THE COMMON RANDOMNESS CAPACITY OF A PAIR OF INDEPENDENT BINARY SYMMETRIC CHANNELS

by

S. Venkatesan and V. Anantharam

Memorandum No. UCB/ERL M95/68

13 August 1995

THE COMMON RANDOMNESS CAPACITY OF A PAIR OF INDEPENDENT BINARY SYMMETRIC CHANNELS

by

S. Venkatesan and V. Anantharam

Memorandum No. UCB/ERL M95/68

13 August 1995

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

The Common Randomness Capacity of a Pair of Independent Binary Symmetric Channels*

S. Venkatesan† V. Anantharam‡§

13 August 1995

Abstract

We study the following problem: two agents are connected to each other by independent binary symmetric channels of crossover probabilities p and q. They wish to generate common randomness by communicating interactively over the two channels. Neither agent has access to any external random sources, so that any randomness generated must come from the *noise* on the two channels. We show that it is possible to generate common randomness in this situation at a rate (in bits per step) of $R^*(p,q) = \min\{h(p)+h(q), 2-h(p)-h(q)\}$. We also prove a strong converse which establishes $R^*(p,q)$ as the common randomness "capacity" of this pair of channels.

*Research supported by NSF IRI 9005849, IRI 9310670, NCR 9422513, and the AT&T Foundation.

†Cornell University and U.C. Berkeley.

‡Univ. of California, Berkeley.

§Address all correspondence to the second author: 570 Cory Hall, Dept. of EECS, U.C. Berkeley, Berkeley, CA 94720.

1 Introduction

There are several situations in which common randomness available at distant terminals plays a significant role. For example, in identification theory ([2], [3], [4]), the amount of common randomness available to both transmitter and receiver essentially determines the maximum achievable identification rate. Also, in the theory of communication complexity ([7], [9]), it is known that common randomness available to two communicating agents can significantly reduce the complexity of computing certain functions. And, in cryptography ([1], [8]), 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).

For these reasons, Ahlswede and Csiszar ([1]) proposed a systematic study of the role of common randomness in information theory and cryptography. In ([1]), they addressed the problem of secret sharing, i.e., generating common randomness at two terminals without giving information about it to an eavesdropper. In the "channel-type" model introduced there, the two terminals are connected by a discrete memoryless channel 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 sources of randomness, to begin with. Ahlswede and Csiszar 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.

In this paper, we study a different but related problem, viz., of generating common randomness at two distant terminals that are connected to each other by independent binary symmetric channels, when neither terminal has access to any external sources of randomness. The objective is to determine the maximum rate (in bits per step of communication) at which common randomness can be extracted from the *noise* on the two channels. However, no secrecy constraints are imposed, i.e., the random outputs generated need not be kept secret from any eavesdroppers. The results proved here are not implied by those of [1], because both channels here are noisy and constrained in capacity.

Imagine an agent Alice at one terminal, and an agent Bob at the other. Let p be the crossover probability of the channel from Alice to Bob, and q

that of the channel from Bob to Alice. To see how the two agents could generate common randomness even in the absence of any external random sources, suppose that $p = 0$ and $q = 1/2$ (this is the simplest case). Consider the following communication between them: Bob transmits n zeros across his channel in n successive steps, indexed $1, 2, \ldots, n$. Since $q = 1/2$, Alice receives a totally random 0-1 sequence $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ (uniformly distributed over $\{0,1\}^n$). In step k, $2 \leq k \leq n$, Alice echoes $\hat{Y}_{k-1}$ back to Bob, who receives it accurately since $p = 0$. Thus, after n steps, both Alice and Bob know $\left(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_{n-1}\right)$, a random variable uniformly distributed over a set of size 2^{n-1}; they have generated "common randomness."

Note that this common randomness is derived from the "noise" on the channel from Bob to Alice. Moreover, both agents act deterministically, in the sense that their decisions in each step are determined solely by all past receptions.

The *rate* at which common randomness is generated by the above procedure is $\frac{n-1}{n}$ bits per step, which can be made arbitrarily close to 1 by making n large enough. It is not hard to prove that no deterministic procedure can yield common randomness at rates higher than 1 (the notions of "deterministic procedure" and "rate" will be made precise later). Thus, the common randomness "capacity" of this pair of channels is 1 bit per step.

For general p and q, 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. The precise formulation of the problem appears in Section 1.2. The main result, stated in Section 1.3, is the determination of the common randomness capacity for arbitrary p and q.

1.1 Preliminaries

We will refer to the binary symmetric channel with crossover probability r as the $BSC(r)$. The two-way channel consisting of a $BSC(p)$ from Alice to Bob and an independent $BSC(q)$ from Bob to Alice will be called the $BSC(p, q)$.

For the $BSC(r)$, $W_r(\hat{z}|z)$ will denote the probability that the ouput is $\hat{z}$ given that the input is z (this equals r if $\hat{z} \neq z$, and $1 - r$ if $\hat{z} = z$). As usual,

$W_r^n(\hat{z}^n|z^n) = \prod_{k=1}^n W_r(\hat{z}_k|z_k)$.

Throughout, $[L]$ will denote the set of integers $\{1, 2, \ldots, L\}$. $h(\cdot)$ and $D(\cdot\|\cdot)$ will denote the binary entropy function and the binary discrimination function respectively:

$$\begin{aligned} h(r) &= -r\log r - (1-r)\log(1-r), \\ D(r\|s) &= r\log\left(\frac{r}{s}\right) + (1-r)\log\left(\frac{1-r}{1-s}\right). \end{aligned}$$

All logarithms and exponentials will be to the base two.

1.2 Definition of a deterministic protocol

In order to generate common randomness, Alice and Bob communicate with each other for, say, n steps. In each step, Alice transmits a bit to Bob across the $BSC(p)$ and, simultaneously, Bob transmits a bit to Alice across the $BSC(q)$. These bits are determined by an agreed-upon strategy which specifies the bits to be transmitted in each step as functions only of all the past receptions available to the respective senders.

Formally, an n-step strategy is a pair (f, g), with $f = (f_1, f_2, \ldots, f_n)$ and $g = (g_1, g_2, \ldots, g_n)$. Here, $(f_1, g_1) \in \{0,1\} \times \{0,1\}$ and, for $2 \leq k \leq n$, f_k and g_k are both maps from $\{0,1\}^{k-1}$ to $\{0,1\}$.

f and g have the following interpretation: let X_k and Y_k denote the bits transmitted by Alice and Bob respectively in the k^{th} step ($1 \leq k \leq n$), 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 $2 \leq k \leq n$, $X_k = f_k(\hat{Y}^{k-1})$, $Y_k = g_k(\hat{X}^{k-1})$.

