The Common Randomness Capacity of a Pair of Independent Discrete Memoryless Channels

Sivarama Venkatesan, *Student Member, IEEE*, and Venkat Anantharam, *Member, IEEE*

***Abstract*—We study the following problem: two agents Alice and Bob are connected to each other by independent discrete memoryless channels. They wish to generate *common randomness*, i.e., agree on a common random variable, by communicating interactively over the two channels. Assuming that Alice and Bob are allowed access to independent external random sources at rates (in bits per step of communication) of H_A and H_B, respectively, we show that they can generate common randomness at a rate of $\max\{\min[H_A + H(W \mid Q), I(P; V)] + \min[H_B + H(V \mid P), I(Q; W)]\}$ bits per step, by exploiting the *noise* on the two channels. Here, V is the channel from Alice to Bob, and W is the channel from Bob to Alice. The maximum is over all probability distributions P and Q on the input alphabets of V and W, respectively. We also prove a strong converse which establishes the above rate as the highest attainable in this situation.**

***Index Terms*—Common randomness capacity, generating randomness from noise, interactive communication.**

I. Introduction

As pointed out by Ahlswede and Csiszár in [2], there are several situations in which common randomness available to communicating agents plays a significant role. For example, in the theory of identification via noisy channels ([3]–[5]), the maximum achievable identification rate is essentially determined by the amount of common randomness that the transmitter and receiver can set up. Common randomness available to transmitter and receiver also allows them to use random codes for data transmission, which can be far superior to deterministic codes in certain situations, e.g., with arbitrarily varying channels ([1], [7]). Finally, in the theory of communication complexity, it is known that common randomness can significantly reduce the amount of interprocessor communication required to perform certain computations in a distributed setting ([8], [11]). For these and other reasons, Ahlswede and Csiszár [2] proposed a systematic study of the role of common randomness in information theory.

Now, in a situation where the communicating agents only have access to private random sources, they must set up common randomness by exchanging the outputs of their respective sources. Typically, this exchange must take place over noisy channels. However, in this situation, it is conceivable that they could generate *additional* common randomness by somehow exploiting the noise on those channels. In fact, intuition suggests that the agents should be able to generate common randomness even in the absence of any external random sources, simply by communicating over the noisy channels and making use of the randomness in the channel outputs. It is then naturally of interest to determine the maximum rate, in bits per step of communication, at which common randomness could be extracted from channel noise in this way.

In this paper, we answer the above question for the case of two agents Alice and Bob connected to each other in both directions by independent discrete memoryless channels (DMC's). To illustrate the problem, consider the simple case where the channel from Alice to Bob is binary symmetric with crossover probability $1/2$, and that from Bob to Alice is a noiseless binary channel. Suppose the following communication takes place between them: Alice transmits 0 in n successive steps; and in step k, $2 \le k \le n$, Bob echoes the bit he received in step $k-1$ back to Alice. Then, after n steps, both Alice and Bob know the value of a random variable uniformly distributed over a set of size 2^{n-1}, i.e., they have generated $n-1$ bits of "common randomness."

Note that this common randomness is derived from the noise on the channel from Alice to Bob at a rate of $(n-1)/n$ bits per step, which can be made arbitrarily close to 1 by making n large enough. It is not hard to see that in the absence of external sources no rate higher than 1 bit/step can be achieved here. Thus the common randomness "capacity" of this pair of channels is 1 bit/step, even without external sources.

In the general case of arbitrary discrete memoryless channels both ways, the situation is more complicated; it may not be possible to guarantee perfect agreement between Alice and Bob, or to generate random variables with perfectly uniform distributions. Therefore, we will only require that Alice and Bob generate random variables that agree with high probability, and have distributions close to uniform on some common set. The question of interest, then, is how fast the size of this set can grow with the number of steps of communication. The common randomness capacity measures the maximum achievable rate of this growth.

The precise formulation of the problem appears in Section II-B. The main result, stated in Section II-C, is the determination of the common randomness capacity for an arbitrary pair of channels when no external sources of randomness are available to either agent. We also consider the case where

Manuscript received November 7, 1995; revised June 11, 1997. This work was supported by the National Science Foundation under Grants IRI 9005849 and NCR 9422513, and by AT&T Foundation. The material in this paper was presented in part at the 34th Annual Allerton Conference on Communications and Control, Allerton, IL, October 2–4, 1996.

S. Venkatesan is with Cornell University, Ithaca, NY, USA, and with the University of California at Berkeley, Berkeley, CA 94720 USA.

V. Anantharam is with the University of California at Berkeley, Berkeley, CA 94720 USA.

Publisher Item Identifier S 0018-9448(98)00012-1.

both agents have access to independent i.i.d. (independent and identically distrbuted) sources of randomness, which they can sample once for each step of communication (this constrains the rate at which they receive external randomness). While this problem appears to be more general, we will show in Section II-D that it can actually be reduced to the problem with no external sources.

For the simple special case of a DMC with perfect instantaneous feedback, the common randomness capacity was obtained by Ahlswede and Dueck in [3], as an auxiliary result in the proof of their identification theorem. They considered two situations, one where no external randomness is available, and another where the agent at the transmitting end of the DMC has unlimited randomness (in both situations, the agent at the receiving end of the DMC is essentially "passive" since he cannot control the inputs to the feedback channel in any way). In fact, in all the identification problems studied in [3] and [4], it turns out that the (second-order) identification capacity equals the (first-order) common randomness capacity. These results were an important motivation for our study of the common randomness capacity in the general case, where both agents can play an "active" role and both channels are allowed to be noisy.

Another area in which common randomness has an obvious significance is cryptography: if two agents share a random key about which an eavesdropper has no information, they can use it to achieve secure communication between them, through encryption of messages. In this context, the problem is one of secret sharing, i.e., generating common randomness at two terminals without giving information about it to an eavesdropper. This has recently been addressed by Maurer ([9], [10]), and Ahlswede and Csiszár [2].

In the "channel-type" model introduced in [2], the two terminals are connected by a DMC with one input and two outputs. One terminal governs the input, while the outputs are seen by the other terminal and the wiretapper, respectively. There is also a noiseless public two-way channel of unlimited capacity connecting the two terminals. Both terminals have access to independent and unlimited sources of randomness, to begin with. The authors proved bounds on the maximum rate at which the two terminals could generate a shared secret key, under various restrictions on the use of the public channel.

We remark that our results are not applicable in a cryptographic context, since no secrecy constraints are imposed on the process of generating common randomness, i.e., the random outputs generated by the two agents need not be kept secret from any eavesdroppers. However, the results proved here are not implied by those of [2], because both channels here are allowed to be noisy and constrained in capacity, and no restrictions are imposed on the allowed use of these two channels.

II. Statement of Problem and Result

A. Preliminaries

A discrete memoryless channel (DMC) with input alphabet $\mathcal{Z}$, output alphabet $\hat{\mathcal{Z}}$, and transition probability function U will be denoted by $(\mathcal{Z}, \hat{\mathcal{Z}}, U)$, or just U if no confusion can result.

$\mathcal{P}(\mathcal{Z})$ will denote the set of all probability distributions on the set $\mathcal{Z}$, and $\mathcal{P}_n(\mathcal{Z}) \subseteq \mathcal{P}(\mathcal{Z})$ will denote the set of all n-types, i.e., $P \in \mathcal{P}_n(\mathcal{Z})$ iff $nP(z)$ is an integer for all $z \in \mathcal{Z}$. $\mathcal{W}(\mathcal{Z}, \hat{\mathcal{Z}})$ will denote the set of all DMC's with input alphabet $\mathcal{Z}$ and output alphabet $\hat{\mathcal{Z}}$. The notation for all standard information-theoretic quantities is that of [6].

All logarithms and exponentials will be to the base two. Throughout, $[L]$ will denote the set of integers $\{1, 2, \cdots, L\}$.

B. Definition of a Protocol

We will now formulate precisely the problem of generating common randomness over noisy channels in the absence of external sources. Later, in Section II-D, we will show how external sources can be incorporated into this framework.

Let the DMC from Alice to Bob be $(\mathcal{X}, \hat{\mathcal{X}}, V)$, and that from Bob to Alice be $(\mathcal{Y}, \hat{\mathcal{Y}}, W)$. To generate common randomness, Alice and Bob communicate with each other for a certain number of steps. In each step, Alice transmits a symbol to Bob across V and, simultaneously, Bob transmits a symbol to Alice across W. These symbols are determined by an agreed-upon strategy, as functions only of all the past receptions available to the respective senders.

