant, then $(T_1,T_2)$ satisfying (1)–(3) will be in the time-sharing region. Hence it is the interaction of three auxiliary random variables which yields achievable rates outside the time-sharing region. The true capacity region of the skewed binary broadcast channel is unknown.

## IV. Concluding Remarks

It may seem paradoxical that we have found rate regions larger than $\mathcal{R}_0$ (and hence $\mathcal{R}$) by considering only achievable rate triples in $\mathcal{R}$. The crux of the matter is that the true capacity region must have an internal consistency which we have shown $\mathcal{R}$ lacks. Specifically, if $(R_1,R_2,S_1+S_2)$ is an achievable rate triple, then so is $(R_1+S_1,R_2+S_2,0)$. However, as we demonstrate by example, $(R_1R_2,S_1+S_2)\in\mathcal{R}$ does not imply that $(R_1+S_1,R_2+S_2,0)\in\mathcal{R}$. (This is equivalent to the fact that $\hat{\mathcal{R}}_0$ is larger than $\mathcal{R}_0$.) The larger regions $\hat{\mathcal{R}}_0$ and $\hat{\mathcal{R}}$ are obtained simply by enlarging $\mathcal{R}$ to be internally consistent. The achievability of the region $\mathcal{R}$, then, implies the achievability of the region $\hat{\mathcal{R}}$ (and hence the achievability of $\hat{\mathcal{R}}_0$, which is obtained as the projection of $\hat{\mathcal{R}}$ onto the $R_0=0$ plane).

Our observation that $\mathcal{R}$ is not internally consistent depends on our main mathematical results, which give bounds on test channel alphabets necessary to compute $\mathcal{R}_0$ and $\hat{\mathcal{R}}_0$ for binary-input channels. We have not extended these results to arbitrary discrete memoryless broadcast channels. We conjecture, however, that $\mathcal{R}_0$ can be calculated for the channel $(\mathcal{X},p_1,p_2,\mathcal{Y}_1,\mathcal{Y}_2)$ using only test channel input alphabets $\mathcal{U}_i$ satisfying

$$\|U_i\|\leqslant\min(\|\mathcal{X}\|,\|\mathcal{Y}_i\|),\qquad i=1,2$$

(this concurs with van der Meulen's conjectures [12]) and that $\hat{\mathcal{R}}_0$ can be calculated using test channel input alphabets satisfying

$$\|\mathcal{U}_0\|\leqslant\min(\|\mathcal{X}\|,\max(\|\mathcal{Y}_1\|,\|\mathcal{Y}_2\|))$$

and

$$\|\mathcal{U}_i\|\leqslant 1+\|\mathcal{U}_0\|\cdot(\min(\|\mathcal{X}\|,\|\mathcal{Y}_i\|)-1),\qquad i=1,2.$$

Bounding alphabet cardinalities necessary for computing the region of rate triples $\hat{\mathcal{R}}$ (or, equivalently, $\mathcal{R}$ since $\hat{\mathcal{R}}$ is obtained directly from $\mathcal{R}$ and $\hat{\mathcal{R}}_0$) appears to be considerably more difficult.

## ACKNOWLEDGMENT

We wish to thank T. M. Cover and an anonymous referee for several helpful comments and suggestions on earlier versions of this paper. A discussion of Lemma 1 with D. S. Parker led to our discovery of reference [4].

## APPENDIX A
## PROOF OF PROPOSITION 1

We shall prove that $\hat{\mathcal{R}}_0 = \overline{\text{co}}\ \mathcal{C}$. By definition, $\mathcal{R}$ is the closed convex hull of the collection of all triples $(R_1, R_2, R_0)$ in $\mathbb{R}^3_+$ which, for some $(\boldsymbol{U}, X, \boldsymbol{Y}) \in \hat{\mathcal{P}}\mathcal{K}$, satisfy the following:

1a) $R_1 \leqslant I(U_1; Y_1|U_0)$,
1b) $R_2 \leqslant I(U_2; Y_2|U_0)$,
1c) $R_0 \leqslant I(U_0; Y_1|U_1)$,
1d) $R_0 \leqslant I(U_0; Y_2|U_2)$,
1e) $R_1 + R_0 \leqslant I(U_0, U_1; Y_1)$,
1f) $R_2 + R_0 \leqslant I(U_0, U_2; Y_2)$.

$\hat{\mathcal{R}}_0$ is the convex hull of pairs $(T_1, T_2) \in \mathbb{R}^2_+$ such that $(R_1, R_2, R_0 = S_1 + S_2) \in \mathcal{R}$, for some $R_1, R_2, S_1 \geqslant 0$ and $S_2 \geqslant 0$ with $T_1 = R_1 + S_1$ and $T_2 = R_2 + S_2$. Hence, by 1a)–1f), $\hat{\mathcal{R}}_0$ is the closed convex hull of the set of pairs $(T_1, T_2) \in \mathbb{R}^2_+$ such that there exists $S_1 \geqslant 0$, $S_2 \geqslant 0$, and $(\boldsymbol{U}, X, \boldsymbol{Y}) \in \hat{\mathcal{P}}\mathcal{K}$ satisfying the following:

2a) $T_1 \leqslant I(U_1; Y_1|U_0) + S_1$,
2b) $T_2 \leqslant I(U_2; Y_2|U_0) + S_2$,
2c) $S_1 + S_2 \leqslant I(U_0; Y_1|U_1)$,
2d) $S_1 + S_2 \leqslant I(U_0; Y_2|U_2)$,
2e) $T_1 \leqslant I(U_0, U_1; Y_1) - S_2$,
2f) $T_2 \leqslant I(U_0, U_2; Y_2) - S_1$,
2g) $T_1 \geqslant S_1$,
2h) $T_2 \geqslant S_2$.