We will use $W_{p,q,f,g}^n(\hat{x}^n, \hat{y}^n)$ to denote the probability that $\hat{X}^n = \hat{x}^n$ and $\hat{Y}^n = \hat{y}^n$, when the n-step strategy (f, g) is used. Note that

$$W_{p,q,f,g}^n(\hat{x}^n, \hat{y}^n) = \prod_{k=1}^n \left[W_p\left(\hat{x}_k \mid f_k\left(\hat{y}^{k-1}\right)\right) \cdot W_q\left(\hat{y}_k \mid g_k\left(\hat{x}^{k-1}\right)\right)\right].$$

(The $k = 1$ term is to be understood as $W_p(\hat{x}_1 \mid f_1) \cdot W_q(\hat{y}_1 \mid g_1)$.)

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 decision is based only on $\hat{X}^n$. Their random outputs take values in some

common finite set of size, say, K. Without loss of generality, we may take this set to be $[K]$.

Formally, Alice computes $I = I(\hat{Y}^n)$, and Bob computes $J = J(\hat{X}^n)$, where I and J are both maps from $\{0,1\}^n$ to $\{e\} \cup [K]$. e is a symbol indicating failure to generate common randomness. To avoid trivialities, we will require $I^{-1}(l)$ and $J^{-1}(l)$ to be nonempty for each $l \in [K]$. ($I^{-1}(e)$ and $J^{-1}(e)$ are, of course, allowed to be empty.)

The quadruple (f, g, I, J) defines an (n, K) *deterministic protocol* for generating common randomness (deterministic because neither agent has access to any external sources of randomness; all transmission and computation decisions are based solely on previously received bits).

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 I and J. Ideally, we would like

$$W^n_{p,q,f,g}\left(I(\hat{Y}^n) = J(\hat{X}^n) = l\right) = \frac{1}{K} \quad \text{for each } l \in [K]. \tag{1}$$

If (1) were true, I and J 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 (1) except in the trivial case $K = 1$. Therefore, we will have to settle for *approximate* equality and uniformity of I and J. To this end, we make the following definition: (f, g, I, J) is an (n, K, λ) deterministic protocol for the $BSC(p, q)$ if

$$\frac{1+\lambda}{K} \geq W^n_{p,q,f,g}\left(I(\hat{Y}^n) = J(\hat{X}^n) = l\right) \geq \frac{1-\lambda}{K} \quad \text{for each } l \in [K]. \tag{2}$$

Condition (2) implies that

$$W^n_{p,q,f,g}\left(\bigcup_{l \in [K]} \{I = J = l\}\right) \geq 1 - \lambda,$$

so that, if λ is small, both agents compute the *same* "non-failure" output with high probability. (In particular, the probability that either agent declares failure to generate common randomness is small.) Further, if $\mu_l =$

$W^n_{p,q,f,g}(I = J = l)$ for $l \in [K]$, then

$$\begin{aligned} H(I,J) &\geq \sum_{l=1}^{K} \mu_l \log \frac{1}{\mu_l} \\ &\geq \sum_{l=1}^{K} \frac{1-\lambda}{K} \log\left(\frac{K}{1+\lambda}\right) \\ &= (1-\lambda)\log\left(\frac{K}{1+\lambda}\right). \end{aligned} \tag{3}$$

Also, since $W^n_{p,q,f,g}(I \neq J) \leq \lambda$, Fano's inequality gives

$$\max\{H(I|J), H(J|I)\} \leq h(\lambda) + \lambda \log K. \tag{4}$$

From (3) and (4), it follows that

$$\min\{H(I), H(J)\} \geq (1-2\lambda)\log K - h(\lambda) - (1-\lambda)\log(1+\lambda).$$

Thus, if λ is small, each agent generates a random output whose distribution is close to uniform on $[K]$.

1.3 Main result

Fix $\lambda \in [0,1]$. For each $n \geq 1$, define $K_{p,q}(n,\lambda)$ to be the largest K such that there exists an (n,K,λ) deterministic protocol for the $BSC(p,q)$. The main result proved here is the following:

Theorem 1.1 (Main Theorem) *Let*

$$R^*(p,q) = \min\{h(p) + h(q), 2 - h(p) - h(q)\}.$$

Then:

a) (Direct part)

$$\liminf_{n\to\infty} \frac{1}{n} \log K_{p,q}(n,\lambda) \geq R^*(p,q) \quad \textit{for all } \lambda \in (0,1]. \tag{5}$$

b) (Converse part)

$$\limsup_{n\to\infty} \frac{1}{n} \log K_{p,q}(n,\lambda) \leq R^*(p,q) \quad \textit{for all } \lambda \in [0,1). \tag{6}$$

Thus, $\lim_{n\to\infty} \frac{1}{n} \log K_{p,q}(n,\lambda) = R^*(p,q)$ for all $\lambda \in (0,1)$.

Define rate R of generating common randomness to be achievable over the $BSC(p,q)$ if there exists a sequence of (n, K_n, λ_n) deterministic protocols (for the $BSC(p,q)$) such that

$$\lim_{n\to\infty} \lambda_n = 0 \qquad \text{and} \qquad \liminf_{n\to\infty} \frac{\log K_n}{n} \geq R. \tag{7}$$

Then, (5) is obviously equivalent to the statement that any rate not exceeding $R^*(p,q)$ is achievable over the $BSC(p,q)$. A "weak" converse to (5) would merely assert that rates above $R^*(p,q)$ are not achievable over the $BSC(p,q)$, i.e.,

$$\lim_{\lambda\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log K_{p,q}(n,\lambda) \leq R^*(p,q). \tag{8}$$

However, (6) says much more than (8); in the usual terminology, (6) is a "strong" converse to (5).

Together, (5) and (6) justify the interpretation of $R^*(p,q)$ as the *common randomness capacity* of the $BSC(p,q)$.

Note that $R^*(p,q) = 0$ if and only if $h(p) = h(q) = 0$ 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, $R^*(p,q)$ attains its maximum value of 1 whenever the entropies and capacities balance each other, i.e., $h(p) + h(q) = (1 - h(p)) + (1 - h(q))$. It is somewhat surprising that it is possible to generate common randomness at a rate of 1 bit per step in all these cases.

We will prove the converse part in Section 2, and the direct part in Section 3.

