Broadcast Channels with Confidential Messages

IMRE CSISZÁR AND JÁNOS KÖRNER

Abstract—**Given two discrete memoryless channels (DMC's) with a common input, it is desired to transmit private messages to receiver 1 at rate R_1 and common messages to both receivers at rate R_0, while keeping receiver 2 as ignorant of the private messages as possible. Measuring ignorance by equivocation, a single-letter characterization is given of the achievable triples (R_1, R_e, R_0) where R_e is the equivocation rate. Based on this channel coding result, the related source-channel matching problem is also settled. These results generalize those of Wyner on the wiretap channel and of Körner–Marton on the broadcast channel.**

I. Introduction

We consider a broadcast channel with two receivers, i.e., a pair of discrete memoryless channels (DMC's) with common input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}$ and $\mathcal{Z}$. In his celebrated paper [2], Cover raised the problem of determining the possible rates R_1, R_2, R_0 such that one can send separate messages to receivers located at the two outputs and respective rates R_1 and R_2, and a common message to both at rate R_0. This problem is still open in the general case (i.e., no single-letter characterization of the set of achievable rates is known). If, however, we assume that no separate message is sent to receiver 2, i.e., $R_2 = 0$, a single-letter characterization of the capacity region has been given in [5]. Our model, which we call a broadcast channel with confidential messages (BCC), has the additional feature that the separate message sent to receiver 1 is confidential, i.e., receiver 2 should be kept as ignorant of it as possible. This point of view has been introduced into channel coding by Wyner [8] in his study of wiretap channels. Following Wyner [8], we shall measure confidentiality by equivocation. Our main result is a single-letter characterization of the set of triples (R_1, R_e, R_0) such that, in addition to a common message at rate R_0, a private message can be sent reliably at rate R_1 to receiver 1 with equivocation at least R_e per channel use at receiver 2. This constitutes a generalization of the results of [8], where the above problem is solved if the channel to receiver 2 is a degraded version of that to receiver 1 and no common message is sent (cf. Corollary 4 below). It also constitutes a generalization of the results of [5], where no confidentiality condition is imposed (cf. Corollary 5 below), although we do not prove a strong converse. On the other hand, our converse proof is simpler.

Notation: We designate sets by capital letters and random variables (RV) ranging over these sets by the same italic capitals. All RV's will have finite ranges. The number of elements of a set $\mathcal{X}$ will be denoted by $\|\mathcal{X}\|$. The set of n-length sequences of elements of $\mathcal{X}$ is $\mathcal{X}^n$. X^n is shorthand for $X_1, X_2, \cdots, X_n$.

With a slight abuse of notation, the probability that an input $x^n \in \mathcal{X}^n$ leads to an output $y^n \in \mathcal{B}$ over channel 1 ($\mathcal{B} \subset \mathcal{Y}^n$) will be denoted by $P^n_{Y|X}(\mathcal{B}|x^n)$. A similar notation will be used for channel 2, as well as for the auxiliary DMC's introduced in the text.

The notation $U \to V \to X \to Y$ means that these RV's form a Markov chain in this order. $I(X \wedge Y)$ stands for mutual information. The functions log and exp are taken to the base 2.

II. Statement of the Problem and Main Results

A deterministic block-encoder for the BCC is a mapping $f: \mathcal{S} \times \mathcal{T} \to \mathcal{X}^n$ where $\mathcal{S}$ and $\mathcal{T}$ are arbitrary sets representing the possible private messages and common messages, respectively. For problems not involving secrecy, randomized encoding seldom offers any advantage; hence attention is usually restricted to deterministic encoders. Since randomization can increase secrecy, we allow stochastic encoding.

Definition 1: A (stochastic) encoder f with block length n for the BCC is specified by a matrix of conditional probabilities $f(x^n|s,t)$. Here $x^n \in \mathcal{X}^n$, $s \in \mathcal{S}$, $t \in \mathcal{T}$, $\Sigma_{x^n} f(x^n|s,t) = 1$, and $f(x^n|s,t)$ is the probability that the message pair (s,t) is encoded as channel input x^n.

Our model involves two decoders, i.e., a pair of mappings $\varphi: \mathcal{Y}^n \to \mathcal{S} \times \mathcal{T}$, $\psi: \mathcal{Z}^n \to \mathcal{T}$; there would be no point in considering stochastic decoders. The reliability of transmission achieved by the encoder–decoder (f, φ, ψ) will be defined in terms of maximal error; from the proof of the converse part of Theorem 1 below, it will be clear that use of average error would lead to the same result.

Definition 2: The encoder–decoder (f, φ, ψ) gives rise to (n, ϵ)-transmission over the BCC iff for every $s \in \mathcal{S}$, $t \in \mathcal{T}$, decoder φ gives the correct (s,t) and decoder ψ gives the correct t with probability $\geq 1 - \epsilon$, i.e.,

$$\sum_{x^n \in \mathcal{X}^n} f(x^n|s,t) P^n_{Y|X}(\varphi(y^n) = (s,t)|x^n) \geq 1 - \epsilon$$

$$\sum_{x^n \in \mathcal{X}^n} f(x^n|s,t) P^n_{Z|X}(\psi(z^n) = t|x^n) \geq 1 - \epsilon.$$

If S and T are random messages with given (not necessarily uniform) joint distribution, the corresponding input and output variables of the BCC satisfy $ST \to X^n \to Y^n Z^n$, where the conditional distribution of X^n given ST is f while those of Y^n and of Z^n given X^n are determined by the two DMC's. These conditions do not specify the joint conditional distribution of $Y^n Z^n$ given X^n, but only the marginals of this joint conditional distribution ever enter our considerations.

Manuscript received June 15, 1976; revised November 7, 1977.
The authors are with the Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary.

0018-9448/78/0500-0339$00.75

The level of ignorance of receiver 2 with respect to the private message will be measured by the equivocation $H(S|Z^n)$. This depends on the joint distribution of ST and on the encoder f.

Definition 3: (R_1, R_e, R_0) is an achievable rate triple for the BCC iff there exists a sequence of message sets $\mathcal{S}_n, \mathcal{T}_n$ and encoder–decoders (f_n, φ_n, ψ_n) giving rise to (n, ϵ_n)-transmission with $\epsilon_n \to 0$, such that

$$\lim_{n\to\infty} \frac{1}{n} \log \|\mathcal{S}_n\| = R_1$$

$$\lim_{n\to\infty} \frac{1}{n} \log \|\mathcal{T}_n\| = R_0',$$

$$\lim_{n\to\infty} \frac{1}{n} H(S_n|Z^n) \geqslant R_e$$

where $H(S_n|Z^n)$ is evaluated under the assumption that the pair of random messages $S_n T_n$ is uniformly distributed over $\mathcal{S}_n \times \mathcal{T}_n$. The set of achievable rate triples will be denoted by $\mathcal{R}$. If $(R_1, R_e, R_0) \in \mathcal{R}$, we say that R_1 and R_0 are achievable private and common transmission rates at equivocation rate R_e.

From a mathematical point of view, our main result is the following.

Theorem 1: $\mathcal{R}$ is a closed convex set consisting of those triples (R_1, R_e, R_0) for which there exist RV's $U \to V \to X \to YZ$ such that the conditional distribution of Y (resp. Z) given X is determined by channel 1 (resp. 2) and

$$0 \leqslant R_e \leqslant R_1 \tag{1}$$

$$R_e \leqslant I(V \wedge Y|U) - I(V \wedge Z|U) \tag{2}$$

$$R_1 + R_0 \leqslant I(V \wedge Y|U) + \min\big[I(U \wedge Y), I(U \wedge Z)\big] \tag{3}$$

$$0 \leqslant R_0 \leqslant \min\big[I(U \wedge Y), I(U \wedge Z)\big]. \tag{4}$$

Moreover, the ranges of U and V may be assumed to satisfy

$$\|\mathcal{U}\| \leqslant \|\mathcal{X}\| + 3, \quad \|\mathcal{V}\| \leqslant \|\mathcal{X}\|^2 + 4\|\mathcal{X}\| + 3,$$

and U may be assumed to be a (deterministic) function of V.

Proof: See Sections IV and V. The admissibility of the range constraints will be shown in the Appendix.

So far we have considered a channel coding problem. From a practical point of view, problems of source-channel matching are even more relevant. In multi-user communication, this is often a nontrivial problem. However, in the present case Theorem 1 leads easily to a necessary and sufficient condition for the transmissibility of two sources, one as a private message with prescribed level of secrecy and the other as a common message.

Let us consider two memoryless sources with alphabets $\mathcal{S}, \mathcal{T}$, i.e., let $S_1T_1, S_2T_2, \cdots$ be independent, identically distributed pairs of RV's (but S_i and T_i need not be independent). Let $\tilde{S}$ and $\tilde{T}$ stand for the generic variables of the two sources. We assume that block-to-block encoding is used: a (k,n)-encoder is a (stochastic) encoder in the sense of Definition 1 with block length n and message sets $\mathcal{S}^k, \mathcal{T}^k$.