On the other hand, $\mathcal{C}$ is the collection of pairs $(T_1, T_2) \in \mathbb{R}^2_+$ which, for some $(\boldsymbol{U}', X, \boldsymbol{Y}) \in \hat{\mathcal{P}}\mathcal{K}$, satisfy the following:

3a) $T_1 \leqslant I(U_1', U_0'; Y_1)$,
3b) $T_2 \leqslant I(U_2', U_0'; Y_2)$,
3c) $T_1 + T_2 \leqslant I(U_1'; Y_1|U_0') + I(U_2'; Y_2|U_0') + I(U_0'; Y_1)$,
3d) $T_1 + T_2 \leqslant I(U_1'; Y_1|U_0') + I(U_2'; Y_2|U_0') + I(U_0'; Y_2)$.

Now to prove that $\hat{\mathcal{R}}_0 \subset \overline{\text{co}}\ \mathcal{C}$ we need merely note that if $\boldsymbol{U}' = \boldsymbol{U}$, then 2a)–2h) imply 3a)–3d). (In fact, 2e)⇒3a), 2f)⇒3b), 2b) and 2e)⇒3c), and 2a) and 2f)⇒3d).)

To prove the reverse inclusion, we assume that $(T_1, T_2)$ satisfies 3a)–3d) for some $(\boldsymbol{U}', X, \boldsymbol{Y})$ and show that this implies $(T_1, T_2) \in \hat{\mathcal{R}}_0$ for each of the following three cases.

1) If $T_1 - I(U_1'; Y_1|U_0') \geqslant I(U_0'; Y_2|U_2')$, then substituting into 3d) and applying the basic information identities, we have

$$T_2 \leqslant I(U_2'; Y_2). \tag{A.1}$$

If in 2a)–2h) we let $S_1 = S_2 = 0$, $U_1 = (U_0', U_1')$, $U_2 = U_2'$, and $U_0 \equiv u$ for an arbitrary $u \in \mathcal{U}_0$, then we see that (A.1)⇔2b)⇔2f) and 3a)⇔2a)⇔2c) (2c), 2d), 2g), and 2h) are trivially satisfied). Hence $(T_1, T_2) \in \hat{\mathcal{R}}_0$.

2) If $T_2 - I(U_2'; Y_2|U_0') \geqslant I(U_0'; Y_1|U_1')$, then we can show $(T_1, T_2) \in \hat{\mathcal{R}}_0$ by the same general procedure as in case 1) above.

3) If

$$T_1 - I(U_1'; Y_1|U_0') < I(U_0'; Y_2|U_2') \tag{A.2}$$

and

$$T_2 - I(U_2'; Y_2|U_0') < I(U_0'; Y_1|U_1'), \tag{A.3}$$

then let $\boldsymbol{U} = \boldsymbol{U}'$ and

$$S_i = \max\{0, T_i - I(U_i; Y_i|U_0)\} \tag{A.4}$$

for $i = 1, 2$. This choice of $S_i$ guarantees that 2a), 2b), 2g), and 2h) are satisfied.

If $S_1 > 0$ and $S_2 > 0$, then 2c)–2f) are verified as follows. Equation (A.4) and 3c) imply $S_1 + S_2 \leqslant I(U_0; Y_1)$, and the independence of $U_0$ and $U_1$ implies $I(U_0; Y_1) \leqslant I(U_0; Y_1|U_1)$, which establishes 2c). Similarly, (A.4) and 3d) imply 2d). Finally, in view of (A.4), 3c) implies 2e), and 3d) implies 2f).

If $S_1 = 0$, then 3b) implies 2f) and 2c)–2e) are verified as follows. If $S_2 = 0$, then 2c) and 2d) are trivially true, and 3a) implies 2e). If $S_2 > 0$, (A.4) and (A.3) imply 2c), (A.4) and 3b) imply $S_1 + S_2 \leqslant I(U_0; Y_2)$, which in turn implies 2d) because of the independence of $U_0$ and $U_2$, and (A.4) and 3c) imply 2e).

If $S_2 = 0$, 2c)–2f) are established in the same manner as in the preceding paragraph, since the case $S_2 = 0$ is just the dual of the case $S_1 = 0$.

Since 1)–3) cover all possible situations, we have shown that if $(T_1, T_2)$ satisfies 3a)–3d), then $(T_1, T_2) \in \hat{\mathcal{R}}_0$. Hence $\hat{\mathcal{R}}_0 \supset \overline{\text{co}}\ \mathcal{C}$.

