Optimal Encoding for Discrete Degraded Broadcast Channels

Bike Xie, Thomas A. Courtade, *Member, IEEE*, and Richard D. Wesel, *Senior Member, IEEE*

***Abstract*—Consider a memoryless degraded broadcast channel (DBC) in which the channel output is a single-letter function of the channel input and the channel noise. As examples, for the Gaussian broadcast channel (BC), this single-letter function is real scalar addition and for the binary-symmetric BC, this single-letter function is modulo-two addition. This paper identifies several classes of discrete memoryless DBCs for which a relatively simple encoding scheme, which we call natural encoding, achieves capacity. Natural encoding (NE) combines symbols from independent codebooks (one for each receiver) using the same single-letter function that adds distortion to the channel. The alphabet size of each NE codebook is bounded by that of the channel input. This paper also defines the input-symmetric DBC, introduces permutation encoding for the input-symmetric DBC, and proves its optimality. Because it is a special case of permutation encoding, NE is capacity achieving for the two-receiver group-operation DBC. Combining the broadcast Z channel and group-operation DBC results yields a proof that NE is also optimal for the discrete multiplication DBC. Along the way, the paper also provides explicit parametric expressions for the two-receiver binary-symmetric DBC and broadcast Z channel.**

***Index Terms*—Broadcast Z channel, degraded broadcast channel (DBC), discrete multiplication (DM) degraded broadcast channel, group-operation degraded broadcast channel, input-symmetric (IS) degraded broadcast channel, natural encoding.**

I. Introduction

A. Background

NEARLY four decades ago, Cover [1], Bergmans [2], and Gallager [3] established the capacity region for degraded broadcast channels (DBCs). A common optimal transmission strategy to achieve the capacity region boundary for DBCs is the joint encoding scheme presented in [1] and [2]. Specifically, the information intended for the receiver with the most degraded channel is encoded to produce a first codeword. Conditioned on that first codeword, a codebook is selected for the receiver with the second most degraded channel, and so forth.

There is at least one independent-encoding scheme (in which the codebook for each user is independent of the messages intended for other users) that can achieve the capacity of any DBC [4]. This scheme essentially embeds all symbols from all the needed codebooks for the less-degraded receiver(s) into a single super-symbol (but perhaps with a large alphabet). Then, a single-letter function uses the input symbol from the more degraded receiver to extract the needed symbol from the super symbol provided by the less-degraded receiver. Appendix A describes this encoding scheme in detail.

Cover [5] introduced an independent-encoding scheme for two-receiver broadcast channels (BCs). When applied to two-receiver DBCs, this scheme independently encodes receivers' messages, and then combines these resulting codewords by applying a single-letter function. This scheme does not specify what codebooks to use or what single-letter function to use. It is a general independent-encoding approach, which includes the independent-encoding scheme described in Appendix A.

Consider DBCs in which the received signal of each component channel can be modeled as a single-letter function of the channel input and the channel noise. A simple encoding scheme that is optimal for some of those DBCs is an independent-encoding approach in which symbols from independent codebooks, each with the same alphabet as the channel input, are combined using the same single-letter function that adds distortion to the channel.

We refer to this encoding scheme as the natural encoding (NE) scheme. As an example, the NE scheme for a two-receiver Gaussian BC has as each transmitted symbol the real scalar addition of two real symbols from independent codebooks. The NE scheme is known to achieve the boundary of the capacity region for several BCs including Gaussian BCs [6], binary-symmetric (BS) BCs [2], [7]–[9], discrete additive DBCs [10], and two-receiver broadcast Z channels [11], [12].

In proving the optimality of NE schemes for Gaussian BCs and BS BCs, Shannon's entropy power inequality (EPI) [13] and "Mrs. Gerber's Lemma" [14], respectively, play the same significant role. Shannon's EPI gives a lower bound on the differential entropy of the sum of independent random variables. Bergmans' remarkable paper [6] applies the EPI to establish a converse showing the optimality of the scheme given by [1] and [2] (the NE scheme) for Gaussian BCs.

Similarly, "Mrs. Gerber's Lemma" provides a lower bound on the entropy of a sequence of BS channel outputs. Wyner and Ziv obtained "Mrs. Gerber's Lemma" and applied it to establish a converse showing that the NE scheme for BS BCs suggested

Manuscript received May 08, 2009; revised October 23, 2011; accepted May 25, 2012. Date of publication January 04, 2013; date of current version February 12, 2013. This work was supported by the Defence Advanced Research Project Agency SPAWAR Systems Center, San Diego, CA, USA, under Grant N66001-02-1-8938.

B. Xie is with Marvell Semiconductor Inc., Santa Clara, CA 95054 USA (e-mail: bike@marvell.com).

T. A. Courtade is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: courtade@stanford.edu).

R. D. Wesel is with the Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA 90095 USA (e-mail: wesel@ee.ucla.edu).

Communicated by E. Erkip, Associate Editor for Shannon Theory.

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

Digital Object Identifier 10.1109/TIT.2012.2237095

by Cover [1] and Bergmans [2] achieves the boundary of the capacity region [7].

Witsenhausen and Wyner made two seminal contributions in [8] and [9]: the notion of minimizing one entropy under the constraint that another related entropy is fixed, called the conditional entropy bound, and the use of input symmetry as a way of solving an entire class of channels with a single unifying approach. Witsenhausen and Wyner applied the first idea to establish an outer bound of the capacity region for DBCs [9]. For BS BCs, this outer bound coincides with the capacity region, which proves again that the NE scheme for BS BCs is capacity achieving.

Later, Benzel [10] applied the conditional entropy bound to prove that the capacity regions for discrete additive degraded interference channels and the corresponding discrete additive DBC are the same, which means that NE is capacity achieving for discrete additive DBCs. Recently, Liu and Ulukus [15], [16] extended Benzel's results to include the larger class of discrete degraded interference channels (DDICs). For these DDICs, Liu and Ulukus introduced a capacity-achieving independent encoding scheme for the corresponding DBCs as long as the transmitted signal for the DBC can be appropriately defined.

B. Contributions

The main contributions of this paper are the following.

1) Establishing that NE is capacity achieving for multireceiver broadcast Z channels
2) Introducing permutation encoding for input-symmetric (IS) DBCs and proving its optimality
3) Proving the optimality of the NE scheme for discrete multiplication (DM) DBCs.

This paper begins its investigation by extending ideas from Witsenhausen and Wyner [9] to study a conditional entropy bound for the channel output of a discrete DBC, leading to a representation of the capacity region of discrete DBCs. As an application, explicit parametric expressions for the capacity regions are derived for two-receiver BS BCs and two-receiver broadcast Z channels. For broadcast Z channels, this simplified expression demonstrates that the NE scheme identified as optimal for two-receiver broadcast Z channels in [11] is also optimal for more than two receivers.

This paper then defines what it means for a DBC to be IS (first introduced in [9] for point-to-point channels) and provides an independent-encoding scheme, referred to as permutation encoding, which achieves the capacity region of all IS-DBCs. The group-operation DBC, which includes the discrete additive DBC [10] as a special case, is a class of IS DBCs for which each channel output is a group operation[1] of the channel input and the channel noise. For group-operation DBCs, permutation encoding is equivalent to NE, establishing the optimality of NE for group-operation DBCs.

The DM DBC is a discrete DBC for which each channel output is a DM[2] of the channel input and the channel noise. This paper concludes its investigations by applying the conditional entropy bound to DM DBCs and proving that NE achieves the boundary of the capacity region in this case.

C. Organization

This paper is organized as follows: Section I-D below lays out the notation used in this paper. Section II defines and studies the conditional entropy bound $F^*(\boldsymbol{q}, s)$ for the channel output of a discrete DBC, and represents the capacity region of the discrete DBC using the function $F^*(\boldsymbol{q}, s)$. Section III uses duality to evaluate $F^*(\boldsymbol{q}, s)$ and provides an approach to characterizing optimal transmission strategies for the discrete DBC based on this evaluation. As an example, Section III-B uses the duality-based computation of $F^*(\boldsymbol{q}, s)$ to provide an explicit parametric expression for the capacity region of the two-receiver BS BC. Section IV proves the optimality of the NE scheme for broadcast Z channels with more than two receivers. Section V defines the IS-DBC, introduces the permutation encoding approach, and proves its optimality for IS-DBCs. Section VI studies the DM DBC and shows that NE achieves the boundary of the capacity region for the DM DBC. Section VII delivers the conclusions.

D. Notation

Denote $X \rightarrow Y$ as a discrete memoryless channel with channel input X and output Y. Denote $X \rightarrow Y^{(1)} \rightarrow \cdots \rightarrow Y^{(K)}$ as a K-receiver ($K \geq 2$) discrete memoryless DBC where X is the channel input, and $Y^{(i)} (i = 1, \ldots, K)$ is the ith least degraded output. For simplicity of notation, we also denote $X \rightarrow Y \rightarrow Z$ as a two-receiver DBC where Y is the less degraded output and Z is the more degraded output. Since the capacity region of a statistically degraded BC without feedback is equivalent to that of the corresponding physically degraded BC with the same marginal transition probabilities, we assume that the DBCs in this paper are physically degraded without loss of generality. Thus, $X \rightarrow Y \rightarrow Z$ also denotes a Markov chain $P(Z = z|Y = y, X = x) = P(Z = z|Y = y)$.

Throughout this paper, we use X to represent a scalar random variable at the channel input. Symbols x and $\mathcal{X}$ denote its specific value and its alphabet, respectively. $\boldsymbol{X}$ denotes a sequence of random variables of length N at the channel input, and $\boldsymbol{x}$ denotes sequence of specific values. X_i and x_i denote the ith element of $\boldsymbol{X}$ and $\boldsymbol{x}$, respectively. The same notation applies to channel outputs Y, Z, $Y^{(i)}$, the auxiliary random variable U, and the codeword $\boldsymbol{X}^{(i)}$ for the ith receiver.

Let $X \rightarrow Y \rightarrow Z$ be a two-receiver discrete memoryless DBC where $\mathcal{X} = \{1, 2, \ldots, k\}$, $\mathcal{Y} = \{1, 2, \ldots, n\}$, and $\mathcal{Z} = \{1, 2, \ldots, m\}$. Let T_{YX} be an $n \times k$ stochastic matrix with entries $T_{YX}(j, i) = \Pr(Y = j|X = i)$ and T_{ZX} be an $m \times k$ stochastic matrix with entries $T_{ZX}(j, i) = \Pr(Z = j|X = i)$. Thus, T_{YX} and T_{ZX} are the transition probability matrices of the marginal channels of the DBC.

Column vectors $\boldsymbol{p}$, $\boldsymbol{q}$, and $\boldsymbol{w}$ denote the distributions of discrete random variables. In particular, $\boldsymbol{p}_X$ denotes the distribution of X. Let $\Delta_n = \{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid \sum_{i=1}^n p_n = 1, \text{and} \quad p_i \geq 0 \quad \text{for all} \quad i\}$ denote the unit $(n-1)$-dimensional simplex of probability n-vectors. We denote $h_n : \Delta_n \mapsto \mathbb{R}$ as the entropy function for $n \geq 2$, i.e.,

[1] A group operation is an operation that satisfies the group axioms (Closure, Associativity, Identity element, Inverse element) on a predefined set. The group operation and the set together form a group.

[2] The definition of the DM is given in Section VI. We refer to this operation as DM because it is a generalization of multiplication as defined in a field.

$h_n([p_1,\ldots,p_n]^T)\triangleq h_n(p_1,\ldots,p_n)\triangleq -\sum p_i \ln p_i$. We also denote $h:[0,1]\mapsto \mathbb{R}$ as $h(p)\triangleq h_2([p,1-p]^T)$.

Following the traditional notation, we denote $H(X)$ as the entropy of X, $H(Y|X)$ as the conditional entropy of Y given X, $I(X;Y)$ as the mutual information between X and Y, and $I(X;Y|U)$ as the mutual information between X and Y given U. Since we have defined $h_n(\cdot)$ using the natural logarithm, all information quantities considered in this paper are in units of *nats*, unless explicitly stated otherwise.

II. Conditional Entropy Bound $F^*(\boldsymbol{q}, s)$

The following definition introduces a conditional entropy bound central to our analysis.

Definition 1: $\left(F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)\right)$ Let $\boldsymbol{q}\in\Delta_k$ be the distribution of the channel input X. The function $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ is defined as

$$F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)=\inf_{\substack{p(u,x):H(Y|U)=s,\,\boldsymbol{p}_X=\boldsymbol{q},\\ \text{and } U\to X\to(Y,Z)}} H(Z|U).$$

Thus, $F^*(\boldsymbol{q}, s)$ is essentially the smallest possible value of $H(Z|U)$ given a specified input distribution and a specified value of $H(Y|U)$. We will sometimes abbreviate $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ to $F^*(\boldsymbol{q}, s)$ or even $F^*(s)$ when there is sufficient context to avoid confusion.

The auxiliary random variable U with alphabet size $l\geq 1$ is characterized by its distribution $\boldsymbol{w}=[w_1,\ldots,w_l]^T\in\Delta_l$, and the joint $p(u,x)$ follows from $\boldsymbol{w}$ and the transition probability matrix from U to X, $T_{XU}=[\boldsymbol{t}_1\ldots\boldsymbol{t}_l]$ where $\boldsymbol{t}_j\in\Delta_k$.

The choices of $p(u,x)$ satisfying $H(Y|U)=s$, $\boldsymbol{p}_X=\boldsymbol{q}$, and $U\to X\to(Y,Z)$ in the definition of $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ correspond to the choices of l, $\boldsymbol{w}$ and T_{XU} such that

$$\boldsymbol{q}=\boldsymbol{p}_X=T_{XU}\boldsymbol{w}=\sum_{j=1}^{l} w_j\boldsymbol{t}_j \tag{1}$$

and

$$s=H(Y|U)=\sum_{j=1}^{l} w_j h_n(T_{YX}\boldsymbol{t}_j). \tag{2}$$

The corresponding $H(Z|U)$ is given by

$$\eta=H(Z|U)=\sum_{j=1}^{l} w_j h_m(T_{ZX}\boldsymbol{t}_j). \tag{3}$$

Let $\mathcal{C}$ be the set of all $(\boldsymbol{p}_X, s,\eta)$ satisfying (1)–(3) for some choice of l, $\boldsymbol{w}$, and T_{XU}. Additionally, define the set $\mathcal{S}=\{(\boldsymbol{p}_X, h_n(T_{YX}\boldsymbol{p}_X), h_m(T_{ZX}\boldsymbol{p}_X))\in\Delta_k\times[0,\ln n]\times[0,\ln n]\}$. Each point in $\mathcal{S}$ corresponds to a $\boldsymbol{p}_X\in\Delta_k$. $\mathcal{C}$ and $\mathcal{S}$ are both triples whose first term is $\boldsymbol{p}_X$, but the last two terms of $\mathcal{C}$ are conditional entropies of Y and Z given U, while the last two terms of $\mathcal{S}$ are *marginal* entropies of Y and Z.

Let $\mathcal{C}^*=\{(s,\eta)|(\boldsymbol{p}_X,s,\eta)\in\mathcal{C}\ \text{ for some }\ \boldsymbol{p}_X\}$ be the projection of the set $\mathcal{C}$ onto the (s,η)-plane. Let $\mathcal{C}^*_{\boldsymbol{q}}=\{(s,\eta)|(\boldsymbol{p}_X,s,\eta)\in\mathcal{C}\ ,\boldsymbol{p}_X=\boldsymbol{q}\}$ be the subset of $\mathcal{C}^*$ for which $\boldsymbol{p}_X=\boldsymbol{q}$. By definition, $\mathcal{C}^*=\bigcup_{\boldsymbol{q}\in\Delta_k}\mathcal{C}^*_{\boldsymbol{q}}$.

Note that $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ is the infimum of all η for which $\mathcal{C}^*_{\boldsymbol{q}}$ contains the point (s,η). Thus

$$\begin{aligned}F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)&=\inf_{\eta}\{\eta|(\boldsymbol{p}_X,s,\eta)\in\mathcal{C},\ \boldsymbol{p}_X=\boldsymbol{q}\}\\ &=\inf_{\eta}\left\{\eta|(s,\eta)\in\mathcal{C}^*_{\boldsymbol{q}}\right\}.\end{aligned}$$

The function $F^*(\boldsymbol{q}, s)$ is an extension to DBCs of the function $F(\boldsymbol{q}, s)$ introduced in [9]. Most properties of $F(\boldsymbol{q}, s)$ shown in [9] can be readily extended to apply to $F^*(\boldsymbol{q}, s)$ as well. These properties are stated in the following as propositions. Readers can refer to [9] to see the proofs for $F(\boldsymbol{q}, s)$ corresponding to the propositions for $F^*(\boldsymbol{q}, s)$ given as follows.

Proposition 1: $\mathcal{C}$ is the convex hull of $\mathcal{S}$. $\mathcal{C}$, $\mathcal{C}^*$, and $\mathcal{C}^*_{\boldsymbol{q}}$ are compact, connected, and convex. See [9, Sec. II.A].

Proposition 2: i) Every point of $\mathcal{C}$ can be obtained by applying (1)–(3) with $l\leq k+1$. In other words, only random variables U taking at most $k+1$ values need to be considered. ii) Every extreme point of the intersection of $\mathcal{C}$ with a 2-D plane can be obtained with $l\leq k$ [9, Lemma 2.2].

Proposition 3: For any fixed $\boldsymbol{q}$ as the distribution of X, the domain of $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ in s is the closed interval $[H(Y|X),H(Y)]=[\sum_{i=1}^{k}q_ih_n(T_{YX}\boldsymbol{e}_i),h_n(T_{YX}\boldsymbol{q})]$, where $\boldsymbol{e}_i$ is a vector with ith entry 1 and all other entries zeros.

Proof: For the Markov chain $U\to X\to Y$, the data processing inequality [17] implies $H(Y|U)\geq H(Y|X)$ and equality is achieved when $U=X$. One also has $H(Y|U)\leq H(Y)$ and equality is achieved when U is a constant. ■

Proposition 4: The function $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ is defined and convex on the compact convex domain $\{(\boldsymbol{q},s)|\boldsymbol{q}\in\Delta_k,\sum_{i=1}^{k}q_ih_n(T_{YX}\boldsymbol{e}_i)\leq s\leq h_n(T_{YX}\boldsymbol{q})\}$ and for each $(\boldsymbol{q},s)$ in this domain, the infimum in its definition is a minimum, attainable with U taking at most $k+1$ values. See [9, Th. 2.3].