The random messages to be transmitted are now S^k, T^k, which are blocks of length k of source outputs, and receiver 2's ignorance of the private message will be measured by the equivocation per source letter

$$\frac{1}{k} H(S^k|Z^n).$$

As the criterion of reliability of transmission, we require that the average error frequencies

$$E\frac{1}{k} d_H\big(S^kT^k, \varphi(Y^n)\big) \text{ and } E\frac{1}{k} d_H\big(T^k, \psi(Z^n)\big)$$

both be small, where d_H stands for Hamming distance; several other criteria would lead to the same result.

Definition 4: The source pair $\tilde{S}, \tilde{T}$ is (R, Δ)-transmissible over the BCC, where $R > 0$, $\Delta \geqslant 0$, iff for every $\epsilon > 0$ there exist a (k,n)-encoder f and decoders (φ, ψ) such that

$$\frac{k}{n} \geqslant R - \epsilon \tag{5}$$

$$\frac{1}{k} H(S^k|Z^n) \geqslant \Delta - \epsilon \tag{6}$$

$$E\frac{1}{k} d_H\big(S^kT^k, \varphi(Y^n)\big) \leqslant \epsilon, \; E\frac{1}{k} d_H\big(T^k, \psi(Z^n)\big) \leqslant \epsilon. \tag{7}$$

We shall refer to R as the rate of source-channel matching.

Theorem 2: In order that the source pair $\tilde{S}, \tilde{T}$ be (R, Δ)-transmissible over the BCC, it is necessary and sufficient that

$$\big(RH(\tilde{S}|\tilde{T}), R\Delta, RH(\tilde{T})\big) = (R_1, R_e, R_0) \in \mathcal{R}.$$

Proof: See Sections IV and V.

III. Some Implications of the Main Results

One of the most interesting points of Wyner's paper [8] was that in his model one could reliably transmit information to receiver 1 at a positive rate while keeping receiver 2 (the "wire-tapper") in essentially perfect ignorance. One can ask to what extent this phenomenon was due to the assumption that channel 2 was a degraded version of channel 1. As a consequence of our main results, we show that the above phenomenon is present under quite general conditions (cf. Corollary 3 below).

Suppose that a source pair $\tilde{S}, \tilde{T}$ is (R, Δ)-transmissible over a BCC in the sense of Definition 4. Since receiver 2 can decode the common message, the information about $\tilde{S}$ contained in $\tilde{T}$ is always available to him; hence necessarily $\Delta \leqslant H(\tilde{S}|\tilde{T})$. (Formally, this follows from Fano's lemma; of course, Theorem 2 does contain the necessity of $\Delta \leqslant H(\tilde{S}|\tilde{T})$ through (1).) If $\Delta = H(\tilde{S}|\tilde{T})$, it is reasonable to speak of perfect secrecy; this is even more reasonable if the two sources are independent, in which case the condition becomes $\Delta = H(\tilde{S})$.

In view of Theorem 2, transmission with perfect secrecy in the above sense is possible iff the rate R of source-channel matching satisfies $(RH(\tilde{S}|\tilde{T}), RH(\tilde{T}) \in \mathcal{C}_s$, where the secrecy capacity region $\mathcal{C}_s$ is defined as follows.

Definition 5: The secrecy capacity region $\mathcal{C}_s$ of the BCC is the set of pairs (R_1, R_0) such that $(R_1, R_1, R_0) \in \mathcal{R}$.

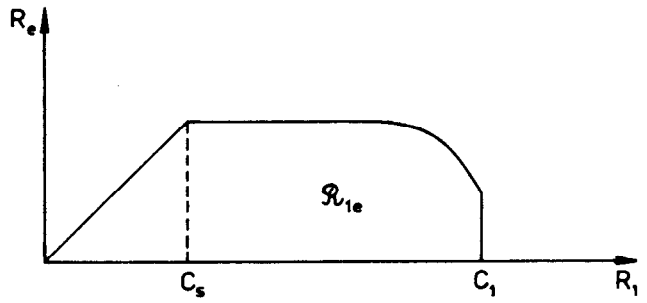

Fig. 1. Typical rate region with no common message ($R_0=0$.)

Corollary 1: $\mathcal{C}_s$ consists of those pairs (R_1,R_0) for which there exist $U\to V\to X\to YZ$ such that

$$0\leqslant R_1\leqslant I(V\wedge Y|U)-I(V\wedge Z|U)$$
$$0\leqslant R_0\leqslant \min\left[I(U\wedge Y),I(U\wedge Z)\right].$$

Now we turn to the special case of no common message ($R_0=0$). We designate by $\mathcal{R}_{1e}$ the set of rate pairs (R_1,R_e) achievable with no common message, i.e., $(R_1,R_e)\in\mathcal{R}_{1e}$ iff $(R_1,R_e,0)\in\mathcal{R}$ (see Fig. 1). Following Wyner [8], we define the secrecy capacity $\mathcal{C}_s$ as the maximum rate at which messages can be sent to receiver 1 in perfect secrecy, i.e.,

$$C_s\triangleq\max_{(R_1,R_1)\in\mathcal{R}_{1e}}R_1=\max_{(R_1,0)\in\mathcal{C}_s}R_1. \tag{8}$$

For the sake of reference, we note a well-known result.

Lemma 1: If $U\to V\to Y$ then

$$I(V\wedge Y)=I(U\wedge Y)+I(V\wedge Y|U).$$

Corollary 2: $(R_1,R_e)\in\mathcal{R}_{1e}$ iff there exist $U\to V\to X\to YZ$ such that $I(U\wedge Y)\leqslant I(U\wedge Z)$ and

$$0\leqslant R_e\leqslant I(V\wedge Y|U)-I(V\wedge Z|U). \tag{9}$$

Thus

$$R_e\leqslant R_1\leqslant I(V\wedge Y). \tag{10}$$

Further,

$$C_s=\max_{V\to X\to YZ}\left[I(V\wedge Y)-I(V\wedge Z)\right]. \tag{11}$$

Proof: Taking $R_0=0$ in Theorem 1 we obtain that $(R_1,R_e)\in\mathcal{R}_{1e}$ iff there exist $U\to V\to X\to YZ$ such that (9) and

$$R_e\leqslant R_1\leqslant I(V\wedge Y|U)+\min\left[I(U\wedge Y),I(U\wedge Z)\right] \tag{10a}$$

hold. If $I(U\wedge Y)\leqslant I(U\wedge Z)$, then (10a) reduces to (10) by Lemma 1. If not, again by Lemma 1, from (9) and (10a) we get

$$0\leqslant R_e\leqslant I(V\wedge Y)-I(V\wedge Z)$$
$$R_e\leqslant R_1\leqslant I(V\wedge Y).$$

In the latter case (9) and (10) are satisfied for $U=$const. From (8), (9), and (10),

$$C_s=\max_{U\to V\to X\to YZ}\left[I(V\wedge Y|U)-I(V\wedge Z|U)\right]$$

where the maximization is subject to $I(U\wedge Y)\leqslant I(U\wedge Z)$. As conditional mutual informations are averages of unconditional ones, the maximum is achieved when $U=$ const. This proves (11).

In [6], two new concepts of partial ordering of channels with common input alphabet were introduced. The single-letter characterization of the relation "channel 1 is *more capable* than channel 2" was that for every input X

$$I(X\wedge Y)\geqslant I(X\wedge Z). \tag{12}$$

The relation "channel 1 is *less noisy* than channel 2" was single-letter characterized by the property that for every $V\to X\to YZ$

$$I(V\wedge Y)\geqslant I(V\wedge Z). \tag{13}$$

It was shown in [6] that the first condition is strictly weaker than the second, which, in turn, is strictly weaker than "channel 2 is a degraded version of channel 1."

In the following corollaries we consider properties (12) and (13) as definitions of the corresponding order relations for channels.

Corollary 3: The secrecy capacity C_s is always positive unless channel 2 is less noisy than channel 1.

Proof: See (11) and (13).

Theorem 3: If channel 1 is less noisy than channel 2 then $(\mathcal{R}_1,\mathcal{R}_e)\in\mathcal{R}_{1e}$ iff there exist X,Y,Z such that

$$0\leqslant R_e\leqslant I(X\wedge Y)-I(X\wedge Z) \tag{14}$$
$$R_e\leqslant R_1\leqslant I(X\wedge Y). \tag{15}$$

In particular,

$$C_s=\max\left[I(X\wedge Y)-I(X\wedge Z)\right].$$

The last assertion holds also under the weaker condition that channel 1 is more capable than channel 2.