Formally, an n-step strategy is a pair (f, g), with

$$f = (f_1, f_2, \cdots, f_n)$$

and

$$g = (g_1, g_2, \cdots, g_n).$$

Here, $f_1 \in \mathcal{X}$, $g_1 \in \mathcal{Y}$, and, for $2 \le k \le n$, $f_k : \hat{\mathcal{Y}}^{k-1} \to \mathcal{X}$ and $g_k : \hat{\mathcal{X}}^{k-1} \to \mathcal{Y}$. Let X_k and Y_k denote the symbols transmitted by Alice and Bob, respectively, in the kth step, $1 \le k \le n$, and let these be received as $\hat{X}_k$ and $\hat{Y}_k$, respectively. Then, $X_1 = f_1$, $Y_1 = g_1$, and, for $2 \le k \le n$, $X_k = f_k(\hat{Y}^{k-1})$ and $Y_k = g_k(\hat{X}^{k-1})$. Note that

$$\Pr[\hat{X}^n = \hat{x}^n, \hat{Y}^n = \hat{y}^n] = \prod_{k=1}^{n} [V(\hat{x}_k \mid f_k(\hat{y}^{k-1})) \cdot W(\hat{y}_k \mid g_k(\hat{x}^{k-1}))]. \quad (1)$$

The $k = 1$ term on the right-hand side is to be understood as $V(\hat{x}_1 \mid f_1) \cdot W(\hat{y}_1 \mid g_1)$.

After n steps, each agent separately decides whether the attempt to generate common randomness was successful or not, and in the former case computes a random output. Alice's decision is based only on $\hat{Y}^n$, and Bob's is based only on $\hat{X}^n$. Their random outputs take values in the common finite set $[K] = \{1, \cdots, K\}$. Formally, Alice computes $S = S(\hat{Y}^n)$ and Bob computes $T = T(\hat{X}^n)$, where $S : \hat{\mathcal{Y}}^n \to \{e\} \cup [K]$ and $T : \hat{\mathcal{X}}^n \to \{e\} \cup [K]$. Here, e is a symbol indicating failure to generate common randomness.

The quadruple (f, g, S, T) defines an (n, K) *protocol* for generating common randomness. Of course, the "amount" of randomness generated by the protocol, and the extent to which it is "common," are determined by the joint distribution of S

and T. Ideally, we would like to have

$$\Pr[S(\hat{Y}^n) = T(\hat{X}^n) = l] = \frac{1}{K}, \qquad \text{for each } l \in [K] \quad (2)$$

with K as large as possible. If (2) were true, S and T would be equal with probability 1, and uniformly distributed over $[K]$. (There would be no "failure" events of positive probability.)

In general, it is not possible to satisfy (2) except in the trivial case $K = 1$. Therefore, we will have to settle for *approximate* equality and uniformity of S and T. To this end, we make the following definition: (f, g, S, T) is an (n, K, λ) protocol if

$$\begin{aligned}\frac{1-\lambda}{K} &\le \Pr[S(\hat{Y}^n) = T(\hat{X}^n) = l] \\ &\le \frac{1+\lambda}{K}, \qquad \text{for each } l \in [K]. \quad (3)\end{aligned}$$

This definition is of interest only for λ in $[0, 1)$. For, if $\lambda \ge 1$, we can satisfy (3) with arbitrarily large K (for any n) simply by taking $S(\hat{y}^n)$ and $T(\hat{x}^n)$ to be e for all $(\hat{x}^n, \hat{y}^n)$, thus making the problem trivial.

To motivate the above definition, note that (3) implies

$$\Pr\left[\bigcup_{l\in[K]} \{S = T = l\}\right] \ge 1 - \lambda \quad (4)$$

so that, if λ is small, both agents compute the *same* random output with high probability. In particular, the probability that either agent declares failure to generate common randomness is small. Further, since $\Pr[S \ne T] \le \lambda$, Fano's inequality gives

$$\max\{H(S \mid T), H(T \mid S)\} \le 1 + \lambda \log K. \quad (5)$$

Also,

$$\begin{aligned}H(S, T) &\ge -\sum_{l=1}^{K} \Pr[S = T = l] \log \Pr[S = T = l] \\ &\ge \sum_{l=1}^{K} \left(\frac{1-\lambda}{K}\right) \log\left(\frac{K}{1+\lambda}\right) \\ &\ge (1-\lambda)\log K - 1, \qquad \text{since } 0 \le \lambda < 1. \quad (6)\end{aligned}$$

From (5) and (6)

$$\min\{H(S), H(T)\} \ge (1 - 2\lambda)\log K - 2. \quad (7)$$

Thus if λ is small, each agent generates a random output whose distribution is close to uniform on $[K]$. For future reference note also that, by (5) and (7)

$$I(S; T) \ge (1 - 3\lambda)\log K - 3. \quad (8)$$

C. Main Result

Fix $\lambda \in [0, 1)$. For each $n \ge 1$, define $K(n, \lambda)$ to be the largest K such that there exists an (n, K, λ) protocol. The main result proved here is the following:

Theorem 2.1 (Main Theorem): Let

$$\begin{aligned}R^*(V, W) = \max_{\substack{P\in\mathcal{P}(\mathcal{X})\\ Q\in\mathcal{P}(\mathcal{Y})}} \{&\min[H(W \mid Q), I(P; V)] \\ &+ \min[H(V \mid P), I(Q; W)]\}. \quad (9)\end{aligned}$$

Then

a) (Direct part)

$$\liminf_{n\to\infty} \frac{1}{n} \log K(n, \lambda) \ge R^*, \qquad \text{for all } \lambda \in (0, 1). \quad (10)$$

b) (Converse part)

$$\limsup_{n\to\infty} \frac{1}{n} \log K(n, \lambda) \le R^*, \qquad \text{for all } \lambda \in [0, 1). \quad (11)$$

Thus

$$\lim_{n\to\infty} \frac{1}{n} \log K(n, \lambda) = R^*, \qquad \text{for all } \lambda \in (0, 1).$$

Define rate R of generating common randomness to be achievable if there exists a sequence of (n, K_n, λ_n) protocols such that

$$\lim_{n\to\infty} \lambda_n = 0 \quad \text{and} \quad \liminf_{n\to\infty} \frac{\log K_n}{n} \ge R. \quad (12)$$

Then, (10) is obviously equivalent to the statement that any rate not exceeding R^* is achievable. A "weak" converse to (10) would merely assert that rates above R^* are not achievable. In the Appendix, we outline the simple proof of the following statement which implies the weak converse:

$$\lim_{\lambda\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log K(n, \lambda) \le R^*. \quad (13)$$

However, (11) says much more than (13); in the usual terminology, (11) is a "strong" converse to (10). Together, (10) and (11) justify the interpretation of R^* as the *common randomness capacity* of the given pair of channels (in the absence of external sources). We will prove (10) in Section III and (11) in Section IV.

D. Incorporating External Random Sources

We will now address the following question: In the above framework, suppose Alice and Bob *do* have access to external sources of randomness at certain rates. What would the common randomness capacity then be, as a function of these rates and the characteristics of the channels?

While this problem appears to be more general, we will show that it is not really so. Suppose Alice and Bob have independent i.i.d. sources with respective entropies H_A and H_B bits per symbol. Let $A_1, A_2, \cdots$ be the sequence of outputs from Alice's source, and $B_1, B_2, \cdots$ that from Bob's source. We will assume that Alice and Bob can sample their sources once for each step of communication, i.e., just before the kth step of communication, Alice learns A_k and Bob learns B_k. (Thus they receive external randomness at the rates of H_A and H_B bits per step.) They are then allowed to choose X_k and Y_k as functions of $(A^k, \hat{Y}^{k-1})$ and $(B^k, \hat{X}^{k-1})$, respectively. Similarly, $S = S(A^n, \hat{Y}^n)$ and $T = T(B^n, \hat{X}^n)$. The requirement for an (n, K, λ) protocol is the same as before, viz., for each $l \in [K]$, $\Pr[S = T = l]$—which now involves an averaging over (A^n, B^n) also—should be within λ/K of $1/K$.