2 Proof of the converse part

Let (f, g, I, J) be any (n, K, λ) deterministic protocol for the $BSC(p,q)$, with $\lambda < 1$. The aim is to prove that

$$K \leq 2^{nR^*(p,q)+o(n)}. \tag{9}$$

Let

$$\mathcal{E}_l = \{(\hat{x}^n, \hat{y}^n) : I(\hat{y}^n) = J(\hat{x}^n) = l\}, \qquad l = 1, 2, \ldots, K.$$

Then, by (2),

$$W^n_{p,q,f,g}(\mathcal{E}_l) \geq \frac{1-\lambda}{K} \qquad \text{for each } l \in [K].$$

The left inequality in (2) will not be needed in the proof. (The converse holds even if this condition is dropped.)

The key idea in the proof is the following lemma. This is a simple generalization, to the "interactive" situation, of the main idea in Kemperman's proof of the strong converse to the coding theorem for DMCs with perfect feedback [6].

Lemma 2.1 *Let*

$$\mathcal{E} = \left\{(\hat{x}^n, \hat{y}^n) : \left|\log W^n_{p,q,f,g}(\hat{x}^n, \hat{y}^n) + n\left(h(p) + h(q)\right)\right| \leq \theta\sqrt{n}\right\}.$$

Then:

a) $|\mathcal{E}| \leq 2^{n(h(p)+h(q))+\theta\sqrt{n}}$.

b) For any $\gamma > 0$, *if* θ *is large enough,*

$$W^n_{p,q,f,g}(\mathcal{E}) \geq 1 - \gamma. \tag{10}$$

Proof: Appendix. □

For the rest of the proof, assume that θ is so large that (10) holds with γ replaced by $(1-\lambda)/4$. We will show that

$$K \leq \min\left\{2 \cdot 2^{n(h(p)+h(q))+\theta\sqrt{n}}, \left(\frac{8}{1-\lambda}\right) 2^{n(2-h(p)-h(q))+\theta\sqrt{n}}\right\}. \tag{11}$$

Obviously, (11) implies (9). Our first goal is to show that many of the decision regions $\mathcal{E}_l$ must intersect significantly with $\mathcal{E}$. More precisely, let

$$\mathcal{L} = \left\{l \in [K] : W^n_{p,q,f,g}(\mathcal{E} \cap \mathcal{E}_l) \geq \frac{1-\lambda}{2K}\right\}.$$

We will prove that $|\mathcal{L}| \geq K/2$. To this end, note that if $l \notin \mathcal{L}$ then

$$\begin{aligned} W^n_{p,q,f,g}(\mathcal{E}^c \cap \mathcal{E}_l) &= W^n_{p,q,f,g}(\mathcal{E}_l) - W^n_{p,q,f,g}(\mathcal{E} \cap \mathcal{E}_l) \\ &> \left(\frac{1-\lambda}{K}\right) - \left(\frac{1-\lambda}{2K}\right) \\ &= \frac{1-\lambda}{2K}. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1-\lambda}{4} &\geq W^n_{p,q,f,g}(\mathcal{E}^c) \\ &\geq \sum_{l \notin \mathcal{L}} W^n_{p,q,f,g}(\mathcal{E}^c \cap \mathcal{E}_l) \\ &\geq (K - |\mathcal{L}|)\left(\frac{1-\lambda}{2K}\right), \end{aligned}$$

which gives

$$|\mathcal{L}| \geq \frac{K}{2}. \tag{12}$$

Next, note that

$$|\mathcal{E} \cap \mathcal{E}_l| \geq \left(\frac{1-\lambda}{2K}\right) 2^{n(h(p)+h(q))-\theta\sqrt{n}} \quad \text{for all } l \in \mathcal{L}, \tag{13}$$

since $W^n_{p,q,f,g}(\hat{x}^n, \hat{y}^n) \leq 2^{-n(h(p)+h(q))+\theta\sqrt{n}}$ for all $(\hat{x}^n, \hat{y}^n) \in \mathcal{E}$. In particular, $\mathcal{E} \cap \mathcal{E}_l$ is nonempty for all $l \in \mathcal{L}$. Consequently,

$$\begin{aligned} 2^{n(h(p)+h(q))+\theta\sqrt{n}} &\geq |\mathcal{E}| \\ &\geq \sum_{l \in \mathcal{L}} |\mathcal{E} \cap \mathcal{E}_l| \\ &\geq \sum_{l \in \mathcal{L}} 1 \\ &\geq \frac{K}{2}, \end{aligned}$$

where the first step is by Part a) of Lemma 2.1 and the last step is by (12). This proves that

$$K \leq 2 \cdot 2^{n(h(p)+h(q))+\theta\sqrt{n}}. \tag{14}$$

Finally, since $2^n \geq \sum_{i=1}^{K} |I^{-1}(i)|$ and $2^n \geq \sum_{j=1}^{K} |J^{-1}(j)|$,

$$\begin{aligned}
2^{2n} &\geq \left(\sum_{i=1}^{K} \left|I^{-1}(i)\right|\right)\left(\sum_{j=1}^{K} \left|J^{-1}(j)\right|\right) \\
&\geq \left(\sum_{l=1}^{K} \sqrt{|I^{-1}(l)|\,|J^{-1}(l)|}\right)^2 \qquad (15) \\
&= \left(\sum_{l=1}^{K} \sqrt{|\mathcal{E}_l|}\right)^2 \\
&\geq \left(\sum_{l\in\mathcal{L}} \sqrt{|\mathcal{E}\cap\mathcal{E}_l|}\right)^2 \\
&\geq \left(\sum_{l\in\mathcal{L}} \left[\left(\frac{1-\lambda}{2K}\right) 2^{n(h(p)+h(q))-\theta\sqrt{n}}\right]^{1/2}\right)^2 \qquad (16) \\
&\geq K\left(\frac{1-\lambda}{8}\right) 2^{n(h(p)+h(q))-\theta\sqrt{n}} \qquad (17)
\end{aligned}$$

where (15) is by the Cauchy-Schwarz inequality, (16) is by (13), and (17) is by (12). This proves that

$$K \leq \left(\frac{8}{1-\lambda}\right) 2^{n(2-h(p)-h(q))+\theta\sqrt{n}}. \qquad (18)$$

The converse follows from (14) and (18).

3 Proof of the direct part

In this section, we will prove that Alice and Bob can generate common randomness over the $BSC(p,q)$ at rates arbitrarily close to $R^*(p,q)$, i.e., for any $R < R^*(p,q)$, we will prove the existence of a sequence of (n, K_n, λ_n) deterministic protocols for the $BSC(p,q)$, satisfying (7). This suffices to prove the direct part of Theorem 1.1.