Proposition 5: $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ is monotonically nondecreasing in s and the infimum in its definition is a minimum. Hence, $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s)$ can be taken as the minimum $H(Z|U)$ with respect to all $p(u,x)$ satisfying the conditions $H(Y|U)=s,\boldsymbol{p}_X=\boldsymbol{q}$, and $U\to X\to(Y,Z)$. See [9, Th. 2.5].

Proposition 6: For $H(Y|X)\leq s\leq H(Y)$ and any fixed $\boldsymbol{q}=\boldsymbol{p}_X$, $F^*(\boldsymbol{q}, s)\geq s+H(Z)-H(Y)$ is a lower bound for $F^*(\boldsymbol{q}, s)$. See [9, Th. 2.6].

Proposition 7: For any given $\boldsymbol{q}=\boldsymbol{p}_X$, and s ranging over the interval $[H(Y|X),H(Y)]$, the attainable region of $F^*(\boldsymbol{q}, s)$ is $H(Z|X)\leq F^*(\boldsymbol{q}, s)\leq H(Z)$.

Proof:

$$\begin{aligned}F^*(\boldsymbol{q}, s)&=\min_{p(u,x)}\{H(Z|U)|\boldsymbol{p}_X=\boldsymbol{q},H(Y|U)=s\}\\ &\overset{(a)}{\geq}\min_{p(u,x)}\{H(Z|U,X)|\boldsymbol{p}_X=\boldsymbol{q},H(Y|U)=s\}\\ &\overset{(b)}{=}H(Z|X)\end{aligned}$$

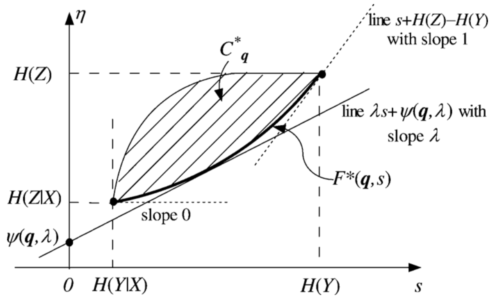

Fig. 1. Illustration of the curve $F^*(\boldsymbol{q}, s)$ for a given $\boldsymbol{q}$ shown in bold, the region $\mathcal{C}^*_{\boldsymbol{q}}$, and the point $(0, \psi(\boldsymbol{q}, \lambda))$.

where (a) follows since conditioning reduces entropy and (b) follows since Z and U are conditionally independent given X. Equality is achieved when $U = X$ and $s = H(Y|X)$. On the other hand

$$\begin{aligned} F^*(\boldsymbol{q}, s) &= \min_{p(u,x)} \{H(Z|U) | \boldsymbol{p}_X = \boldsymbol{q}, H(Y|U) = s\} \\ &\overset{(a)}{\leq} \min_{p(u,x)} \{H(Z) | \boldsymbol{p}_X = \boldsymbol{q}, H(Y|U) = s\} \\ &= H(Z) \end{aligned}$$

where (a) follows since conditioning reduces entropy. Equality is achieved when U is a constant and $s = H(Y)$. ■

Proposition 8: For any given $\boldsymbol{q} = \boldsymbol{p}_X$, $F^*(s) \overset{\triangle}{=} F^*(\boldsymbol{q}, s)$ is differentiable at all but at most countably many points. At differentiable points of $F^*(s)$

$$0 \leq \frac{dF^*(s)}{ds} \leq 1.$$

Proof: Since $F^*(s)$ is convex in s, it is differentiable at all but at most countably many points. As illustrated in Fig. 1, for any $H(Y|X) \leq s \leq H(Y)$ where $F^*(s)$ is differentiable, the slope of the supporting line at the point $(s, F^*(s))$ is less than or equal to the slope of the supporting line $s + H(Z) - H(Y)$ at the point $(H(Y), F^*(H(Y)))$ because of the convexity of $F^*(s)$. Thus, $\frac{dF^*(s)}{ds} \leq 1$ for any $H(Y|X) \leq s \leq H(Y)$ where $F^*(s)$ is differentiable. Also, $\frac{dF^*(s)}{ds} \geq 0$ because $F^*(s)$ is monotonically nondecreasing. ■

Let $\boldsymbol{X} = (X_1, \ldots, X_N)$ be a sequence of channel inputs to the BC $X \to Y \to Z$. The corresponding channel outputs are $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_N)$. Thus, any two channel output pairs (Y_i, Z_i) and (Y_j, Z_j) with $i \neq j$ are conditionally independent given $\boldsymbol{X}$. Note that the channel outputs $\{(Y_i, Z_i)\}_{i=1}^N$ are not necessarily i.i.d. since $X_1, \ldots, X_N$ could be correlated and have different distributions.

Denote $\boldsymbol{q_i}$ as the distribution of X_i for $i \in \{1, \ldots, N\}$. Thus, $\boldsymbol{q} = \sum \boldsymbol{q_i}/N$ is the average of the distributions of the channel inputs. For any $\boldsymbol{q} \in \Delta_k$, define $F^*_{T^{(N)}_{YX}, T^{(N)}_{ZX}}(\boldsymbol{q}, Ns)$ as the infimum of $H(\boldsymbol{Z}|U)$ with respect to all random variables U and all possible channel inputs $\boldsymbol{X}$ such that $H(\boldsymbol{Y}|U) = Ns$, the average of the distributions of the channel inputs is $\boldsymbol{q}$, and $U \to \boldsymbol{X} \to \boldsymbol{Y} \to \boldsymbol{Z}$ is a Markov chain.

Proposition 9: For all $N \in \{1, 2, \ldots, \}$ and all T_{YX}, T_{ZX}, $\boldsymbol{q}$, and $H(Y|X) \leq s \leq H(Y)$, one has $F^*_{T^{(N)}_{YX}, T^{(N)}_{ZX}}(\boldsymbol{q}, Ns) = NF^*_{T_{YX}, T_{ZX}}(\boldsymbol{q}, s)$ See [9, Th. 2.4].

Proposition 9 is the key to the applications in Section IV. It indicates that i.i.d. inputs $\boldsymbol{X}$ achieve the conditional entropy bound $F^*_{T^{(N)}_{YX}, T^{(N)}_{ZX}}(\boldsymbol{q}, Ns)$. Moreover, at each time instant, a single use of the channel achieves the conditional entropy bound $F^*_{T_{YX}, T_{ZX}}(\boldsymbol{q}, s)$.

Theorem 1: The capacity region for the discrete memoryless DBC $X \to Y \to Z$ is the closure of the convex hull of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} &0 \leq R_1 \leq I(X;Y) \\ &0 \leq R_2 \leq H(Z) - F^*_{T_{YX}, T_{ZX}}(\boldsymbol{q}, R_1 + H(Y|X)) \end{aligned}$$

for some $\boldsymbol{p}_X = \boldsymbol{q} \in \Delta_k$.

For a fixed $\boldsymbol{p}_X = \boldsymbol{q}$ and $\lambda \geq 0$, a pareto-optimal rate pair is given by

$$\begin{aligned} \max_{p(u,x): \boldsymbol{p}_X = \boldsymbol{q}} \{R_2 + \lambda R_1\} &= H(Z) - \lambda H(Y|X) \\ &\quad - \min_{s \in [H(Y|X), H(Y)]} \{F^*(\boldsymbol{q}, s) - \lambda s\}. \end{aligned} \tag{4}$$

Proof: The capacity region for the DBC is known in [1], [3], and [18] as

$$\overline{\text{co}} \left[\bigcup_{p(u), p(x|u)} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq I(X;Y|U) \\ R_2 \leq I(U;Z) \end{array} \right\} \right] \tag{5}$$

where $\overline{\text{co}}$ denotes the closure of the convex hull operation, and U is the auxiliary random variable which satisfies the Markov chain $U \to X \to Y \to Z$ and $|\mathcal{U}| \leq \min(|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|)$. Rewrite (5) and we have

$$\begin{aligned} &\overline{\text{co}} \left[\bigcup_{p(u), p(x|u)} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq I(X;Y|U) \\ R_2 \leq I(U;Z) \end{array} \right\} \right] \\ &\overset{(a)}{=} \overline{\text{co}} \left[\bigcup_{\boldsymbol{p}_X = \boldsymbol{q} \in \Delta_k} \left\{ \bigcup_{p(u,x) \text{ s.t. } \boldsymbol{p}_X = \boldsymbol{q}} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq I(X;Y|U) \\ R_2 \leq I(U;Z) \end{array} \right\} \right\} \right] \\ &= \overline{\text{co}} \left[\bigcup_{\boldsymbol{p}_X = \boldsymbol{q} \in \Delta_k} \left\{ \bigcup_{p(u,x) \text{ s.t. } \boldsymbol{p}_X = \boldsymbol{q}} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq H(Y|U) - H(Y|X) \\ R_2 \leq H(Z) - H(Z|U) \end{array} \right\} \right\} \right] \\ &\overset{(b)}{=} \overline{\text{co}} \left[\bigcup_{\boldsymbol{p}_X = \boldsymbol{q} \in \Delta_k} \left\{ \bigcup_{H(Y|X) \leq s \leq H(Y)} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq s - H(Y|X) \\ R_2 \leq H(Z) - F^*(\boldsymbol{q}, s) \end{array} \right\} \right\} \right] \quad (6) \\ &\overset{(c)}{=} \overline{\text{co}} \left[\bigcup_{\boldsymbol{p}_X = \boldsymbol{q} \in \Delta_k} \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1 \leq I(X;Y) \\ R_2 \leq H(Z) - F^*(\boldsymbol{q}, R_1 + H(Y|X)) \end{array} \right\} \right]. \quad (7) \end{aligned}$$

Some of these steps are justified as follows:

1) (a) follows from the equivalence of $\bigcup_{p(u), p(x|u)}$ and $\bigcup_{\boldsymbol{p}_X = \boldsymbol{q} \in \Delta_k} \bigcup_{p(u,x) \text{ s.t. } \boldsymbol{p}_X = \boldsymbol{q}}$;

2) (b) follows from the definition of the conditional entropy bound $F^*(\boldsymbol{q}, s)$;

3) (c) follows from the nondecreasing property of $F^*(s)$ in Proposition 5, which allows the substitution $s = R_1 + H(Y|X)$ in the argument of $F^*(\boldsymbol{q}, s)$.

To see that (4) holds, observe that

$$\begin{aligned}
&\max_{p(u,x):\boldsymbol{p}_X=\boldsymbol{q}} \{R_2+\lambda R_1\}\\
&= \max_{R_1\in[0,I(X;Y)]} \{H(Z) - F^*(\boldsymbol{q}, R_1+H(Y|X))\\
&\qquad\qquad + \lambda R_1 + \lambda H(Y|X) - \lambda H(Y|X)\}\\
&= H(Z) - \lambda H(Y|X) +\\
&\quad \max_{R_1\in[0,I(X;Y)]}\{-F^*(\boldsymbol{q},R_1+H(Y|X))+\lambda(R_1+H(Y|X))\}\\
&= H(Z)-\lambda H(Y|X) - \min_{s\in[H(Y|X),H(Y)]}\{F^*(\boldsymbol{q},s)-\lambda s\}\,.
\end{aligned}$$

■

Note that for a fixed input distribution $\boldsymbol{q}=\boldsymbol{p}_X$, the quantities $I(X;Y)$, $H(Z)$, and $H(Y|X)$ in (7) are constant. This theorem provides the relationship between the capacity region and the conditional entropy bound $F^*(\boldsymbol{q},s)$ for a discrete DBC.

For any given $\boldsymbol{p}_X=\boldsymbol{q}$, Theorem 1 states that maximizing $R_2+\lambda R_1$ is equivalent to minimizing $F^*(\boldsymbol{q},s)-\lambda s$. Propositions 6, 7, and 8 indicate that for every $\lambda>1$, the minimum of $F^*(\boldsymbol{q},s)-\lambda s$ is attained when $s=H(Y)$ and $F^*(\boldsymbol{q},s)=H(Z)$, i.e., U is a constant. Thus, the nontrivial range of λ is $0\le\lambda\le 1$.

III. Evaluation of $F^*(\boldsymbol{q},s)$

In this section, we evaluate $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q},s)$ for a given $\boldsymbol{q}$ via a duality technique, which is also used for evaluating $F(\cdot)$ in [9]. This duality technique also provides the optimal transmission strategy for the DBC $X\rightarrow Y\rightarrow Z$ to achieve the maximum of $R_2+\lambda R_1$ for any $\lambda\ge 0$. The section concludes with an application to the BS BC.

A. Duality Technique

Proposition 4 shows $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q},s)=\min_\eta\{\eta|(s,\eta)\in\mathcal{C}^*_{\boldsymbol{q}}\}$. Thus, the function $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q},s)$ is determined by the lower boundary of $\mathcal{C}^*_{\boldsymbol{q}}$ as illustrated in Fig. 1. Since $\mathcal{C}^*_{\boldsymbol{q}}$ is convex, its lower boundary can be described by the lines supporting the boundary from below. The line with slope λ in the (s,η)-plane supporting $\mathcal{C}^*_{\boldsymbol{q}}$ as shown in Fig. 1 is given by

$$\eta=\lambda s+\psi(\boldsymbol{q},\lambda)$$

where $\psi(\boldsymbol{q},\lambda)$ is the η-intercept of the tangent line with slope λ for the function $F^*_{T_{YX},T_{ZX}}(\boldsymbol{q},s)$. Thus

$$\begin{aligned}
&\psi(\boldsymbol{q},\lambda)\\
&= \min_s\{F^*(\boldsymbol{q},s)-\lambda s\big|H(Y|X)\le s\le H(Y)\} &(8)\\
&= \min_{s,\eta}\{\eta-\lambda s\big|(s,\eta)\in\mathcal{C}^*_{\boldsymbol{q}}\} &(9)\\
&= \min_{s,\eta}\{\eta-\lambda s\big|(\boldsymbol{q},s,\eta)\in\mathcal{C}\} &(10)\\
&= \min_{U\rightarrow X\rightarrow Y,Z \text{ s.t. } \boldsymbol{p}_X=\boldsymbol{q}}\{H(Z|U)-\lambda H(Y|U)\}\,.
\end{aligned}$$

For any given $\boldsymbol{q}$, and $H(Y|X)\le s\le H(Y)$, the function $F^*(\boldsymbol{q},s)$ can be represented as

$$\begin{aligned}
F^*(\boldsymbol{q},s)&=\max_\lambda\{\psi(\boldsymbol{q},\lambda)+\lambda s|-\infty<\lambda<\infty\}\\
&=\max_\lambda\{\psi(\boldsymbol{q},\lambda)+\lambda s|0\le\lambda\le 1\} \qquad (11)
\end{aligned}$$

where (11) follows from Proposition 8.

Let L_λ be the linear transformation $(\boldsymbol{q},s,\eta)\mapsto(\boldsymbol{q},\eta-\lambda s)$. L_λ maps $\mathcal{C}$ and $\mathcal{S}$ onto the sets

$$\mathcal{C}_\lambda=\{(\boldsymbol{q},\eta-\lambda s)|(\boldsymbol{q},s,\eta)\in\mathcal{C}\}$$

and

$$\mathcal{S}_\lambda=\{(\boldsymbol{q},h_m(T_{ZX}\boldsymbol{q})-\lambda h_n(T_{YX}\boldsymbol{q}))|\boldsymbol{q}\in\Delta_k\}.$$

Define $\phi(\boldsymbol{q},\lambda)=h_m(T_{ZX}\boldsymbol{q})-\lambda h_n(T_{YX}\boldsymbol{q})$. The lower boundaries of $\mathcal{C}_\lambda$ and $\mathcal{S}_\lambda$ are the graphs of $\psi(\boldsymbol{q},\lambda)$ and $\phi(\boldsymbol{q},\lambda)$, respectively. Since $\mathcal{C}$ is the convex hull of $\mathcal{S}$, $\mathcal{C}_\lambda$ is the convex hull of $\mathcal{S}_\lambda$, and thus, $\psi(\boldsymbol{q},\lambda)$ is the lower convex envelope of $\phi(\boldsymbol{q},\lambda)$ with respect to $\boldsymbol{q}\in\Delta_k$.

For each λ, we conclude that $\psi(\boldsymbol{q},\lambda)$ can be obtained by forming the lower convex envelope of $\phi(\boldsymbol{q},\lambda)$ with respect to $\boldsymbol{q}$. $F^*(\boldsymbol{q},s)$ can be reconstructed from $\psi(\boldsymbol{q},\lambda)$ by (11). This is the dual approach to the evaluation of $F^*(\boldsymbol{q},s)$.

Theorem 1 describes the capacity region for a DBC in terms of the function $F^*(\boldsymbol{q},s)$. Since $\psi(\boldsymbol{q},\lambda)$ and $F^*(\boldsymbol{q},s)$ can be constructed from each other by (8) and (11) for any $\lambda\ge 0$, the associated point on the boundary of the capacity region may be found (from its unique value of $R_2+\lambda R_1$) as follows:

$$\begin{aligned}
&\max_{p(u,x)}\{R_2+\lambda R_1\}\\
&=\max_{\boldsymbol{q}\in\Delta_k}\left\{\max_{p(u,x)\text{ s.t. }\boldsymbol{p}_X=\boldsymbol{q}}\{R_2+\lambda R_1\}\right\}\\
&=\max_{\boldsymbol{q}\in\Delta_k}\left\{\max_{s\in[H(Y|X),H(Y)],\boldsymbol{p}_X=\boldsymbol{q}}\{H(Z)-F^*(\boldsymbol{q},s)+\lambda s-\lambda H(Y|X)\}\right\}\\
&=\max_{\boldsymbol{q}\in\Delta_k}\left\{H(Z)-\lambda H(Y|X)-\min_s\{F^*(\boldsymbol{q},s)-\lambda s\}\big|\boldsymbol{p}_X=\boldsymbol{q}\right\}\\
&=\max_{\boldsymbol{q}\in\Delta_k}\left\{H(Z)-\lambda H(Y|X)-\psi(\boldsymbol{q},\lambda)\big|\boldsymbol{p}_X=\boldsymbol{q}\right\}\,.
\end{aligned}$$

We have shown the relationship among $F^*(\boldsymbol{q},s)$, $\psi(\boldsymbol{q},\lambda)$ and the capacity region for the DBC. Now, we state a theorem that provides the relationship among $F^*(\boldsymbol{q},s)$, $\psi(\boldsymbol{q},\lambda)$, $\phi(\boldsymbol{q},\lambda)$, and the optimal transmission strategies $p(u,x)$ for the DBC. This theorem is a straightforward extension of Theorem 4.1 in [9].