## APPENDIX B
## PROOF OF LEMMA 1

Notice that if $x_i = y_i$, then the $i$th term of (16) and (17) has no effect on the inequality. Hence we can assume $x_j \neq y_j$, for $0 \leqslant j \leqslant N$, without loss of generality. For a given concave function $\Lambda$, define a slope function $s$ by

$$s(x, y) = (x - y)^{-1}[\Lambda(x) - \Lambda(y)] \tag{B.1}$$

for $x \neq y$. The concavity of $\Lambda$ guarantees that

$$s(x_{k-1}, y_{k-1}) \geqslant s(x_k, y_k) \tag{B.2}$$

for $1 \leqslant k \leqslant N$. It follows from (B.2) and (16) that

$$\sum_{j=k}^{N} \xi_j (x_j - y_j)[s(x_{k-1}, y_{k-1}) - s(x_k, y_k)] \geqslant 0 \tag{B.3}$$

for $1 \leqslant k \leqslant N$. In fact, if we define $x_{-1} = x_0$ and $y_{-1} = y_0$, then (B.3) holds (with equality) for $k = 0$ as well. Thus we have

$$\sum_{k=0}^{N} \sum_{j=k}^{N} \xi_j (x_j - y_j)[s(x_{k-1}, y_{k-1}) - s(x_k, y_k)] \geqslant 0$$

which is equivalent to

$$\sum_{k=-1}^{N-1} \sum_{j=k+1}^{N} \xi_j (x_j - y_j) s(x_k, y_k) - \sum_{k=0}^{N} \sum_{j=k}^{N} \xi_j (x_j - y_j) s(x_k, y_k) \geqslant 0. \tag{B.4}$$

Because equality holds in (16) for $k = 0$, the $k = -1$ term in the left sum of (B.4) is zero, so (B.4) is equivalent to

$$-\sum_{k=0}^{N} \xi_k (x_k - y_k) s(x_k, y_k) \geqslant 0. \tag{B.5}$$

Using (B.1) in (B.5) we obtain (17).

## Appendix C
## Proof of Theorem 3

Let $\mathcal{S}$ be the set of $\boldsymbol{S}\in\mathbb{R}_+^4$ such that, for some $(\boldsymbol{U},X,\boldsymbol{Y})\in\hat{\mathcal{P}}\mathcal{K}$,

$$S_1\leqslant I(U_0,U_1;Y_1)$$
$$S_2\leqslant I(U_0,U_2;Y_2)$$
$$S_3\leqslant I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)+I(U_0;Y_1)$$
$$S_4\leqslant I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)+I(U_0;Y_2).$$

Let $T$ be the linear map from $\mathbb{R}^2$ to $\mathbb{R}^4$ defined by

$$T(x_1,x_2)=(x_1,x_2,x_1+x_2,x_1+x_2).$$

Referring to the definition of $\mathcal{C}$ in Section I, we see that $\mathcal{C}=T^{-1}\mathcal{S}$. It is easy to check directly that since $T$ is linear (continuous) and one-to-one,

$$\hat{\mathcal{R}}_0=\overline{\text{co}}\ \mathcal{C}=\overline{\text{co}}\ T^{-1}\mathcal{S}=T^{-1}\ \overline{\text{co}}\ \mathcal{S}.$$

Hence to compute $\hat{\mathcal{R}}_0$ it suffices to compute $\overline{\text{co}}\ \mathcal{S}$. To this end define for each $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\lambda_3)\in\mathbb{R}_+^3$ and each $(\boldsymbol{U},X)\in\hat{\mathcal{P}}$

$$\begin{aligned}\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)=&\lambda_1 I(U_0,U_1;Y_1)+\lambda_2 I(U_0,U_2;Y_2)\\&+\lambda_3[I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)+I(U_0;Y_1)]\\&+I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)+I(U_0;Y_2)\end{aligned}$$

where $\boldsymbol{Y}$ is an output of $\mathcal{K}$ corresponding to $(\boldsymbol{U},X)$. For each $\boldsymbol{\lambda}\in\mathbb{R}_+^3$ let

$$\hat{c}(\boldsymbol{\lambda})=\sup\{\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)|(\boldsymbol{U},X)\in\hat{\mathcal{P}}\}. \tag{C.1}$$

Using the characterization of convex hulls as in Section II, $\overline{\text{co}}\ \mathcal{S}$ is the collection of points in $\mathbb{R}_+^4$ dominated by the family of hyperplanes

$$\{\lambda_1x_1+\lambda_2x_2+\lambda_3x_3+x_4=\hat{c}(\boldsymbol{\lambda})|\boldsymbol{\lambda}\in\mathbb{R}_+^3\}.$$

Thus to compute $\overline{\text{co}}\ \mathcal{S}$ (and hence $\hat{\mathcal{R}}_0$) one need only compute $\hat{c}(\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda}\in\mathbb{R}_+^3$. Thus Theorem 3 will follow if we demonstrate that $\hat{c}(\boldsymbol{\lambda})$ remains unchanged when $\hat{\mathcal{P}}$ is replaced by $\hat{\mathcal{P}}_b$ in (C.1).