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

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

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 $\lim_{t\to\infty} \frac{t^2}{(t+1)^2} = 1$.)

Accordingly, in all that follows, we will restrict ourselves to describing protocols with t^2 steps, where t is suitably large.

Without loss of generality, we may assume that $0 \leq p \leq q \leq 1/2$. Since there is nothing to prove if $R^*(p,q) = 0$, we will assume from now on that $(p,q) \neq (0,0)$ and $(p,q) \neq (1/2, 1/2)$. The remaining values of p and q will be classified under three cases:

1) $0 < p \leq q < 1/2$,
2) $p = 0$, $0 < q \leq 1/2$,
3) $0 < p < 1/2$, $q = 1/2$.

The main case, viz. Case 1, is handled in Section 3.2. The two "boundary" cases, viz. Cases 2 and 3, are disposed of in Sections 3.3 and 3.4.

But first we state two lemmas (Lemma 3.1 and Lemma 3.2) which will be needed in the proof.

3.1 Preliminary results

Definition 3.1 *A* (t, L, γ) **block code** *for the* $BSC(r)$ *is a collection,*

$$\{(\mathbf{u}_l, \mathcal{U}_l) : l = 1, 2, \ldots, L\},$$

where $\mathbf{u}_l \in \{0,1\}^t$ *for each* $l \in [L]$, $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_L$ *partition* $\{0,1\}^t$, *and*

$$W_r^t\left(\mathcal{U}_l \,\middle|\, \mathbf{u}_l\right) \geq 1 - \gamma \quad \text{for each } l \in [L].$$

Lemma 3.1 *Suppose* $0 < r < 1/2$, *and* $\eta > 0$ *is so small that*

$$r + \eta \leq \frac{\sqrt{r}}{\sqrt{r} + \sqrt{1-r}}.$$

Then, for any $t \geq 1$ *and* $L \leq 2^{t(1-h(r+\eta))}$, *there exists a* $\left(t, L, 4 \cdot 2^{-tD(r+\eta\|r)}\right)$ *block code for the* $BSC(r)$.

Proof: Standard. See, e.g., [5]. □

Definition 3.2 *Let* $\mathbf{u} \in \{0,1\}^t$ *and* $\mathcal{U} \subseteq \{0,1\}^t$. *An* (r, L, τ) **equipartition** *of* $\mathcal{U}$ *w.r.t.* $\mathbf{u}$ *is a partition of* $\mathcal{U}$ *into* $L+1$ *subsets* $\mathcal{U}(e), \mathcal{U}(1), \ldots, \mathcal{U}(L)$ *such that*

$$W_r^t\left(\mathcal{U}(l) \,\middle|\, \mathbf{u}\right) = \frac{1}{L}\left[W_r^t\left(\mathcal{U} \,\middle|\, \mathbf{u}\right) - W_r^t\left(\mathcal{U}(e) \,\middle|\, \mathbf{u}\right)\right] \quad \text{for all } l \in [L], \tag{19}$$

and $W_r^t\left(\mathcal{U}(e) \,\middle|\, \mathbf{u}\right) \leq \tau$.

Lemma 3.2 *Suppose* $0 < r - \eta < r \leq 1/2$, *and* $t \geq 1$. *Let* $L \leq 2^{th(r-\eta)}$ *and*

$$\tau = \min_{r-\eta \leq \mu \leq r} \left\{ t \cdot 2^{-t[h(\mu)-h(r-\eta)]} + 2 \cdot 2^{-tD(\mu\|r)} \right\}. \tag{20}$$

Then, for any $\mathbf{u} \in \{0,1\}^t$ *and* $\mathcal{U} \subseteq \{0,1\}^t$, *there exists an* (r, L, τ) *equipartition of* $\mathcal{U}$ *w.r.t.* $\mathbf{u}$.

Proof: Appendix. □

3.2 Case 1: $0 < p \leq q < 1/2$

Choose $\delta > 0$ and $\epsilon > 0$ small enough that

$$p + \delta \leq \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}} \quad \text{and} \quad q + \epsilon \leq \frac{\sqrt{q}}{\sqrt{q} + \sqrt{1-q}}. \tag{21}$$

Next, choose $\mu \in (0, p)$ and $\nu \in (0, q)$ such that:

a) If $h(p) + h(q) \leq 1$, then

$$\max\{h(p+\delta) + h(q-\nu),\ h(p-\mu) + h(q+\epsilon)\} \leq 1. \tag{22}$$

b) If $h(p) + h(q) > 1$, then

$$\min\{h(p+\delta) + h(q-\nu),\ h(p-\mu) + h(q+\epsilon)\} > 1. \tag{23}$$

The sequence of protocols to be described will achieve the rate

$$\min\{h(p-\mu) + h(q-\nu),\ 2 - h(p+\delta) - h(q+\epsilon)\}, \tag{24}$$

which can be made arbitrarily close to $R^*(p,q)$ by choosing δ, ϵ, μ, and ν sufficiently small.

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. These are described next.

3.2.1 Block codes and equipartitions of the decoding regions

Let

$$M = \min\left\{\left\lfloor 2^{t(1-h(p+\delta))}\right\rfloor, \left\lfloor 2^{th(q-\nu)}\right\rfloor\right\} - 1, \tag{25}$$

$$N = \min\left\{\left\lfloor 2^{t(1-h(q+\epsilon))}\right\rfloor, \left\lfloor 2^{th(p-\mu)}\right\rfloor\right\} - 1, \tag{26}$$

and let

$$\rho = \min_{p-\mu\le\tau\le p}\left\{t\cdot 2^{-t[h(\tau)-h(p-\mu)]} + 2\cdot 2^{-tD(\tau\|p)}\right\}, \tag{27}$$

$$\sigma = \min_{q-\nu\le\tau\le q}\left\{t\cdot 2^{-t[h(\tau)-h(q-\nu)]} + 2\cdot 2^{-tD(\tau\|q)}\right\}. \tag{28}$$

Pick arbitrary bitstrings $\mathbf{a}$ and $\mathbf{b}$ of length t. Then, Lemma 3.2 guarantees that there exist a (p, N, ρ) equipartition of $\{0,1\}^t$ w.r.t. $\mathbf{a}$ into subsets $\mathcal{A}(e), \mathcal{A}(1), \ldots, \mathcal{A}(N)$, and a (q, M, σ) equipartition of $\{0,1\}^t$ w.r.t. $\mathbf{b}$ into subsets $\mathcal{B}(e), \mathcal{B}(1), \ldots, \mathcal{B}(M)$. From the definition of equipartition, it follows that

$$\frac{1}{N} \ge W_p^t\left(\mathcal{A}(j)\,\middle|\,\mathbf{a}\right) \ge \frac{1-\rho}{N} \quad \text{for each } j\in[N], \tag{29}$$

$$\frac{1}{M} \ge W_q^t\left(\mathcal{B}(i)\,\middle|\,\mathbf{b}\right) \ge \frac{1-\sigma}{M} \quad \text{for each } i\in[M]. \tag{30}$$

Alice and Bob agree upon such bitstrings and equipartitions before communication begins.