If $K'(n, \lambda)$ is the largest K such that there exists an (n, K, λ) protocol in this situation, then for any $\lambda \in (0, 1)$ we claim that $\lim_{n\to\infty} (1/n) \log K'(n, \lambda)$ equals

$$\max_{\substack{P\in\mathcal{P}(\mathcal{X})\\ Q\in\mathcal{P}(\mathcal{Y})}} \{\min[H_A + H(W \mid Q), I(P; V)] + \min[H_B + H(V \mid P), I(Q; W)]\}. \quad (14)$$

To see this, consider the original problem (without external sources) but with the channels $(\mathcal{X}, \hat{\mathcal{X}}, V)$ and $(\mathcal{Y}, \hat{\mathcal{Y}}, W)$ replaced by $(\mathcal{X}, \hat{\mathcal{X}} \times \mathcal{B}, \tilde{V})$ and $(\mathcal{Y}, \hat{\mathcal{Y}} \times \mathcal{A}, \tilde{W})$, respectively. Here, $\mathcal{A}$ and $\mathcal{B}$ are the alphabets in which A_k and B_k, respectively, take values

$$\tilde{V}(\hat{x}, b \mid x) = V(\hat{x} \mid x)\Pr[B_k = b]$$

and

$$\tilde{W}(\hat{y}, a \mid y) = W(\hat{y} \mid y)\Pr[A_k = a].$$

Then, we may think of the channel $\tilde{V}$ as providing the sequence $B_1, B_2, \cdots$ to Bob, and the channel $\tilde{W}$ as providing $A_1, A_2, \cdots$ to Alice, in addition to behaving like V and W, respectively.

Clearly, to any (n, K, λ) protocol in the problem with external sources over V and W, there corresponds an $(n + 1, K, \lambda)$ protocol in the problem without external sources over $\tilde{V}$ and $\tilde{W}$ (an extra step is required at the beginning just to provide A_1 to Alice and B_1 to Bob); and to any (n, K, λ) protocol in the latter problem, there corresponds an (n, K, λ) protocol in the former. Therefore, by Theorem 2.1,

$$\lim_{n\to\infty} (1/n) \log K'(n, \lambda) = R^*(\tilde{V}, \tilde{W}), \qquad \text{for any } \lambda \in (0, 1).$$

It only remains to observe that $R^*(\tilde{V}, \tilde{W})$ reduces to (14).

In fact, the capacity remains the same even if A^n and B^n are revealed to Alice and Bob right at the start of an n-step protocol. The proof of this only requires some simple modifications to the proof of the converse in Section IV.

E. Examples

Let

$$H^V = \max_x H(V(\cdot \mid x))$$

and

$$H^W = \max_y H(W(\cdot \mid y)).$$

Let

$$C^V = \max_P I(P; V)$$

and

$$C^W = \max_Q I(Q; W)$$

be the Shannon capacities of V and W.

1) Suppose $H(V(\cdot \mid x)) = H^V$ for all x, and $H(W(\cdot \mid y)) = H^W$ for all y (this holds, e.g., for all symmetric channels). Then, the maximum in (14) is attained by the P and Q that achieve the Shannon capacities of V and W, respectively, and the common randomness capacity reduces to

$$\min[H_A + H^W, C^V] + \min[H_B + H^V, C^W].$$

In fact, this is always an upper bound on the capacity.

In particular, consider the case where V and W are both binary-symmetric channels, with crossover probabilities p and q, respectively. Then, assuming $H_A = H_B = 0$, the common randomness capacity equals

$$\min\{h(p) + h(q), 2 - h(p) - h(q)\}$$

where $h(\cdot)$ is the binary entropy function. Note that this is 0 iff $h(p) = h(q) = 0$ or $h(p) = h(q) = 1$. In the first case, the two channels do not provide any randomness (zero entropy), although they allow for perfect agreement between the two agents (high capacity). In the second case, the situation is reversed; a transmission by either agent provides a totally random bit to the other (high entropy), but the randomness generated this way cannot be reliably communicated back to the sender (zero capacity).

On the other hand, the capacity attains its maximum value of 1 whenever the entropies and capacities balance each other, i.e., when

$$h(p) + h(q) = (1 - h(p)) + (1 - h(q)).$$

It is somewhat surprising that it is possible to generate common randomness at a rate of 1 bit/step in all these cases.

In the binary-symmetric case, a much simpler proof of Theorem 2.1 is given in [12]. This proof can easily be extended to the case where V and W are symmetric DMC's.

2) Suppose $H^W = 0$, i.e., W is completely noiseless. Then, (14) reduces to

$$\max_P \{\min[H_A, I(P; V)] + \min[H_B + H(V \mid P), C^W]\}.$$

In addition, a) if $H_A = 0$, then the capacity is $\min[H_B + H^V, C^W]$; b) if $C^W \geq H^V$, $H_A \geq C^V$, and $H_B \leq C^W - H^V$, then the capacity is $H_B + \max_P H(PV)$. These are slight generalizations of the results of Ahlswede and Dueck [3] for DMC's with feedback, mentioned in the Introduction. On the other hand, if $C^W = 0$ (i.e., W is completely noisy), then (14) reduces to $\min[H_A + H^W, C^V]$.

3) If $H_A \geq C^V$ and $H_B \geq C^W$, then (14) just equals $C^V + C^W$. Thus if Alice and Bob have external randomness available at sufficiently large rates, then there is nothing to be gained from channel noise (they can simply exchange their source outputs, using the usual channel coding techniques, in order to generate common randomness at the optimal rate). Note, however, that the proof of the strong converse is nontrivial even in this extreme case!

III. Proof of the Direct Part

We will now prove that Alice and Bob can generate common randomness at rates arbitrarily close to R^*, i.e., for any $R < R^*$, we will prove the existence of a sequence of (n, K_n, λ_n) protocols satisfying (12). This suffices to prove the direct part of Theorem 2.1.

Actually, to prove that rate R is achievable, it is sufficient to exhibit a $(t^2, K_{t^2}, \lambda_{t^2})$ protocol for each t, such that

$$\lim_{t\to\infty} \lambda_{t^2} = 0 \quad \text{and} \quad \liminf_{t\to\infty} \frac{\log K_{t^2}}{t^2} \geq R.$$

For, given any n satisfying $t^2 < n < (t+1)^2$, Alice and Bob could execute the $(t^2, K_{t^2}, \lambda_{t^2})$ protocol and fill the remaining $n - t^2$ steps arbitrarily, without affecting the rate achieved. (Essentially, this is because $\lim_{t\to\infty} t^2/(t+1)^2 = 1$.) We will therefore restrict attention to protocols with t^2 steps, in all that follows. But, first, we state some results that will be needed in the proof.

A. Preliminary Results

Definition 3.1: A (t, L, γ) *block code* for the DMC $(\mathcal{Z}, \hat{\mathcal{Z}}, U)$ is a collection $\{(\boldsymbol{c}_l, \mathcal{C}_l) : l = 1, 2, \cdots, L\}$, where $\boldsymbol{c}_l \in \mathcal{Z}^t$ for each $l \in [L]$, $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_L$ partition $\hat{\mathcal{Z}}^t$, and $U^t(\mathcal{C}_l \mid \boldsymbol{c}_l) \geq 1 - \gamma$ for each $l \in [L]$.

Lemma 3.1: Let $t \geq 1$, $R \geq 0$, and $L \leq 2^{tR}$. Then, for any $P \in \mathcal{P}_t(\mathcal{Z})$, there exists a $(t, L, \exp[-tE(R, P, U) + o(t)])$ block code for the DMC$(\mathcal{Z}, \hat{\mathcal{Z}}, U)$, all of whose codewords have type P. Here

$$E(R, P, U) = \min_{U' \in \mathcal{W}(\mathcal{Z}, \hat{\mathcal{Z}})} [D(U' \mid |U|P) + (I(P; U') - R)^+].$$

$E(R, P, U)$ is a continuous function of R and P, which is positive if $R < I(P; U)$ and zero otherwise.

Proof: Standard. See [6, p. 165, Theorem 5.2]. □

Definition 3.2: Let $\boldsymbol{c} \in \mathcal{Z}^t$ and $\mathcal{C} \subseteq \hat{\mathcal{Z}}^t$. An (L, τ) *equipartition* of $\mathcal{C}$ w.r.t. $\boldsymbol{c}$, over the DMC $(\mathcal{Z}, \hat{\mathcal{Z}}, U)$, is a partition of $\mathcal{C}$ into $L+1$ subsets $\mathcal{C}(e), \mathcal{C}(1), \cdots, \mathcal{C}(L)$ such that $U^t(\mathcal{C}(l) \mid \boldsymbol{c})$ is the same for all $l \in [L]$, and $U^t(\mathcal{C}(e) \mid \boldsymbol{c}) \leq \tau$.

Lemma 3.2: Let $t \geq 1$, $R \geq 0$, and $L \leq 2^{tR}$. Then, for any $\boldsymbol{c} \in \mathcal{Z}^t$ and $\mathcal{C} \subseteq \hat{\mathcal{Z}}^t$, there exists an $(L, \exp[-tF(R, Q, U) + o(t)])$ equipartition of $\mathcal{C}$ w.r.t. $\boldsymbol{c}$, over the DMC $(\mathcal{Z}, \hat{\mathcal{Z}}, U)$. Here, $Q \in \mathcal{P}_t(\mathcal{Z})$ is the type of $\boldsymbol{c}$, and

$$F(R, Q, U) = \min_{U' \in \mathcal{W}(\mathcal{Z}, \hat{\mathcal{Z}})} [D(U' \mid |U|Q) + (H(U' \mid Q) - R)^+].$$

$F(R, Q, U)$ is a continuous function of R and Q, which is positive if $R < H(U \mid Q)$ and zero otherwise.