Theorem 2: i) For any $0\le\lambda\le 1$, if a point of the graph of $\psi(\cdot,\lambda)$ is a convex combination of l points of the graph of $\phi(\cdot,\lambda)$ with arguments $\boldsymbol{t}_j$ and weights w_j, $j=1,\ldots,l$, then

$$F^*_{T_{YX},T_{ZX}}\left(\sum_j w_j\boldsymbol{t}_j,\sum_j w_j h_n(T_{YX}\boldsymbol{t}_j)\right)=\sum_j w_j h_m(T_{ZX}\boldsymbol{t}_j).$$

ii) For a predetermined channel input distribution $\boldsymbol{q}$, if the transmission strategy $|\mathcal{U}|=l$, $\Pr(U=j)=w_j$, and $\boldsymbol{p}_{X|U=j}=\boldsymbol{t}_j$ achieves $\max\{R_2+\lambda R_1|\sum_j w_j\boldsymbol{t}_j=\boldsymbol{q}\}$, then the point $(\boldsymbol{q},\psi(\boldsymbol{q},\lambda))$ is the convex combination of l points of the graph of $\phi(\cdot,\lambda)$ with arguments $\boldsymbol{t}_j$ and weights w_j, $j=1,\ldots,l$.

A few remarks are in order. The representation of a point in $\psi(\cdot,\lambda)$ as a convex combination implies that for the fixed channel input distribution $\boldsymbol{q}=\sum_j w_j\boldsymbol{t}_j$, an optimal transmission strategy to achieve the maximum of $R_2+\lambda R_1$ is determined by l, w_j, and $\boldsymbol{t}_j$. In particular, an optimal transmission strategy has $|\mathcal{U}|=l$, $\Pr(U=j)=w_j$ and $\boldsymbol{p}_{X|U=j}=\boldsymbol{t}_j$,

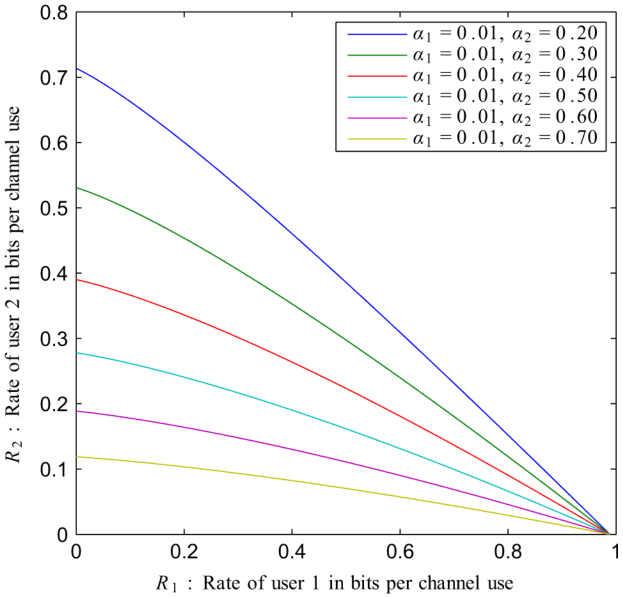

Fig. 2. BS BC capacity regions (in bits per channel use) obtained using the explicit parametric expressions given in Theorem 3 for $\alpha_1 = 0.001$ and a variety of α_2 values.

where $\boldsymbol{p}_{X|U=j}$ denotes the conditional distribution of X given $U = j$.

Note that if for some pair $(\boldsymbol{q}, \lambda)$, $\psi(\boldsymbol{q}, \lambda) = \phi(\boldsymbol{q}, \lambda)$, then the corresponding optimal transmission strategy has $l = 1$, which means U is a constant. For such a $(\boldsymbol{q}, \lambda)$ pair, the line $\eta = \lambda s + \psi(\boldsymbol{q}, \lambda)$ supports the graph of $F^*(s)$ at its endpoint $(H(Y), H(Z)) = (h_n(T_{YX}\boldsymbol{q}), h_m(T_{ZX}\boldsymbol{q}))$.

B. Example: Application to the BS BC

Consider the BS BC $X \to Y \to Z$ with

$$T_{YX} = \begin{bmatrix} 1-\alpha_1 & \alpha_1 \\ \alpha_1 & 1-\alpha_1 \end{bmatrix}, \quad T_{ZX} = \begin{bmatrix} 1-\alpha_2 & \alpha_2 \\ \alpha_2 & 1-\alpha_2 \end{bmatrix}$$

where $0 < \alpha_1 < \alpha_2 < 1/2$. The following theorem, which is proved by the duality technique, provides an explicit parameterized characterization of the capacity region.

Theorem 3: Consider the BS BC with crossover probabilities $0 < \alpha_1 < \alpha_2 < 1/2$. For $\lambda \geq 0$, the achievable rate pair (R_1, R_2) that maximizes $\lambda R_1 + R_2$ is given by

$$\begin{aligned} R_1 &= h\left(\alpha_1 + (1-2\alpha_1)p_\lambda\right) - h(\alpha_1) \\ R_2 &= \ln(2) - h\left(\alpha_2 + (1-2\alpha_2)p_\lambda\right) \end{aligned}$$

where λ, R_1, and R_2 are parameterized by $0 \leq p_\lambda \leq 1/2$ satisfying

$$\lambda = \frac{1-2\alpha_2}{1-2\alpha_1} \cdot \frac{\ln \frac{1-\alpha_2-(1-2\alpha_2)p_\lambda}{\alpha_2+(1-2\alpha_2)p_\lambda}}{\ln \frac{1-\alpha_1-(1-2\alpha_1)p_\lambda}{\alpha_1+(1-2\alpha_1)p_\lambda}}. \tag{12}$$

Moreover, NE achieves all points in the capacity region.

Fig. 2 shows several example capacity region boundaries computed using Theorem 3.

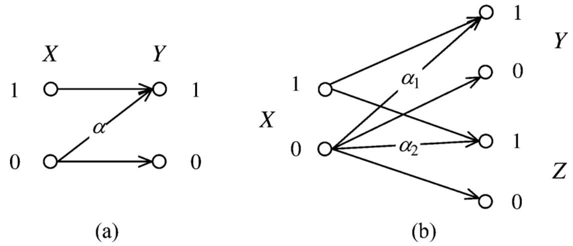

Fig. 3. (a) Z channel and (b) broadcast Z channel.

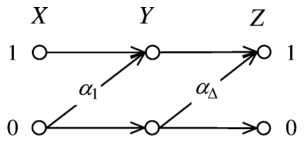

Fig. 4. Physically degraded broadcast Z channel.

The proof of Theorem 3 and the detailed analysis of $\phi(p, \lambda)$, $\psi(p, \lambda)$, and $F^*(\boldsymbol{q}, s)$ for the BS BC are given in Appendix B.

IV. Broadcast Z Channels

The Z channel, shown in Fig. 3(a), is a binary-asymmetric channel that is noiseless when symbol 1 is transmitted but noisy when symbol 0 is transmitted. The channel output Y is the binary OR of the channel input X and Bernoulli distributed noise with parameter α. The capacity of the Z channel was studied in [19]. The broadcast Z channel is a class of discrete memoryless BCs whose component channels are Z channels. A two-receiver broadcast Z channel with marginal transition probability matrices

$$T_{YX} = \begin{bmatrix} 1 & \alpha_1 \\ 0 & 1-\alpha_1 \end{bmatrix}, \quad T_{ZX} = \begin{bmatrix} 1 & \alpha_2 \\ 0 & 1-\alpha_2 \end{bmatrix}$$

where $0 < \alpha_1 \leq \alpha_2 < 1$, is shown in Fig. 3(b). The two-receiver broadcast Z channel is stochastically degraded and can be modeled as a physically DBC as shown in Fig. 4, where $\alpha_\Delta = (\alpha_2 - \alpha_1)/(1-\alpha_1)$[11]. NE for broadcast Z channels uses the binary OR function to combine each receiver's independently encoded message. As shown in [11] and [12], NE achieves the entire boundary of the capacity region for the two-receiver broadcast Z channel. In this section, we will show that NE also achieves the entire boundary of the capacity region for broadcast Z channels with more than two receivers.

A. Capacity Region for the Two-Receiver Broadcast Z Channel

Similar to Theorem 3 for the BS BC, we can apply our analysis of F^* to obtain a parametric expression for the capacity region of the broadcast Z channel.

Theorem 4: Consider the broadcast Z channel with crossover probabilities $0 < \alpha_1 \leq \alpha_2 < 1$. Define $\beta_i = 1 - \alpha_i$ for $i = 1, 2$. For $\lambda \geq 0$, the achievable rate pair (R_1, R_2) which maximizes $\lambda R_1 + R_2$ is given by

$$R_1 = \frac{q_\lambda}{p_\lambda} h(\beta_1 p_\lambda) - q_\lambda h(\beta_1) \tag{13}$$

$$R_2 = h(q_\lambda \beta_2) - \frac{q_\lambda}{p_\lambda} h(\beta_2 p_\lambda) \tag{14}$$

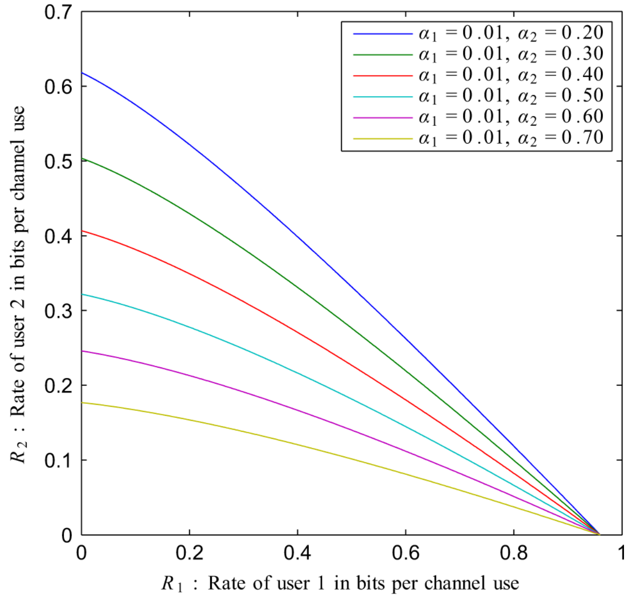

Fig. 5. Broadcast Z channel capacity regions (in bits per channel use) obtained using the explicit parametric procedure for $\alpha_1 = 0.01$ and a variety of α_2 values.

where λ, q_λ, R_1, and R_2 are parameterized by $0 \leq p_\lambda \leq 1$ satisfying

$$\lambda = \frac{\ln(1-\beta_2 p_\lambda)}{\ln(1-\beta_1 p_\lambda)} \tag{15}$$

and

$$q_\lambda = \min\left(p_\lambda, \frac{1}{\beta_2\left(1+\exp\left(\frac{1}{\beta_2 p_\lambda}(h(\beta_2 p_\lambda)-\lambda h(\beta_1 p_\lambda)+\lambda p_\lambda h(\beta_1))\right)\right)}\right). \tag{16}$$

Moreover, NE achieves all points in the capacity region.

Thus, Theorem 4 implies that for a specified α_1 and α_2, the capacity region for the two-receiver broadcast Z channel can be determined parametrically for each λ as follows.

1) Use (15) to compute p_λ from λ.
2) Use (16) to compute q_λ from p_λ.
3) Use q_λ and p_λ in (13) and (14) to find R_1 and R_2 that maximize $R_2 + \lambda R_1$.

Fig. 5 shows several example capacity region boundaries found using this procedure.

Proof: For the broadcast Z channel $X \to Y \to Z$ shown in Figs. 3(b) and 4 with

$$T_{YX} = \begin{bmatrix} 1 & \alpha_1 \\ 0 & \beta_1 \end{bmatrix}, \ T_{ZX} = \begin{bmatrix} 1 & \alpha_2 \\ 0 & \beta_2 \end{bmatrix}$$

where $0 < \alpha_1 \leq \alpha_2 < 1$, $\beta_1 = 1 - \alpha_1$, and $\beta_2 = 1 - \alpha_2$, one has

$$\phi(p,\lambda) \overset{\triangle}{=} \phi\left([1-p, p]^T, \lambda\right) = h(p\beta_2) - \lambda h(p\beta_1).$$

Taking the second derivative of $\phi(p,\lambda)$ with respect to p, we have

$$\phi''(p,\lambda) = \frac{-\beta_2}{(1-p\beta_2)p} + \frac{\lambda\beta_1}{(1-p\beta_1)p}. \tag{17}$$

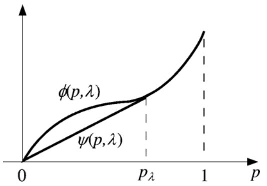

Fig. 6. Illustration of $\phi(p,\lambda)$ and $\psi(p,\lambda)$ for the broadcast Z channel with a given λ.

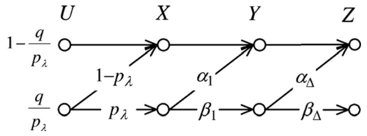

Fig. 7. Optimal transmission strategy for the two-receiver broadcast Z channel.

Multiplying $\phi''(p,\lambda)$ in (17) by the positive quantity $(1-p\beta_1)(1-p\beta_2)p$ produces

$$\begin{aligned}\rho(p,\lambda) &= \phi''(p,\lambda)\cdot(1-p\beta_1)(1-p\beta_2)p \\ &= p\beta_1\beta_2(1-\lambda) + \lambda\beta_1 - \beta_2\end{aligned}$$

which has the same sign as $\phi''(p,\lambda)$.

Let $\beta_\Delta \overset{\triangle}{=} \beta_2/\beta_1$. For the case of $\beta_\Delta \leq \lambda \leq 1$, $\phi''(p,\lambda) \geq 0$ for all $0 \leq p \leq 1$. Hence, $\phi(p,\lambda)$ is convex in p, and thus, $\phi(p,\lambda) = \psi(p,\lambda)$ for all $0 \leq p \leq 1$. In this case, the transmission strategy that maximizes R_1 also maximizes $R_2 + \lambda R_1$. Thus, the optimal transmission strategy has $l = 1$, i.e., U is a constant. Note that the transmission strategy with $l = 1$ is a special case of the NE scheme in which the only codeword for the second receiver is an all-ones codeword.

For the case of $0 \leq \lambda < \beta_\Delta$, $\phi(p,\lambda)$ is concave in p on $[0, \frac{\beta_2-\lambda\beta_1}{\beta_1\beta_2(1-\lambda)}]$ and convex on $[\frac{\beta_2-\lambda\beta_1}{\beta_1\beta_2(1-\lambda)}, 1]$. Fig. 6 illustrates the graph in this case. Since $\phi(0,\lambda) = 0$, $\psi(\cdot,\lambda)$, the lower convex envelope of $\phi(\cdot,\lambda)$, is constructed using the tangent of $\phi(\cdot,\lambda)$ that passes through the origin as shown in Fig. 6. Let $(p_\lambda, \phi(p_\lambda,\lambda))$ be the point of contact. The value of p_λ is determined by $\phi'_p(p_\lambda,\lambda) = \phi(p_\lambda,\lambda)/p_\lambda$, which implies (15).

Let $\boldsymbol{q} = [1-q, q]^T$ be the distribution of the channel input X. For $q \leq p_\lambda$, $\psi(q,\lambda)$ is obtained as a convex combination of points $(0,0)$ and $(p_\lambda, \phi(p_\lambda,\lambda))$ with weights $(p_\lambda - q)/p_\lambda$ and q/p_λ. By Theorem 2, it corresponds to $s = [(p_\lambda - q)/p_\lambda] \cdot 0 + [q/p_\lambda] \cdot h(\beta_1 p_\lambda) = q h(\beta_1 p_\lambda)/p_\lambda$ and $F^*(q,s) \overset{\triangle}{=} F^*(\boldsymbol{q}, s) = q/p_\lambda \cdot h(\beta_2 p_\lambda)$. Hence, for the broadcast Z channel

$$F^*_{T_{YX},T_{ZX}}(q, qh(\beta_1 p)/p) = qh(\beta_2 p)/p \tag{18}$$

for $p \in [q, 1]$, which defines $F^*_{T_{YX},T_{ZX}}(q,\cdot)$ on its entire domain $[qh(\beta_1), h(q\beta_1)]$. Also by Theorem 2, the optimal transmission strategy $p(u,x)$ to maximize $(R_2+\lambda R_1)$ given the constraint $\boldsymbol{p}_X = \boldsymbol{q}$ is determined by $l = 2$, $w_1 = (p_\lambda - q)/p_\lambda$, $w_2 = q/p_\lambda$, $\boldsymbol{t}_1 = [1,0]^T$, and $\boldsymbol{t}_2 = [1-p_\lambda, p_\lambda]^T$. Since the optimal transmission strategy $p(u,x)$ can be modeled as a Z channel as shown in Fig. 7, the random variable X can be constructed as the OR of two Bernoulli random variables with parameters $(p_\lambda - q)/p_\lambda$ and $1 - p_\lambda$, respectively. Hence, an optimal transmission strategy for the broadcast Z channel is NE.

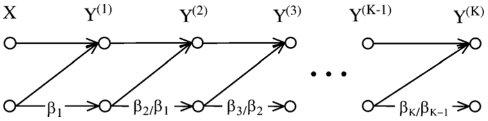

Fig. 8. K-receiver broadcast Z channel.

For $q > p_\lambda$, $\psi(q,\lambda) = \phi(q,\lambda)$, and an optimal strategy has $l = 1$, i.e., U is a constant.

For practical implementations, nonlinear trellis codes [20] and nonlinear turbo codes [11], [12] provide encoded binary streams with the desired ones densities $(p_\lambda - q)/p_\lambda$ and $1 - p_\lambda$ that can be combined through natural encoding by an OR operation.