Next, let

$$\alpha = 4\cdot 2^{-tD(p+\delta\|p)} \quad \text{and} \quad \beta = 4\cdot 2^{-tD(q+\epsilon\|q)}. \tag{31}$$

Then, Lemma 3.1, along with (21), (25), (26), and (31), guarantees the existence of a $(t, M+1, \alpha)$ block code, $\{(\mathbf{a}_i, \mathcal{A}_i) : i = e, 1, \ldots, M\}$, for the $BSC(p)$, and a $(t, N+1, \beta)$ block code, $\{(\mathbf{b}_j, \mathcal{B}_j) : j = e, 1, \ldots, N\}$, for the $BSC(q)$.

Further, by Lemma 3.2, (25), (26), (27), and (28), for each $i \in \{e\}\cup[M]$ and $j \in \{e\}\cup[N]$, there exist a (p, N, ρ) equipartition of $\mathcal{A}_i$ w.r.t. $\mathbf{a}_i$ into subsets $\mathcal{A}_i(e), \mathcal{A}_i(1), \ldots, \mathcal{A}_i(N)$, and a (q, M, σ) equipartition of $\mathcal{B}_j$ w.r.t. $\mathbf{b}_j$ into subsets $\mathcal{B}_j(e), \mathcal{B}_j(1), \ldots, \mathcal{B}_j(M)$.

Since $1 \geq W_p^t\left(\mathcal{A}_i \,\middle|\, \mathbf{a}_i\right) \geq 1-\alpha$ and $1 \geq W_q^t\left(\mathcal{B}_j \,\middle|\, \mathbf{b}_j\right) \geq 1-\beta$, we get

$$\frac{1}{N} \geq W_p^t\left(\mathcal{A}_i(j') \,\middle|\, \mathbf{a}_i\right) \geq \frac{1-\alpha-\rho}{N} \quad \text{for each } j' \in [N], \tag{32}$$

$$\frac{1}{M} \geq W_q^t\left(\mathcal{B}_j(i') \,\middle|\, \mathbf{b}_j\right) \geq \frac{1-\beta-\sigma}{M} \quad \text{for each } i' \in [M]. \tag{33}$$

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

3.2.2 Outline of the protocol

The protocol proceeds in t *rounds*, indexed $0, 1, \ldots, t-1$. In each round, Alice and Bob send each other bitstrings 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 bitstrings $\mathbf{a}$ and $\mathbf{b}$ respectively. Alice defines I_1 to be the (unique) $i \in \{e\} \cup [M]$ such that the bitstring she receives in round 0 falls in $\mathcal{B}(i)$. Similarly, Bob defines J_1 to be the (unique) $j \in \{e\} \cup [N]$ such that the bitstring he receives in round 0 falls in $\mathcal{A}(j)$. This completes round 0.

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

Based on the bitstrings 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 unique $(i,j) \in (\{e\} \cup [M]) \times (\{e\} \cup [N])$ such that the bitstring she receives in round k falls in $\mathcal{B}_j(i)$. She then estimates J_k as $\hat{J}_k = j$, and takes $I_{k+1} = i$. Similarly, Bob finds the unique $(i,j) \in (\{e\} \cup [M]) \times (\{e\} \cup [N])$ such that the bitstring he receives in round k falls in $\mathcal{A}_i(j)$. He then estimates I_k as $\hat{I}_k = i$, and takes $J_{k+1} = j$. This completes round k.

Let

$$I^* = \left(\left(I_1, \hat{J}_1\right), \left(I_2, \hat{J}_2\right) \ldots, \left(I_{t-1}, \hat{J}_{t-1}\right) \right),$$

$$J^* = \left(\left(\hat{I}_1, J_1\right), \left(\hat{I}_2, J_2\right) \ldots, \left(\hat{I}_{t-1}, J_{t-1}\right) \right).$$

Both I^* and J^* 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 takes her random output to be $I = \mathcal{R}(I^*)$ and Bob takes his random output to be $J = \mathcal{R}(J^*)$. Thus, both I and J take values in $\{e\} \cup [K]$, where $K = (MN)^{t-1}$.

3.2.3 Analysis

We will now prove that the sequence of protocols just described does achieve the rate promised in (24).

Claim 3.1 *a) For each $k \in \{1, 2, \ldots, t-1\}$, choose any $(i_k, j_k) \in [M] \times [N]$. Then,*

$$\frac{1}{(MN)^{t-1}} \geq Pr\left[I^* = J^* = ((i_k, j_k))_{k=1}^{t-1}\right] \geq \frac{1-\lambda}{(MN)^{t-1}}$$

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

$$b) \quad \lim_{t\to\infty} \frac{\log (MN)^{t-1}}{t^2} = \begin{cases} h(p-\mu) + h(q-\nu) & \text{if } h(p)+h(q) \leq 1; \\ 2 - h(p+\delta) - h(q+\epsilon) & \text{if } h(p)+h(q) > 1. \end{cases}$$

Proof: For convenience, let

$$G_k = \{I_k = i_k, J_k = j_k\} \text{ and } \hat{G}_k = \left\{\hat{I}_k = i_k, \hat{J}_k = j_k\right\}.$$

Then, note that

$$Pr\left[I^* = J^* = ((i_k, j_k))_{k=1}^{t-1}\right] = Pr\left[\bigcap_{k=1}^{t-1} \left(G_k \cap \hat{G}_k\right)\right]. \tag{34}$$

Now, for each $k \geq 1$, $\left(\hat{I}_k, \hat{J}_k, I_{k+1}, J_{k+1}\right)$ is conditionally independent of $\left(I_1^{k-1}, J_1^{k-1}, \hat{I}_1^{k-1}, \hat{J}_1^{k-1}\right)$, given (I_k, J_k). Therefore,

$$Pr\left[\bigcap_{k=1}^{t-1} \left(G_k \cap \hat{G}_k\right)\right] = Pr\left[G_1\right] \left(\prod_{k=1}^{t-2} Pr\left[\hat{G}_k \cap G_{k+1} \middle| G_k\right]\right) Pr\left[\hat{G}_{t-1} \middle| G_{t-1}\right]. \tag{35}$$