Basically our approach now will be to first consider maximizing $\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)$ over all $(\boldsymbol{U},X)\in\hat{\mathcal{P}}$ with $(U_0,X)$ having a specified distribution. We will show that if the cardinalities of the ranges of $U_1$ and $U_2$ are restricted to be at most one plus the cardinality of the range of $U_0$, then the same maximum results. Removing the constraint on the distribution of $(U_0,X)$, we will then show that $U_0$ can be restricted to be binary.

Let us rewrite $\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)$. Define, for $(\boldsymbol{U},X)\in\hat{\mathcal{P}}$,

$$\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U},X)=(\lambda_1+\lambda_3+1)I(U_1;Y_1|U_0)+(\lambda_2+\lambda_3+1)I(U_2;Y_2|U_0)$$
$$\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{U},X)=(-\lambda_1-\lambda_3)H(Y_1|U_0)+(-\lambda_2-1)H(Y_2|U_0)$$
$$\hat{\Phi}_{\boldsymbol{\lambda}}^{(3)}(\boldsymbol{U},X)=(\lambda_1+\lambda_3)H(Y_1)+(\lambda_2+1)H(Y_2)$$

so that

$$\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)=\sum_{i=1}^{3}\hat{\Phi}_{\boldsymbol{\lambda}}^{(i)}(\boldsymbol{U},X). \tag{C.2}$$

Define the set $\hat{\mathcal{P}}'\subset\hat{\mathcal{P}}$ to be the collection of test channels $((U_1,U_2,U_0),X)\in\hat{\mathcal{P}}$ such that $X$ is a deterministic function of $(U_1,U_2,U_0)$ and $U_1$ and $U_2$ are each distributed on $\{0,1,\cdots,r\}$, where $r$ is the cardinality of the range set of $U_0$.

Let $\mathcal{U}_0$ be any finite set, let $p_0(u)$ be a probability distribution on $\mathcal{U}_0$, and let $q_0(1|u)$ be such that $0\leqslant q_0(1|u)\leqslant 1$, for $u\in\mathcal{U}_0$. If $U_0$ and $X$ are random variables such that $X$ is binary, $P[U_0=u]=p_0(u)$ and $P[X=1|U_0=u]=q_0(1|u)$, we shall write $(U_0,X)\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)$.

Finally, define for $0\leqslant p\leqslant 1$

$$\Psi_{\boldsymbol{\lambda}}^{(1)}(p)=c((\lambda_1+\lambda_3+1,\lambda_2+\lambda_3+1),p). \tag{C.3}$$

(See the remark after Theorem 2 for the definition of $c(\cdot,p)$, which is continuous in $p$.) The next lemma deals with $\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}$.

*Lemma 3:* The following are equal:

$$a=\sup\{\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U},X)|(\boldsymbol{U},X)\in\hat{\mathcal{P}}',(U_0,X)\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)\} \tag{C.4}$$
$$b=\sup\{\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U},X)|(\boldsymbol{U},X)\in\hat{\mathcal{P}},(U_0,X)\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)\}$$
$$c=\sum_{u\in\mathcal{U}_0}\Psi_{\boldsymbol{\lambda}}^{(1)}(q_0(1|u))p_0(u).$$

Furthermore, the supremum in (C.4) is actually a maximum.

*Proof:* We shall prove that $a\leqslant b\leqslant c\leqslant a$. Since $\hat{\mathcal{P}}'\subset\hat{\mathcal{P}}$, $a\leqslant b$ is immediate. To see that $b\leqslant c$, note that for $((U_1,U_2,U_0),X)\in\hat{\mathcal{P}}$ with $U_0\sim\boldsymbol{p}_0$,

$$\begin{aligned}\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U},X)=\sum_{u\in\mathcal{U}_0}p_0(u)\{&(\lambda_1+\lambda_2+1)I(U_1;Y_1|U_0=u)\\&+(\lambda_2+\lambda_3+1)I(U_2;Y_2|U_0=u)\}.\end{aligned}$$

The terms of this sum are the same as the information quantities of Section II. (Notice that $U_1$ and $U_2$ are conditionally independent given $U_0$ since $U_1$, $U_2$, and $U_0$ are mutually independent.) Hence the sum may be bounded term by term to yield $b\leqslant c$. To prove the final inequality $c\leqslant a$, we will construct $(\boldsymbol{U}',X')\in\hat{\mathcal{P}}'$ with $(U_0',X')\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)$ and $\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U},X')=c$, which will also demonstrate that the supremum in (C.4) is actually a maximum.