Thus, the two-receiver broadcast Z channel capacity region is the convex hull of the rate pairs (R_1, R_2) satisfying

$$0 \le R_1 \le \frac{q}{p_\lambda} h(\beta_1 p_\lambda) - q h(\beta_1)$$
$$0 \le R_2 \le h(q\beta_2) - \frac{q}{p_\lambda} h(\beta_2 p_\lambda)$$

for some $q \in [0,1]$ and $p_\lambda \in [q,1]$. For a fixed input distribution $\boldsymbol{p}_X = [1-q, q]^T$, the rate pair (R_1, R_2) of

$$R_1 = \frac{q}{p_\lambda} h(\beta_1 p_\lambda) - q h(\beta_1)$$
$$R_2 = h(q\beta_2) - \frac{q}{p_\lambda} h(\beta_2 p_\lambda)$$

maximizes $R_2 + \lambda R_1$ for each pair of λ and p_λ satisfying (15). Among all possible input distributions $q \in [0,1]$, only one will finally maximize $R_2 + \lambda R_1$ over all rate pairs in the capacity region. Let q_λ be the input distribution that maximizes $R_2 + \lambda R_1$ and thus

$$\begin{aligned} q_\lambda &= \arg\max_{0 \le q \le p_\lambda} (R_2 + \lambda R_1) \\ &= \arg\max_{0 \le q \le p_\lambda} \left(h(q\beta_2) - \frac{q}{p_\lambda} h(\beta_2 p_\lambda) + \lambda \left(\frac{q}{p_\lambda} h(\beta_1 p_\lambda) - q h(\beta_1) \right) \right) \\ &= \min \left(p_\lambda, \frac{1}{\beta_2 \left(1 + \exp\left(\frac{1}{\beta_2 p_\lambda} (h(\beta_2 p_\lambda) - \lambda h(\beta_1 p_\lambda) + \lambda p_\lambda h(\beta_1)) \right) \right)} \right). \end{aligned}$$

■

B. Broadcast Z Channel With More Than Two Receivers

Consider a K-receiver broadcast Z channel $X \to Y^{(1)} \to \cdots \to Y^{(K)}$ with marginal transition probability matrices

$$T_{Y_j X} = \begin{bmatrix} 1 & \alpha_j \\ 0 & \beta_j \end{bmatrix}$$

where $0 < \alpha_1 \le \cdots \le \alpha_K < 1$, and $\beta_j = 1 - \alpha_j$ for $j = 1, \ldots, K$. The K-receiver broadcast Z channel is stochastically degraded and can be modeled as a physically DBC as shown in Fig. 8. NE for the K-receiver broadcast Z channel combines the K independently generated codewords (one for each receiver) using the binary OR operation. The $j\text{th}$ receiver then successively decodes the messages for Receiver K, Receiver $K-1$, …, and, finally, for Receiver j. The codebook for the $j\text{th}$ receiver is a random codebook drawn according to the binary random variable $X^{(j)}$ with $\Pr\{X^{(j)} = 0\} = q^{(j)}$. Denote $X^{(i)} \circ X^{(j)}$ as the binary OR of $X^{(i)}$ and $X^{(j)}$. Hence, the channel input X is the OR of $X^{(j)}$ for all $1 \le j \le K$, i.e., $X = X^{(1)} \circ \cdots \circ X^{(K)}$. From the analysis of successive decoding in the proof of the coding theorem for DBCs [2], [3], the achievable region of NE for the K-receiver broadcast Z channel is determined by

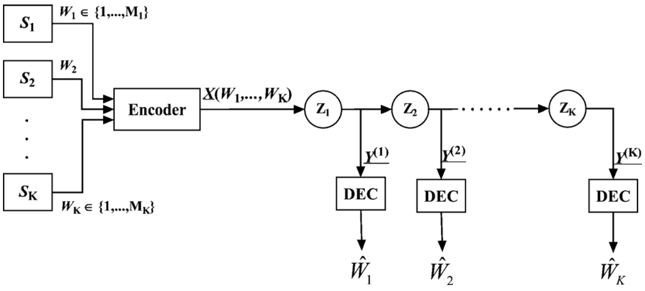

Fig. 9. Communication system for a K-receiver broadcast Z channel.

$$\begin{aligned} R_j &\le I\left(Y^{(j)}, X^{(j)} | X^{(j+1)}, \ldots, X^{(K)}\right) \\ &= H\left(Y^{(j)} | X^{(j+1)}, \ldots, X^{(K)}\right) \\ &\quad - H\left(Y^{(j)} | X^{(j)}, X^{(j+1)}, \ldots, X^{(K)}\right) \\ &= \left(\prod_{i=j+1}^{K} q^{(i)} \right) \cdot h\left(\beta_j \prod_{i=1}^{j} q^{(i)} \right) \\ &\quad - \left(\prod_{i=j}^{K} q^{(i)} \right) \cdot h\left(\beta_j \prod_{i=1}^{j-1} q^{(i)} \right) \\ &= \frac{q}{t_j} h(\beta_j t_j) - \frac{q}{t_{j-1}} h(\beta_j t_{j-1}), \end{aligned} \tag{19}$$

where $t_j = \prod_{i=1}^{j} q^{(i)}$ for $j = 1, \ldots, K$, and $q = \Pr(X=0) = \prod_{i=1}^{K} q^{(i)}$. Denote $t_0 = 1$. Since $0 \le q^{(1)}, \ldots, q^{(K)} \le 1$, one has

$$1 = t_0 \ge t_1 \ge \cdots \ge t_K = q. \tag{20}$$

Theorem 5 below states that NE achieves the entire boundary of the capacity region for broadcast Z channels with any finite number of receivers. Consider the communication system for the K-receiver broadcast Z channel in Fig. 9. $\boldsymbol{X} = (X_1, \ldots, X_N)$ is a length-N codeword determined by the messages $W_1, \ldots, W_K$. $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(K)}$ are the channel outputs corresponding to the channel input $\boldsymbol{X}$.

Theorem 5: If $\sum_{i=1}^{N} \Pr\{X_i = 0\}/N = q$, then no point $(R_1, \ldots, R_K)$ such that

$$R_j \ge \frac{q}{t_j} h(\beta_j t_j) - \frac{q}{t_{j-1}} h(\beta_j t_{j-1}), \quad j = 1, \ldots, K \tag{21}$$

$$R_d = \frac{q}{t_d} h(\beta_d t_d) - \frac{q}{t_{d-1}} h(\beta_d t_{d-1}) + \delta, \ \exists d \in \{1, \ldots, K\}, \ \delta > 0 \tag{22}$$

is achievable, where t_j are as in (19) and (20).

Theorem 5 indicates that no rate point $(R_1, \ldots, R_K)$ outside the achievable region of the NE scheme is achievable because if there exists an achievable rate point $(R_1, \ldots, R_K)$ outside the NE scheme's achievable region determined by (19), then there must exist a boundary point $(R_1^*, \ldots, R_K^*)$ on the NE scheme's achievable region such that $R_j \geq R_j^*$ for all $j = 1, \ldots, K$, and $R_d > R_d^*$ for some $d \in \{1, \ldots, K\}$.

The proof of Theorem 5 uses the same basic approach as the proof of the converse of the coding theorem for Gaussian BCs [2]. Lemma 1 below plays the same role in this proof as the EPI does in the proof for Gaussian BCs. We state and prove Lemma 1 and then proceed with the proof of Theorem 5.

Lemma 1: Consider the Markov chain $U \to \boldsymbol{X} \to \boldsymbol{Y} \to \boldsymbol{Z}$ with $\sum_{i=1}^{N} \Pr(X_i = 0)/N = q$. If

$$H(\boldsymbol{Y}|U) \geq N \cdot \frac{q}{p} \cdot h(\beta_1 p)$$

for some $p \in [q, 1]$, then

$$\begin{aligned} H(\boldsymbol{Z}|U) &\geq N \cdot \frac{q}{p} \cdot h(\beta_2 p) \\ &= N \cdot \frac{q}{p} \cdot h(\beta_1 p \beta_\Delta). \end{aligned}$$

The proof of Lemma 1 is given in Appendix C.

Proof of Theorem 5: The proof is by contradiction. To this end, suppose that the rates of (22) are achievable, which means that the probability of decoding error for each receiver can be upper bounded by an arbitrarily small ϵ for sufficiently large N

$$\Pr\{\hat{W}_j \neq W_j | \boldsymbol{Y}^{(j)}\} < \epsilon, \quad j = 1, \ldots, K.$$

By Fano's inequality, this implies that

$$H(W_j|\boldsymbol{Y}^{(j)}) \leq h(\epsilon) + \epsilon \ln(M_j - 1), \; j = 1, \ldots, K. \tag{23}$$

Let $o(\epsilon)$ represent any function of ϵ such that $o(\epsilon) \geq 0$ and $o(\epsilon) \to 0$ as $\epsilon \to 0$. Equation (23) implies that $H(W_j|\boldsymbol{Y}^{(j)}), \; j = 1, \ldots, K$, are all $o(\epsilon)$. Therefore

$$\begin{aligned} H(W_j) &= H(W_j|W_{j+1}, \ldots, W_K) && (24) \\ &= I(W_j; \boldsymbol{Y}^{(j)}|W_{j+1}, \ldots, W_K) \\ &\quad + H(W_j|\boldsymbol{Y}^{(j)}, W_{j+1}, \ldots, W_K) \\ &\leq I(W_j; \boldsymbol{Y}^{(j)}|W_{j+1}, \ldots, W_K) + H(W_j|\boldsymbol{Y}^{(j)}) \\ &= H(\boldsymbol{Y}^{(j)}|W_{j+1}, \ldots, W_K) \\ &\quad - H(\boldsymbol{Y}^{(j)}|W_j, W_{j+1}, \ldots, W_K) + o(\epsilon), && (25) \end{aligned}$$

where (24) follows from the independence of the W_j, $j = 1, \ldots, K$. From (22), (25) and the fact that $NR_j \leq H(W_j)$

$$\begin{aligned} &H(\boldsymbol{Y}^{(j)}|W_{j+1}, \ldots, W_K) - H(\boldsymbol{Y}^{(j)}|W_j, W_{j+1}, \ldots, W_K) \\ &\geq N\frac{q}{t_j} h(\beta_j t_j) - N\frac{q}{t_{j-1}} h(\beta_j t_{j-1}) - o(\epsilon). \end{aligned} \tag{26}$$

Next, using Lemma 1 and (26), we show in Appendix D that

$$H(\boldsymbol{Y}^{(K)}) \geq Nh(\beta_K q) + N\delta - o(\epsilon) \tag{27}$$

where $q = t_K = \sum_{i=1}^{N} \Pr(X_i = 0)/N$. Since ϵ can be arbitrarily small for sufficient large N, $o(\epsilon) \to 0$ as $N \to \infty$. For sufficiently large N, $H(\boldsymbol{Y}^{(K)}) \geq Nh(\beta_K q) + N\delta/2$. However, this contradicts

$$\begin{aligned} H(\boldsymbol{Y}^{(K)}) &\overset{(a)}{\leq} \sum_{i=1}^{N} H(Y_i^{(K)}) \\ &= \sum_{i=1}^{N} h\left(\beta_K \cdot \Pr(X_i = 0)\right) \\ &\overset{(b)}{\leq} Nh\left(\beta_K \cdot \sum_{i=1}^{N} \Pr(X_i = 0)/N\right) \\ &\overset{(c)}{=} Nh(\beta_K q). \end{aligned}$$

Some of these steps are justified as follows:

(a) follows from $\boldsymbol{Y}^{(K)} = (Y_1^{(K)}, \ldots, Y_N^{(K)})$;

(b) is obtained by applying Jensen's inequality to the concave function $h(\cdot)$;

(c) follows from $q = \sum_{i=1}^{N} \Pr(X_i = 0)/N$.

The desired contradiction has been obtained, so the theorem is proved. ■

V. IS DBCs

The IS channel was first introduced in [9] and studied further in [15], [16], and [21]. The definition of the IS channel is as follows: let Φ_n denote the symmetric group of permutations of n objects by $n \times n$ permutation matrices. An n-input m-output channel with transition probability matrix $T_{m \times n}$ is IS if the set

$$\mathcal{G}_T = \{G \in \Phi_n | \exists \Pi \in \Phi_m, \text{ s.t. } TG = \Pi T\}$$

is transitive, which means for any $i, j \in \{1, \ldots, n\}$, there exists a permutation matrix $G \in \mathcal{G}_T$ which maps the ith row to the jth row [9]. An important property of IS channels is that the uniform distribution achieves capacity. We extend the definition of the IS channel to the IS DBC as follows.

Definition 2: (IS DBC): A discrete memoryless DBC $X \to Y \to Z$ with $|\mathcal{X}| = k$, $|\mathcal{Y}| = n$, and $|\mathcal{Z}| = m$ is IS if the set $\mathcal{G}_{T_{YX}, T_{ZX}}$ is transitive where

$$\begin{aligned} \mathcal{G}_{T_{YX}, T_{ZX}} &\overset{\triangle}{=} \mathcal{G}_{T_{YX}} \cap \mathcal{G}_{T_{ZX}} \\ &= \big\{G \in \Phi_k | \exists \Pi_{YX} \in \Phi_n, \Pi_{ZX} \in \Phi_m, \\ &\qquad \text{s.t. } T_{YX}G = \Pi_{YX}T_{YX}, T_{ZX}G = \Pi_{ZX}T_{ZX}\big\}. \end{aligned}$$

Lemmas 2 and 3 below establish basic properties of $\mathcal{G}_{T_{YX}, T_{ZX}}$.

Lemma 2: $\mathcal{G}_{T_{YX}, T_{ZX}}$ is a group under matrix multiplication.

The proof of Lemma 2 is given in Appendix E.

Lemma 3: Let $l = |\mathcal{G}_{T_{YX}, T_{ZX}}|$ so that $\mathcal{G}_{T_{YX}, T_{ZX}} \overset{\triangle}{=} \mathcal{G}_{T_{YX}} \cap \mathcal{G}_{T_{ZX}} = \{G_1, \ldots, G_l\}$. Also, let $k = |\mathcal{X}|$. Then, $\sum_{i=1}^{l} G_i = \frac{l}{k}\mathbf{1}\mathbf{1}^T$, where $\frac{l}{k}$ is an integer and $\mathbf{1}$ is an all-ones vector.

The proof of Lemma 3 is given in Appendix F.

Definition 3: (Smallest Transitive Set): A subset of $\mathcal{G}_{T_{YX}, T_{ZX}}$, $\{G_{i_1}, \ldots, G_{i_{l_s}}\}$, is a smallest transitive subset of $\mathcal{G}_{T_{YX}, T_{ZX}}$ if

$$\sum_{j=1}^{l_s} G_{i_j} = \frac{l_s}{k}\mathbf{1}\mathbf{1}^T \tag{28}$$

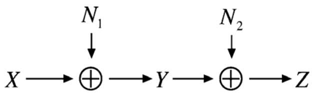

Fig. 10. Group-operation DBC.

where $\frac{l_s}{k}$ is the smallest possible integer for which (28) is satisfied.

A. Examples: BS BCs and Binary-Erasure BCs

The class of IS DBCs includes most of the common discrete memoryless DBCs. For example, the BS BC $X \to Y \to Z$ with marginal transition probability matrices

$$T_{YX} = \begin{bmatrix} 1-\alpha_1 & \alpha_1 \\ \alpha_1 & 1-\alpha_1 \end{bmatrix} \text{and } T_{ZX} = \begin{bmatrix} 1-\alpha_2 & \alpha_2 \\ \alpha_2 & 1-\alpha_2 \end{bmatrix}$$

where $0 \le \alpha_1 \le \alpha_2 \le 1/2$ is IS since

$$\mathcal{G}_{T_{YX},T_{ZX}} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \tag{29}$$

is transitive.

Another interesting example is the binary-erasure BC with marginal transition probability matrices

$$T_{YX} = \begin{bmatrix} 1-a_1 & 0 \\ a_1 & a_1 \\ 0 & 1-a_1 \end{bmatrix} \text{and } T_{ZX} = \begin{bmatrix} 1-a_2 & 0 \\ a_2 & a_2 \\ 0 & 1-a_2 \end{bmatrix}$$

where $0 \le a_1 \le a_2 \le 1$. It is IS since its $\mathcal{G}_{T_{YX},T_{ZX}}$ is the same as that of the BS BC shown in (29).

B. Group-Operation DBCs Are IS

We now define group-operation DBCs and show that they are IS.

Definition 4: (Group-Operation DBC): A discrete DBC $X \to Y \to Z$ with $\mathcal{X}, \mathcal{Y}, \mathcal{Z} = \{1, \dots, n\}$ is a group-operation DBC if there exist two n-ary random variables N_1 and N_2 such that $Y \sim X \oplus N_1$ and $Z \sim Y \oplus N_2$ as shown in Fig. 10, where $\sim$ denotes identical distribution and $\oplus$ denotes a group operation which is an operation that satisfies the group axioms on the set $\{1, \dots, n\}$.

Group-operation DBCs include the BS BC and the discrete additive DBC of [10] as special cases. It is also a channel model for Gaussian broadcast communication systems with phase-shift-keying modulation at the transmitter and direct hard decisions on modulated symbols at the receivers.

Theorem 6: Group-operation DBCs are IS.

Proof: For the group-operation DBC $X \to Y \to Z$ with $\mathcal{X}, \mathcal{Y}, \mathcal{Z} = \{1, \dots, n\}$, let G_x for $x = 1, \dots, n$ be 0–1 matrices with entries

$$G_x(i,j) = \begin{cases} 1, & \text{if } j \oplus x = i \\ 0, & \text{otherwise} \end{cases} \quad \text{for } i,j = 1, \dots, n.$$

G_x for $x = 1, \dots, n$ are actually permutation matrices and have the property that $G_{x_1} \cdot G_{x_2} = G_{x_2} \cdot G_{x_1} = G_{x_1 \oplus x_2}$. Let $[\gamma_1, \dots, \gamma_n]^T$ be the distribution of N_1. Since Y has the same distribution as $X \oplus N_1$, one has

$$T_{YX} = \sum_{x=1}^{n} \gamma_x G_x.$$

Hence, $T_{YX} G_x = G_x T_{YX}$ for all $x = 1, \dots, n$. Similarly, we have $T_{ZX} G_x = G_x T_{ZX}$ for all $x = 1, \dots, n$, and so

$$\{G_1, \dots, G_n\} \subseteq \mathcal{G}_{T_{YX},T_{ZX}}.$$

Since the set $\{G_1, \dots, G_n\}$ is transitive by definition, $\mathcal{G}_{T_{YX},T_{ZX}}$ is also transitive, and hence, the group-operation DBC is IS. ■

By definition, $\sum_{j=1}^{n} G_j = \mathbf{1}\mathbf{1}^T$, and hence, $\{G_1, \dots, G_n\}$ is a smallest transitive subset of $\mathcal{G}_{T_{YX},T_{ZX}}$ for the group-operation DBC.