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

$$\begin{aligned} Pr\left[G_1\right] &= Pr\left[J_1 = j_1\right] \cdot Pr\left[I_1 = i_1\right] \\ &= W_p^t\left(\mathcal{A}(j_1) \,\middle|\, \mathbf{a}\right) \cdot W_q^t\left(\mathcal{B}(i_1) \,\middle|\, \mathbf{b}\right). \end{aligned}$$

From (29) and (30), it follows that

$$\frac{1}{MN} \geq Pr\left[G_1\right] \geq \left(\frac{1-\rho}{N}\right)\left(\frac{1-\sigma}{M}\right). \tag{36}$$

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

$$\begin{aligned} Pr\left[\hat{G}_k \cap G_{k+1} \middle| G_k\right] &= Pr\left[\hat{I}_k = i_k, J_{k+1} = j_{k+1} \middle| I_k = i_k\right] \cdot \\ &\qquad \cdot Pr\left[\hat{J}_k = j_k, I_{k+1} = i_{k+1} \middle| J_k = j_k\right] \\ &= W_p^t\left(\mathcal{A}_{i_k}(j_{k+1}) \,\middle|\, \mathbf{a}_{i_k}\right) \cdot W_q^t\left(\mathcal{B}_{j_k}(i_{k+1}) \,\middle|\, \mathbf{b}_{j_k}\right). \end{aligned}$$

From (32) and (33), it follows that

$$\frac{1}{MN} \geq Pr\left[\hat{G}_k \cap G_{k+1} \middle| G_k\right] \geq \left(\frac{1-\alpha-\rho}{N}\right)\left(\frac{1-\beta-\sigma}{M}\right), \quad 1 \leq k \leq t-2. \tag{37}$$

Finally,

$$Pr\left[\hat{G}_{t-1} \middle| G_{t-1}\right] = W_p^t\left(\mathcal{A}_{i_{t-1}} \,\middle|\, \mathbf{a}_{i_{t-1}}\right) \cdot W_q^t\left(\mathcal{B}_{j_{t-1}} \,\middle|\, \mathbf{b}_{j_{t-1}}\right),$$

so that

$$1 \geq Pr\left[\hat{G}_{t-1} \middle| G_{t-1}\right] \geq (1-\alpha)(1-\beta). \tag{38}$$

By (34), (35), (36), (37), and (38),

$$\frac{1}{(MN)^{t-1}} \geq Pr\left[I^* = J^* = ((i_k, j_k))_{k=1}^{t-1}\right]$$

and

$$\begin{aligned} &Pr\left[I^* = J^* = ((i_k, j_k))_{k=1}^{t-1}\right] \\ &\geq \left(\frac{1-\rho}{N}\right)\left(\frac{1-\sigma}{M}\right)\left[\frac{(1-\alpha-\rho)(1-\beta-\sigma)}{MN}\right]^{t-2}(1-\alpha)(1-\beta) \\ &\geq \frac{[(1-\alpha-\rho)(1-\beta-\sigma)]^t}{(MN)^{t-1}} \\ &\geq \frac{1-t(\alpha+\rho+\beta+\sigma)}{(MN)^{t-1}}, \end{aligned}$$

which proves Part a).

Next, note that by (22), (23), (25), and (26),

$$\begin{aligned} MN &= \left(\left\lfloor 2^{th(q-\nu)} \right\rfloor - 1\right)\left(\left\lfloor 2^{th(p-\mu)} \right\rfloor - 1\right) && \text{if } h(p)+h(q) \le 1; \\ &= \left(\left\lfloor 2^{t(1-h(p+\delta))} \right\rfloor - 1\right)\left(\left\lfloor 2^{t(1-h(q+\epsilon))} \right\rfloor - 1\right) && \text{if } h(p)+h(q) > 1. \end{aligned}$$

Part b) of the claim follows easily from this. □

3.3 Case 2: $p = 0$, $0 < q \le 1/2$

Note that $R^*(p,q) = h(q)$ in this case. Pick any $\nu \in (0,q)$. The sequence of protocols to be described will achieve the rate $h(q-\nu)$, which can be made arbitrarily close to $R^*(p,q)$ by choosing ν small enough.

Let $M = \left\lfloor 2^{th(q-\nu)} \right\rfloor - 1$, and let σ be given by (28). By Lemma 3.2, for any $\mathbf{b} \in \{0,1\}^t$, there exists a (q, M, σ) equipartition of $\{0,1\}^t$ w.r.t. $\mathbf{b}$ into subsets $\mathcal{B}(e), \mathcal{B}(1), \ldots, \mathcal{B}(M)$. This equipartition satisfies (30). Alice and Bob agree on some such bitstring and corresponding equipartition. They also agree upon $M+1$ distinct (but otherwise arbitrary) bitstrings, $\mathbf{a}_e, \mathbf{a}_1, \ldots, \mathbf{a}_M$, each of length t.

As before, the t^2-step protocol proceeds in t rounds, indexed $0, 1, \ldots, t-1$. In each round, Bob simply transmits the bitstring $\mathbf{b}$. As for Alice, she transmits a dummy bitstring (say, $00\cdots0$) in round 0, and the bitstring $\mathbf{a}(I_k)$ in round k $(1 \le k < t)$. Here, I_k is the unique $i \in \{e\} \cup [M]$ such that the bitstring she receives from Bob in round $k-1$ falls in $\mathcal{B}(i)$.

Since $p = 0$, Bob receives all transmissions from Alice without any errors. Thus, after round $t-1$, both Alice and Bob know $I^* = (I_1, I_2, \ldots, I_{t-1})$. Let $\mathcal{R}$ be a function that maps onto $[M^{t-1}]$ the M^{t-1} values of I^* in which no component is e, and maps all the remaining values to e. Both Alice and Bob take $\mathcal{R}(I^*)$ to be their random output.