Proof: Clearly pairs satisfying (14) and (15) belong to $\mathcal{R}_{1e}$ (take $U=$const, $V=X$ in Corollary 2). Further, if $(R_1,R_e)\in\mathcal{R}_{1e}$, then (9) gives, applying Lemma 1 repeatedly,

$$\begin{aligned}R_e&\leqslant I(V\wedge Y|U)-I(V\wedge Z|U)\\&=I(V\wedge Y)-I(V\wedge Z)-\left[I(U\wedge Y)-I(U\wedge Z)\right]\\&=I(X\wedge Y)-I(X\wedge Z)-\left[I(X\wedge Y|V)-I(X\wedge Z|V)\right]\\&\quad-\left[I(U\wedge Y)-I(U\wedge Z)\right].\end{aligned}$$

Since channel 1 is less noisy than channel 2, both brackets are nonnegative (for the first this follows even from (12)). Also, by (10),

$$R_e\leqslant R_1\leqslant I(V\wedge Y)\leqslant I(X\wedge Y).$$

Thus (R_1,R_e) satisfies (14) and (15).

Similarly, from (11)

$$\begin{aligned}C_s&=\max_{V\to X\to YZ}\left[I(V\wedge Y)-I(V\wedge Z)\right]\\&=\max_{V\to X\to YZ}\left[I(X\wedge Y)-I(X\wedge Z)\right.\\&\quad\left.-(I(X\wedge Y|V)-I(X\wedge Z|V))\right]\\&=\max\left[I(X\wedge Y)-I(X\wedge Z)\right],\end{aligned}$$

since, as noted above, $I(X\wedge Y|V)-I(X\wedge Z|V)\geqslant 0$, and by taking $V=X$ we can make this term 0.

Corollary 4: If channel 1 is less noisy than channel 2, a source $\tilde{S}$ is (R,Δ)-transmissible over the BCC (with no common message) iff $(RH(\tilde{S}),R\Delta)\in\mathcal{R}_{1e}$ where $\mathcal{R}_{1e}$ is given by Theorem 3.

Proof: See the special case of Theorem 2.

Remark: In the case when channel 2 is a degraded version of channel 1, Wyner [8] characterized the set of those pairs (R,d) for which, in our terminology, the source $\tilde{S}$ is $(R/(H(\tilde{S})), d)$-transmissible. Corollary 4 gives for this the condition $(R,\ Rd/(H(\tilde{S})))\in\mathcal{R}_{1e}$, which means that there exist X,Y,Z such that

$$0\leqslant\frac{Rd}{H(\tilde{S})}\leqslant R\leqslant I(X\wedge Y);$$

$$\frac{Rd}{H(\tilde{S})}\leqslant I(X\wedge Y)-I(X\wedge Z).$$

Clearly this is the same as

$$0\leqslant d\leqslant H(\tilde{S});\qquad 0\leqslant R\leqslant \max I(X\wedge Y);$$

$$Rd\leqslant H(\tilde{S})\cdot\max_{I(X\wedge Y)\geqslant R}\left[I(X\wedge Y)-I(X\wedge Z)\right]$$

which was Wyner's characterization. He observed that the set of such (R,d) pairs was not convex. Of course, this does not contradict the convexity of our region $\mathcal{R}_{1e}$.

Corollary 5: The pair (R_1,R_0) is achievable in the absence of a secrecy constraint ($R_e=0$) iff there exist $U\to X\to YZ$ such that

$$0\leqslant R_1+R_0\leqslant I(X\wedge Y|U)+\min\left[I(U\wedge Y),I(U\wedge Z)\right]$$

$$0\leqslant R_0\leqslant\min\left[(U\wedge Y),I(U\wedge Z)\right].$$

Proof: It suffices to observe in Theorem 1 that $I(V\wedge Y|U)\leqslant I(X\wedge Y|U)$.

Remark: In [5], the above capacity region was characterized somewhat differently. The equivalence of the two characterizations can be easily shown by some algebra. Still, Corollary 5 does not fully contain the Theorem of [5], since here we have only the weak converse.

IV. Proof of the Direct Parts of Theorems 1 and 2

We shall prove Theorem 1 in a slightly stronger form. Namely, in definition 3, instead of uniformly distributed random messages, we shall also allow "conditionally nearly uniform" distributions in the sense that

$$\max_{s_1,s_2,t}\frac{\Pr\{S=s_1|T=t\}}{\Pr\{S=s_2|T=t\}}\leqslant\exp n\delta_n \tag{16}$$

for an arbitrary but fixed sequence $\delta_n\to 0$.

Definition 3:* A rate triple (R_1,R_e,R_0) is stably achievable for the BCC if in Definition 3

$$\liminf_{n\to\infty}\frac{1}{n}H(S_n|Z^n)\geqslant R_e$$

holds whenever the pair S_n,T_n satisfies (16) for every n.

We shall show that the rate region specified in Theorem 1 is stably achievable. This will be done by first determining a subset of the stably achievable rate triples (Lemma 3 below) and then applying this result for the cascade of the given DMC's with an arbitrary prefixed DMC.

Our main tool will be Lemma 2, stated below, which asserts the existence of a code of block length n for channel 1 with a specific structure; there is an equipartition of the codeword set so that the class index of the transmitted codeword can be decoded by receiver 2. Moreover, there is a finer equipartition splitting each class into codes for channel 2. The number of codes into which each class is partitioned will characterize, intuitively, how much additional information would be needed for receiver 2 to decode the sent codeword rather than to only find its class. We shall lower bound the equivocation at receiver 2 by the logarithm of this number.

Before stating Lemma 2, we digress to note some facts on typical sequences.

Digression on Typical Sequences

Given two RV's X and Y with ranges $\mathcal{X}$ and $\mathcal{Y}$, a sequence $x^n\in\mathcal{X}^n$ will be called X-typical iff

$$|N(a|x^n)-n\Pr\{X=a\}|\leqslant r_n,\qquad\text{for all } a\in\mathcal{X} \tag{17}$$

where $N(a|x^n)$ is the number of occurrences of letter a in the sequence x^n, and $\{r_n\}$ is a fixed sequence of positive numbers such that $r_n\cdot n^{-1/2}\to\infty$, $r_n\cdot n^{-1}\to 0$. Moreover, a pair of sequences $(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n$ will be called $Y|X$-typical, or y^n will be called $Y|X$-generated by x^n iff x^n is $\mathcal{X}$-typical and

$$|N(a,b|x^n,y^n)-N(a|x^n)\Pr\{Y=b|X=a\}|\leqslant r_n \tag{18}$$

for all $a\in\mathcal{X},b\in\mathcal{Y}$.

Let $\mathcal{T}_{Y|X}(x^n)$ denote the set of all $y^n\in\mathcal{Y}^n$, $Y|X$-generated by the X-typical $x^n\in\mathcal{X}^n$. Consider a DMC which has a matrix equal to the conditional distribution of Y given X.

Fact A:

$$\frac{1}{n}\log P^n_{Y|X}(y^n|x^n)\to -H(Y|X),\qquad\text{as } n\to\infty,$$

uniformly in $Y|X$-typical pairs x^n,y^n.

Fact B:

$$P^n_{Y|X}(\mathcal{T}_{Y|X}(x^n)|x^n)\to 1,\qquad\text{as } n\to\infty,$$

uniformly in X-typical sequences x^n.

Fact C:

$$\frac{1}{n}\log\|(\mathcal{T}_{Y|X}(x^n))\|\to H(Y|X),$$

uniformly in X-typical x^n. For RV's $U\to X\to Y$, we shall speak both of X-typical and U-typical sequences. If the former are defined with constants r_n, the latter will be understood with constants $r'_n=r_n/2\|U\|$, cf. (17). A similar convention applies to $Y|X$-typical pairs and $X|U$-typical pairs, so that we have the following.

Fact D: If (u^n,x^n) is an $X|U$-typical pair, then x^n is X-typical.