Proof: See the Appendix. □

B. Block Codes and Equipartitions of Their Decoding Regions

Let $P^* \in \mathcal{P}(\mathcal{X})$ and $Q^* \in \mathcal{P}(\mathcal{Y})$ be the distributions achieving the maximum in (9). Assume that V and W are such that both $\min[H(W \mid Q^*), I(P^*; V)]$ and $\min[H(V \mid P^*), I(Q^*; W)]$ are positive. The degenerate cases where one of these terms is zero can be handled with obvious modifications, and will therefore not be considered. Let

$$R_A = \min[H(W \mid Q^*), I(P^*; V)] - \epsilon$$
$$R_B = \min[H(V \mid P^*), I(Q^*; W)] - \epsilon$$

where $\epsilon > 0$ is small enough that R_A and R_B are positive. Then, $E(R_A, P^*, V)$, $E(R_B, Q^*, W)$, $F(R_B, P^*, V)$, and $F(R_A, Q^*, W)$ are all positive. Choose types $P_t \in \mathcal{P}_t(\mathcal{X})$ and $Q_t \in \mathcal{P}_t(\mathcal{Y})$, $t = 1, 2, \cdots$, such that $P_t \to P^*$ and $Q_t \to Q^*$ as $t \to \infty$. By continuity, $E(R_A, P_t, V)$, $E(R_B, Q_t, W)$, $F(R_B, P_t, V)$, and $F(R_A, Q_t, W)$ are all bounded away from zero if t is sufficiently large.

The sequence of protocols to be described will achieve the rate $R_A + R_B = R^* - 2\epsilon$. The protocol with t^2 steps requires two block codes of blocklength t (one for each channel), and equipartitions of their decoding regions w.r.t. the corresponding codewords. We will describe these now.

Let

$$M = \lfloor 2^{tR_A} \rfloor - 1$$
$$N = \lfloor 2^{tR_B} \rfloor - 1$$
$$\rho = \exp\{-tF(R_B, P_t, V) + o(t)\}$$

and

$$\sigma = \exp\{-tF(R_A, Q_t, W) + o(t)\}.$$

Pick arbitrary sequences $\boldsymbol{a} \in \mathcal{X}^t$ and $\boldsymbol{b} \in \mathcal{Y}^t$ of types P_t and Q_t, respectively. Then, Lemma 3.2 guarantees that

1) there exists an (N, ρ) equipartition of $\hat{\mathcal{X}}^t$ w.r.t. $\boldsymbol{a}$ over V, into subsets $\mathcal{A}(e), \mathcal{A}(1), \cdots, \mathcal{A}(N)$;
2) there exists an (M, σ) equipartition of $\hat{\mathcal{Y}}^t$ w.r.t. $\boldsymbol{b}$ over W, into subsets $\mathcal{B}(e), \mathcal{B}(1), \cdots, \mathcal{B}(M)$.

From the definition of equipartition, it follows that

$$\frac{1}{N} \geq V^t(\mathcal{A}(j) \mid \boldsymbol{a}) \geq \frac{1-\rho}{N}, \qquad \text{for each } j \in [N] \quad (15)$$
$$\frac{1}{M} \geq W^t(\mathcal{B}(i) \mid \boldsymbol{b}) \geq \frac{1-\sigma}{M}, \qquad \text{for each } i \in [M]. \quad (16)$$

Before communication begins, Alice and Bob agree upon such sequences and equipartitions.

Next, let

$$\alpha = \exp\{-tE(R_A, P_t, V) + o(t)\}$$

and

$$\beta = \exp\{-tE(R_B, Q_t, W) + o(t)\}.$$

Then, by Lemmas 3.1 and 3.2

1) there exists a $(t, M+1, \alpha)$ block code $\{(\boldsymbol{a}_i, \mathcal{A}_i) : i \in \{e\} \cup [M]\}$ for V, all of whose codewords have type P_t. Further, for each $i \in \{e\} \cup [M]$, there exists an (N, ρ) equipartition of $\mathcal{A}_i$ w.r.t. $\boldsymbol{a}_i$ over V, into subsets $\mathcal{A}_i(e), \mathcal{A}_i(1), \cdots, \mathcal{A}_i(N)$;
2) there exists a $(t, N+1, \beta)$ block code $\{(\boldsymbol{b}_j, \mathcal{B}_j) : j \in \{e\} \cup [N]\}$ for W, all of whose codewords have type Q_t. Further, for each $j \in \{e\} \cup [N]$, there exists an (M, σ) equipartition of $\mathcal{B}_j$ w.r.t. $\boldsymbol{b}_j$ over W, into subsets $\mathcal{B}_j(e), \mathcal{B}_j(1), \cdots, \mathcal{B}_j(M)$.

Since

$$1 \geq V^t(\mathcal{A}_i \mid \boldsymbol{a}_i) \geq 1 - \alpha$$

and

$$1 \geq W^t(\mathcal{B}_j \mid \boldsymbol{b}_j) \geq 1 - \beta$$

we have

$$\frac{1}{N} \geq V^t(\mathcal{A}_i(j') \mid \boldsymbol{a}_i) \geq \frac{1-\alpha-\rho}{N}, \qquad \text{for each } j' \in [N] \tag{17}$$

$$\frac{1}{M} \geq W^t(\mathcal{B}_j(i') \mid \boldsymbol{b}_j) \geq \frac{1-\beta-\sigma}{M}, \qquad \text{for each } i' \in [M]. \tag{18}$$

Before communication begins, Alice and Bob agree upon such block codes and equipartitions of their decoding regions.

C. Outline of the Protocol

The protocol proceeds in t *rounds*, indexed $0, 1, \cdots, t-1$. In each round, Alice and Bob send each other sequences of length t, so that the total number of steps is t^2. We will describe these t rounds recursively.

In round 0, Alice and Bob transmit the sequences $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. Alice defines S_1 to be the $i \in \{e\} \cup [M]$ such that the sequence she received from Bob falls in $\mathcal{B}(i)$. Similarly, Bob defines T_1 to be the $j \in \{e\} \cup [N]$ such that the sequence he received from Alice falls in $\mathcal{A}(j)$. This completes round 0.

Now let $1 \leq k < t$. Assume that Alice and Bob have computed $S_k \in \{e\} \cup [M]$ and $T_k \in \{e\} \cup [N]$, respectively, based on the sequences they received in round $k-1$. Then, in round k, Alice transmits the codeword $\boldsymbol{a}(S_k)$ and Bob transmits the codeword $\boldsymbol{b}(T_k)$. (The indices are written in parentheses, rather than as subscripts, for typographical convenience.)

Based on the sequences they receive from each other, Alice and Bob try to guess the index of the codeword sent by the other, and also decide the index of the codeword to transmit in the *next* round. This is done as follows: Alice finds the $(i, j) \in (\{e\} \cup [M]) \times (\{e\} \cup [N])$ such that the sequence she received falls in $\mathcal{B}_j(i)$. She then estimates T_k as $\hat{T}_k = j$, and takes $S_{k+1} = i$. Similarly, Bob finds the $(i, j) \in (\{e\} \cup [M]) \times (\{e\} \cup [N])$ such that the sequence he received falls in $\mathcal{A}_i(j)$. He then estimates S_k as $\hat{S}_k = i$, and takes $T_{k+1} = j$. This completes round k.

Let

$$\begin{aligned} \boldsymbol{S} &= ((S_1, \hat{T}_1), (S_2, \hat{T}_2) \cdots, (S_{t-1}, \hat{T}_{t-1})) \\ \boldsymbol{T} &= ((\hat{S}_1, T_1), (\hat{S}_2, T_2) \cdots, (\hat{S}_{t-1}, T_{t-1})). \end{aligned}$$

Both $\boldsymbol{S}$ and $\boldsymbol{T}$ can take on $[(M+1)(N+1)]^{t-1}$ different values. Of these, there are $(MN)^{t-1}$ in which none of the $2(t-1)$ components is e. Let $\mathcal{R}$ be an arbitrary function that maps these $(MN)^{t-1}$ possibilities onto $[(MN)^{t-1}]$, and maps all the remaining possibilities to e.

Then, after round $t-1$, Alice and Bob take their random outputs to be $S = \mathcal{R}(\boldsymbol{S})$ and $T = \mathcal{R}(\boldsymbol{T})$, respectively. Thus both S and T take values in $\{e\} \cup [K]$, where $K = (MN)^{t-1}$.