Claim 3.2 *a) For each $k \in \{1, 2, \ldots, t-1\}$, choose any $i_k \in [M]$. Then,*

$$\frac{1}{M^{t-1}} \ge Pr\left[I^* = (i_1, i_2, \ldots, i_{t-1})\right] \ge \frac{1-\lambda}{M^{t-1}}$$

where $\lambda = t\sigma \to 0$ as $t \to \infty$.

b)

$$\lim_{t\to\infty} \frac{\log M^{t-1}}{t^2} = h(q-\nu).$$

Proof:

$$\begin{aligned} Pr\left[I^* = (i_1, i_2, \ldots, i_{t-1})\right] &= \prod_{k=1}^{t-1} Pr\left[I_k = i_k\right] \\ &= \prod_{k=1}^{t-1} W_q^t\left(\mathcal{B}(i_k) \,\middle|\, \mathbf{b}\right). \end{aligned}$$

Part a) now follows easily from (30). Part b) is obvious from the definition of M. □

3.4 Case 3: $0 < p < 1/2$, $q = 1/2$

Note that $R^*(p,q) = 1 - h(p)$ in this case. Pick any $\delta > 0$ satisfying $p + \delta \leq \frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}$. Then, pick any $\nu \in (0, q)$ such that $1 - h(p+\delta) \leq h(q-\nu)$. The sequence of protocols to be described will achieve the rate $1 - h(p+\delta)$, which can be made arbitrarily close to $R^*(p,q)$ by choosing δ small enough.

Let $M = \left\lfloor 2^{t(1-h(p+\delta))} \right\rfloor - 1$, and $\alpha = 4 \cdot 2^{-tD(p+\delta\|p)}$. By Lemma 3.1, there exists a $(t, M+1, \alpha)$ block code, $\{(\mathbf{a}_i, \mathcal{A}_i) : i = e, 1, \ldots, M\}$, for the $BSC(p)$. Alice and Bob agree upon such a code.

Let σ be given by (28). Since $M \leq 2^{th(q-\nu)}$, Lemma 3.2 guarantees that, for any $\mathbf{b} \in \{0,1\}^t$, there exists a (q, M, σ) equipartition of $\{0,1\}^t$ w.r.t. $\mathbf{b}$ into subsets $\mathcal{B}(e), \mathcal{B}(1), \ldots, \mathcal{B}(M)$. As before, this equipartition satisfies (30). Alice and Bob also agree on some such bitstring and corresponding equipartition.

Again, the t^2-step protocol proceeds in t rounds, indexed $0, 1, \ldots, t-1$. In each round, Bob simply transmits the bitstring $\mathbf{b}$. Alice transmits a dummy bitstring (say, $00\cdots0$) in round 0, and the codeword $\mathbf{a}(I_k)$ in round k $(1 \leq k < t)$. Here, I_k is the unique $i \in \{e\} \cup [M]$ such that the bitstring she receives from Bob in round $k-1$ falls in $\mathcal{B}(i)$.

Bob estimates I_k, $1 \leq k < t$, as the unique $i \in \{e\} \cup [M]$ such that the bitstring he receives from Alice in round k falls in $\mathcal{A}_i$. Call the estimate $\hat{I}_k$.

Let $I^* = (I_1, I_2, \ldots, I_{t-1})$ and $J^* = \left(\hat{I}_1, \hat{I}_2, \ldots, \hat{I}_{t-1}\right)$. After round $t-1$, Alice's random output is $I = \mathcal{R}(I^*)$ and Bob's random output is $J = \mathcal{R}(J^*)$. Here, $\mathcal{R}$ is a map exactly as in Case 2.

Claim 3.3 *a) For each $k \in \{1,2,\ldots,t-1\}$, choose any $i_k \in [M]$. Then,*

$$\frac{1}{M^{t-1}} \geq Pr\left[I^* = J^* = (i_1, i_2, \ldots, i_{t-1})\right] \geq \frac{1-\lambda}{M^{t-1}}$$

where $\lambda = t(\alpha+\sigma) \to 0 \quad as\ t \to \infty$.

b)

$$\lim_{t\to\infty} \frac{\log M^{t-1}}{t^2} = 1 - h(p+\delta).$$

Proof:

$$\begin{aligned} Pr\left[I^* = J^* = (i_1, i_2, \ldots, i_{t-1})\right] &= \prod_{k=1}^{t-1} \left(Pr\left[I_k = i_k\right] \cdot Pr\left[\hat{I}_k = i_k \middle| I_k = i_k\right]\right) \\ &= \prod_{k=1}^{t-1} \left[W_q^t\left(\mathcal{B}(i_k) \,\middle|\, \mathbf{b}\right) \cdot W_p^t\left(\mathcal{A}_{i_k} \,\middle|\, \mathbf{a}_{i_k}\right)\right]. \end{aligned} \tag{39}$$

Part a) follows easily from (39), (30), and the fact that $1 \geq W_p^t\left(\mathcal{A}_{i_k} \,\middle|\, \mathbf{a}_{i_k}\right) \geq 1-\alpha$. Part b) is obvious from the definition of M. □

4 Appendix

Proof of Lemma 2.1: Part a) is obvious from the definition of $\mathcal{E}$. To prove Part b), first note that

$$W_{p,q,f,g}^n(\mathcal{E}^c) = Pr\left[\left|\log W_{p,q,f,g}^n\left(\hat{X}^n, \hat{Y}^n\right) + n\left(h(p)+h(q)\right)\right| > \theta\sqrt{n}\right] \tag{40}$$

where $\hat{X}_k$ and $\hat{Y}_k$ are respectively the (random) bits that Bob and Alice receive in the k^{th} step of the given protocol. Let

$$Z_k = \log W_p\left(\hat{X}_k \,\middle|\, f_k\left(\hat{Y}^{k-1}\right)\right) + \log W_q\left(\hat{Y}_k \,\middle|\, g_k\left(\hat{X}^{k-1}\right)\right).$$

The right hand side of (40) then equals

$$Pr\left[\left|\sum_{k=1}^{n} Z_k + n\left(h(p)+h(q)\right)\right| > \theta\sqrt{n}\right] \tag{41}$$

so that it suffices to prove that (41) can be made arbitrarily small by choosing θ large enough. This will be done using Chebyshev's inequality. To this end, we will now estimate the mean and variance of $\sum_{k=1}^{n} Z_k$. It is easy to see that

$$E\left[Z_k \mid \hat{X}^{k-1}, \hat{Y}^{k-1}\right] = -h(p) - h(q), \tag{42}$$

so that

$$\begin{aligned} E[Z_k] &= E\left[E\left[Z_k \mid \hat{X}^{k-1}, \hat{Y}^{k-1}\right]\right] \\ &= -h(p) - h(q), \qquad \text{for } 1 \le k \le n. \end{aligned} \tag{43}$$