Code Construction

Lemma 2: If $U\to X\to YZ$ and $I(X\wedge Y|U)>I(X\wedge Z|U)$, then for every n there exists a set $\{x^n_{jlm}\}\subset\mathcal{X}^n$ where j, l, m run over index sets J_n,L_n,M_n, with the following properties.

a) For each $m \in \mathfrak{M}_n$, there exists a U-typical sequence $u_m^n \in \mathfrak{U}^n$ such that every x_{jlm}^n is $X|U$-generated by u_m^n. Moreover, there exist pairwise disjoint subsets $\mathfrak{B}_m \subset \mathfrak{T}_{Y|U}(u_m^n)$ of $\mathfrak{Y}^n$ resp. $\mathfrak{C}_m \subset \mathfrak{T}_{Z|U}(u_m^n)$ of $\mathfrak{Z}^n$ such that

$$P_{Y|X}(\mathfrak{B}_m|x_{jlm}^n) \geq 1-\epsilon_n; \qquad P_{Z|X}(\mathfrak{C}_m|x_{jlm}^n) \geq 1-\epsilon_n$$

for all $j \in \mathfrak{J}_n, l \in \mathfrak{L}_n, m \in \mathfrak{M}_n$.

b) There exist pairwise disjoint subsets $\mathfrak{B}_{jlm} \subset \mathfrak{T}_{Y|X}(x_{jlm}^n)$ of $\mathfrak{B}_m$ and subsets $\mathfrak{C}_{jlm} \subset \mathfrak{T}_{Z|X}(x_{jlm}^n)$ of $\mathfrak{C}_m$, of which those with the same middle index l are pairwise disjoint, such that

$$P_{Y|X}(\mathfrak{B}_{jlm}|x_{jlm}^n) \geq 1-\epsilon_n, \qquad P_{Z|X}(\mathfrak{C}_{jlm}|x_{jlm}^n) \geq 1-\epsilon_n.$$

c) Also,

$$\frac{1}{n}\log \|\mathfrak{J}_n\| \to I(X \wedge Z|U)$$

$$\frac{1}{n}\log \|\mathfrak{L}_n\| \to I(X \wedge Y|U) - I(X \wedge Z|U)$$

$$\frac{1}{n}\log \|\mathfrak{M}_n\| \to \min\left[I(U \wedge Y), I(U \wedge Z)\right],$$

$$\epsilon_n \to 0.$$

Lemma 2 will be proved in the Appendix. Based on Lemma 2, we now determine a subset of $\mathfrak{R}$.

Lemma 3: Under the hypothesis of Lemma 2, the rate triples (R_1, R_e, R_0) satisfying

$$R_e = I(X \wedge Y|U) - I(X \wedge Z|U) \leq R_1$$

$$R_1 + R_0 \leq I(X \wedge Y|U) + \min\left[I(U \wedge Y), I(U \wedge Z)\right]$$

$$0 \leq R_0 \leq \min\left[I(U \wedge Y), I(U \wedge Z)\right]$$

are stably achievable.

Proof: Let (R_1, R_e, R_0) satisfy the above conditions. We construct message sets $\mathfrak{S}_n$ and $\mathfrak{T}_n$ and encoders f_n with block length n giving rise—with suitable decoders—to (n, ϵ_n)-transmission, achieving

$$\frac{1}{n}\log \|\mathfrak{S}_n\| \to R_1, \tag{19}$$

$$\frac{1}{n}\log \|\mathfrak{T}_n\| \to R_0, \tag{20}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} H(S_n|Z^n) \geq I(X \wedge Y|U) - I(X \wedge Z|U), \tag{21}$$

provided that the message RV's S_n, T_n satisfy (16).

If $R_1 \geq I(X \wedge Y|U)$, we take $\mathfrak{S}_n = \mathfrak{J}_n \times \mathfrak{L}_n \times \mathfrak{K}_n$ where $\mathfrak{J}_n$ and $\mathfrak{L}_n$ are the same as in Lemma 2 and $\mathfrak{K}_n$ is an arbitrary set such that (19) holds. $\mathfrak{T}_n$ can be arbitrary—satisfying (20)—and we consider a one-to-one mapping g_n of $\mathfrak{K}_n \times \mathfrak{T}_n$ into $\mathfrak{M}_n$. We define a deterministic encoder f_n which maps $(j,l,k) \in \mathfrak{S}_n$, $t \in \mathfrak{T}_n$ into the codeword x_{jlm}^n with $m = g_n(k,t)$.

If $R_1 < I(X \wedge Y|U)$, we take $\mathfrak{S}_n = \tilde{\mathfrak{K}}_n \times \mathfrak{L}_n$ where $\tilde{\mathfrak{K}}_n$ is an arbitrary set such that (19) holds, and we take $\mathfrak{T}_n \subset \mathfrak{M}_n$. Let $\tilde{g}_n$ be a mapping of $\mathfrak{J}_n$ into $\tilde{\mathfrak{K}}_n$ partitioning $\mathfrak{J}_n$ into subsets of nearly equal size in the sense that $\|\tilde{g}_n^{-1}(k')\| \leq 2\|\tilde{g}_n^{-1}(k'')\|$ for all $k', k'' \in \tilde{\mathfrak{K}}_n$. We define a stochastic encoder f_n associating with $(k,l) \in \mathfrak{S}_n$, $m \in \mathfrak{T}_n$ a codeword x_{jlm} where the index j is drawn at random (with uniform distribution) from the set $\tilde{g}_n^{-1}(k) \subset \mathfrak{J}_n$.

With the help of the sets $\mathfrak{B}_{jlm}$ resp. $\mathfrak{C}_m$ of Lemma 2, it is obvious in both cases how to define decoders φ_n, ψ_n giving rise to (n, ϵ_n)-transmission.

It remains to check (21). Let X^n be the input RV of the channel corresponding to the random messages S_n, T_n (satisfying (16)) and encoder f_n. The possible values of X^n are codewords x_{jlm}^n. Let M_n be the RV defined as the third coordinate of the actual value of X^n. Then

$$\begin{aligned} H(S_n|Z^n) &\geq H(S_n|Z^nM_n) = H(S_nZ^n|M_n) - H(Z^n|M_n) \\ &= H(S_nZ^nX^n|M_n) - H(X^n|S_nM_nZ^n) - H(Z^n|M_n) \\ &= H(S_nX^n|M_n) + H(Z^n|S_nM_nX^n) \\ &\quad - H(X^n|S_nM_nZ^n) - H(Z^n|M_n) \\ &\geq H(X^n|M_n) + H(Z^n|X^n) \\ &\quad - H(X^n|S_nM_nZ^n) - H(Z^n|M_n). \end{aligned} \tag{22}$$

We bound the terms in the last line separately. Given $M_n = m$, X^n has $\|\mathfrak{J}_n\| \cdot \|\mathfrak{L}_n\|$ possible values. Moreover, it follows from (16) that its conditional distribution is nearly uniform over this range in the sense that

$$\frac{\Pr\{X^n = x^n | M_n = m\}}{\Pr\{X^n = \bar{x}^n | M_n = m\}} \leq 2\exp\{n\delta_n\}$$

for any pair x^n, x^{-n} of codewords with last index m. Hence

$$H(X^n|M_n) \geq \log \|J_n\| + \log \|L_n\| - n\delta_n - 1. \tag{23}$$

Further, we have

$$H(Z^n|X^n) = -\sum_{X^n} \Pr\{X^n = x^n\} \sum_{a \in \mathfrak{X}} N(a|x^n) \cdot \sum_{z \in Z} P_{Z|X}(z|a) \log P_{Z|X}(z|a).$$

Thus, using Fact D,

$$\frac{1}{n} H(Z^n|X^n) \to H(Z|X). \tag{24}$$

The next term vanishes in the case $R_1 \geq I(X \wedge Y|U)$, and it is negligible also in the case $R_1 < I(X \wedge Y|U)$, since S_n, M_n, Z^n determine X^n with error probability $\leq \epsilon_n$. Formally, set

$$\delta(k,l,m,z^n) = \begin{cases} x_{klm}^n, & \text{if } z^n \in \mathfrak{C}_{jlm},\ \tilde{g}_n(j) = k \\ \text{arbitrary}, & \text{otherwise} \end{cases}$$

then $\Pr\{X^n \neq \delta(S_nM_nZ^n)\} \leq \epsilon_n$. Thus by Fano's lemma,

$$\frac{1}{n} H(X^n|S_nM_nZ^n) \to 0. \tag{25}$$

Finally, define an RV $\hat{Z}^n$ with values in $\mathfrak{C}_{M_n}$ by

$$\hat{Z}^n = \begin{cases} Z^n, & \text{if } Z^n \in \mathfrak{C}_{M_n} \\ \text{arbitrary}, & \text{if } Z^n \notin \mathfrak{C}_{M_n}. \end{cases}$$

Then

$$H(Z^n|M_n) \leq H(Z^n|\hat{Z}^n) + H(\hat{Z}_n|M_n)$$

where the first term is negligible by Fano's lemma, since $\Pr\{Z^n \neq \hat{Z}^n\} \leq \epsilon_n$.

Moreover,

$$\frac{1}{n}H(\hat{Z}^n|M_n)\leqslant\frac{1}{n}\max_m \log\|\mathcal{C}_m\|.$$

However, the right side of this tends to $H(Z|U)$ because $P_{Z|U}(\mathcal{C}_m|u_m^n)\geqslant 1-\epsilon_n$ and $\mathcal{C}_m\subset\mathcal{T}_{Z|U}(u_m^n)$ (see Fact A). Hence $(1/n)H(Z^n|M_n)$ can be upper bounded by a term tending to $H(Z|U)$. Using this and substituting (23)–(25) into (22), we obtain (21). Thus Lemma 3 is proved.

In order to obtain the whole region $\mathcal{R}$, we introduce additional randomization by prefixing an arbitrary DMC to the given ones.