D. Analysis

We will now prove that the sequence of protocols just described does achieve the rate $R^* - 2\epsilon$.

Claim 3.1:

a) For each $k \in \{1, 2, \cdots, t-1\}$, choose any $(i_k, j_k) \in [M] \times [N]$. Then

$$\frac{1}{(MN)^{t-1}} \geq \Pr[\boldsymbol{S} = \boldsymbol{T} = ((i_k, j_k))_{k=1}^{t-1}] \geq \frac{1-\lambda}{(MN)^{t-1}}$$

where

$$\lambda = t(\alpha + \rho + \beta + \sigma) \to 0 \text{ as } t \to \infty.$$

b) $\lim_{t\to\infty} (1/t^2) \log(MN)^{t-1} = R^* - 2\epsilon.$

Proof: Let $G_k = \{S_k = i_k, T_k = j_k\}$ and $\hat{G}_k = \{\hat{S}_k = i_k, \hat{T}_k = j_k\}$. Then

$$\Pr[\boldsymbol{S} = \boldsymbol{T} = ((i_k, j_k))_{k=1}^{t-1}] = \Pr\left[\bigcap_{k=1}^{t-1} (G_k \cap \hat{G}_k)\right]. \tag{19}$$

Now, for each $k \geq 1$, $(\hat{S}_k, \hat{T}_k, S_{k+1}, T_{k+1})$ is conditionally independent of $(S_1^{k-1}, T_1^{k-1}, \hat{S}_1^{k-1}, \hat{T}_1^{k-1})$, given (S_k, T_k). Therefore,

$$\begin{aligned} &\Pr\left[\bigcap_{k=1}^{t-1} (G_k \cap \hat{G}_k)\right] \\ &\quad = \Pr[G_1] \left(\prod_{k=1}^{t-2} \Pr[\hat{G}_k \cap G_{k+1} \mid G_k]\right) \Pr[\hat{G}_{t-1} \mid G_{t-1}]. \end{aligned} \tag{20}$$

We will bound each of the terms in the above product separately. To begin with

$$\begin{aligned} \Pr[G_1] &= \Pr[T_1 = j_1] \cdot \Pr[S_1 = i_1] \\ &= V^t(\mathcal{A}(j_1) \mid \boldsymbol{a}) \cdot W^t(\mathcal{B}(i_1) \mid \boldsymbol{b}). \end{aligned}$$

From (15) and (16), it follows that

$$\frac{1}{MN} \geq \Pr[G_1] \geq \left(\frac{1-\rho}{N}\right)\left(\frac{1-\sigma}{M}\right). \tag{21}$$

Next, for $1 \leq k \leq t-2$,

$$\begin{aligned} &\Pr[\hat{G}_k \cap G_{k+1} \mid G_k] \\ &\quad = \Pr[\hat{S}_k = i_k, T_{k+1} = j_{k+1} \mid S_k = i_k] \\ &\qquad \cdot \Pr[\hat{T}_k = j_k, S_{k+1} = i_{k+1} \mid T_k = j_k] \\ &\quad = V^t(\mathcal{A}_{i_k}(j_{k+1}) \mid \boldsymbol{a}_{i_k}) \cdot W^t(\mathcal{B}_{j_k}(i_{k+1}) \mid \boldsymbol{b}_{j_k}). \end{aligned}$$

From (17) and (18), it follows that

$$\frac{1}{MN} \geq \Pr[\hat{G}_k \cap G_{k+1} \mid G_k] \geq \left(\frac{1-\alpha-\rho}{N}\right)\left(\frac{1-\beta-\sigma}{M}\right), \quad 1 \leq k \leq t-2. \tag{22}$$

Finally,

$$\Pr[\hat{G}_{t-1} \mid G_{t-1}] = V^t(\mathcal{A}_{i_{t-1}} \mid \boldsymbol{a}_{i_{t-1}}) \cdot W^t(\mathcal{B}_{j_{t-1}} \mid \boldsymbol{b}_{j_{t-1}})$$

so that

$$1 \geq \Pr[\hat{G}_{t-1} \mid G_{t-1}] \geq (1-\alpha)(1-\beta). \tag{23}$$

By (19)–(23)

$$1/(MN)^{t-1} \ge \Pr[\boldsymbol{S} = \boldsymbol{T} = ((i_k, j_k))_{k=1}^{t-1}]$$

and

$$\begin{aligned}\Pr&[\boldsymbol{S} = \boldsymbol{T} = ((i_k, j_k))_{k=1}^{t-1}]\\ &\ge \left(\frac{1-\rho}{N}\right)\left(\frac{1-\sigma}{M}\right)\\ &\quad\times\left[\frac{(1-\alpha-\rho)(1-\beta-\sigma)}{MN}\right]^{t-2}(1-\alpha)(1-\beta)\\ &\ge \frac{[(1-\alpha-\rho)(1-\beta-\sigma)]^t}{(MN)^{t-1}}\\ &\ge \frac{1-t(\alpha+\rho+\beta+\sigma)}{(MN)^{t-1}}\end{aligned}$$

which proves Part a). That $\lambda \to 0$ as $t \to \infty$ follows from the fact that $E(R_A, P_t, V)$, $E(R_B, Q_t, W)$, $F(R_B, P_t, V)$, and $F(R_A, Q_t, W)$ converge to positive numbers as $t \to \infty$.

Part b) is obvious from the definitions of M and N. □

IV. Proof of the Converse Part

Let (f, g, S, T) be any (n, K, λ) protocol, with $\lambda < 1$. The aim is to prove that $K \le 2^{nR^* + o(n)}$. This will be done by exhibiting a $P \in \mathcal{P}(\mathcal{X})$ and a $Q \in \mathcal{P}(\mathcal{Y})$ such that $(1/n)\log K$ is bounded above by

$$\begin{aligned}\min[H(PV), H(QW), H(V \mid P) + H(W \mid Q), I(P;V)\\ + I(Q;W)] + o(1).\end{aligned} \tag{24}$$

The minimum in (24) equals

$$\min[H(W \mid Q), I(P;V)] + \min[H(V \mid P), I(Q;W)].$$

For convenience, let $U_{f,g}(\hat{x}^n, \hat{y}^n)$ denote the right-hand side of (1), i.e., the probability that $(\hat{X}^n, \hat{Y}^n) = (\hat{x}^n, \hat{y}^n)$ under the strategy (f, g). If $\mathcal{C} \subseteq \hat{\mathcal{X}}^n \times \hat{\mathcal{Y}}^n$, then let

$$U_{f,g}(\mathcal{C}) = \sum_{(\hat{x}^n, \hat{y}^n) \in \mathcal{C}} U_{f,g}(\hat{x}^n, \hat{y}^n).$$

We will also need some notation for various empirical distributions induced by an $(\hat{x}^n, \hat{y}^n)$ pair. Let

1) $N(\hat{x} \mid \hat{x}^n) = |\{k : \hat{x}_k = \hat{x}\}|$
   $N(\hat{y} \mid \hat{y}^n) = |\{k : \hat{y}_k = \hat{y}\}|$.
2) $N(x \mid \hat{y}^n) = |\{k : f_k(\hat{y}^{k-1}) = x\}|$
   $N(y \mid \hat{x}^n) = |\{k : g_k(\hat{x}^{k-1}) = y\}|$.
3) $N(x, \hat{x} \mid \hat{x}^n, \hat{y}^n) = |\{k : f_k(\hat{y}^{k-1}) = x, \hat{x}_k = \hat{x}\}|$.
4) $N(y, \hat{y} \mid \hat{x}^n, \hat{y}^n) = |\{k : g_k(\hat{x}^{k-1}) = y, \hat{y}_k = \hat{y}\}|$.

Finally, let $H^V(x) = H(V(\cdot|x))$ and $H^W(y) = H(W(\cdot|y))$.

The key idea in the proof is the following lemma, which helps identify a suitable high-probability subset of "jointly typical" $(\hat{x}^n, \hat{y}^n)$ sequences.

Lemma 4.1: Let $\mathcal{E}$ be the set of all $(\hat{x}^n, \hat{y}^n) \in \hat{\mathcal{X}}^n \times \hat{\mathcal{Y}}^n$ satisfying the following conditions:

a) $|N(\hat{x} \mid \hat{x}^n) - \sum_x N(x \mid \hat{y}^n)V(\hat{x} \mid x)| \le \theta\sqrt{n}$ for all $\hat{x} \in \hat{\mathcal{X}}$.

b) $|N(\hat{y} \mid \hat{y}^n) - \sum_y N(y \mid \hat{x}^n)W(\hat{y} \mid y)| \le \theta\sqrt{n}$ for all $\hat{y} \in \hat{\mathcal{Y}}$.

c)
$$\begin{aligned}|\log U_{f,g}(\hat{x}^n, \hat{y}^n) + \sum_x N(x \mid \hat{y}^n)H^V(x)\\ + \sum_y N(y \mid \hat{x}^n)H^W(y)| \le \theta\sqrt{n}.\end{aligned}$$

Then, $1 - U_{f,g}(\mathcal{E}) \le \gamma/\theta^2$, where γ is a constant determined by the channels V and W.