C. Note on DDICs

We briefly note that while DDICs and their related DBCs are closely related to IS-DBCs, the class of IS-DBCs is not addressed by [15] or [16]. The class of DDICs and the corresponding DBCs studied in [15] and [16] have to satisfy the condition that the transition probability matrix T_{ZY} is IS, i.e., $\mathcal{G}_{T_{ZY}}$ is transitive. The IS DBC, however, does not have to satisfy this condition. The following example provides an IS-DBC, which is not covered in [15] and [16]. Consider a binary-input DBC $X \to Y \to Z$ with transition probability matrices

$$T_{YX} = \begin{bmatrix} a & c \\ b & d \\ c & a \\ d & b \end{bmatrix}, \; T_{ZY} = \begin{bmatrix} e & f & g & h \\ g & h & e & f \end{bmatrix}$$

and

$$T_{ZX} = T_{ZY} T_{YX} = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}$$

where $a+c = b+d = 1, e+f+g+h = 1, \alpha = ae+bf+cg+dh$, and $\beta = ag + bh + ce + df$. This DBC is IS since its $\mathcal{G}_{T_{YX},T_{ZX}}$ is the same as that of the broadcast BS channel shown in (29). It is not covered by the results of [15] and [16] because

$$\mathcal{G}_{T_{ZY}} = \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \right\}$$

is *not* transitive.

D. IS-DBC Optimal Input Distributions and Capacity Regions

Consider the IS DBC $X \to Y \to Z$ with the marginal transition probability matrices T_{YX} and T_{ZX}. Recall that the set $\mathcal{C}$ is the set of all $(\boldsymbol{p}_X, s, \eta)$ satisfying (1)–(3) for some choice of l, $\boldsymbol{w}$, and T_{XU}, the set $\mathcal{C}^* = \{(s, \eta) | (\boldsymbol{p}_X, s, \eta) \in \mathcal{C} \text{ for some } \boldsymbol{p}_X\}$ is the projection of the set $\mathcal{C}$ on the (s, η)-plane, and the set $\mathcal{C}^*_{\boldsymbol{q}}$ is the subset of $\mathcal{C}^*$ for which $\boldsymbol{p}_X = \boldsymbol{q}$.

Lemma 4: For any permutation matrix $G \in \mathcal{G}_{T_{YX},T_{ZX}}$ and $(\boldsymbol{p}, s, \eta) \in \mathcal{C}$, $(G\boldsymbol{p}, s, \eta) \in \mathcal{C}$.

The proof of Lemma 4 is given in Appendix G.

Corollary 1: $\forall \boldsymbol{p} \in \Delta_k$ and $G \in \mathcal{G}_{T_{YX},T_{ZX}}$, one has $\mathcal{C}^*_{G\boldsymbol{p}} = \mathcal{C}^*_{\boldsymbol{p}}$, and so, $F^*(G\boldsymbol{p}, s) = F^*(\boldsymbol{p}, s)$ for any $H(Y|X) \le s \le H(Y)$.

Lemma 5: For any IS DBC, $\mathcal{C}^* = \mathcal{C}^*_{\boldsymbol{u}}$, where $\boldsymbol{u}$ denotes the uniform distribution.

The proof of Lemma 5 is given in Appendix H.

Now, we state and prove that the uniformly distributed X is optimal for IS DBCs.

Theorem 7: For any IS DBC, its capacity region can be achieved by using the transmission strategies such that the broadcast signal X is uniformly distributed. As a consequence, the capacity region is

$$\overline{\text{co}}\left\{(R_1,R_2)\colon \begin{array}{c} R_1 \le s - h_n(T_{YX}\boldsymbol{e_1}) \\ R_2 \le h_m(T_{ZX}\boldsymbol{u}) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{u}, s) \\ h_n(T_{YX}\boldsymbol{e_1}) \le s \le \ln(n) \end{array}\right\} \tag{30}$$

where $\boldsymbol{e_1} = [1, 0, \ldots, 0]^T$, $n = |\mathcal{Y}|$, and $m = |\mathcal{Z}|$.

Proof: Let $\boldsymbol{q} = [q_1, \ldots, q_k]^T$ be the distribution of the channel input X for the IS DBC $X \to Y \to Z$. Since $\mathcal{G}_{T_{YX}}$ is transitive, the columns of T_{YX} are permutations of each other

$$\begin{aligned} H(Y|X) &= \sum_{i=1}^{k} q_i H(Y|X=i) \\ &= \sum_{i=1}^{k} q_i h_n(T_{YX}\boldsymbol{e_i}) \\ &= \sum_{i=1}^{k} q_i h_n(T_{YX}\boldsymbol{e_1}) \\ &= h_n(T_{YX}\boldsymbol{e_1}), \end{aligned} \tag{31}$$

which is independent of $\boldsymbol{q}$. Let $l = |\mathcal{G}_{T_{YX},T_{ZX}}|$ and $\mathcal{G}_{T_{YX},T_{ZX}} = \{G_1, \ldots, G_l\}$

$$\begin{aligned} H(Z) &= h_m(T_{ZX}\boldsymbol{q}) \\ &= \frac{1}{l}\sum_{i=1}^{l} h_m(T_{ZX}G_i\boldsymbol{q}) \\ &\le h_m\left(T_{ZX}\sum_{i=1}^{l}\frac{1}{l}G_i\boldsymbol{q}\right) &\quad (32) \\ &= h_m(T_{ZX}\boldsymbol{u}) &\quad (33) \end{aligned}$$

where (32) follows from Jensen's inequality. Since $\mathcal{C}^* = \mathcal{C}^*_{\boldsymbol{u}}$ for the IS DBC

$$F^*(\boldsymbol{q}, s) \ge F^*(\boldsymbol{u}, s). \tag{34}$$

Plugging (31), (33), and (34) into (6), the expression of the capacity region for the DBC, the capacity region for IS DBCs is

$$\begin{aligned} &\overline{\text{co}}\left[\bigcup_{\boldsymbol{p}_X=\boldsymbol{q}\in\Delta_k}\left\{(R_1,R_2)\colon \begin{array}{l} R_1 \le s - H(Y|X) \\ R_2 \le H(Z) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s) \end{array}\right\}\right] &\quad (35) \\ &\subseteq \overline{\text{co}}\left[\bigcup_{\boldsymbol{p}_X=\boldsymbol{q}\in\Delta_k}\left\{(R_1,R_2)\colon \begin{array}{l} R_1 \le s - h_n(T_{YX}\boldsymbol{e_1}) \\ R_2 \le h_m(T_{ZX}\boldsymbol{u}) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{u}, s) \end{array}\right\}\right] \\ &= \overline{\text{co}}\left\{(R_1,R_2) : \begin{array}{l} R_1 \le s - h_n(T_{YX}\boldsymbol{e_1}) \\ R_2 \le h_m(T_{ZX}\boldsymbol{u}) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{u}, s) \end{array}\right\} &\quad (36) \\ &= \overline{\text{co}}\left\{(R_1,R_2) : \begin{array}{l} \boldsymbol{p}_X = \boldsymbol{u} \\ R_1 \le s - H(Y|X) \\ R_2 \le H(Z) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{u}, s) \end{array}\right\} \\ &\subseteq \overline{\text{co}}\left[\bigcup_{\boldsymbol{p}_X=\boldsymbol{q}\in\Delta_k}\left\{(R_1,R_2)\colon \begin{array}{l} R_1 \le s - H(Y|X) \\ R_2 \le H(Z) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{q}, s) \end{array}\right\}\right]. &\quad (37) \end{aligned}$$

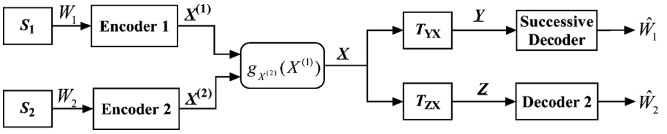

Fig. 11. Block diagram of the permutation encoding approach.

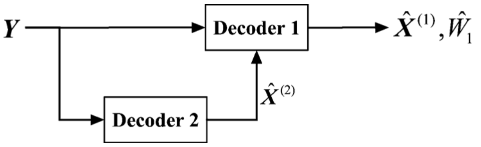

Fig. 12. Structure of the successive decoder for IS DBCs.

Note that (35) and (37) are identical expressions; hence, (35)–(37) are all equal. Therefore, (30) and (36) express the capacity region for the IS DBC, which also means that the capacity region can be achieved by using transmission strategies where the broadcast signal X is uniformly distributed. ■

E. Permutation Encoding and its Optimality for IS-DBCs

The permutation encoding approach is an independent-encoding scheme that achieves the capacity region for IS DBCs. The block diagram of this approach is shown in Fig. 11. In Fig. 11, W_1 is the message for Receiver 1, which sees the less degraded channel T_{YX}, and W_2 is the message for Receiver 2, which sees the more degraded channel T_{ZX}. The permutation encoding approach is first to independently encode these two messages into two codewords $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$, and then to combine these two independent codewords using a single-letter operation.

Let $\mathcal{G}_s$ be a smallest transitive subset of $\mathcal{G}_{T_{YX},T_{ZX}}$. Denote $k = |\mathcal{X}|$ and $l_s = |\mathcal{G}_s|$. Use a random coding technique to design the codebook for Receiver 1 according to the k-ary random variable $X^{(1)}$ with distribution $\boldsymbol{p_1}$ and the codebook for Receiver 2 according to the l_s-ary random variable $X^{(2)}$ with uniform distribution. Let $\mathcal{G}_s = \{G_1, \ldots, G_{l_s}\}$. Define the permutation function $g_{x^{(2)}}(x^{(1)}) = x$ if the permutation matrix $G_{x^{(2)}}$ maps the $x^{(1)}$th column to the xth column, where $x^{(2)} \in \{1, \ldots, l_s\}$ and $x, x^{(1)} \in \{1, \ldots, k\}$. Hence, $g_{x^{(2)}}(x^{(1)}) = x$ if and only if the $x^{(1)}$th row, xth column entry of $G_{x^{(2)}}$ is 1. The permutation encoding approach is then to broadcast $\boldsymbol{X}$ which is obtained by applying the single-letter permutation function $X = g_{X^{(2)}}(X^{(1)})$ on symbols of codewords $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$. Since $X^{(2)}$ is uniformly distributed and $\sum_{j=1}^{l_s} G_j = \frac{l_s}{k}\boldsymbol{1}\boldsymbol{1}^T$, the broadcast signal X is also uniformly distributed.

Receiver 2 receives $\boldsymbol{Z}$ and decodes the desired message directly. Receiver 1 receives $\boldsymbol{Y}$ and successively decodes the message for Receiver 2 and then for Receiver 1. The structure of the successive decoder is shown in Fig. 12. Note that Decoder 1 in Fig. 12 is *not* a joint decoder even though it has two inputs $\boldsymbol{Y}$ and $\hat{\boldsymbol{X}}^{(2)}$

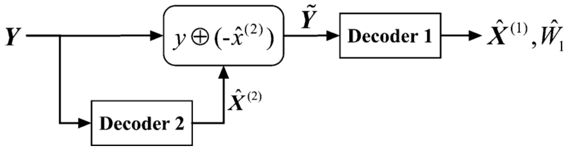

Fig. 13. Structure of the successive decoder for degraded group-operation DBCs.

In particular, for the group-operation DBC with $Y \sim X \oplus N_1$ and $Z \sim Y \oplus N_2$, the permutation function $g_{x^{(2)}}(x^{(1)})$ is the group operation $x^{(2)} \oplus x^{(1)}$. Hence, the permutation encoding approach for the group-operation DBC is the NE scheme for the group-operation DBC. The successive decoder for the group-operation DBC is shown in Fig. 13, where

$$\tilde{y} = y \oplus (-\hat{x}^{(2)}).$$

From the analysis of successive decoding in the proof of the coding theorem for DBCs [2], [3], the achievable region of the permutation encoding approach for the IS DBC is determined by

$$\begin{aligned}
R_1 &\le I(X;Y|X^{(2)})\\
&= H(Y|X^{(2)}) - H(Y|X)\\
&= \sum_{x^{(2)}=1}^{l_s} \Pr(X^{(2)} = x^{(2)}) H(Y|X^{(2)} = x^{(2)})\\
&\quad - \sum_{x=1}^{k} \Pr(X = x) H(Y|X = x)\\
&= \sum_{x^{(2)}=1}^{l_s} \Pr(X^{(2)} = x^{(2)}) h_n(T_{YX} G_{x^{(2)}} \boldsymbol{p_1})\\
&\quad - \sum_{x=1}^{k} \Pr(X = x) h_n(T_{YX} \boldsymbol{e}_x)\\
&= \sum_{x^{(2)}=1}^{l_s} \Pr(X^{(2)} = x^{(2)}) h_n(\Pi_{YX,x^{(2)}} T_{YX} \boldsymbol{p_1})\\
&\quad - \sum_{x=1}^{k} \Pr(X = x) h_n(T_{YX} \boldsymbol{e}_1)\\
&= h_n(T_{YX}\boldsymbol{p_1}) - h_n(T_{YX}\boldsymbol{e}_1)
\end{aligned}$$

and

$$\begin{aligned}
R_2 &\le I(X^{(2)};Z)\\
&= H(Z) - H(Z|X^{(2)})\\
&= h_m(T_{ZX}\boldsymbol{u}) - \sum_{x^{(2)}=1}^{l_s} \Pr(X^{(2)} = x^{(2)}) h_m(T_{ZX} G_{x^{(2)}} \boldsymbol{p_1})\\
&= h_m(T_{ZX}\boldsymbol{u}) - \sum_{x^{(2)}=1}^{l_s} \Pr(X^{(2)} = x^{(2)}) h_m(\Pi_{ZX,x^{(2)}} T_{ZX} \boldsymbol{p_1})\\
&= h_m(T_{ZX}\boldsymbol{u}) - h_m(T_{ZX}\boldsymbol{p_1})
\end{aligned}$$

where $\boldsymbol{u}$ is the k-ary uniform distribution, $\boldsymbol{p}_1$ is the distribution of $X^{(1)}$, and $\boldsymbol{e}_x$ is a 0–1 vector such that the xth entry is 1 and all other entries are 0. Hence, the achievable region is

$$\overline{\text{co}}\left[\bigcup_{\boldsymbol{p_1}\in\Delta_k}\left\{(R_1,R_2)\colon \begin{array}{l} R_1 \le h_n(T_{YX}\boldsymbol{p_1}) - h_n(T_{YX}\boldsymbol{e_1}) \\ R_2 \le h_m(T_{ZX}\boldsymbol{u}) - h_m(T_{ZX}\boldsymbol{p_1}) \end{array}\right\}\right]. \tag{38}$$

Define $\tilde{F}(s)$ as the infimum of $h_m(T_{ZX}\boldsymbol{p_1})$ with respect to all distributions $\boldsymbol{p_1}$ such that $h_n(T_{YX}\boldsymbol{p_1}) = s$. Hence, the achievable region (38) can be expressed as

$$\left\{(R_1,R_2)\colon \begin{array}{c} R_1 \le s - h_n(T_{YX}\boldsymbol{e_1}) \\ R_2 \le h_m(T_{ZX}\boldsymbol{u}) - \underline{\text{env}}\tilde{F}(s) \\ h_n(T_{YX}\boldsymbol{e_1}) \le s \le h_n(T_{YX}\boldsymbol{u}) \end{array}\right\} \tag{39}$$

where $\underline{\text{env}}\tilde{F}(s)$ denotes the lower convex envelope of $\tilde{F}(s)$.

Theorem 8: The permutation encoding approach achieves the capacity region for IS DBCs, which is expressed in (30), (38), and (39).

Proof: In order to show that the achievable region (39) is the same as the capacity region (30) for the IS DBC, it suffices to show that

$$\underline{\text{env}}\tilde{F}(s) \le F^*(\boldsymbol{u}, s).$$

For any $p(u,x)$ with uniformly distributed X

$$\begin{aligned}
H(Z|U) &= \sum_u \Pr(U = u) H(Z|U = u)\\
&= \sum_u \Pr(U = u) h_m(T_{ZX}\boldsymbol{p}_{X|U=u})\\
&\overset{(a)}{\ge} \sum_u \Pr(U = u) \tilde{F}(h_n(T_{YX}\boldsymbol{p}_{X|U=u}))\\
&\ge \sum_u \Pr(U = u) \underline{\text{env}}\tilde{F}\left(h_n(T_{YX}\boldsymbol{p}_{X|U=u})\right)\\
&\overset{(b)}{\ge} \underline{\text{env}}\tilde{F}\left(\sum_u \Pr(U = u) h_n(T_{YX}\boldsymbol{p}_{X|U=u})\right)\\
&= \underline{\text{env}}\tilde{F}(H(Y|U))
\end{aligned} \tag{40}$$

where $\boldsymbol{p}_{X|U=u}$ is the conditional distribution of X given $U = u$. Some of these steps are justified as follows:

1) (a) follows from the definition of $\tilde{F}(s)$;
2) (b) follows from Jensen's inequality.

Combining (40) and the definition of F^*, one has $\underline{\text{env}}\tilde{F}(s) \le F^*(\boldsymbol{u}, s)$. ■

Corollary 2: The NE scheme achieves the capacity region for group-operation DBCs.

Conjecture 1: The alphabet size of the code for Receiver 2 l_s is equal to the alphabet size of the channel input k in a permutation encoding approach for any IS DBC. In other words, a smallest transitive subset $\{G_1, \ldots, G_{l_s}\}$ of $\mathcal{G}_{T_{YX},T_{ZX}}$ for any IS DBC has

$$\sum_{j=1}^{l_s} G_j = \mathbf{1}\mathbf{1}^T.$$

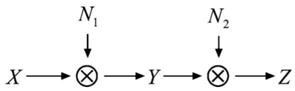

Fig. 14. DM DBC.

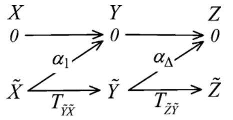

Fig. 15. Channel structure of a DBC with erasures.

VI. DM DBCs

Definition 5: (DM): A commutative operation on two inputs from the set $\{0, 1, \dots, n\}$ is a DM if it satisfies the group axioms on $\{1, \dots, n\}$, and also produces zero if either input is zero. Use $\otimes$ to denote DM.

Definition 6: (DM DBC): A discrete DBC $X \to Y \to Z$ with $\mathcal{X}, \mathcal{Y}, \mathcal{Z} = \{0, 1, \dots, n\}$ is a DM DBC if there exist two $(n+1)$-ary random variables N_1 and N_2 such that $Y \sim X \otimes N_1$ and $Z \sim Y \otimes N_2$ as shown in Fig. 14.