The construction proceeds as follows. Let $r=\|\mathcal{U}_0\|$. For each $u\in\mathcal{U}_0$, there is a test channel $((V_1^u,V_2^u),X^u)\in\mathcal{P}_b$ such that (see (22)) $P[X^u=1]=q_0(1|u)$ and

$$\begin{aligned}\Psi_{\boldsymbol{\lambda}}^{(1)}(q_0(1|u))&=\Phi_{(\lambda_1+\lambda_3+1,\lambda_2+\lambda_3+1)}(\boldsymbol{V}^u,X^u)\\&=(\lambda_1+\lambda_3+1)I(V_1^u,Y_1^u)+(\lambda_2+\lambda_3+1)I(V_2^u,Y_2^u)\end{aligned} \tag{C.5}$$

where $\boldsymbol{Y}^u$ is an output of the channel corresponding to input $(\boldsymbol{V}^u,X^u)$. For $i=1,2$, let $p_i^u\triangleq P[V_i^u=1]$, let $\alpha_i^{(1)}\leqslant\alpha_i^{(2)}\leqslant\cdots\leqslant\alpha_i^{(r)}$ be an ordering of the set $\{p_i^u|u\in\mathcal{U}_0\}$, and let $\alpha_i^{(0)}=0$ and $\alpha_i^{(r+1)}=1$. Let $\sigma_i(u)$ denote the rank of $p_i^u$ in $\{\alpha_i^{(j)}\}$; that is,

$$\sigma_i(u)=\min\{k|p_i^u\leqslant\alpha_i^{(k)}\}.$$

Define $U_0'$, $U_1'$, and $U_2'$ to be mutually independent random variables with $U_0'\sim\boldsymbol{p}_0$ (i.e., $P[U_0'=u]=p_0(u)$, $u\in\mathcal{U}_0$) and

$$P[U_i'=j]=\alpha_i^{(j+1)}-\alpha_i^{(j)},\qquad\text{for } 0\leqslant j\leqslant r,\ i=1,2.$$

To complete the construction we must choose $X'$ as a binary function of $(U_0',U_1',U_2')$ so that $(U_0',X')\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)$ and $\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U}',X')=c$. The trick is to make the conditional distribution of $(U_1',U_2',X')$ given $U_0'=u$ imitate in an information-theoretic sense the distribution of $(V_1^u,V_2^u,X^u)$.

For $u\in\mathcal{U}_0$, $i=1$ or 2, and $0\leqslant j\leqslant r$, let $\chi_i^u(j)=\chi_{\{j<\sigma_i(u)\}}$. Then

$$\begin{aligned}P[\chi_i^u(U_i')=1|U_0=u]&=P[\chi_i^u(U_i')=1]\\&=\sum_{j=0}^{\sigma_i(u)-1}\alpha_i^{(j+1)}-\alpha_i^{(j)}=p_i^u=P[V_i^u=1].\end{aligned}$$

That is, $\chi_i^u(U_i')\sim V_i^u$. Recall by the definition of $\mathcal{P}_b$ that either $X^u\equiv V_1^u\vee V_2^u$ or $X^u\equiv V_1^u\wedge V_2^u$. When $U_0=u$, let $X'=\chi_1^u(U_1')\wedge\chi_2^u(U_2')$ (respectively, $\chi_1^u(U_1')\vee\chi_1^u(U_2')$) if $X^u\equiv V_1^u\wedge V_2^u$ (respectively, $V_1^u\vee V_2^u$). Then conditioned on $U_0=u$,

$$(V_1^u,V_2^u,X^u)\sim(\chi_1^u(U_1'),\chi_2^u(U_2'),X'). \tag{C.6}$$

One consequence of (C.6) is that

$$P[X'=1|U_0'=u]=P[X^u=1]=q_0(1|u)$$

which implies that $(U_0',X')\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)$. Another consequence of (28) is that if $(Y_1',Y_2')$ is an output of the broadcast channel corresponding to input $(\boldsymbol{U}',X')$, then, for $i=1,2$ and all $u$,

$$\begin{aligned} I(V_i^u,Y_i^u)&=I(\chi_i^u(U_i'),Y_i'|U_0'=u)\\ &=I(U_i';Y_i'|U_0'=u). \end{aligned} \tag{C.7}$$

The second equality follows from the fact that $U_i'$ and $Y_i'$ are conditionally independent given $U_0'=u$ and $\chi_i^u(U_i')$. Substitution of (C.7) into (C.5), multiplying by $p_0(u)$, and summing over $u$ yields that $\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U}',X')=c$, completing the proof of Lemma 3.

Note that if $((U_1,U_2,U_0),X)\in\hat{\mathcal{P}}$ and $(U_0,X)\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)$, then $\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{U},X)$ may be expressed in terms of $(\boldsymbol{p}_0,\boldsymbol{q}_0)$. Indeed, by (10),

$$\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{U},X)=\sum_u p_0(u)\Psi_{\boldsymbol{\lambda}}^{(2)}(q_0(1|u)) \tag{C.8}$$

where

$$\Psi_{\boldsymbol{\lambda}}^{(2)}(\alpha)\triangleq(-\lambda_1-\lambda_3)\Lambda_1(\alpha)+(-\lambda_2-1)\Lambda_2(\alpha).$$

Note that $\Psi_{\boldsymbol{\lambda}}^{(2)}(\alpha)$ is continuous in $\alpha$. Since the entropies of the channel outputs depend only on the input probability distribution, $\Phi_{\boldsymbol{\lambda}}^{(3)}(\boldsymbol{U},X)$ is a function of $p=P[X=1]=\sum_u q_0(1|u)p_0(u)$. We shall denote these facts by writing $\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{p}_0,\boldsymbol{q}_0)$ and $\hat{\Phi}_{\boldsymbol{\lambda}}^{(3)}(p)$ for $\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{U},X)$ and $\hat{\Phi}_{\boldsymbol{\lambda}}^{(3)}(\boldsymbol{U},X)$, respectively.