Lemma 4: If $U\to V\to X\to YZ$ and $I(V\wedge Y|U)\geqslant I(V\wedge Z|U)$, then all rate triples (R_1,R_e,R_0) satisfying

$$0\leqslant R_e=I(V\wedge Y|U)-I(V\wedge Z|U)\leqslant R_1 \tag{26}$$

$$R_1+R_0\leqslant I(V\wedge Y|U)+\min\left[I(U\wedge Y),I(U\wedge Z)\right] \tag{27}$$

$$0\leqslant R_0\leqslant \min\left[I(U\wedge Y),I(U\wedge Z)\right] \tag{28}$$

are stably achievable.

Proof: Consider the DMC's with input alphabet $\mathcal{V}$, output alphabets $\mathcal{Y}$, $\mathcal{Z}$, and transition probability matrices defined by the conditional distribution of Y resp. Z given V. Any encoder f' for this new BCC determines an encoder f for the original BCC by the matrix product of f' with the conditional distribution of X given V. Both encoders yield the same stochastic connection of messages and received sequences, so the assertion follows by applying Lemma 3 to the new BCC.

It is obvious from the definition of (stably) achievable rate triples that, if (R_1,R_e,R_0) is such a triple, then so is (R_1,R_e',R_0) for every $0\leqslant R_e'\leqslant R_e$.

Let $\tilde{\mathcal{R}}$ be the set of all triples (R_1,R_e,R_0) for which there exist $U\to V\to X\to YZ$ such that (26)–(28) are fulfilled, the equality in (26) being replaced by $\leqslant$. It remains to prove that $\tilde{\mathcal{R}}$ coincides with the rate region claimed by Theorem 1 and that the latter is convex.

Lemma 5: $\tilde{\mathcal{R}}$ is convex.

Proof: Let (R_1',R_e',R_0') and (R_1'',R_e'',R_0'') satisfy (26)–(28) with RV's $U_1\to V_1\to X_1\to Y_1Z_1$ and $U_2\to V_2\to X_2\to Y_2Z_2$, respectively. Let J be a RV independent of all the others and taking the values 1 and 2 with probabilities α, $1-\alpha$. Define

$$U\triangleq U_JJ,\ V\triangleq V_J,\ X\triangleq X_J,\ Y\triangleq Y_J,\ Z\triangleq Z_J.$$

Then $U\to V\to X\to YZ$, the conditional distributions of Y and Z given X are right, and

$$I(V\wedge Y|U)=\alpha I(V_1\wedge Y_1|U_1)+(1-\alpha)I(V_2\wedge Y_2|U_2)$$

$$I(U\wedge Y)\geqslant I(U\wedge Y|J)=\alpha I(U_1\wedge Y_1)+(1-\alpha)I(U_2\wedge Y_2).$$

This and the analogous relations for Z mean that also

$$\alpha(R_1',R_e',R_0')+(1-\alpha)(R_1'',R_e'',R_0'')\in\tilde{\mathcal{R}}.$$

Lemma 6: $\tilde{\mathcal{R}}$ equals the rate region described by (1)–(4).

Proof: Clearly, $\tilde{\mathcal{R}}$ is contained in the other region. To see the reverse inclusion, consider any (R_1,R_e,R_0) satisfying (1)–(4) for some $U\to V\to X\to YZ$. Let

$$R_1^*=I(V\wedge Y|U)+\min\left[I(U\wedge Y),I(U\wedge Z)\right]-R_0$$

$$R_e^*=I(V\wedge Y|U)-I(V\wedge Z|U).$$

Then $R_e^*\leqslant R_1\leqslant R_1^*$, $R_e\leqslant R_e^*$, and $(R_1^*,R_e^*,R_0)\in\tilde{\mathcal{R}}$. It follows from the definition of $\tilde{\mathcal{R}}$ that (R_e^*,R_e^*,R_0), $(R_1^*,0,R_0)$, and $(0,0,R_0)$ also belong to $\tilde{\mathcal{R}}$. Hence $(R_1,R_e,R_0)\in\tilde{\mathcal{R}}$ by the convexity of $\tilde{\mathcal{R}}$.

The proof of the direct part of Theorem 1 is complete. Moreover, we have shown that all $(R_1,R_e,R_0)\in\mathcal{R}$ are stably achievable. The proof of the direct half of Theorem 2 is now a matter of standard technique; it is relegated to the Appendix.

V. Proof of the Converse Parts of Theorems 1 and 2

We prove both converses at once, using the technique of "single-letterization" of information quantities.

For reference we state: *Fano's Lemma.* If an RV S with values in $\mathcal{S}$ can be reproduced as a function of another RV W with error probability

$$\Pr\{S\neq g(W)\}=\delta,$$

then

$$H(S|W)\leqslant\delta\log\|\mathcal{S}\|-\delta\log\delta-(1-\delta)\log(1-\delta). \tag{29}$$

Moreover, if for a sequence $S^k=S_1\cdots S_k$

$$E\frac{1}{k}d_H(S^k,g(W))=\delta,$$

then the right side of (29) is an upper bound for $(1/k)H(S^k|W)$.

For a proof see e.g., [8], Appendix A.

Let us consider RV's $ST\to X^n\to Y^nZ^n$ corresponding to the transmission of random messages over the BCC with encoder f; the ranges of S and T are supposed to be of size $\leqslant\exp(nK)$ for some constant K. More precisely, for the purpose of Theorem 1 (case A), we assume that S and T are independent and uniformly distributed over their ranges $\mathcal{S}_n$ and $\mathcal{T}_n$, while for Theorem 2 (case B) we take $S=S^k$, $T=T^k$, the source outputs of length k. Also, we assume the existence of decoders φ,ψ such that (f,φ,ψ) gives rise to (n,ϵ)-transmission in the sense of Definition 2 (case A) or achieves (5)–(7) (case B).

In both cases, we have by Fano's Lemma

$$\frac{1}{n}H(ST|Y^n)\leqslant\eta(\epsilon),\qquad\frac{1}{n}H(T|Z^n)\leqslant\eta(\epsilon) \tag{30}$$

where $\eta(\epsilon)$ does not depend on n and $\lim_{\epsilon\to 0}\eta(\epsilon)=0$.

We shall show the existence of RV's $U\to V\to X\to YZ$ such that

$$\frac{1}{n}H(S|Z^n)=I(V\wedge Y|U)-I(V\wedge Z|U)+\delta_1 \tag{31}$$

$$\frac{1}{n}\left[H(S|T)+H(T)\right]\leqslant I(V\wedge Y|U)+\min\left[I(U\wedge Y),I(U\wedge Z)\right]+\delta_2 \tag{32}$$

$$\frac{1}{n}H(T)\leqslant\min\left[I(U\wedge Y),I(U\wedge Z)\right]+\delta_3 \tag{33}$$

where $0 \leq \delta_i \leq 2\eta(\epsilon)$, $i=1,2,3$, and also

$$\frac{1}{n}H(S|Z^n) \leq \frac{1}{n}H(S|T)+\delta_1. \tag{34}$$

Since in case A

$$H(S|T)=\log \|\mathcal{S}_n\|, \qquad H(T)=\log \|\mathcal{T}_n\|,$$

and in case B

$$H(S|T)=kH(\tilde{S}|\tilde{T}), \qquad H(T)=kH(\tilde{T}),$$

this will prove the converse parts of both Theorems 1 and 2.

The starting point of the proof is three identities.

$$H(S|T)=I(S\wedge Y^n|T)+H(S|Y^nT) \tag{35}$$

$$\begin{aligned} H(S|Z^n) &= H(S|Z^nT)+I(S\wedge T|Z^n) \\ &= H(S|T)-I(S\wedge Z^n|T)+I(S\wedge T|Z^n) \\ &= [I(S\wedge Y^n|T)-I(S\wedge Z^n|T)] \\ &\quad +[H(S|Y^nT)+I(S\wedge T|Z^n)] \end{aligned} \tag{36}$$

$$H(T)=I(T\wedge Y^n)+H(T|Y^n)=I(T\wedge Z^n)+H(T|Z^n), \tag{37}$$

where the terms furthest on the right are negligible by (30); actually, we shall take

$$\delta_1=\frac{1}{n}(H(S|Y^nT)+I(S\wedge T|Z^n)),$$

$$\delta_2=\frac{1}{n}(H(S|Y^nT)+\max[H(T|Y^n),H(T|Z^n)]),$$

$$\delta_3=\frac{1}{n}\max[H(T|Y^n),H(T|Z^n)]$$

in (35)–(37). Note that (34) is already established by (35) and (36).