As an example, the DM DBC with $n = 1$ is the broadcast Z channel, which is studied in Section IV. By the definition of DM, the DM DBC $X \to Y \to Z$ has the channel structure as shown in Fig. 15. The subchannel $\tilde{X} \to \tilde{Y} \to \tilde{Z}$ is a group-operation DBC with transition matrices $T_{\tilde{Y}\tilde{X}}$ and $T_{\tilde{Z}\tilde{X}} = T_{\tilde{Z}\tilde{Y}} T_{\tilde{Y}\tilde{X}}$, where $\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Z}} = \{1, \dots, n\}$. For the DM DBC $X \to Y \to Z$, if the channel input X is zero, the channel outputs Y and Z are also zeros. If the channel input is a nonzero symbol, the channel output Y is zero with probability α_1 and Z is zero with probability α_2, where $\alpha_2 = \alpha_1 + (1 - \alpha_1)\alpha_\Delta$. Therefore, the transition matrices for $X \to Y \to Z$ are

$$T_{YX} = \begin{bmatrix} 1 & \alpha_1 \mathbf{1}^T \\ \mathbf{0} & (1-\alpha_1) T_{\tilde{Y}\tilde{X}} \end{bmatrix}$$
$$T_{ZY} = \begin{bmatrix} 1 & \alpha_\Delta \mathbf{1}^T \\ \mathbf{0} & (1-\alpha_\Delta) T_{\tilde{Z}\tilde{Y}} \end{bmatrix}$$

and

$$\begin{aligned} T_{ZX} &= T_{ZY} T_{YX} \\ &= \begin{bmatrix} 1 & \alpha_\Delta \mathbf{1}^T \\ \mathbf{0} & (1-\alpha_\Delta) T_{\tilde{Z}\tilde{Y}} \end{bmatrix} \begin{bmatrix} 1 & \alpha_1 \mathbf{1}^T \\ \mathbf{0} & (1-\alpha_1) T_{\tilde{Y}\tilde{X}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \alpha_2 \mathbf{1}^T \\ \mathbf{0} & (1-\alpha_2) T_{\tilde{Z}\tilde{X}} \end{bmatrix} \end{aligned}$$

where $\mathbf{1}$ is an all-ones vector and $\mathbf{0}$ is an all-zeros vector.

A. Optimal Input Distribution

The subchannel $\tilde{X} \to \tilde{Y} \to \tilde{Z}$ is a group-operation DBC, and hence, $\mathcal{G}_{T_{\tilde{Y}\tilde{X}}, T_{\tilde{Z}\tilde{X}}}$ is transitive. For any $n \times n$ permutation matrix $\tilde{G} \in \mathcal{G}_{T_{\tilde{Y}\tilde{X}}, T_{\tilde{Z}\tilde{X}}}$ with $T_{\tilde{Y}\tilde{X}} \tilde{G} = \tilde{\Pi}_{\tilde{Y}\tilde{X}} T_{\tilde{Y}\tilde{X}}$ and $T_{\tilde{Z}\tilde{X}} \tilde{G} = \tilde{\Pi}_{\tilde{Z}\tilde{X}} T_{\tilde{Z}\tilde{X}}$, the $(n+1) \times (n+1)$ permutation matrix

$$G = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \tilde{G} \end{bmatrix}$$

has

$$T_{YX} G = \begin{bmatrix} 1 & \alpha_1 \mathbf{1}^T \\ \mathbf{0} & (1-\alpha_1) T_{\tilde{Y}\tilde{X}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \tilde{G} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \tilde{\Pi}_{\tilde{Y}\tilde{X}} \end{bmatrix} T_{YX}$$

and so $G \in \mathcal{G}_{T_{YX}}$. Similarly, $G \in \mathcal{G}_{T_{ZX}}$, and hence, $G \in \mathcal{G}_{T_{YX}, T_{ZX}}$. Therefore, for any $i, j \in \{1, \dots, n\}$, there exists a permutation matrix $G \in \mathcal{G}_{T_{YX}, T_{ZX}}$ that maps the $(i+1)$th row (corresponding to the element i) to the $(j+1)$th row (corresponding to the element j). However, there is no matrix in $\mathcal{G}_{T_{YX}, T_{ZX}}$ that maps the first row (corresponding to the element 0) to other rows (corresponding nonzero elements) or vice versa. Hence, any permutation matrix $G \in \mathcal{G}_{T_{YX}, T_{ZX}}$ has

$$G = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \tilde{G} \end{bmatrix}$$

for some $\tilde{G} \in \mathcal{G}_{T_{\tilde{Y}\tilde{X}}, T_{\tilde{Z}\tilde{X}}}$. These results may be summarized in the following lemma.

Lemma 6: Let $\mathcal{G}_{T_{\tilde{Y}\tilde{X}}, T_{\tilde{Z}\tilde{X}}} = \{\tilde{G}_1, \dots, \tilde{G}_l\}$. Hence, $\mathcal{G}_{T_{YX}, T_{ZX}} = \{G_1, \dots, G_l\}$, where

$$G_j = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \tilde{G}_j \end{bmatrix}$$

for $j = 1, \dots, l$.

Lemma 7 states that the uniformly distributed $\tilde{X}$ is optimal for the DM DBC.

Lemma 7: Let $\boldsymbol{p}_X = [1 - q, q\boldsymbol{p}_{\tilde{X}}^T]^T \in \Delta_{n+1}$ be the distribution of channel input X, where $\boldsymbol{p}_{\tilde{X}}$ is the distribution of $\tilde{X}$. For any DM DBC, $\mathcal{C}^*_{\boldsymbol{p}_X} \subseteq \mathcal{C}^*_{[1-q, q\boldsymbol{u}^T]^T}$ and $\mathcal{C}^* = \bigcup_{q \in [0,1]} \mathcal{C}^*_{[1-q, q\boldsymbol{u}^T]^T}$, where $\boldsymbol{u} \in \Delta_n$ denotes the uniform distribution.

The proof of Lemma 7 is given in Appendix I.

Theorem 9: The capacity region of the DM DBC can be achieved by using transmission strategies where $\tilde{X}$ is uniformly distributed, i.e., the distribution of X has $\boldsymbol{p}_X = [1 - q, q\boldsymbol{u}^T]^T$ for some $q \in [0, 1]$. As a consequence, the capacity region is

$$\begin{aligned} \overline{\text{co}}\Big[\bigcup_{q \in [0,1]} \big\{ (R_1, R_2) : R_1 &\le s - q h_n(T_{\tilde{Y}\tilde{X}} \boldsymbol{e_1}) \\ R_2 &\le h((1-\alpha_2)q) + (1-\alpha_2) q \ln(n) \\ &\quad - F^*_{T_{YX}, T_{ZX}}([1-q, q\boldsymbol{u}^T]^T, s) \big\} \Big]. \end{aligned}$$

Proof: Let $\boldsymbol{p}_X = [1 - q, q\boldsymbol{p}_{\tilde{X}}]^T$ be the distribution of the channel input X, where $\boldsymbol{p}_{\tilde{X}} = [p_1, \dots, p_n]^T$. Since $\mathcal{G}_{T_{\tilde{Y}\tilde{X}}}$ is transitive and the columns of $T_{\tilde{Y}\tilde{X}}$ are permutations of each other

$$\begin{aligned} H(Y|X) &= \sum_{i=0}^{n} \Pr(X = i) H(Y|X = i) \\ &= (1-q) H(Y|X = 0) + \sum_{i=1}^{n} q p_i h_n(T_{\tilde{Y}\tilde{X}} \boldsymbol{e_i}) \\ &= \sum_{i=1}^{n} q p_i h_n(T_{\tilde{Y}\tilde{X}} \boldsymbol{e_1}) \\ &= q h_n(T_{\tilde{Y}\tilde{X}} \boldsymbol{e_1}) \end{aligned} \quad (41)$$

which is independent of $\boldsymbol{p}_X$. Let $\mathcal{G}_{T_{YX},T_{ZX}} = \{G_1, \ldots, G_l\}$

$$\begin{aligned} H(Z) &= h_{n+1}(T_{ZX}\boldsymbol{p}_X) \\ &= \frac{1}{l}\sum_{i=1}^{l} h_{n+1}\left(T_{ZX}G_i\boldsymbol{p}_X\right) \\ &\overset{(a)}{\le} h_{n+1}\left(T_{ZX}\frac{1}{l}\sum_{i=1}^{l} G_i\boldsymbol{p}_X\right) \\ &= h_{n+1}\left(T_{ZX}[1-q, q\boldsymbol{u}^T]^T\right) \\ &= h_{n+1}\left([1-q+\alpha_2 q, (1-\alpha_2)q\boldsymbol{u}]^T\right) \\ &\overset{(b)}{=} h((1-\alpha_2)q) + (1-\alpha_2)q\ln(n) \end{aligned} \tag{42}$$

where (a) follows from Jensen's inequality and (b) follows from the grouping rule for entropy [18, Problem 2.27]. By Lemma 7, $\mathcal{C}^*_{\boldsymbol{p}_X} \subseteq \mathcal{C}^*_{[1-q,q\boldsymbol{u}^T]^T}$ for the DM DBC. Hence

$$F^*(\boldsymbol{p}_X, s) \ge F^*([1-q, q\boldsymbol{u}^T]^T, s). \tag{43}$$

Plugging (41)–(43) into (6), the capacity region for DM DBCs is

$$\begin{aligned} &\overline{\text{co}}\Big[\bigcup_{\boldsymbol{p}_X \in \Delta_k} \{(R_1, R_2) : R_1 \le s - H(Y|X) \\ &\qquad R_2 \le H(Z) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{p}_X, s)\}\Big] \qquad (44) \\ &\subseteq \overline{\text{co}}\Big[\bigcup_{\boldsymbol{p}_X \in \Delta_k} \{(R_1, R_2) : R_1 \le s - h_n(T_{\tilde{Y}\tilde{X}}\boldsymbol{e}_1), \\ &\qquad R_2 \le h((1-\alpha_2)q) + (1-\alpha_2)q\ln(n) \\ &\qquad\quad - F^*_{T_{YX},T_{ZX}}([1-q, q\boldsymbol{u}^T]^T, s)\}\Big] \\ &= \overline{\text{co}}\Big[\bigcup_{q\in[0,1]} \{(R_1, R_2) : R_1 \le s - qh_n(T_{\tilde{Y}\tilde{X}}\boldsymbol{e}_1), \\ &\qquad R_2 \le h((1-\alpha_2)q) + (1-\alpha_2)q\ln(n) \\ &\qquad\quad - F^*_{T_{YX},T_{ZX}}([1-q, q\boldsymbol{u}^T]^T, s)\}\Big] \qquad (45) \\ &= \overline{\text{co}}\Big[\bigcup_{\boldsymbol{p}_X = [1-q, q\boldsymbol{u}^T]^T} \{(R_1, R_2) : R_1 \le s - H(Y|X), \\ &\qquad R_2 \le H(Z) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{p}_X, s)\}\Big] \\ &\subseteq \overline{\text{co}}\Big[\bigcup_{\boldsymbol{p}_X \in \Delta_k} \{(R_1, R_2) : R_1 \le s - H(Y|X), \\ &\qquad R_2 \le H(Z) - F^*_{T_{YX},T_{ZX}}(\boldsymbol{p}_X, s)\}\Big] \qquad (46) \end{aligned}$$

where $\overline{\text{co}}$ denotes the convex hull of the closure. Note that (44) and (46) are identical expressions; hence, (44)–(46) are all equal. Therefore, (45) expresses the capacity region for the DM DBC, which also means that the capacity region can be achieved by using transmission strategies where the broadcast signal X has distribution $\boldsymbol{p}_X = [1-q, q\boldsymbol{u}^T]^T$ for some $q \in [0, 1]$. ■

B. Optimality of the NE Scheme for DM DBCs

The NE scheme for the DM DBC is shown in Fig. 16. W_1 is the message for Receiver 1 who sees the less degraded channel T_{YX} and W_2 is the message for Receiver 2 who sees the more degraded channel T_{ZX}. The NE scheme is first to independently encode these two messages into two codewords $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$, respectively, where $\mathcal{X}^{(1)}, \mathcal{X}^{(2)} = \{0, 1, \ldots, n\}$, and then to broadcast $\boldsymbol{X}$ which is obtained by applying the single-letter function $X = X^{(2)} \otimes X^{(1)}$ on symbols of codewords $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$. The distribution of $X^{(2)}$ is constrained to be $\boldsymbol{p}_{X^{(2)}} = [1-q, q\boldsymbol{u}^T]^T$ for some $q \in [0, 1]$, and hence, the distribution of the broadcast signal X also has $\boldsymbol{p}_X = [1-q, q\boldsymbol{u}^T]^T$ for some $q \in [0, 1]$, which was proved to be the optimal input distribution for the DM DBC. Receiver 2 receives $\boldsymbol{Z}$ and decodes the desired message directly. Receiver 1 receives $\boldsymbol{Y}$ and successively decodes the message for Receiver 2 and then for Receiver 1.

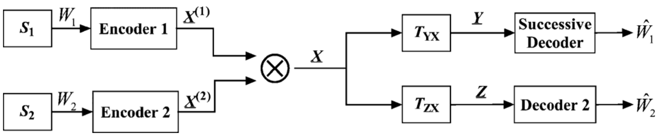

Fig. 16. Block diagram of the NE scheme for the DM DBC.

Let $\boldsymbol{p}_X = [1-q, q\boldsymbol{p}_{\tilde{X}}]^T$ be the distribution of the channel input X, where $\boldsymbol{p}_{\tilde{X}}$ is the distribution of subchannel input $\tilde{X}$. For the DM DBC $X \to Y \to Z$, the ϕ function is

$$\begin{aligned} \phi(\boldsymbol{p}_X, \lambda) &= h_{n+1}(T_{ZX}\boldsymbol{p}_X) - \lambda h_{n+1}(T_{YX}\boldsymbol{p}_X) \\ &= h_{n+1}\left(\begin{bmatrix} 1-q+q\alpha_2 \\ q(1-\alpha_2)T_{\tilde{Z}\tilde{X}}\boldsymbol{p}_{\tilde{X}} \end{bmatrix}\right) \\ &\quad - \lambda h_{n+1}\left(\begin{bmatrix} 1-q+q\alpha_1 \\ q(1-\alpha_1)T_{\tilde{Y}\tilde{X}}\boldsymbol{p}_{\tilde{X}} \end{bmatrix}\right) \\ &= h(q(1-\alpha_2)) - q(1-\alpha_2)h_n\left(T_{\tilde{Z}\tilde{X}}\boldsymbol{p}_{\tilde{X}}\right) \\ &\quad - \lambda\left(h(q(1-\alpha_1)) - q(1-\alpha_1)h_n\left(T_{\tilde{Y}\tilde{X}}\boldsymbol{p}_{\tilde{X}}\right)\right) \\ &= h(q\beta_2) - \lambda h(q\beta_1) \\ &\quad + q\beta_2\left(h_n\left(T_{\tilde{Z}\tilde{X}}\boldsymbol{p}_{\tilde{X}}\right) - \frac{\lambda}{1-\alpha_\Delta}h_n\left(T_{\tilde{Y}\tilde{X}}\boldsymbol{p}_{\tilde{X}}\right)\right) \\ &= h(q\beta_2) - \lambda h(q\beta_1) + q\beta_2\tilde{\phi}\left(\boldsymbol{p}_{\tilde{X}}, \frac{\lambda}{1-\alpha_\Delta}\right) \end{aligned}$$

where $\beta_1 = 1-\alpha_1$, $\beta_2 = 1-\alpha_2$, and $\tilde{\phi}(\boldsymbol{q}, \lambda) \overset{\triangle}{=} h_n(T_{\tilde{Z}\tilde{X}}\boldsymbol{q}) - \lambda h_n(T_{\tilde{Y}\tilde{X}}\boldsymbol{q})$ is the ϕ function defined on the group-operation degraded broadcast subchannel $\tilde{X} \to \tilde{Y} \to \tilde{Z}$.

Define $\tilde{\psi}(\boldsymbol{q}, \lambda) \overset{\triangle}{=} \underline{\text{env}}_{\boldsymbol{q}}\tilde{\phi}(\boldsymbol{q}, \lambda)$ as the ψ function for group-operation degraded broadcast subchannel $\tilde{X} \to \tilde{Y} \to \tilde{Z}$ where the lower envelope is taken with respect to $\boldsymbol{q}$.

For the channel $X \to Y \to Z$, define the lower envelope of $\phi(\boldsymbol{p}_X, \lambda)$ with respect to $\boldsymbol{p}_{\tilde{X}}$ (not with respect to $\boldsymbol{p}_X$) as

$$\begin{aligned} \varphi(q, \boldsymbol{p}_{\tilde{X}}, \lambda) &\overset{\triangle}{=} \underline{\text{env}}_{\boldsymbol{p}_{\tilde{X}}}\phi(\boldsymbol{p}_X, \lambda) \\ &= h(q\beta_2) - \lambda h(q\beta_1) + q\beta_2\tilde{\psi}\left(\boldsymbol{p}_{\tilde{X}}, \frac{\lambda}{1-\alpha_\Delta}\right). \end{aligned}$$

Therefore, the ψ function for $X \to Y \to Z$ has

$$\begin{aligned} \psi(\boldsymbol{p}_X, \lambda) &= \underline{\text{env}}_{\boldsymbol{p}_X}\phi(\boldsymbol{p}_X, \lambda) \\ &= \underline{\text{env}}_{\boldsymbol{p}_X}\varphi(q, \boldsymbol{p}_{\tilde{X}}, \lambda). \end{aligned}$$

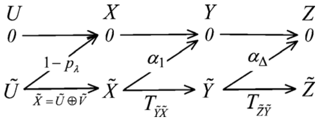

Fig. 17. Optimal transmission strategy for the DM DBC.

Lemma 8: $\psi([1-q, q\boldsymbol{u}^T]^T, \lambda)$ is the lower envelope of $\varphi(q, \boldsymbol{u}, \lambda)$ with respect to q, i.e.,

$$\begin{aligned}\psi([1-q, q\boldsymbol{u}^T]^T, \lambda) &= \underline{\text{env}}_q \varphi(q, \boldsymbol{u}, \lambda)\\ &= \underline{\text{env}}_q \left(\underline{\text{env}}_{\boldsymbol{p}_{\tilde{X}}} \phi(\boldsymbol{p}_X, \lambda)\right)\Big|_{\boldsymbol{p}_X = [1-q, q\boldsymbol{u}^T]^T}.\end{aligned}$$

The proof is given in Appendix J.