We now reconsider $\hat{c}(\boldsymbol{\lambda})$. Observe that

$$\hat{c}(\boldsymbol{\lambda})=\sup\{\hat{c}(\boldsymbol{\lambda},p)|0\leqslant p\leqslant 1\}$$

where

$$\hat{c}(\boldsymbol{\lambda},p)\triangleq\sup\{\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)|(\boldsymbol{U},X)\in\hat{\mathcal{P}},P[X=1]=p\}.$$

For $0\leqslant p\leqslant 1$, let $F_p$ denote the collection of distributions $(\boldsymbol{p}_0,\boldsymbol{q}_0)$ such that

$$p=\sum_u p_0(u)q_0(1|u). \tag{C.9}$$

Then using (C.3), Lemma 3, and (C.8),

$$\begin{aligned} \hat{c}(\boldsymbol{\lambda},p)&=\sup_{\boldsymbol{p}_0,\boldsymbol{q}_0\in F_p}\left[\sup\{\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U},X)|(U,X)\in\hat{\mathcal{P}},(U_0,X)\sim(\boldsymbol{p}_0,\boldsymbol{q}_0)\}\right.\\ &\quad\left.+\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{p}_0,q_0)\right]+\hat{\Phi}_{\boldsymbol{\lambda}}^{(3)}(p)\\ &=\sup_{\boldsymbol{p}_0,\boldsymbol{q}_0\in F_p}\left[\sum_u p_0(u)\left(\Psi_{\boldsymbol{\lambda}}^{(1)}(q_0(1|u))+\Psi_{\boldsymbol{\lambda}}^{(2)}(q_0(1|u))\right)\right]\\ &\quad+\hat{\Phi}_{\boldsymbol{\lambda}}^{(3)}(p). \end{aligned} \tag{C.10}$$

Let

$$\Psi_{\boldsymbol{\lambda}}(\alpha)=\Psi_{\boldsymbol{\lambda}}^{(1)}(\alpha)+\Psi_{\boldsymbol{\lambda}}^{(2)}(\alpha),\qquad 0\leqslant\alpha\leqslant 1.$$

Since $\Psi_{\boldsymbol{\lambda}}(\alpha)$ is continuous in $\alpha$, the mapping $t\to(t,\Psi_{\boldsymbol{\lambda}}(t))$ is also continuous and defines a compact arc $\Gamma$ in $\mathbb{R}^2$ as $t$ ranges over the compact set $[0,1]$. Given any distribution $(\boldsymbol{p}_0,\boldsymbol{q}_0)$, we have

$$\left(\sum_u p_0(u)q_0(1|u),\sum_u p_0(u)\Psi_{\boldsymbol{\lambda}}(q_0(1|u))\right)\in\text{co }\Gamma,$$

and any point of co $\Gamma$ may be expressed in this way. Therefore, the supremum, let us say $M$, of the bracketed term in (C.10) over $(\boldsymbol{p}_0,\boldsymbol{q}_0)\in F_p$ is the supremum of $y$ such that $(p,y)\in\text{co }\Gamma$. Since co $\Gamma$ is closed, $(p,M)\in\text{co }\Gamma$. Clearly any point of co $\Gamma$ can be expressed as the convex combination of two points of $\Gamma$ so that there exists $p_0^*(0)+p_0^*(1)=1$ and $0\leqslant q^*(1|0),\ q^*(1|1)\leqslant 1$ such that

$$(p,M)=\left(\sum_{u=0}^{1}p_0^*(u)q_0^*(1|u),\sum_{u=0}^{1}p_0^*(u)\Psi_{\boldsymbol{\lambda}}(q_0^*(1|u))\right).$$

It follows that (using (C.10) and Lemma 3)

$$\begin{aligned} \hat{c}(\boldsymbol{\lambda},p)&=\sum_{u=0}^{1}p_0^*(u)\Psi_{\boldsymbol{\lambda}}^{(1)}(q_0^*(1|u))+\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{p}_0^*,\boldsymbol{q}_0^*)+\hat{\Phi}_{\boldsymbol{\lambda}}^{(3)}(p)\\ &=\max\left\{\hat{\Phi}_{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{U},X)|(\boldsymbol{U},X)\in\hat{\mathcal{P}}',(U_0X)\sim(\boldsymbol{p}_0^*,\boldsymbol{q}_0^*)\right\}\\ &\quad+\hat{\Phi}_{\boldsymbol{\lambda}}^{(2)}(\boldsymbol{p}_0^*,\boldsymbol{q}_0^*)+\hat{\Phi}_{\boldsymbol{\lambda}}^{(3)}(p). \end{aligned}$$

Thus

$$\hat{c}(\boldsymbol{\lambda},p)=\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U}^*,X^*)$$

for some $(\boldsymbol{U}^*,X^*)\in\hat{\mathcal{P}}'$ with $(U_0^*,X^*)\sim(\boldsymbol{p}_0^*,\boldsymbol{q}_0^*)$. The fact that $U_0^*$ is binary and $(\boldsymbol{U}^*,X^*)\in\hat{\mathcal{P}}'$ implies that $(\boldsymbol{U}^*,X^*)\in\hat{\mathcal{P}}_b$. Hence we have

$$\begin{aligned} \hat{c}(\boldsymbol{\lambda})&=\sup\{\hat{c}(\boldsymbol{\lambda},p)|0\leqslant p\leqslant 1\}\\ &=\sup\{\max\{\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)|(\boldsymbol{U},X)\in\hat{\mathcal{P}}_b,P[X=1]=p\}|0\leqslant p\leqslant 1\}\\ &=\sup\{\hat{\Phi}_{\boldsymbol{\lambda}}(\boldsymbol{U},X)|(\boldsymbol{U},X)\in\hat{\mathcal{P}}_b\} \end{aligned}$$

completing the proof of Theorem 3.