In addition to the notation $Y^i=Y_1\cdots Y_i$, we designate $\tilde{Z}^i=Z_i\cdots Z_n$. Expand $I(S\wedge Y^n|T)$ and $I(S\wedge Z^n|T)$ starting with $I(S\wedge Y_1|T)$ and $I(S\wedge \tilde{Z}^n|T)$, respectively:

$$I(S\wedge Y^n|T)=\sum_{i=1}^n I(S\wedge Y_i|Y^{i-1}T)$$

$$I(S\wedge Z^n|T)=\sum_{i=1}^n I(S\wedge Z_i|\tilde{Z}^{i+1}T).$$

Using the identity

$$\begin{aligned} I(S\wedge Y_i|Y^{i-1}T) &= I(S\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1}T) \\ &\quad -I(\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1}TS) \\ &= I(S\wedge Y_i|Y^{i-1}\tilde{Z}^{i+1}T) \\ &\quad +I(\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1}T) \\ &\quad -I(\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1}TS), \end{aligned}$$

and its analog for $I(S\wedge Z_i|\tilde{Z}^{i+1}T)$, we obtain

$$I(S\wedge Y^n|T)=\sum_{i=1}^n I(S\wedge Y_i|Y^{i-1}\tilde{Z}^{i+1}T)+\Sigma_1-\Sigma_2 \tag{38}$$

$$I(S\wedge Z^n|T)=\sum_{i=1}^n I(S\wedge Z_i|Y^{i-1}\tilde{Z}^{i+1}T)+\Sigma_1^*-\Sigma_2^* \tag{39}$$

where

$$\Sigma_1=\sum_{i=1}^n I(\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1}T)$$

$$\Sigma_1^*=\sum_{i=1}^n I(Y^{i-1}\wedge Z_i|\tilde{Z}^{i+1}T)$$

and Σ_2, Σ_2^* are the analogous sums with TS instead of T.

Further, we have

$$\begin{aligned} I(T\wedge Y^n) &= \sum_{i=1}^n I(T\wedge Y_i|Y^{i-1}) \\ &= \sum_{i=1}^n (T\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1})-\sum_{i=1}^n I(\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1}T) \\ &\leq \sum_{i=1}^n (Y^{i-1}\tilde{Z}^{i+1}T\wedge Y_i)-\Sigma_1, \end{aligned} \tag{40}$$

and similarly

$$I(T\wedge Z^n) \leq \sum_{i=1}^n I(Y^{i-1}\tilde{Z}^{i+1}T\wedge Z_i)-\Sigma_1^*. \tag{41}$$

The key observation is the following.

Lemma 7: $\Sigma_1=\Sigma_1^*, \Sigma_2=\Sigma_2^*$.

Proof: Since

$$I(\tilde{Z}^{i+1}\wedge Y_i|Y^{i-1}T)=\sum_{j=i+1}^n I(Z_j\wedge Y_i|Y^{i-1}\tilde{Z}^{j+1}T),$$

$$I(Y^{i-1}\wedge Z_i|\tilde{Z}^{i+1}T)=\sum_{j=1}^{i-1} I(Y_j\wedge Z_i|Y^{j-1}\tilde{Z}^{i+1}T),$$

both Σ_1 and Σ_1^* split into terms of the form $I(Y_i\wedge Z_j|Y^{i-1}\tilde{Z}^{j+1}T)$ with $i<j$. $\Sigma_2=\Sigma_2^*$ follows similarly.

To conclude the proof, let us introduce an RV J independent of $STX^nY^nZ^n$ and uniformly distributed over $\{1,\cdots,n\}$. Set

$$U \triangleq Y^{J-1}\tilde{Z}^{J+1}TJ, \quad V \triangleq US, \quad X \triangleq X_J, \quad Y \triangleq Y_J, \quad Z \triangleq Z_J.$$

Then

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n I(S\wedge Y_i|Y^{i-1}\tilde{Z}^{i+1}T) &= I(S\wedge Y|U) \\ &= I(V\wedge Y|U), \end{aligned}$$

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n I(S\wedge Z_i|Y^{i-1}\tilde{Z}^{i+1}T) &= I(S\wedge Z|U) \\ &= I(V\wedge Z|U), \end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^n I(Y^{i-1}\tilde{Z}^{i+1}T\wedge Y_i)=I(U\wedge Y|J) \leq I(U\wedge Y),$$

$$\frac{1}{n}\sum_{i=1}^n I(Y^{i-1}\tilde{Z}^{i+1}T\wedge Z_i)=I(U\wedge Z|J) \leq I(U\wedge Z).$$

Substituting this into (38)–(41) and utilizing Lemma 7, (35)–(37) gives rise to (31)–(34). Using the memoryless

character of the channel, it is straightforward to verify that $U \to V \to X \to YZ$ and that the conditional distributions of Y and Z given X coincide with the corresponding channel matrices. The proof is complete.

VI. Discussion

We have considered a model for simultaneously broadcasting both messages for common use and confidential messages. For this model we characterized the achievable rates in terms of information quantities, so that the rate region is, in principle, computable. This is the commonly accepted criterion of a "solution" in multi-user Shannon theory. Of course, the actual computation may be very difficult. A possible approach is to look at the tangent planes to the rate region, as in [3], although we did not pursue this here. In some simple cases, however, numerical results are readily obtained. In the absence of effective coding-decoding techniques, our results are mainly of theoretical interest. We feel, however, that certain ideas of the proof—such as randomization via a prefixed DMC—might also be useful for practical code construction. It should be pointed out that our restriction to finite alphabets was a matter of convenience; the generalization to arbitrary alphabets is immediate by the standard technique of discrete approximations, cf. [4, Ch. 7]. Note that the only effect of eventual input constraints on our Theorems 1 and 2 is that the RV X appearing in the characterization of the rate region should satisfy the constraints. Finally, in Theorem 2, no essential use was made of the memoryless character of the source pair $\tilde{S}, \tilde{T}$; the generalization for stationary ergodic source pairs is straightforward, as in the Appendix to [8].

Appendix

Proof of Lemma 2: Given $U \to X \to YZ$, consider the DMC with input alphabet $\mathcal{U}$ and output alphabet $\mathcal{X}$ determined by the conditional distribution of X given U, as well as the cascade of this DMC with the given ones.

Let $\{u_m^n, \mathcal{B}_m, \mathcal{C}_m\}_{m \in \mathcal{M}_n}$ be an ϵ-code of block length n for simultaneous use for the cascaded channels, where the codewords $u_m^n \in \mathcal{U}^n$ are U-typical sequences, and $\mathcal{B}_m \subset \mathcal{F}_{Y|U}(u_m^n) \subset \mathcal{Y}^n$, $\mathcal{C}_m \subset \mathcal{F}_{Z|U}(u_m^n) \subset \mathcal{Z}^n$ are decoding sets such that

$$P_{Y|U}^n(\mathcal{B}_m|u_m^n) \geqslant 1-\epsilon, \qquad P_{Z|U}^n(\mathcal{C}_m|u_m^n) \geqslant 1-\epsilon \tag{A1}$$

where $0 < \epsilon < \frac{1}{5}$ is arbitrary and for $n \to \infty$

$$\frac{1}{n} \log \|\mathcal{M}_n\| \to \min [I(U \wedge Y), I(U \wedge Z)].$$

The existence of such codes is well-known and easily shown by standard methods (see e.g., [7, Ch. 4]).

Let $\mathcal{A}_m \subset \mathcal{X}^n$ consist of those sequences $x^n \in \mathcal{X}^n$, $X|U$-generated by u_m^n for which

$$P_{Y|X}^n(\mathcal{B}_m|x^n) \geqslant 1-3\epsilon, \qquad P_{Z|X}^n(\mathcal{C}_m|x^n) \geqslant 1-3\epsilon. \tag{A2}$$

The set of $x^n \in \mathcal{X}^n$ violating the first condition in (A2) has $P_{X|U}^n$-probability $\leqslant \frac{1}{3}$, by assumption (A1); this is similar for the second condition in (A2). Hence, using Fact B,

$$P_{X|U}^n(\mathcal{A}_m|u_m^n) \geqslant \frac{1}{4}, \qquad \text{for } m \in \mathcal{M}_n \tag{A3}$$

and for sufficiently large n. For $x^n \in \mathcal{A}_m$, let us write for brevity

$$\mathcal{B}(x^n) \triangleq \mathcal{F}_{Y|X}(x^n) \cap \mathcal{B}_m, \qquad \mathcal{C}(x^n) \triangleq \mathcal{F}_{Z|X}(x^n) \cap \mathcal{C}_m.$$

Then (A2) implies

$$P_{Y|X}^n(\mathcal{B}(x^n)|x^n) > 1-4\epsilon, \qquad P_{Z|X}^n(\mathcal{C}(x^n)|x^n) > 1-4\epsilon \tag{A4}$$

for sufficiently large n. We shall select the codewords x_{jlm}^n from $\mathcal{A}_m$ by a two-cycle maximal code construction, together with decoding sets $\mathcal{B}_{jlm} \subset \mathcal{B}(x_{jlm}^n)$ $\mathcal{C}_{jlm} \subset \mathcal{C}(x_{jlm}^n)$ satisfying

$$P_{Y|X}^n(\mathcal{B}_{jlm}|x_{jlm}^n) \geqslant 1-5\epsilon$$

$$P_{Z|X}^n(\mathcal{C}_{jlm}|x_{jlm}^n) \geqslant 1-5\epsilon.$$

Here all $\mathcal{B}_{jlm}$ must be disjoint. Of the sets $\mathcal{C}_{jlm}$, however, only those having the same middle index are required to be disjoint.