Proof: See the Appendix. □

For the rest of the proof, assume that θ is so large that $U_{f,g}(\mathcal{E}) \ge (1 + \lambda)/2$. For $l \in [K]$, let

$$\mathcal{E}_l = \{\hat{x}^n : T(\hat{x}^n) = l\} \times \{\hat{y}^n : S(\hat{y}^n) = l\}.$$

Then

$$\begin{aligned}U_{f,g}\left(\mathcal{E} \cap \bigcup_{l \in [K]} \mathcal{E}_l\right) &\ge U_{f,g}(\mathcal{E}) - U_{f,g}\left(\bigcap_{l \in [K]} \mathcal{E}_l^c\right)\\ &\ge \left(\frac{1+\lambda}{2}\right) - \lambda\\ &= \left(\frac{1-\lambda}{2}\right).\end{aligned} \tag{25}$$

For $\sigma' \in \mathcal{P}_n(\mathcal{X} \times \hat{\mathcal{X}})$ and $\tau' \in \mathcal{P}_n(\mathcal{Y} \times \hat{\mathcal{Y}})$, let $\mathcal{E}_{\sigma', \tau'}$ be the set of $(\hat{x}^n, \hat{y}^n) \in \mathcal{E}$ such that $N(x, \hat{x} \mid \hat{x}^n, \hat{y}^n) = n\sigma'(x, \hat{x})$ and $N(y, \hat{y} \mid \hat{x}^n, \hat{y}^n) = n\tau'(y, \hat{y})$ for all $x, \hat{x}, y, \hat{y}$. Since the number of such (σ', τ') pairs can be bounded above by n^c, for a suitably large c, there must exist (σ, τ) such that

$$U_{f,g}\left(\mathcal{E}_{\sigma,\tau} \cap \bigcup_{l \in [K]} \mathcal{E}_l\right) \ge \frac{1-\lambda}{2n^c}. \tag{26}$$

From now on, we will focus only on those $(\hat{x}^n, \hat{y}^n)$ that belong to the subset $\mathcal{E}_{\sigma,\tau}$. Let

$$\begin{aligned}P_\sigma(x) &= \sum_{\hat{x}} \sigma(x, \hat{x})\\ Q_\tau(y) &= \sum_{\hat{y}} \tau(y, \hat{y})\\ \hat{P}_\sigma(\hat{x}) &= \sum_x \sigma(x, \hat{x})\end{aligned}$$

and

$$\hat{Q}_\tau(\hat{y}) = \sum_y \tau(y, \hat{y})$$

be the marginals of σ and τ. We will in fact prove (24) with P_σ and Q_τ in place of P and Q.

Now, by condition a) in the definition of $\mathcal{E}$, we have $|\hat{P}_\sigma(\hat{x}) - P_\sigma V(\hat{x})\ | \le \theta/\sqrt{n}$ for all $\hat{x}$. It follows by the continuity of the entropy function (see, e.g., [6, Lemma 2.7, p. 33]) that, for all large n

$$|H(\hat{P}_\sigma) - H(P_\sigma V)| \le |\hat{\mathcal{X}}|(\theta/\sqrt{n})\log(\sqrt{n}/\theta). \tag{27}$$

Similarly, by condition b) in the definition of $\mathcal{E}$, $|\hat{Q}_\tau(\hat{y}) - Q_\tau W(\hat{y})| \le \theta/\sqrt{n}$ for all $\hat{y}$, and

$$|H(\hat{Q}_\tau) - H(Q_\tau W)| \le |\hat{\mathcal{Y}}|(\theta/\sqrt{n})\log(\sqrt{n}/\theta) \tag{28}$$

for all large n.

If $(\hat{x}^n, \hat{y}^n) \in \mathcal{E}_{\sigma,\tau}$, then $N(x \mid \hat{y}^n) = nP_\sigma(x)$ and $N(y \mid \hat{x}^n) = nQ_\tau(y)$. Hence,

$$U_{f,g}(\hat{x}^n, \hat{y}^n) \ge \exp\{-n[H(V \mid P_\sigma) + H(W \mid Q_\tau)] - \theta\sqrt{n}\}$$

by condition c) in the definition of $\mathcal{E}$, so that

$$|\mathcal{E}_{\sigma,\tau}| \le 2^{n[H(V|P_\sigma)+H(W|Q_\tau)]+\theta\sqrt{n}}. \tag{29}$$

Also, the type of $\hat{x}^n$ must be $\hat{P}_\sigma$, so that $\hat{P}^n_\sigma(\hat{x}^n) = \exp\{-nH(\hat{P}_\sigma)\}$. Therefore,

$$\begin{aligned} |\{\hat{x}^n : \exists \hat{y}^n \text{ s.t. } (\hat{x}^n, \hat{y}^n) \in \mathcal{E}_{\sigma,\tau}\}| &\le \exp\{nH(\hat{P}_\sigma)\} \\ &\le \exp\{nH(P_\sigma V) + o(n)\} \end{aligned} \tag{30}$$

by (27). Similarly, $\hat{y}^n$ must have type $\hat{Q}_\tau$, so that $\hat{Q}^n_\tau(\hat{y}^n) = \exp\{-nH(\hat{Q}_\tau)\}$, and

$$|\{\hat{y}^n : \exists \hat{x}^n \text{ s.t. } (\hat{x}^n, \hat{y}^n) \in \mathcal{E}_{\sigma,\tau}\}| \le \exp\{nH(Q_\tau W) + o(n)\} \tag{31}$$

by (28). Finally,

$$\begin{aligned} n^{-1} \log&\left(\frac{U_{f,g}(\hat{x}^n \hat{y}^n)}{\hat{P}^n_\sigma(\hat{x}^n)\hat{Q}^n_\tau(\hat{y}^n)}\right) \\ &\le -(H(V \mid P_\sigma) + H(W \mid Q_\tau) - \theta/\sqrt{n}) \\ &\quad + (H(P_\sigma V) + H(Q_\tau W) + o(1)) \\ &= I(P_\sigma; V) + I(Q_\tau; W) + o(1) \end{aligned} \tag{32}$$

where the inequality is by condition c) in the definition of $\mathcal{E}$, (27), and (28).

We will now show that many of the decision regions $\mathcal{E}_l$ must intersect significantly with $\mathcal{E}_{\sigma,\tau}$. More precisely, let

$$\mathcal{L} = \left\{ l \in [K] : U_{f,g}(\mathcal{E}_{\sigma,\tau} \cap \mathcal{E}_l) \ge \frac{1-\lambda}{4Kn^c} \right\}.$$

We will prove that $|\mathcal{L}| \ge K(1-\lambda)/8n^c$. To this end, note that

$$\begin{aligned} \frac{1-\lambda}{2n^c} &\le U_{f,g}\left(\mathcal{E}_{\sigma,\tau} \cap \bigcup_{l \in [K]} \mathcal{E}_l\right) \\ &= U_{f,g}\left(\mathcal{E}_{\sigma,\tau} \cap \bigcup_{l \in \mathcal{L}} \mathcal{E}_l\right) + U_{f,g}\left(\mathcal{E}_{\sigma,\tau} \cap \bigcup_{l \notin \mathcal{L}} \mathcal{E}_l\right) \\ &\le U_{f,g}\left(\bigcup_{l \in \mathcal{L}} \mathcal{E}_l\right) + (K - |\mathcal{L}|)\left(\frac{1-\lambda}{4Kn^c}\right) \\ &\le |\mathcal{L}|\left(\frac{2}{K}\right) + \frac{1-\lambda}{4n^c} \end{aligned}$$

from which the desired bound on $|\mathcal{L}|$ follows.