## Appendix D
## Proof of Proposition 2

We will prove Proposition 2 of Section III. Using (4) and (23), we note here for convenience that

$$c(\boldsymbol{\lambda},p)=\max\{\lambda_1I(U_1;Y_1)+\lambda_2I(U_2;Y_2)|(\boldsymbol{U},X,\boldsymbol{Y})\in\mathcal{P}_b\mathcal{H},\ P[X=1]=p\}.$$

The fact that $c(\boldsymbol{\lambda},p)\geqslant R_{sb}^p\cdot\max(\lambda_1,\lambda_2)$ may be proven by considering test channels $(\boldsymbol{U},X)\in\mathcal{P}_b$ such that either $U_1\equiv X$ or $U_2\equiv X$ and $P[X=1]=p$. To prove the reverse inequality, note that it suffices to consider the case $\boldsymbol{\lambda}=(1,1)$ since

$$c(\boldsymbol{\lambda},p)\leqslant c((1,1),p)\max(\lambda_1,\lambda_2),\qquad 0\leqslant p\leqslant 1,\ \boldsymbol{\lambda}\in\mathbb{R}_+^2.$$

Hence all we need to show is that if $(\boldsymbol{U},X,\boldsymbol{Y})\in\mathcal{P}_b\mathcal{H}$ and $P[X=1]=p$, then

$$I(U_1;Y_1)+I(U_2;Y_2)\leqslant\max(I(X;Y_1),I(X,Y_2)). \tag{D.1}$$

Recall that $(\boldsymbol{U},X)\in\mathcal{P}_b$ implies that either $X\equiv U_1\wedge U_2$ or $X\equiv U_1\vee U_2$. Suppose that $P[U_1=1]=x$. Then in order that $P[X=1]=p$, it is necessary that $p\leqslant x\leqslant 1$ and $P[U_2=1]=px^{-1}$. Note that when $x=p$, $I(U_1;Y_1)=I(X;Y_1)$, and when $x=1$, $I(U_2;Y_2)=I(X;Y_2)$. Thus if we define

$$\begin{aligned} f(p,x)&=I(U_1;Y_1)+I(U_2;Y_2)\\ &=h\left(\frac{p}{2}\right)-xh\left(\frac{p}{2x}\right)+h\left(\frac{1+p}{2}\right)-1+\frac{p}{x}\left[1-h\left(\frac{1+x}{2}\right)\right], \end{aligned}$$

(D.1) may be rewritten as

$$f(p,x)\leqslant\max[f(p,p),f(p,1)],\qquad\text{for }0\leqslant p\leqslant x\leqslant 1. \tag{D.2}$$

We will now prove that

$$f(p,x)\leqslant f(p,p),\qquad\text{for }0\leqslant p\leqslant\max\left(x,\tfrac{1}{2}\right),\ x\leqslant 1 \tag{D.3}$$

and

$$f(p,x)\leqslant f(p,1),\qquad\text{for }\tfrac{1}{2}\leqslant p\leqslant x\leqslant 1. \tag{D.4}$$

Clearly (D.3) and (D.4) imply (D.2) so that establishing (D.3) and (D.4) will complete the proof of the proposition.

To prove (D.3) we will note that the function $g_x(p) \triangleq f(p,p) - f(p,x)$, as a function of $p$, is concave on $[0, \min(x, \frac{1}{2})]$ and satisfies $g_x(0)=0$ and $g_x(\min(x,\frac{1}{2})) \geqslant 0$. It is easy to check that $g_x(0)=g_x(x)=0$. It can be seen from the graph of $g_x(\frac{1}{2})$ as a function of $x$ that $g_x(\frac{1}{2})>0$, for $\frac{1}{2}<x<1$ (we were unable to prove this analytically). Finally, $g_x(p)$ is concave in $p$ since its second derivative

$$g_x''(p) = -\left[x(1-2p+p^2)+p^2\right]\left[p(1-p^2)(2x-p)\right]^{-1}$$

is negative for $p \in [0, \min(x, \frac{1}{2})]$.