Let us pick sequences $x_{jlm} \in \mathcal{A}_m, j = 1, 2, \cdots$ and corresponding decoding sets with the above properties successively, in an arbitrary manner, and suppose that after the Nth step this procedure cannot be continued. Then for all $x^n \in \mathcal{A}_m$ either

$$P_{Y|X}^n\left(\mathcal{B}(x^n) \setminus \bigcup_{j=1}^{N} \mathcal{B}_{j1m}|x^n\right) < 1-5\epsilon, \tag{A5}$$

or

$$P_{Z|X}^n\left(\mathcal{C}(x^n) \setminus \bigcup_{j=1}^{N} \mathcal{C}_{j1m}|x^n\right) < 1-5\epsilon. \tag{A6}$$

Let $\mathcal{A}_m'$ resp. $\mathcal{A}_m''$ be the subsets of $\mathcal{A}_m$ where (A5) resp. (A6) holds. Then $\mathcal{A}_m = \mathcal{A}_m' \cup \mathcal{A}_m''$ and hence at least one of the relations

$$P_{X|U}^n(\mathcal{A}_m'|u_m^n) > \frac{1}{8}, \qquad P_{X|U}^n(\mathcal{A}_m''|u_m^n) > \frac{1}{8} \tag{A7}$$

holds (see (A3)). We have from (A4), (A5), and (A6)

$$P_{Y|X}^n\left(\bigcup_{j=1}^{N} \mathcal{B}_{j1m}|x^n\right) > \epsilon, \qquad \text{if } x^n \in \mathcal{A}_m'$$

$$P_{Z|X}^n\left(\bigcup_{j=1}^{N} \mathcal{C}_{j1m}|x^n\right) > \epsilon, \qquad \text{if } x^n \in \mathcal{A}_m''.$$

It follows by (A7) that either

$$P_{Y|U}^n\left(\bigcup_{j=1}^{N} \mathcal{B}_{j1m}|u_m^n\right) > \frac{\epsilon}{8} \tag{A8}$$

or

$$P_{Z|U}^n\left(\bigcup_{j=1}^{N} \mathcal{C}_{j1m}|u_m^n\right) > \frac{\epsilon}{8}. \tag{A9}$$

But because of Facts A and C we have

$$\begin{aligned} P_{Y|U}^n(\mathcal{B}_{j1m}|u_m^n) &\leqslant \|\mathcal{B}_{j1m}\| \exp\{-n(H(Y|U)-\eta_n)\} \\ &\leqslant \exp\{nH(Y|X)+n\eta_n\} \exp\{-nH(Y|U)+n\eta_n\} \\ &= \exp\{-nI(X \wedge Y|U)+2n\eta_n\} \end{aligned}$$

where $\eta_n \to 0$ as $n \to \infty$. Hence, if (A8) holds, we arrive at

$$N \exp\{-nI(X \wedge Y|U)+2n\eta_n\} > \frac{\epsilon}{8}. \tag{A10}$$

Similarly, if (A9) holds, we obtain

$$N \exp\{-nI(X \wedge Z|U)+2n\eta_n\} > \frac{\epsilon}{8}. \tag{A11}$$

Since by assumption $I(X \wedge Y|U) > I(X \wedge Z|U)$ for arbitrary $\eta > 0$, we have

$$N > \exp\{n(I(X \wedge Z|U)-\eta)\} \tag{A12}$$

if n is large enough. Thus we can select $N_1=[\exp\{n(I(X\wedge Z|U)-\eta)\}]$ codewords x_{j1m}^n with the required properties. Now we proceed to $l\geqslant 2$.

Suppose that we have already selected codewords x_{jlm}^n, $1\leqslant j\leqslant N_1$, $1\leqslant l\leqslant N_2-1$ with corresponding decoding sets $\mathcal{B}_{jlm}$ and $\mathcal{C}_{jlm}$, and start selecting x_{jlm}^n with $l=N_2$ in an arbitrary manner. If this procedure cannot be continued after the Nth step (including the possibility $N=0$), then for all $x^n\in\mathcal{A}_m$ either

$$P_{Y|X}^n\left(\mathcal{B}(x^n)\setminus\bigcup_{j=1}^{N_1}\bigcup_{l=1}^{N_2-1}\mathcal{B}_{jlm}\setminus\bigcup_{j=1}^{N}\mathcal{B}_{jN_2m}|x^n\right)<1-5\epsilon \quad \text{(A13)}$$

or

$$P_{Z|X}^n\left(\mathcal{C}(x^n)\setminus\bigcup_{j=1}^{N}\mathcal{C}_{jN_2m}|x^n\right)<1-5\epsilon. \quad \text{(A14)}$$

(Recall that of the sets $\mathcal{C}_{jlm}$, only those having the same middle index have to be disjoint.) By the same argument as above, it follows that either (cf. (A10), (A11))

$$(N+N_1(N_2-1))\exp\{-nI(X\wedge Y|U)+2n\eta_n\}>\frac{\epsilon}{8} \quad \text{(A15)}$$

or

$$N\exp\{-nI(X\wedge Z|U)+2n\eta_n\}>\frac{\epsilon}{8}. \quad \text{(A16)}$$

Supposing that $N_2\leqslant\exp\{n(I(X\wedge Y|U)-I(X\wedge Z|U))\}$ we still obtain (A12) which means that the N_2th cycle of codeword selection can be completed.

This proves the existence of a class of codewords x_{j1m}^n,

$$1\leqslant j\leqslant\exp\{n(I(X\wedge Z|U)-\eta)\}$$

$$1\leqslant l\leqslant\exp\{n(I(X\wedge Y|U)-I(X\wedge Z|U)\}$$

with corresponding decoding sets, where $\eta>0$ is arbitrary. It was shown earlier that the range of m is of the right size, so the proof of Lemma 2 is complete.

Proof of the direct part of Theorem 2: Given two memoryless sources $\tilde{S}$ and $\tilde{T}$ with alphabets $\mathcal{S}$ and $\mathcal{T}$, a standard argument with typical sequences shows the existence of subsets $\mathcal{A}_k\subset\mathcal{T}^k$ and for each $t^k\in\mathcal{A}_k$, of subsets $\mathcal{B}_k(t^k)\in\mathcal{S}^k$ such that

$$\Pr\{T^k\in\mathcal{A}_k\}\geqslant 1-\eta_k,$$

$$\Pr\{S^k\in\mathcal{B}_k(t^k)|T^k=t^k\}\geqslant 1-\eta_k$$

where for each $t^k\in\mathcal{A}_k$

$$\exp[-k(H(\tilde{T})+\eta_k)]\leqslant\Pr\{T^k=t^k\}\leqslant\exp[-k(H(\tilde{T})-\eta_k)]$$

and for each $s^k\in\mathcal{B}_k(t^k)$, $t^k\in\mathcal{A}_k$,

$$\exp[-k(H(\tilde{S}|\tilde{T})+\eta_k)]\leqslant\Pr\{S^k=s^k|T^k=t^k\}\leqslant\exp[-k(H(\tilde{S}|\tilde{T})-\eta_k)]$$

where $\eta_k\to 0$. Moreover, one may assume that $\|\mathcal{B}_k(t^k)\|$ is constant for $t^k\in\mathcal{A}_k$. Of course, we have

$$\frac{1}{k}\log\|\mathcal{A}_k\|\to H(\tilde{T}) \quad \text{(A17)}$$

$$\frac{1}{k}\log\|\mathcal{B}_k(t^k)\|\to H(\tilde{S}|\tilde{T}),\qquad (t^k\in\mathcal{A}_k). \quad \text{(A18)}$$

Let us suppose that for some $R>0$, $\Delta\geqslant 0$

$$RH(\tilde{S}|\tilde{T})=R_1,\qquad R\Delta=R_e,\qquad RH(\tilde{T})=R_0 \quad \text{(A19)}$$

where $(R_1,R_e,R_0)\in\mathcal{R}$. Consider a sequence of encoders f_n with block length $n=1,2,\cdots$ for the BCC, with message sets $\mathcal{S}_n$ and $\mathcal{T}_n$, such that

$$\frac{1}{n}\log\|\mathcal{S}_n\|\to R_1\qquad \frac{1}{n}\log\|\mathcal{T}_n\|\to R_0 \quad \text{(A20)}$$

and

$$\liminf_{n\to\infty}\frac{1}{n}H(S_n'|Z^n)\geqslant R_e \quad \text{(A21)}$$

for all pairs of RV's S_n', T_n' satisfying (16), and let (φ_n,ψ_n) be suitable decoders giving rise to (n,ϵ_n)-transmission, with $\epsilon_n\to 0$. On account of (A17)–(A19), for any $\epsilon>0$ and for sufficiently large n, there exists k such that

$$R-\epsilon\leqslant\frac{k}{n}\leqslant R,$$

$$\|\mathcal{A}_k\|\leqslant\|\mathcal{T}_n\|-1,\quad \|\mathcal{B}_k(t^k)\|=\|\mathcal{S}_n\|\qquad (t^k\in\mathcal{A}_k).$$

(More exactly, $\mathcal{B}_k(t^k)$ can be chosen to satisfy the latter condition.) Let $g_n:\mathcal{S}^k\times\mathcal{T}^k\to\mathcal{S}_n\times\mathcal{T}_n$ map different pairs (s^k,t^k) $t^k\in\mathcal{A}_k$, $s^k\in\mathcal{B}_k(t^k)$ into different pairs $(s,t)\in\mathcal{S}_n\times\mathcal{T}_n$, with t depending only on t^k (if $S^k\in\mathcal{B}_k(t^k)$); moreover, all the remaining (s^k,t^k) are mapped into the same $(s_0,t_0)\in\mathcal{S}_n\times\mathcal{T}_n$ where t_0 differs from all the previous t's.