If $l \in \mathcal{L}$, then $U_{f,g}(\mathcal{E}_{\sigma,\tau} \cap \mathcal{E}_l) > 0$, so that $\mathcal{E}_{\sigma,\tau} \cap \mathcal{E}_l$ is nonempty. But this means that

1) $$\begin{aligned} |\mathcal{L}| &\le |\mathcal{E}_{\sigma,\tau}| \\ &\le \exp\{n[H(V \mid P_\sigma) + H(W \mid Q_\tau)] + o(n)\}. \end{aligned}$$

2) $$\begin{aligned} |\mathcal{L}| &\le |\{\hat{x}^n : \exists \hat{y}^n \text{ s.t. } (\hat{x}^n, \hat{y}^n) \in \mathcal{E}_{\sigma,\tau}\}| \\ &\le \exp\{nH(P_\sigma V) + o(n)\}. \end{aligned}$$

3) $$\begin{aligned} |\mathcal{L}| &\le |\{\hat{y}^n : \exists \hat{x}^n \text{ s.t. } (\hat{x}^n, \hat{y}^n) \in \mathcal{E}_{\sigma,\tau}\}| \\ &\le \exp\{nH(Q_\tau W) + o(n)\}. \end{aligned}$$

The right inequalities in 1), 2), and 3) are by (29), (30), and (31), respectively. Together with the bound $|\mathcal{L}| \ge K(1-\lambda)/8n^c$ just proved, these inequalities yield three of the desired terms in the minimum occurring in (24) For the fourth term, note that

$$\begin{aligned} 1 &\ge \left[\sum_{j=1}^{K} \hat{P}^n_\sigma(T^{-1}(j))\right]\left[\sum_{i=1}^{K} \hat{Q}^n_\tau(S^{-1}(i))\right] \\ &\ge \left(\sum_{l=1}^{K} \sqrt{\hat{P}^n_\sigma(T^{-1}(l))\hat{Q}^n_\tau(S^{-1}(l))}\right)^2 \\ &\ge \left(\sum_{l \in \mathcal{L}} \sqrt{\hat{P}^n_\sigma(T^{-1}(l))\hat{Q}^n_\tau(S^{-1}(l))}\right)^2 \end{aligned} \tag{33}$$

where the second step is by the Cauchy–Schwarz inequality. Now

$$\begin{aligned} &\hat{P}^n_\sigma(T^{-1}(l))\hat{Q}^n_\tau(S^{-1}(l)) \\ &\quad = \sum_{(\hat{x}^n, \hat{y}^n) \in \mathcal{E}_l} \hat{P}^n_\sigma(\hat{x}^n)\hat{Q}^n_\tau(\hat{y}^n) \\ &\quad \ge \sum_{(\hat{x}^n, \hat{y}^n) \in \mathcal{E}_l \cap \mathcal{E}_{\sigma,\tau}} \hat{P}^n_\sigma(\hat{x}^n)\hat{Q}^n_\tau(\hat{y}^n) \\ &\quad \ge \sum_{(\hat{x}^n, \hat{y}^n) \in \mathcal{E}_l \cap \mathcal{E}_{\sigma,\tau}} U_{f,g}(\hat{x}^n, \hat{y}^n) 2^{-n[I(P_\sigma;V)+I(Q_\tau;W)]-o(n)} \\ &\quad \ge \left(\frac{1-\lambda}{4Kn^c}\right) 2^{-n[I(P_\sigma;V)+I(Q_\tau;W)]-o(n)}, \qquad \text{if } l \in \mathcal{L}. \end{aligned}$$

The second inequality is by (32). Substituting back in (33), and using the fact that $|\mathcal{L}| \ge K(1-\lambda)/8n^c$, we get

$$K \le \exp\{n[I(P_\sigma; V) + I(Q_\tau; W)] + o(n)\}.$$

This completes the proof. □

APPENDIX

Proof of the Weak Converse

We will now sketch the proof of (13). Let (f, g, S, T) be any (n, K, λ) protocol. Let

$$\begin{aligned} P_k(x) &= \Pr[X_k = x] \\ Q_k(y) &= \Pr[Y_k = y] \\ P(x) &= n^{-1} \sum_k P_k(x) \end{aligned}$$

and

$$Q(y) = n^{-1} \sum_k Q_k(y).$$

Then

$$\begin{aligned} H(\hat{X}^n) &\le \sum_{k=1}^{n} H(\hat{X}_k) \\ &= \sum_{k=1}^{n} H(P_k V) \\ &\le nH(PV). \end{aligned} \tag{34}$$

Similarly

$$H(\hat{Y}^n) \le nH(QW). \tag{35}$$

Also,

$$\begin{aligned} H(\hat{X}^n, \hat{Y}^n) &= \sum_{k=1}^{n} H(\hat{X}_k, \hat{Y}_k \mid \hat{X}^{k-1}, \hat{Y}^{k-1}) \\ &= \sum_{k=1}^{n} H(\hat{X}_k, \hat{Y}_k \mid \hat{X}^{k-1}, \hat{Y}^{k-1}, X_k, Y_k) \quad (36) \\ &= \sum_{k=1}^{n} [H(\hat{X}_k \mid X_k) + H(\hat{Y}_k \mid Y_k)] \quad (37) \\ &= \sum_{k=1}^{n} [H(V \mid P_k) + H(W \mid Q_k)] \\ &= n[H(V \mid P) + H(W \mid Q)] \quad (38) \end{aligned}$$

where (36) is because (X_k, Y_k) is a function of $(\hat{X}^{k-1}, \hat{Y}^{k-1})$, and (37) is because $\hat{X}_k$ is independent of $(\hat{X}^{k-1}, \hat{Y}^k, Y_k)$ given X_k, and $\hat{Y}_k$ is independent of $(\hat{Y}^{k-1}, \hat{X}^k, X_k)$ given Y_k.

By (34), (35) and (38)

$$\begin{aligned} I(\hat{X}^n; \hat{Y}^n) &= H(\hat{X}^n) + H(\hat{Y}^n) - H(\hat{X}^n, \hat{Y}^n) \\ &\le n[I(P; V) + I(Q; W)]. \end{aligned} \tag{39}$$

But

1) $(1-\lambda)\log K - 1 \le H(S, T) \le H(\hat{X}^n, \hat{Y}^n)$

2) $(1-2\lambda)\log K - 2 \le \min\{H(S), H(T)\} \le \min\{H(\hat{Y}^n), H(\hat{X}^n)\}$

3) $(1-3\lambda)\log K - 3 \le I(S; T) \le I(\hat{Y}^n; \hat{X}^n)$.

The left inequalities in 1), 2), and 3) are by (6), (7), and (8), respectively. The right inequalities hold because S is a function of $\hat{Y}^n$ and T is a function of $\hat{X}^n$. Combining these with (34), (35), (38), and (39), we have

$$\begin{aligned} &n^{-1}[(1-3\lambda)\log K - 3] \\ &\quad \le \min\{H(PV), H(QW), H(V \mid P) \\ &\qquad + H(W \mid Q), I(P; V) + I(Q; W)\} \\ &\quad = \min\{H(W \mid Q), I(P; V)\} + \min\{H(V \mid P), I(Q; W)\} \\ &\quad \le R^*. \end{aligned}$$

The weak converse follows from this. □

Proof of Lemma 3.2

Let $\mathcal{W}_t(Q)$ be the set of those $U' \in \mathcal{W}(\mathcal{Z}, \hat{\mathcal{Z}})$ such that $tQ(z)U'(\hat{z} \mid z)$ is an integer, for all $z, \hat{z}$. For any such U', define $\mathcal{T}_{U'}(\boldsymbol{c})$ to be the set of $\hat{z}^t$ such that

$$N(z, \hat{z} \mid \boldsymbol{c}, \hat{z}^t) = tQ(z)U'(\hat{z} \mid z) \text{ for all } z, \hat{z}.$$

Here, $N(z, \hat{z} \mid \boldsymbol{c}, \hat{z}^t)$ is the number of occurences of the pair $(z, \hat{z})$ in $(\boldsymbol{c}, \hat{z}^t)$. (Recall that Q is the type of $\boldsymbol{c}$.)