To prove (D.4) we note that, for $x > \frac{1}{2}$, the function $t_x(p) \triangleq f(p,1) - f(p,x)$, as a function of $p$, is concave on $[\frac{1}{2}, x]$ and nonnegative at the endpoints. It turns out that $t_x(\frac{1}{2}) = g_x(\frac{1}{2})$, which as noted in the preceding paragraph, we have found to be positive. One finds that

$$t_x(x) = h\left(\frac{1+x}{2}\right) - h\left(\frac{x}{2}\right) + 2x - 1$$

which satisfies $t_x(\frac{1}{2}) = t_x(1) = 0$ and $(t_x(x))'' \leqslant 0$, for $\frac{1}{2} \leqslant x \leqslant 1$, implying that $t_x(x) \geqslant 0$, for $x \in [\frac{1}{2}, 1]$. Finally, the function $t_x(p)$ is concave in $p$ for $p \in [\frac{1}{2}, x]$ since

$$t_x''(p) = -p^2(1-x)\left[(2p-p^2)(2px-p^2)\right]^{-1}$$

which is negative on that interval.

## Appendix E
## Proof of Proposition 3

The procedure used in Appendix C to prove Theorem 3—that of fixing the distribution of $(U_0, X)$ and then "reducing" the class of variables $(U_1, U_2)$—may be adapted to prove the present result. Indeed, note that (26) implies that $X = U_1$ or $X = U_2$, depending on whether $U_0 = 1$ or $U_0 = 0$. That is, formally, given $U_0 = u$, the variables $U_1$ and $U_2$ mimic the random variables which, as the proposition shows, are sufficient to attain

$$\max \{\lambda_1 I(U_1; Y_1 | U_0 = u) + \lambda_2 I(U_2; Y_2 | U_0 = u) | P[X = 1 | U_0 = u] = p\}.$$

A specific approach to adapting the proof of Theorem 3 to prove that $\hat{\mathcal{R}}_0$ may be calculated using only test channels $(U, X)$ satisfying (26) will now be outlined. Define $\hat{\mathcal{P}}'_{sb}$ to be the collection of $(U,X) \in \hat{\mathcal{P}}'$ such that $\|\mathcal{U}_1\| + \|\mathcal{U}_2\| \leqslant \|\mathcal{U}_0\| + 2$ (where $\mathcal{U}_i$ is the range of $U_i$), and for each $u \in \mathcal{U}_0$, either $I(U_1; Y_1 | U_0 = u) = 0$ or $I(U_2; Y_2 | U_0 = u) = 0$ (i.e., given $U_0 = u$, either $U_1$ or $U_2$ is independent of $X$). We will indicate a proof of Lemma 3 when $\hat{\mathcal{P}}'$ is replaced by $\hat{\mathcal{P}}'_{sb}$. This being done, the remaining portion of the proof of Theorem 3, which shows that $U_0$ can be chosen to be binary, still applies. The special form of $(U,X) \in \hat{\mathcal{P}}'_{sb}$ then insures that $\hat{\mathcal{R}}_0$ can be chosen using only $(U,X)$ satisfying (26).

When $\hat{\mathcal{P}}'$ is replaced by $\hat{\mathcal{P}}'_{sb}$ in Lemma 3, the inequalities $a \leqslant b \leqslant c$ are proven as before. However, the proof of $c \leqslant a$ must be modified. The essential change is that the test channels $(V_1^u, V_2^u, X^u)$ should now be chosen so that either $V_1^u \equiv X^u$ or $V_2^u \equiv X^u$. In fact, by the proposition, this can be done so that equality still holds in (C.5) (see (C.3)). The construction of the variables $(U', X')$, starting with the $(V^u, X^u)$ chosen here, can then be modified so that $(U', X') \in \hat{\mathcal{P}}'_{sb}$ replaces the condition $(U', X') \in \hat{\mathcal{P}}'$. The construction, as before, implies that $c \leqslant a$, completing the proof of Lemma 3 with $\hat{\mathcal{P}}'$ replaced by $\hat{\mathcal{P}}_b$.

## References


[1] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629–637, 1975.

[2] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[3] ——, "An achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 399–404, July 1975.

[4] L. Fuchs, "A new proof of an inequality of Hardy–Littlewood–Pólya," *Matematisk Tidsskrift B*, pp. 53–54, 1947.

[5] R. G. Gallager, "Capacity and coding for degraded broadcast channels" (in Russian), *Probl. Peredach. Inform.*, vol. 10, no. 3, pp. 3–14, 1974; available in English in *Problems of Information Transmission*, Jan. 1976.

[6] S. I. Gelfand, "Capacity of one broadcast channel," (in Russian), *Probl. Peredach. Inform.*, vol. 13, no. 3, pp. 106–108, July–September 1977. English translation available in *Problems of Information Transmission*, January 1978.

[7] S. I. Gelfand and M. S. Pinsker, unpublished manuscript, 1977.

[8] B. E. Hajek and M. B. Pursley, "Representation theory for a class of multivariate distributions with applications to broadcast channels," part II of "On the evaluation of achievable rate regions for broadcast channels," Univ. Illinois, Urbana, Coordinated Science Lab. Rep. R-786, Aug. 1977.

[9] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed. London: Cambridge Univ., 1952.

[10] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 60–64, Jan. 1977.

[11] K. Marton, "An achievable rate region for the broadcast channel problem," submitted to *IEEE Trans. Inform. Theory*.

[12] E. C. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 180–190, Mar. 1975.

[13] ——, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 1–37, Jan. 1977.

[14] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 772–777, Nov. 1973.