Consider the (k,n)-encoder defined as the superposition of the source encoder g_n and the channel encoder f_n. Let X^n,Y^n,Z^n be the corresponding channel input and output RV's. It is obvious from the construction that receiver 1 can reproduce S^k,T^k and receiver 2 can reproduce T^k with probability of error less than $\epsilon_n+\eta_k\to 0$. This is even stronger than the error frequency criterion of reliability used in Theorem 2.

Moreover, the RV's $S_nT_n=g_n(S^k,T^k)$ satisfy the condition (16), hence by (A21) we have

$$\liminf_{n\to\infty}\frac{1}{n}H(S_n|Z^n)\geqslant R_e=R\Delta.$$

But

$$\frac{1}{n}H(S^k|Z^n)\geqslant\frac{1}{n}H(S_n|Z^n)-\frac{1}{n}H(S_n|S^kZ^n)$$

where S^kZ^n determines S_n with error probability at most $\epsilon_n+\eta_k\to 0$ (since $\Pr\{\psi(Z^n)\neq T_n\}\leqslant\epsilon_n$). Hence by Fano's lemma $(1/n)H(S_n|S^kZ^n)\to 0$. Thus for some sequence $\delta_n\to 0$

$$\frac{1}{n}H(S^k|Z^n)\geqslant R\Delta-\delta_n\geqslant\frac{k}{n}\Delta-\delta_n,$$

and so

$$\liminf_{\substack{n\to\infty\\ k\to\infty}}\frac{1}{k}H(S^k|Z^n)\geqslant\Delta.$$

This completes the proof.

Size Constraints

Proof of the admissibility of the constraints:

$$\|\mathcal{U}\|\leqslant\|\mathcal{X}\|+3,\quad \|\mathcal{V}\|\leqslant\|\mathcal{X}\|^2+4\|\mathcal{X}\|+3. \quad \text{(A22)}$$

We use Lemma 3 of [1], noting that the number $k+1$ there can be replaced by k by using instead of Caratheodory's theorem its strengthening by Fenchel and Eggleston referred to in [9]. It suffices to show that if $U\to V\to X\to YZ$ the RV's U and V may be replaced by new ones, preserving the Markovity and the mutual information $I(U\wedge Y)$, $I(U\wedge Z)$, $I(V\wedge Y|U)$, $I(V\wedge Z|U)$, such that for the range of the new U and V (A22) holds. If $\bar{p}=\{p(v);\ v\in\mathcal{V}\}$ ranges over the distributions on $\mathcal{V}$, the distribution

$$\bar{p}^X=\left\{\sum_{v\in\mathcal{V}}p(v)p_{X|V}(x|v);\ x\in\mathcal{X}\right\}$$

on $\mathcal{X}$ as well as the corresponding output distributions $\bar{p}^Y$ resp. $\bar{p}^Z$ of channel 1 resp. 2 are continuous vector functions of $\bar{p}$. By means of them we define $\|\mathcal{X}\|+3$ continuous scalar functions of $\bar{p}$:

$$f_x(\bar{p}) \triangleq p^X(x),\ f_Y(\bar{p}) \triangleq H(p^Y),\ f_Z(\bar{p}) \triangleq H(\bar{p}^Z)$$

$$f_{YV}(\bar{p}) \triangleq \sum_{v\in\mathcal{V}} p(v)H(Y|V=v),$$

$$f_{ZV}(\bar{p}) \triangleq \sum_{v\in\mathcal{V}} p(v)H(Z|V=v),$$

where of the functions $f_x(\bar{p})$, $x\in\mathcal{X}$ only $\|\mathcal{X}\|-1$ are to be considered. Notice that $H(Y|V=v)$ and $H(Z|V=v)$ do not depend on $\bar{p}$. Now let $\bar{p}_u(v)=\Pr\{V=v|U=u\}$. With these distributions $\bar{p}_u$ on $\mathcal{V}$, we have

$$\Pr(X=x)=\sum_{u\in\mathcal{U}} \Pr(U=u)f_x(\bar{p}_u) \tag{A23}$$

$$I(U\wedge Y)=H(Y)-\sum_{u\in\mathcal{U}} \Pr(U=u)f_Y(\bar{p}_u) \tag{A24}$$

$$I(U\wedge Z)=H(Z)-\sum_{u\in\mathcal{U}} \Pr(U=u)f_Z(\bar{p}_u) \tag{A25}$$

$$I(V\wedge Y|U)=\sum_{u\in\mathcal{U}} \Pr(U=u)\left[f_Y(\bar{p}_u)-f_{YV}(\bar{p}_u)\right] \tag{A26}$$

$$I(V\wedge Z|U)=\sum_{u\in\mathcal{U}} \Pr(U=u)\left[f_Z(\bar{p}_u)-f_{ZV}(\bar{p}_u)\right]. \tag{A27}$$

It follows from Lemma 3 of [1] that the RV's U and V may be replaced by new ones such that the new U takes at most $\|\mathcal{X}\|+3$ different values and the expressions (A23)–(A27) are preserved. (X, Y, Z need not be changed as the distribution of X has been fixed by (A23).)

Starting from the new U and V, we now turn to the range constraint on V and look at the following $\|\mathcal{X}\|+1$ continuous functions of a distribution $\bar{p}=\{p(x);\ x\in\mathcal{X}\}$ on $\mathcal{X}$:

$$f_x(\bar{p}) \triangleq p(\bar{x}),\qquad f_Y(\bar{p}) \triangleq H(\bar{p}^Y),\qquad f_Z(\bar{p}) \triangleq H(\bar{p}^Z),$$

where of the functions $f_x(\bar{p})$, only $\|\mathcal{X}\|-1$ are considered. Then, for every fixed u,

$$\Pr(X=x|U=u)=\sum_{v\in\mathcal{V}} \Pr(V=v|U=u)f_x(\bar{p}_v) \tag{A28}$$

$$I(V\wedge Y|U=u)=H(Y|U=u)-\sum_{v\in\mathcal{V}} \Pr(V=v|U=u)f_Y(\bar{p}_v) \tag{A29}$$

$$I(V\wedge Z|U=u)=H(Z|U=u)-\sum_{v\in\mathcal{V}} \Pr(V=v|U=u)f_Z(\bar{p}_v) \tag{A30}$$

where $\bar{p}_v$ is the conditional distribution of X given $V=v$. If the probabilities $\Pr(X=x|U=u)$ are fixed, then so are $H(Y|U=u)$ and $H(Z|U=u)$. Hence by Lemma 3 of [1], for every fixed u there exists a RV V_u with no more than $\|\mathcal{X}\|+1$ values such that, under the condition $U=u$, we have $V_u\to X\to XZ$ and the expressions (A28)–(A30) are preserved. Again, X, Y, Z need not be changed since the conditional distribution of X has been fixed by (A28). Choosing UV_U for our new V completes the proof.

References

[1] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 6, pp. 629–638, Nov. 1975.

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

[3] R. G. Gallager, "Coding for degraded broadcast channels," *Probl. Peredači Informacii*, vol. 10, no. 3, pp. 3–14, July-Sept. 1974.

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

[5] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 1, pp. 60–64.

[6] J. Körner and K. Marton, "Comparison of two noisy channels," *Topics in Information Theory*, Keszthely (Hungary) 1975, Colloquia Math. Soc. János Bolyai, Amsterdam: North-Holland Publ., 1977, pp. 411–423.

[7] J. Wolfowitz, *Coding Theorems of Information Theory*, 2nd Ed. Berlin: Springer-Verlag, 1964.

[8] A. D. Wyner, "The wire-tap channel," *Bell. Syst. Tech. J.* vol. 54, no. 8, pp. 1355–1387, 1975.

[9] A. D. Wyner, and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–11, Jan. 1976.