For each $U' \in \mathcal{W}_t(Q)$, construct L pairwise disjoint subsets of $\mathcal{C} \cap \mathcal{T}_{U'}(\boldsymbol{c})$, say $\mathcal{C}_{U'}(1), \cdots, \mathcal{C}_{U'}(L)$, each of size exactly $\lfloor |\mathcal{C} \cap \mathcal{T}_{U'}(\boldsymbol{c})|/L \rfloor$ (the subsets are otherwise arbitrary). Let

$$\mathcal{C}(l) = \bigcup_{U' \in \mathcal{W}_t(Q)} \mathcal{C}_{U'}(l), \qquad l \in [L].$$

It is then clear that $U^t(\mathcal{C}(l) \mid \boldsymbol{c})$ is the same for all $l \in [L]$, since every $\mathcal{C}(l)$ has the same number of sequences of any given conditional type w.r.t. $\boldsymbol{c}$. It remains to upper-bound $U^t(\mathcal{C}(e) \mid \boldsymbol{c})$, where $\mathcal{C}(e)$ is the set of those sequences in $\mathcal{C}$ that are not in any of the $\mathcal{C}(l)$. To this end, note that

$$\begin{aligned} |\mathcal{C} \cap \mathcal{T}_{U'}(\boldsymbol{c})| \bmod L &\le \min\{L, |\mathcal{T}_{U'}(\mathbf{c})|\} \\ &\le \min\{2^{tR}, 2^{tH(U'|Q)}\} \\ &= 2^{t[H(U'|Q) - (H(U'|Q) - R)^+]}. \end{aligned}$$

Therefore,

$$\begin{aligned} &U^t(\mathcal{C}(e) \mid \boldsymbol{c}) \\ &\quad = \sum_{U' \in \mathcal{W}_t(Q)} [|\mathcal{C} \cap \mathcal{T}_{U'}(\mathbf{c})| \bmod L] \cdot 2^{-t[H(U'|Q) + D(U'||U|Q)]} \\ &\quad \le |\mathcal{W}_t(Q)| \max_{U' \in \mathcal{W}_t(Q)} \{2^{-t[D(U'||U|Q) + (H(U'|Q) - R)^+]}\} \\ &\quad \le 2^{-tF(R,Q,U) + o(t)} \end{aligned}$$

since $|\mathcal{W}_t(Q)|$ can be bounded above by a polynomial in t. The stated properties of $F(R, Q, U)$ are easy to establish. □

Proof of Lemma 4.1 Let

$$\begin{aligned} C_k = \log V(\hat{X}_k \mid f_k(\hat{Y}^{k-1})) \\ + \log W(\hat{Y}_k \mid g_k(\hat{X}^{k-1})), \qquad 1 \le k \le n. \end{aligned}$$

Then

$$E[C_k | \hat{X}^{k-1}, \hat{Y}^{k-1}] = -H^V(f_k(\hat{Y}^{k-1})) - H^W(g_k(\hat{X}^{k-1})).$$

Therefore,

$$\begin{aligned} &\log U_{f,g}(\hat{X}^n, \hat{Y}^n) + \sum_x N(x \mid \hat{Y}^n) H^V(x) \\ &\qquad + \sum_y N(y \mid \hat{X}^n) H^W(y) \\ &\quad = \sum_{k=1}^{n} [\log V(\hat{X}_k \mid f_k(\hat{Y}^{k-1})) + \log W(\hat{Y}_k \mid g_k(\hat{X}^{k-1}))] \\ &\qquad + \sum_{k=1}^{n} [H^V(f_k(\hat{Y}^{k-1})) + H^W(g_k(\hat{X}^{k-1}))] \\ &\quad = \sum_{k=1}^{n} (C_k - E[C_k \mid \hat{X}^{k-1}, \hat{Y}^{k-1}]) \end{aligned}$$

so that

$$\begin{aligned}\Pr\Bigg\{\Bigg|&\log U_{f,g}(\hat{X}^n,\hat{Y}^n)+\sum_x N(x\mid \hat{Y}^n)H^V(x)\\&+\sum_y N(y\mid \hat{X}^n)H^W(y)\Bigg|>\theta\sqrt{n}\Bigg\}\\&=\Pr\left\{\left|\sum_{k=1}^n (C_k-E[C_k\mid \hat{X}^{k-1},\hat{Y}^{k-1}])\right|>\theta\sqrt{n}\right\}\\&\le\left(\frac{1}{\theta^2 n}\right)\mathrm{Var}\left[\sum_{k=1}^n \tilde{C}_k\right] &&(40)\\&\le \delta/\theta^2. &&(41)\end{aligned}$$

Here, $\tilde{C}_k=C_k-E[C_k\mid \hat{X}^{k-1},\hat{Y}^{k-1}]$, and (40) is by Chebyshev's inequality. To justify (41), observe that the $\tilde{C}_k$'s are pairwise uncorrelated, and that there exists a constant δ, determined by V and W only, such that $\mathrm{Var}\,(\tilde{C}_k)\le\delta$ for all k.

Next, for any $\hat{x}$, define the random variable A_k to be 1 if $\hat{X}_k=\hat{x}$ and 0 otherwise $(1\le k\le n)$. Then

$$E[A_k\mid \hat{X}^{k-1},\hat{Y}^{k-1}]=V(\hat{x}\mid f_k(\hat{Y}^{k-1})).$$

Therefore,

$$\begin{aligned}N(\hat{x}\mid \hat{X}^n)&-\sum_x N(x\mid \hat{Y}^n)V(\hat{x}\mid x)\\&=\sum_{k=1}^n[\mathbf{1}(\hat{X}_k=\hat{x})-V(\hat{x}\mid f_k(\hat{Y}^{k-1}))]\\&=\sum_{k=1}^n(A_k-E[A_k\mid \hat{X}^{k-1},\hat{Y}^{k-1}])\end{aligned}$$

so that

$$\begin{aligned}\Pr&\left\{\left|N(\hat{x}\mid \hat{X}^n)-\sum_x N(x\mid \hat{Y}^n)V(\hat{x}\mid x)\right|>\theta\sqrt{n}\right\}\\&\le\left(\frac{1}{\theta^2 n}\right)\mathrm{Var}\left[\sum_{k=1}^n \tilde{A}_k\right]\\&\le 1/\theta^2.\end{aligned}$$

Here, $\tilde{A}_k=A_k-E[A_k\mid \hat{X}^{k-1},\hat{Y}^{k-1}]$. The last inequality holds because the $\tilde{A}_k$'s are pairwise uncorrelated, and their variances cannot exceed 1. By taking a union bound over all $\hat{x}$, we have

$$\begin{aligned}\Pr&\left\{\exists\hat{x}\text{ s.t. }\left|N(\hat{x}\mid \hat{X}^n)-\sum_x N(x\mid \hat{Y}^n)V(\hat{x}\mid x)\right|>\theta\sqrt{n}\right\}\\&\le|\hat{\mathcal{X}}|/\theta^2. \qquad (42)\end{aligned}$$

A similar argument proves that

$$\begin{aligned}\Pr&\left\{\exists\hat{y}\text{ s.t. }\left|N(\hat{Y}\mid \hat{Y}^n)-\sum_y N(y\mid \hat{X}^n)W(\hat{Y}\mid y)\right|>\theta\sqrt{n}\right\}\\&\le|\hat{\mathcal{Y}}|/\theta^2. \qquad (43)\end{aligned}$$

By (41), (42), and (43), the probability that $(\hat{X}^n,\hat{Y}^n)$ violates any of the conditions in the definition of $\mathcal{E}$ is at most γ/θ^2, where $\gamma=\delta+|\hat{\mathcal{X}}|+|\hat{\mathcal{Y}}|$. □

References

[1] R. Ahlswede, “Elimination of correlation in random codes for arbitrarily varying channels,” *Z. Wahrsch. Verw. Gebiete*, vol. 33, pp. 159–175, 1978.

[2] R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography, Part 1: Secret sharing,” *IEEE Trans. Inform. Theory*, vol. 39, no. 4, July 1993.

[3] R. Ahlswede and G. Dueck, “Identification in the presence of feedback—A Discovery of new capacity formulas,” *IEEE Trans. Inform. Theory*, vol. 35, no. 1, Jan. 1989.

[4] ——, “Identification via channels, *IEEE Trans. Inform. Theory*, vol. 35, no. 1, Jan. 1989.

[5] R. Ahlswede and B. Verboven, “On identification via multiway channels with feedback,” *IEEE Trans. Inform. Theory*, vol. 37, no. 5, Sept. 1991.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[7] I. Csiszár and P. Narayan, “The capacity of the arbitrarily varying channel revisited: Positivity constraints,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, Mar. 1988.

[8] L. Lovász, “Communication complexity: A survey,” in *Paths, Flows and VLSI Layout*, B. H. Korte *et al.*, Eds. Berlin, Germany: Springer-Verlag, 1990.

[9] U. M. Maurer, “Perfect cryptographic security from partially independent channels,” in *Proc. 23rd Annual ACM Symp. on the Theory of Computing*, 1991.

[10] ——, “Secret key agreement by public discussion from common information,” *IEEE Trans. Inform. Theory*, vol. 39, no. 3, May 1993.

[11] A. Orlitsky and A. El Gamal, “Communication complexity,” in *Complexity in Information Theory*, Y. Abu-Mostafa, Ed. Berlin, Germany: Springer-Verlag, 1988.

[12] S. Venkatesan and V. Anantharam, “The common randomness capacity of a pair of independent binary symmetric channels,” Tech. Rep. UCB/ERL M95/68, Electronics Research Lab., Univ. of California, Berkeley, Aug. 1995.