Now, we state and prove that NE is optimal for the DM DBC.

Theorem 10: NE achieves the capacity region for the DM DBC.

Proof: This proof shows that combining NE for the broadcast Z channel with NE for the group-operation DBC achieves the capacity region of the DM DBC. This encoding is also the NE for this channel.

Theorem 9 shows that the capacity region for the DM DBC can be achieved by using transmission strategies with uniformly distributed $\tilde{X}$, i.e., the input distribution $\boldsymbol{p}_X = [1-q, q\boldsymbol{u}^T]^T$. By Lemma 8, for such a $\boldsymbol{p}_X$, $\psi([1-q, q\boldsymbol{u}^T]^T, \lambda)$ can be attained by the convex combination of points on the graph of $\varphi(q, \boldsymbol{u}, \lambda)$. Recall that

$$\begin{aligned}\varphi(q, \boldsymbol{u}, \lambda) &= h(q\beta_2) - \lambda h(q\beta_1) + q\beta_2 \tilde{\psi}\left(\boldsymbol{u}, \frac{\lambda}{1-\alpha_\Delta}\right)\\ &= \phi_Z(q, \lambda) + q\beta_2 \tilde{\psi}\left(\boldsymbol{u}, \frac{\lambda}{1-\alpha_\Delta}\right)\end{aligned}$$

where ϕ_Z is ϕ for the broadcast Z channel and $\tilde{\psi}$ is ψ for the group-operation DBC.

Hence, by a discussion analogous to Section IV, $\psi([1-q, q\boldsymbol{u}^T]^T, \lambda)$ can be attained by the convex combination of two points on the graph of $\varphi(q, \boldsymbol{u}, \lambda)$. One point is at $q = 0$ and $\varphi(0, \boldsymbol{u}, \lambda) = 0$. The other point is at $q = p_\lambda$, determined by solving $\ln(1-\beta_2 p_\lambda) = \lambda \ln(1-\beta_1 p_\lambda)$ for p_λ.

Note that the point (0,0) on the graph of $\varphi(q, \boldsymbol{u}, \lambda)$ is also on the graph of $\phi(\boldsymbol{p}_X, \lambda)$. By Theorem 2, the point $(p_\lambda, \varphi(p_\lambda, \boldsymbol{u}, \lambda))$ is the convex combination of n points on the graph of $\phi(\boldsymbol{p}_X, \lambda)$, which corresponds to the group-operation encoding approach for the subchannel $\tilde{X} \to \tilde{Y} \to \tilde{Z}$ because the group-operation encoding approach is the optimal NE scheme for the group-operation DBC $\tilde{X} \to \tilde{Y} \to \tilde{Z}$. Therefore, by Theorem 2, an optimal transmission strategy for the DM DBC $X \to Y \to Z$ is NE as shown in Fig. 17. ■

If the auxiliary random variable U is 0, then the channel input X equals 0 with probability 1. If U is nonzero, then X equals 0 with probability $1 - p_\lambda$. In the case where U and X are both nonzero, $\tilde{X}$ can be obtained as $\tilde{X} = \tilde{U} \oplus \tilde{V}$, where $\oplus$ is the group operation defined in the group-operation degraded broadcast subchannel $\tilde{X} \to \tilde{Y} \to \tilde{Z}$. Here, $\tilde{U}$ is uniformly distributed and $\tilde{V}$ is an n-ary random variable. In order to achieve a pareto-optimal rate pair that maximizes $(R_2 + \lambda R_1)$ for the DM DBC $X \to Y \to Z$, the crossover probability $1 - p_\lambda$ is determined by $\ln(1-\beta_2 p_\lambda) = \lambda \ln(1-\beta_1 p_\lambda)$, and the distribution of $\tilde{V}$ should be the one which also maximizes $(\tilde{R}_2 + \frac{\lambda}{1-\alpha_\Delta}\tilde{R}_1)$ for the group-operation DBC $\tilde{X} \to \tilde{Y} \to \tilde{Z}$.

Since the NE scheme is optimal for DM DBCs, its achievable rate region is the capacity region for DM DBCs. Hence, the capacity region for the DM DBC in Fig. 14 is

$$\begin{aligned}\overline{\text{co}}\Big[\bigcup_{p_U = [1-q, q\boldsymbol{u}^T]^T, p_V \in \Delta_{n+1}} &\{(R_1, R_2) :\\ R_1 &\le H(U \otimes V \otimes N_1 | U) - H(U \otimes V \otimes N_1 | U \otimes V)\\ R_2 &\le H(U \otimes V \otimes N_2) - H(U \otimes V \otimes N_2 | U)\}\Big]. \qquad (47)\end{aligned}$$

VII. Conclusion

This paper proves that simple approaches such as natural encoding and permutation encoding achieve the capacity region of DBCs much more often than has been previously known. Specifically, we show that this is the case for the broadcast Z channel with any number of receivers, the two-receiver group-operation DBC, and (by combining the two previous results) the two-receiver DM DBC.

It would seem that there are more settings where natural encoding achieves the DBC capacity region waiting to be identified. It remains an open problem to prove a general theorem establishing the optimality of natural encoding over a suitably large class of DBCs. The results of this paper also open interesting problems in channel coding to find practical channel codes that use permutation encoding or natural encoding to approach the channel capacity region for the DBCs studied in this paper.

The capacity-region characterization approach that we use has the potential to provide explicit characterizations of DBC capacity regions. As examples, we provide explicit capacity regions for the two-receiver BS DBC and the two-receiver broadcast Z channel.

Appendix A
Simple Independent Encoding Scheme

This appendix presents a simple independent encoding scheme made known to us by Telatar [4] which achieves the capacity region for DBCs. The scheme generalizes to any number of receivers, but showing the two-receiver case suffices to explain the approach. It indicates that any achievable rate pair (R_1, R_2) for a DBC can be achieved by combining symbols from independent encoders with a single-letter function. The independent encoders operate using two codebooks $\{v^n(i) : i = 1, \ldots, 2^{nR_1}\}$, $\{u^n(j) : j = 1, \ldots, 2^{nR_2}\}$ and a single-letter function $f(v, u)$. In order to transmit the message pair (i, j), the transmitter sends the sequence $f(v_1(i), u_1(j)), \ldots, f(v_n(i), u_n(j))$. The scheme is described in the following.

Lemma 9: Suppose U and X are discrete random variables with joint distribution $p_{U,X}(u, x)$. There exists a random vector V independent of U and a deterministic function f such that the pair $(U, f(V, U))$ has joint distribution $p_{U,X}(u, x)$. [4]

Proof: Suppose U and X take values in $\{1,\ldots,l\}$ and $\{1,\ldots,k\}$ respectively. Let $V=(V_1,\ldots,V_l)$, independent of U, be a random variable taking values in $\{1,\ldots,k\}^l$ with $\Pr(V_j=i)=p_{X|U}(i|j)$. Set $f((v_1,\ldots,v_l),u)=v_u$. Then, we have

$$\begin{aligned}\Pr(U=u,f(V,U)=x) &= \Pr(U=u,V_u=x)\\ &= \Pr(U=u)Pr(V_u=x)\\ &= p_U(u)p_{X|U}(x|u)\\ &= p_{U,X}(u,x).\end{aligned}$$

■

If the rate pair (R_1,R_2) is achievable for a DBC $X\to Y\to Z$, there exists an auxiliary random variable U such that

$$\begin{aligned}&\text{(a) } U\to X\to Y\to Z\\ &\text{(b) } I(X;Y|U)\geq R_1\\ &\text{(c) } I(U;Z)\geq R_2.\end{aligned}$$

Apply Lemma 9 to find V independent of U and the deterministic function $f(v,u)$ such that the pair $(U,f(V,U))$ has the same joint distribution as that of (U,X). Randomly and independently choose codewords $\{v^n(1),\ldots,v^n(2^{nR_1})\}$ according to $p(v^n)=p_V(v_1)\cdots p_V(v_n)$, and choose codewords $\{u^n(1),\ldots,u^n(2^{nR_2})\}$ according to $p(u^n)=p_U(u_1)\cdots p_U(u_n)$. To send message pair (i,j), the encoder transmits $f(v_1(i),u_1(j)),\ldots,f(v_n(i),u_n(j))$.

Using a typical-set-decoding random-coding argument, the weak decoder, given z^n, searches for the unique j' such that $(z^n,u^n(j'))$ is jointly typical. The error probability converges to zero as $n\to\infty$ since $R_2\leq I(U;Z)$. The strong decoder, given y^n, also searches for the unique j' such that $(y^n,u^n(j'))$ is jointly typical, and then searches for the unique i' such that $(y^n,v^n(i'))$ is jointly typical given $u^n(j')$. The error probability converges to zero as $n\to\infty$ since

$$R_2\leq I(U;Z)\leq I(U;Y)$$

and

$$\begin{aligned}R_1\leq I(X;Y|U) &= H(Y|U)-H(Y|f(V,U),U)\\ &= H(Y|U)-H(Y|f(V,U),U,V)\\ &= H(Y|U)-H(Y|U,V)\\ &= I(V;Y|U).\end{aligned}$$

APPENDIX B
PROOF OF THEOREM 3

Proof of Theorem 3: For the BS BC $X\to Y\to Z$ with $0<\alpha_1<\alpha_2<1/2$, one has

$$\begin{aligned}\phi(p,\lambda)&\triangleq\phi\left([p,1-p]^T,\lambda\right)\\ &= h_m\left(T_{ZX}\boldsymbol{q}\right)-\lambda h_n\left(T_{YX}\boldsymbol{q}\right)\\ &= h\left((1-\alpha_2)p+\alpha_2(1-p)\right)\\ &\quad -\lambda h\left((1-\alpha_1)p+\alpha_1(1-p)\right).\end{aligned}$$

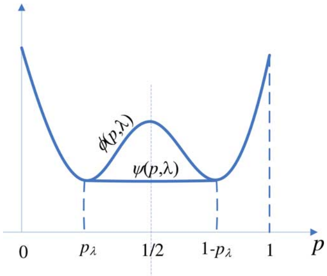

Fig. 18. Illustration of $\psi(p,\lambda)$ and $\phi(p,\lambda)$ for the BS BC with $\lambda<(1-2\alpha_2)^2/(1-2\alpha_1)^2$.

Taking the second derivative of $\phi(p,\lambda)$ with respect to p, we have

$$\begin{aligned}&\phi''(p,\lambda)\\ &=\frac{-(1-2\alpha_2)^2}{(\alpha_2p+(1-\alpha_2)(1-p))\left((1-\alpha_2)p+\alpha_2(1-p)\right)}\\ &+\frac{\lambda(1-2\alpha_1)^2}{(\alpha_1p+(1-\alpha_1)(1-p))\left((1-\alpha_1)p+\alpha_1(1-p)\right)}.\end{aligned}\tag{48}$$

In (48), $\phi''(p,\lambda)=-A+\lambda B$ where A and B are both positive. Thus, $\phi''(p,\lambda)$ has the sign of

$$\begin{aligned}\rho(p,\lambda)&=\frac{\phi''(p,\lambda)}{AB}\\ &=-\left(\frac{1-\alpha_1}{1-2\alpha_1}-p\right)\left(\frac{\alpha_1}{1-2\alpha_1}+p\right)\\ &\quad+\lambda\left(\frac{1-\alpha_2}{1-2\alpha_2}-p\right)\left(\frac{\alpha_2}{1-2\alpha_2}+p\right).\end{aligned}$$

For any $0\leq\lambda\leq1$, $p=1/2$ minimizes ρ so that

$$\min_p\rho(p,\lambda)=\frac{\lambda}{4(1-2\alpha_2)^2}-\frac{1}{4(1-2\alpha_1)^2}.$$

Thus, for $\lambda\geq(1-2\alpha_2)^2/(1-2\alpha_1)^2$, $\phi''(p,\lambda)\geq0$ for all $0\leq p\leq1$, and so, $\psi(p,\lambda)=\phi(p,\lambda)$. In this case, the transmission strategy that maximizes R_1 also maximizes $R_2+\lambda R_1$. Thus, the optimal transmission strategy has $l=1$, which means U is a constant.

Note that $\phi(1/2+p,\lambda)=\phi(1/2-p,\lambda)$. For $\lambda<(1-2\alpha_2)^2/(1-2\alpha_1)^2$, $\phi(p,\lambda)$ has negative second derivative on an interval symmetric about $p=1/2$. Let $p_\lambda=\arg\min_p\phi(p,\lambda)$ with $p_\lambda\leq1/2$. Thus, p_λ satisfies $\phi_{p'}(p_\lambda,\lambda)=0$.

By symmetry, the envelope $\psi(\cdot,\lambda)$ is obtained by replacing $\phi(p,\lambda)$ on the interval $(p_\lambda,1-p_\lambda)$ by its minimum over p, as shown in Fig. 18. Therefore, the lower envelope of $\phi(p,\lambda)$ for the BS BC is

$$\psi(p,\lambda)=\begin{cases}\phi(p_\lambda,\lambda), & \text{for } p_\lambda\leq p\leq1-p_\lambda\\ \phi(p,\lambda), & \text{otherwise.}\end{cases}$$

For a predetermined distribution of X, $\boldsymbol{p}_X=\boldsymbol{q}=[q,1-q]^T$ with $p_\lambda<q<1-p_\lambda$, the pair $(q,\psi(q,\lambda))$ is the convex combination of the points $(p_\lambda,\phi(p_\lambda,\lambda))$ and $(1-p_\lambda,\phi(1-p_\lambda,\lambda))$.

Therefore, by Theorem 2, the optimal transmission strategy with $\boldsymbol{p}_X = \boldsymbol{q}$ is NE with

$$\boldsymbol{p}_U = \begin{bmatrix} \frac{1-p_\lambda-q}{1-2p_\lambda} \\ \frac{q-p_\lambda}{1-2p_\lambda} \end{bmatrix} \quad \text{and} \quad T_{XU} = \begin{bmatrix} p_\lambda & 1-p_\lambda \\ 1-p_\lambda & p_\lambda \end{bmatrix}. \tag{49}$$

The conditional entropy bound $F^*(\boldsymbol{q}, s) = h_2(T_{ZX} \cdot [p_\lambda, 1-p_\lambda]^T) = h(\alpha_2 + (1-2\alpha_2)p_\lambda)$ for $s = h_2(T_{YX} \cdot [p_\lambda, 1-p_\lambda]^T) = h(\alpha_1 + (1-2\alpha_1)p_\lambda)$, and $p_\lambda \le q \le 1-p_\lambda$. For the given $\boldsymbol{q}$, this defines $F^*(s) \overset{\triangle}{=} F^*(\boldsymbol{q}, s)$ on its entire domain $s \in [h(\alpha_1), h(\alpha_1 + (1-2\alpha_1)q)]$, i.e., $s \in [H(Y|X), H(Y)]$.

Note that for a predetermined distribution of X, $\boldsymbol{p}_X = \boldsymbol{q} = [q, 1-q]^T$ with the suboptimal choices of $q < p_\lambda$ or $q > 1-p_\lambda$, one has $\phi(q, \lambda) = \psi(q, \lambda)$, which means that a line with slope λ supports $F^*(\boldsymbol{q}, \cdot)$ at point $s = H(Y) = h(\alpha_1 + (1-2\alpha_1)q)$, and thus, the optimal transmission strategy under the constraint that $q < p_\lambda$ or $q > 1-p_\lambda$ has $l = 1$, which means U is a constant.

The boundary of the capacity region for the BS BC is always achieved when $\boldsymbol{p}_X = [1/2, 1/2]^T$ (see [2]). Hence, the optimal transmission strategy to achieve the boundary of the capacity region always has $l = 2$ and follows from (49) with $q = 1/2$. This leads to the following explicit parametric expression for the boundary of the capacity region of the two-receiver BS BC:

$$R_1 = h\left(\alpha_1 + (1-2\alpha_1)p_\lambda\right) - h(\alpha_1) \tag{50}$$

$$R_2 = \ln(2) - h\left(\alpha_2 + (1-2\alpha_2)p_\lambda\right) \tag{51}$$

where the parameter p_λ is ranging from 0 to 1/2. In addition, the rate pair (R_1, R_2) in (50) and (51) maximizes $R_2 + \lambda R_1$ for each pair of λ and p_λ satisfying $\phi_{p'}(p_\lambda, \lambda) = 0$, which implies (12). ■

Appendix C
Proof of Lemma 1

Proof of Lemma 1: Lemma 1 is the consequence of Proposition 9 for the broadcast Z channel. Since $H(\boldsymbol{Y}|U) \ge N \cdot q/p \cdot h(\beta_1 p)$

$$\begin{aligned} H(\boldsymbol{Z}|U) &\overset{(a)}{\ge} F^*_{T^{(N)}_{YX}, T^{(N)}_{ZX}}(q, N \cdot q/p \cdot h(\beta_1 p)) \\ &\overset{(b)}{=} N \cdot F^*_{T_{YX}, T_{ZX}}(q, q/p \cdot h(\beta_1 p)) \\ &\overset{(c)}{=} N \cdot \frac{q}{p} \cdot h(\beta_2 p) \\ &\overset{(d)}{=} N \cdot \frac{q}{p} \cdot h(\beta_1 p \beta_\triangle). \end{aligned}$$

These steps are justified as follows:

1) (a) follows from the definition of $F^*_{T^{(N)}_{YX}, T^{(N)}_{ZX}}(\boldsymbol{q}, s)$;
2) (b) follows from Proposition 9;
3) (c) follows from the expression of the function F^* for the broadcast Z channel in (18);
4) (d) follows from $\beta_\triangle = \Pr\{Z = 0|Y = 0\} = \beta_2/\beta_1$. ■

Appendix D
Proof of (27)

Proof of (27): Plugging $j = 1$ in (26), we have

$$\begin{aligned} &H(\boldsymbol{Y}^{(1)}|W_2, \ldots, W_K) - H(\boldsymbol{Y}^{(1)}|W_1, \ldots, W_K) \\ &\ge N\frac{q}{t_1}h(\beta_1 t_1) - Nqh(\beta_1) - o(\epsilon) \end{aligned}$$

or

$$H(\boldsymbol{Y}^{(1)}|W_2, \ldots, W_K) \ge N\frac{q}{t_1}h(\beta_1 t_1) - o(\epsilon), \tag{52}$$

since

$$\begin{aligned} H(\boldsymbol{Y}^{(1)}|W_1, \ldots, W_K) &\overset{(a)}{=} H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}) \\ &\overset{(b)}{=} \sum_{i=1}^{N} H(Y_i^{(1)}|\boldsymbol{X}) \\ &\overset{(c)}{=} \sum_{i=1}^{N} H(Y_i^{(1)}|X_i) \\ &= \sum_{i=1}^{N} \Pr(X_i = 0)h(\beta_1) \\ &= Nqh(\beta_1). \end{aligned}$$

Some of these steps are justified as follows:

1) (a) follows since $\boldsymbol{X}$ is a function of $(W_1, \ldots, W_K)$;
2) (b) follows from the conditional independence of $Y_i^{(1)}$, $i = 1, \ldots, N$, given $\boldsymbol{X}$;
3) (c) follows from the conditional independence of $Y_i^{(1)}$ and $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N)$ given X_i.

Inequality (52) indicates that

$$H(\boldsymbol{Y}^{(j)}|W_{j+1}, \ldots, W_K) \ge N\frac{q}{t_j}h(\beta_j t_j) - o(\epsilon) \tag{53}$$

is true for $j = 1$. The rest of the proof is by induction. We assume that (53) is true for j, which means

$$\begin{aligned} H(\boldsymbol{Y}^{(j)}|W_{j+1}, \ldots, W_K) &\ge N\left[\frac{q}{t_j}h(\beta_j t_j) - \frac{o(\epsilon)}{N}\right] \qquad (54) \\ &= N\frac{q}{t_j + \frac{\tau(\epsilon)}{N}}h(\beta_j(t_j + \frac{\tau(\epsilon)}{N})) \end{aligned}$$

where the function $\tau(\epsilon) \to 0$ as $\epsilon \to 0$, since $\frac{q}{t_j}h(\beta_j t_j)$ is continuous in t_j. Applying Lemma 1 to the Markov chain $(W_{j+1}, \ldots, W_K) \to \boldsymbol{X} \to \boldsymbol{Y}^{(j)} \to \boldsymbol{Y}^{(j+1)}$, we have

$$\begin{aligned} H(\boldsymbol{Y}^{(j+1)}|W_{j+1}, \ldots, W_K) &\ge N\frac{q}{t_j + \frac{\tau(\epsilon)}{N}}h(\beta_{j+1}(t_j + \frac{\tau(\epsilon)}{N})) \\ &= N\frac{q}{t_j}h(\beta_{j+1}t_j) + o(\epsilon). \qquad (55) \end{aligned}$$

Considering (26) for $j+1$, we have

$$\begin{aligned} &H(\boldsymbol{Y}^{(j+1)}|W_{j+2}, \ldots, W_K) - H(\boldsymbol{Y}^{(j+1)}|W_{j+1}, \ldots, W_K) \\ &\ge N\frac{q}{t_{j+1}}h(\beta_{j+1}t_{j+1}) - N\frac{q}{t_j}h(\beta_{j+1}t_j) - o(\epsilon). \qquad (56) \end{aligned}$$

Substituting (55) into (56) yields

$$H(\boldsymbol{Y}^{(j+1)}|W_{j+2},\ldots,W_K) \geq N\frac{q}{t_{j+1}}h(\beta_{j+1}t_{j+1}) - o(\epsilon)$$

which establishes the induction. Finally, for $j \geq d$, $N\delta$ should be added to the right-hand side of (54) because of the presence of δ in (22) for $j = d$ and, hence, of $N\delta$ in (26). ■

Appendix E
Proof of Lemma 2

Proof of Lemma 2: Every closed subset of a group is a group. Since $\mathcal{G}_{T_{YX},T_{ZX}}$ is a subset of Φ_k, which is a group under matrix multiplication, it suffices to show that $\mathcal{G}_{T_{YX},T_{ZX}}$ is closed under matrix multiplication. Suppose $G_1, G_2 \in \mathcal{G}_{T_{YX},T_{ZX}}$ such that $T_{YX}G_1 = \Pi_{YX,1}T_{YX}$, $T_{ZX}G_1 = \Pi_{ZX,1}T_{ZX}$, $T_{YX}G_2 = \Pi_{YX,2}T_{YX}$ and $T_{ZX}G_2 = \Pi_{ZX,2}T_{ZX}$. Thus

$$T_{YX}G_1G_2 = \Pi_{YX,1}\Pi_{YX,2}T_{YX}$$

and

$$T_{ZX}G_1G_2 = \Pi_{ZX,1}\Pi_{ZX,2}T_{ZX}.$$

Therefore, $G_1G_2 \in \mathcal{G}_{T_{YX},T_{ZX}}$. ■

Appendix F
Proof of Lemma 3

Proof of Lemma 3: For all $j = 1, \ldots, l$

$$G_j\left(\sum_{i=1}^{l} G_i\right) \overset{(a)}{=} \sum_{i=1}^{l} G_jG_i \overset{(b)}{=} \sum_{i=1}^{l} G_i$$

where (a) follows from the distributive law for the field of rational matrices and (b) follows from the closure axiom and the inverse element axiom for the group $\mathcal{G}_{T_{YX},T_{ZX}}$.

Hence, $\sum_{i=1}^{l} G_i$ has k identical columns and k identical rows since $\mathcal{G}_{T_{YX},T_{ZX}}$ is transitive. Therefore, $\sum_{i=1}^{l} G_i = \frac{l}{k}\boldsymbol{1}\boldsymbol{1}^T$. ■

Appendix G
Proof of Lemma 4

Proof of Lemma 4: Since $(\boldsymbol{p}, s, \eta)$ satisfies (1)–(3) for some choice of l, $\boldsymbol{w}$, and $T_{XU} = [\boldsymbol{t}_1 \ldots \boldsymbol{t}_l]$

$$GT_{XU}\boldsymbol{w} = G\boldsymbol{p}$$
$$\sum_{j=1}^{l} w_jh_n(T_{YX}G\boldsymbol{t}_j) = \sum_{j=1}^{l} w_jh_n(\Pi_{YX}T_{YX}\boldsymbol{t}_j) = s$$
$$\sum_{j=1}^{l} w_jh_m(T_{ZX}G\boldsymbol{t}_j) = \sum_{j=1}^{l} w_jh_m(\Pi_{ZX}T_{ZX}\boldsymbol{t}_j) = \eta.$$

Hence, $(G\boldsymbol{p}, s, \eta)$ satisfies (1)–(3) for the choice of l, $\boldsymbol{w}$, and GT_{XU}. ■

Appendix H
Proof of Lemma 5

Proof of Lemma 5: For any $(s, \eta) \in \mathcal{C}^*$, there exists a distribution $\boldsymbol{p}$ such that $(\boldsymbol{p}, s, \eta) \in \mathcal{C}$. Let $\mathcal{G}_{T_{YX},T_{ZX}} = \{G_1, \ldots, G_l\}$. By Corollary 1, $(G_j\boldsymbol{p}, s, \eta) \in \mathcal{C}$ for all $j = 1, \ldots, l$. By the convexity of the set $\mathcal{C}$

$$(\boldsymbol{q}, s, \eta) = \left(\sum_{j=1}^{l} \frac{1}{l}G_j\boldsymbol{p}\ , s, \eta\right) \in \mathcal{C}$$

where $\boldsymbol{q} = \sum_{j=1}^{l} \frac{1}{l}G_j\boldsymbol{p}$. Since $\mathcal{G}_{T_{YX},T_{ZX}}$ is a group, for any permutation matrix $G' \in \mathcal{G}_{T_{YX},T_{ZX}}$

$$G'\boldsymbol{q} = \sum_{j=1}^{l}\frac{1}{l}G'G_j\boldsymbol{p} = \sum_{j=1}^{l}\frac{1}{l}G_j\boldsymbol{p} = \boldsymbol{q}.$$

Since $G'\boldsymbol{q} = \boldsymbol{q}$, the ith entry and the jth entry of $\boldsymbol{q}$ are the same if G' permutes the ith row to the jth row. Since the set $\mathcal{G}_{T_{YX},T_{ZX}}$ for an IS DBC is transitive, all the entries of $\boldsymbol{q}$ are the same, and so, $\boldsymbol{q} = \boldsymbol{u}$. This implies that $(s, \eta) \in \mathcal{C}^*_{\boldsymbol{u}}$. Since (s, η) is arbitrarily taken from $\mathcal{C}^*$, one has $\mathcal{C}^* \subseteq \mathcal{C}^*_{\boldsymbol{u}}$. On the other hand, by definition, $\mathcal{C}^* \supseteq \mathcal{C}^*_{\boldsymbol{u}}$. Therefore, $\mathcal{C}^* = \mathcal{C}^*_{\boldsymbol{u}}$. ■

Appendix I
Proof of Lemma 7

Proof of Lemma 7: Let $\mathcal{G}_{T_{YX},T_{ZX}} = \{G_1, \ldots, G_l\}$. For any $(s, \eta) \in \mathcal{C}^*_{\boldsymbol{p}_X}$, where $\boldsymbol{p}_X = [1 - q, q\boldsymbol{p}^T_{\tilde{X}}]^T$, one has $(\boldsymbol{p}_X, s, \eta) \in \mathcal{C}$. Since Lemma 4 and Corollary 1 also hold for the DM DBC, $(G_j\boldsymbol{p}_X, s, \eta) \in \mathcal{C}$ for all $j = 1, \ldots, l$. By the convexity of the set $\mathcal{C}$

$$(\boldsymbol{q}, s, \eta) = \left(\sum_{j=1}^{l} \frac{1}{l}G_j\boldsymbol{p}_X\ , s, \eta\right) \in \mathcal{C}$$

where $\boldsymbol{q} = \sum_{j=1}^{l} \frac{1}{l}G_j\boldsymbol{p}_X$. Since $\mathcal{G}_{T_{YX},T_{ZX}}$ is a group, for any permutation matrix $G' \in \mathcal{G}_{T_{YX},T_{ZX}}$

$$G'\boldsymbol{q} = \sum_{j=1}^{l}\frac{1}{l}G'G_j\boldsymbol{p}_X = \sum_{j=1}^{l}\frac{1}{l}G_j\boldsymbol{p}_X = \boldsymbol{q}.$$

Hence, the $(i + 1)$th entry and the $(j + 1)$th entry of $\boldsymbol{q}$ are the same if G' permutes the $(i + 1)$th row to the $(j + 1)$th row for $i, j \in \{1, \ldots, n\}$. Therefore, the second to the $(n + 1)$th entries of $\boldsymbol{q}$ are all the same because the set $\mathcal{G}_{T_{YX},T_{ZX}}$ for the DM DBC permutes the $(i + 1)$th row to the $(j + 1)$th row for all $i, j \in \{1, \ldots, n\}$. Furthermore, no matrix in $\mathcal{G}_{T_{YX},T_{ZX}}$ maps the first row to other rows; hence, the first entry of $\boldsymbol{q}$ is the same as the first entry of $\boldsymbol{p}_X$. Therefore, $\boldsymbol{q} = [1 - q, q\boldsymbol{u}^T]^T$. This implies that $(s, \eta) \in \mathcal{C}^*_{[1-q,q\boldsymbol{u}^T]^T}$, and hence, $\mathcal{C}^*_{\boldsymbol{p}_X} \subseteq \mathcal{C}^*_{[1-q,q\boldsymbol{u}^T]^T}$. Therefore, $\mathcal{C}^* = \bigcup_{q\in[0,1]} \mathcal{C}^*_{[1-q,q\boldsymbol{u}^T]^T}$. ■

Appendix J
Proof of Lemma 8

Proof of Lemma 8: $\psi(\boldsymbol{p}_X, \lambda)$ is the lower envelope of $\varphi(q, \boldsymbol{p}_{\tilde{X}}, \lambda)$ with respect to $\boldsymbol{p}_X$. For $\boldsymbol{p}_X = [1 - q, q\boldsymbol{u}^T]^T$, suppose the point $(\boldsymbol{p}_X, \psi(\boldsymbol{p}_X, \lambda))$ is the convex combination of

$n+1$ points $((q_i,\boldsymbol{t}_i),\varphi(q_i,\boldsymbol{t}_i,\lambda))$ on the graph of $\varphi(q,\boldsymbol{p}_{\tilde{X}},\lambda)$ with weights w_i for $i=1,\ldots,n+1$. Therefore

$$q=\sum_{i=1}^{n+1} w_i q_i$$

$$\boldsymbol{u}=\sum_{i=1}^{n+1} w_i \boldsymbol{t}_i$$

$$\psi(\boldsymbol{p}_X,\lambda)=\sum_{i=1}^{n+1} w_i \varphi(q_i,\boldsymbol{t}_i,\lambda).$$

By Lemma 5, for the group-operation degraded broadcast subchannel, one has $\mathcal{C}^*_{\boldsymbol{t}} \subseteq \mathcal{C}^*_{\boldsymbol{u}}$ for any $\boldsymbol{t}$. Hence, from (9), $\tilde{\psi}(\boldsymbol{t},\lambda) \geq \tilde{\psi}(\boldsymbol{u},\lambda)$ for any $\boldsymbol{t}$, and so

$$\varphi(q_i,\boldsymbol{t}_i,\lambda) \geq \varphi(q_i,\boldsymbol{u},\lambda).$$

Therefore, the convex combination of $n+1$ points $((q_i,\boldsymbol{u}),\varphi(q_i,\boldsymbol{u},\lambda))$ with weights w_i has

$$\sum_{i=1}^{n+1} w_i q_i = q$$

and

$$\sum_{i=1}^{n+1} w_i \varphi(q_i,\boldsymbol{u},\lambda) \leq \sum_{i=1}^{n+1} w_i \varphi(q_i,\boldsymbol{t}_i,\lambda)=\psi(\boldsymbol{p}_X,\lambda).$$

On the other hand, since $\psi(\boldsymbol{p}_X,\lambda)$ is the lower envelope of $\varphi(q,\boldsymbol{p}_{\tilde{X}},\lambda)$ with respect to $\boldsymbol{p}_X$, $\sum_{i=1}^{n+1} w_i\varphi(q_i,\boldsymbol{u},\lambda) \geq \psi(\boldsymbol{p}_X,\lambda)$, and hence, $\sum_{i=1}^{n+1} w_i\varphi(q_i,\boldsymbol{u},\lambda) = \psi(\boldsymbol{p}_X,\lambda)$. Therefore, $\psi([1-q,q\boldsymbol{u}^T]^T,\lambda)$ can be attained as the convex combination of points on the graph of $\varphi(q,\boldsymbol{u},\lambda)$ only in the dimension of q. ■

References

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[2] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 197–207, Mar. 1973.

[3] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Pered. Inf.*, vol. 10, pp. 3–14, Jul.–Sep. 1974.

[4] E. Telatar, Private Communication With the Authors Feb. 2009.

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

[6] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 279–280, Mar. 1974.

[7] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 772–777, Nov. 1973.

[8] H. Witsenhausen, "Entropy inequalities for discrete channels," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 5, pp. 610–616, Sep. 1974.

[9] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 5, pp. 493–501, Sep. 1975.

[10] R. Benzel, "The capacity region of a class of discrete additive degraded interference channels," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 228–231, Mar. 1979.

[11] B. Xie, M. Griot, A. I. V. Casado, and R. D. Wesel, "Optimal transmission strategy and capacity region for broadcast Z channels," in *IEEE Inf. Theory Workshop*, Lake Tahoe, NV, USA, Sep. 2007, pp. 390–395.

[12] B. Xie, M. Griot, A. I. V. Casado, and R. D. Wesel, "Optimal transmission strategy and explicit capacity region for broadcast Z channels," *IEEE Trans. Inf. Theory*, vol. IT-54, no. 9, pp. 4296–4304, Sep. 2008.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, Oct. 1948.

[14] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications—I," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov. 1973.

[15] N. Liu and S. Ulukus, "The capacity region of a class of discrete degraded interference channels," presented at the Inf. Theory Appl., San Diego, CA, USA, Jan. 2, 2007.

[16] N. Liu and S. Ulukus, "The capacity region of a class of discrete degraded interference channels," *IEEE Trans. Inf. Theory*, vol. IT-54, no. 9, pp. 4372–4378, Sep. 2008.

[17] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 2006.

[19] S. W. Golomb, "The limiting behavior of the Z-channel," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 3, p. 372, May 1980.

[20] M. Griot, A. I. V. Casado, W.-Y. Weng, H. Chan, J. Wang, and R. D. Wesel, "Nonlinear trellis codes for binary-input binary-output multiple access channels with single-user decoding," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 364–374, Feb. 2012.

[21] B. Xie and R. D. Wesel, "A mutual information invariance approach to symmetry in discrete memoryless channels," in *Inf. Theory Appl.*, San Diego, CA, USA, Jan. 1, 2008, pp. 444–448.

Bike Xie received his B.S. degree in electronic engineering from Tsinghua University, Beijing, in 2005. He received his M.S. and the Ph.D. degrees in electrical engineering from the University of California, Los Angeles in 2006 and 2010 respectively. At UCLA, he worked on a broad range of research topics including capacity regions and encoding schemes for broadcast channels, download-time regions for peer-to-peer networks, packet coding and exchange for broadcast networks, universal turbo codes for space-time channels, and channel code design for optical communications. Dr. Xie then joined Marvell Semiconductor Inc., Santa Clara in 2010. At Marvell, he is and has been working on physical layer modeling and development for SerDes systems and Zigbee applications. His current research interests include signal processing, information theory, and channel coding for communication systems.

Thomas A. Courtade (S'06–M'12) received the B.S. degree in Electrical Engineering from Michigan Technological University in 2007, and the M.S. and Ph.D. degrees in Electrical Engineering from UCLA in 2008 and 2012, respectively. In 2012, he was awarded a Postdoctoral Research Fellowship at the Center for Science of Information. He currently holds this position, and resides at Stanford University. His recent honors include a Distinguished Ph.D. Dissertation award and an Excellence in Teaching award from the UCLA Department of Electrical Engineering, and a Best Student Paper Award at the 2012 International Symposium on Information Theory.

Richard D. Wesel (S'91–M'96–SM'01) is a Professor with the UCLA Electrical Engineering Department and is the Associate Dean for Academic and Student Affairs for the UCLA Henry Samueli School of Engineering and Applied Science. He joined UCLA in 1996 after receiving his Ph.D. in electrical engineering from Stanford. His B.S. and M.S. degrees in electrical engineering are from MIT. His research is in the area of communication theory with particular interest in channel coding. He has received the National Science Foundation CAREER Award, an Okawa Foundation award for research in information and telecommunications, and the Excellence in Teaching Award from the Henry Samueli School of Engineering and Applied Science. He has authored or co-authored over 130 conference and journal publications.