Next, note that if $k' < k$ then $Z_{k'}$ is a function of $\left(\hat{X}^{k-1}, \hat{Y}^{k-1}\right)$. Therefore,

$$\begin{aligned} E[Z_{k'} Z_k] &= E\left[E\left[Z_{k'} Z_k \mid \hat{X}^{k-1}, \hat{Y}^{k-1}\right]\right] \\ &= E\left[Z_{k'} \cdot E\left[Z_k \mid \hat{X}^{k-1}, \hat{Y}^{k-1}\right]\right] \\ &= E[Z_{k'} \cdot E[Z_k]] \\ &= E[Z_{k'}] \cdot E[Z_k], \end{aligned} \tag{44}$$

where (44) is by (42) and (43). But this means that

$$\mathrm{Var}\left(\sum_{k=1}^{n} Z_k\right) = \sum_{k=1}^{n} \mathrm{Var}(Z_k). \tag{45}$$

Now, let σ^2 be a uniform upper bound on $\mathrm{Var}(Z_k)$, $k = 1, 2, \ldots, n$. (Such a bound obviously exists.) By (43), (45), and Chebyshev's inequality, (41) is upper bounded by σ^2/θ^2, which can be made arbitrarily small by choosing θ large enough. □

Proof of Lemma 3.2: For $0 \le w \le t$, let

$$\mathcal{T}_w = \left\{\mathbf{z} \in \{0,1\}^t : \sum_{k=1}^{t} z_k = w\right\},$$

and let

$$\mathbf{u} \oplus \mathcal{T}_w = \{\mathbf{u} \oplus \mathbf{z} : \mathbf{z} \in \mathcal{T}_w\}.$$

($\mathbf{u} \oplus \mathbf{z}$ is the bitstring obtained by bitwise mod 2 addition of $\mathbf{u}$ and $\mathbf{z}$.)

Pick any $\mu \in [r-\eta, r]$. For each w satisfying $h(\frac{w}{t}) > h(\mu)$, construct L pairwise disjoint subsets of $\mathcal{U} \cap (\mathbf{u} \oplus \mathcal{T}_w)$, say $\mathcal{U}_w(1), \mathcal{U}_w(2), \ldots, \mathcal{U}_w(L)$, each of size exactly $\lfloor |\mathcal{U} \cap (\mathbf{u} \oplus \mathcal{T}_w)| / L \rfloor$ (the subsets are otherwise arbitrary).

Let

$$\mathcal{U}(l) = \bigcup_{h(\frac{w}{t}) > h(\mu)} \mathcal{U}_w(l), \quad l = 1, 2, \ldots, L$$

and let

$$\mathcal{U}(e) = \mathcal{U} \cap \left[\bigcup_{l=1}^{L} \mathcal{U}(l) \right]^c .$$

By construction, $W_r^t\left(\mathcal{U}(l) \,\middle|\, \mathbf{u}\right) = W_r^t\left(\mathcal{U}(l') \,\middle|\, \mathbf{u}\right)$ for all $l, l' \in [L]$. Therefore, for any $l' \in [L]$,

$$\begin{aligned} W_r^t\left(\mathcal{U}(l') \,\middle|\, \mathbf{u}\right) &= \frac{1}{L} W_r^t\left(\bigcup_{l=1}^{L} \mathcal{U}(l) \,\middle|\, \mathbf{u}\right) \\ &= \frac{1}{L}\left[W_r^t\left(\mathcal{U} \,\middle|\, \mathbf{u}\right) - W_r^t\left(\mathcal{U}(e) \,\middle|\, \mathbf{u}\right) \right], \end{aligned}$$

which proves (19). To upper bound $W_r^t\left(\mathcal{U}(e) \,\middle|\, \mathbf{u}\right)$, note that

$$\begin{aligned} W_r^t\left(\mathcal{U}(e) \,\middle|\, \mathbf{u}\right) &= \sum_{h(\frac{w}{t}) > h(\mu)} \left[|\mathcal{U} \cap (\mathbf{u} \oplus \mathcal{T}_w)| \bmod L \right] r^w (1-r)^{t-w} \\ &\qquad + \sum_{h(\frac{w}{t}) \le h(\mu)} |\mathcal{U} \cap (\mathbf{u} \oplus \mathcal{T}_w)| \, r^w (1-r)^{t-w} \\ &\le \sum_{h(\frac{w}{t}) > h(\mu)} L \cdot r^w (1-r)^{t-w} \\ &\qquad + \sum_{h(\frac{w}{t}) \le h(\mu)} |\mathcal{T}_w| \, r^w (1-r)^{t-w}. \end{aligned} \tag{46}$$

The first term in (46) can be upper bounded as follows:

$$\begin{aligned} \sum_{h(\frac{w}{t}) > h(\mu)} L \cdot r^w (1-r)^{t-w} &= \sum_{t\mu < w < t(1-\mu)} L \cdot 2^{-t\left[h(\frac{w}{t}) + D(\frac{w}{t} \| r)\right]} \\ &\le L \cdot 2^{-th(\mu)} \sum_{t\mu < w < t(1-\mu)} 2^{-tD(\frac{w}{t} \| r)} \\ &\le 2^{th(r-\eta)} 2^{-th(\mu)} t. \end{aligned} \tag{47}$$

The second term in (46) can be upper bounded using Chernoff's theorem:

$$\begin{aligned}
\sum_{h(\frac{w}{t}) \le h(\mu)} |\mathcal{T}_w| \, r^w (1-r)^{t-w} &= \sum_{w \le t\mu} \binom{t}{w} r^w (1-r)^{t-w} \\
&\quad + \sum_{w \ge t(1-\mu)} \binom{t}{w} r^w (1-r)^{t-w} \\
&\le 2^{-tD(\mu \| r)} + 2^{-tD(1-\mu \| r)} \\
&\le 2 \cdot 2^{-tD(\mu \| r)} \qquad \text{since } \mu < r < 1/2. \quad (48)
\end{aligned}$$

From (46), (47), and (48), it follows that

$$\begin{aligned}
W_r^t \left(\mathcal{U}(e) \,\middle|\, \mathbf{u} \right) &\le t \cdot 2^{-t[h(\mu) - h(r-\eta)]} + 2 \cdot 2^{-tD(\mu \| r)} \\
&= \tau
\end{aligned}$$

if μ is chosen to attain the minimum in (20). □

References

[1] R. Ahlswede and I. Csiszàr. Common randomness in information theory and cryptography - part I: Secret sharing. *IEEE Transactions on Information Theory*, Vol. 39(No. 4), July 1993.

[2] R. Ahlswede and G. Dueck. Identification in the presence of feedback - a discovery of new capacity formulas. *IEEE Transactions on Information Theory*, Vol. 35(No. 1), January 1989.

[3] R. Ahlswede and G. Dueck. Identification via channels. *IEEE Transactions on Information Theory*, Vol. 35(No. 1), January 1989.

[4] R. Ahlswede and B. Verboven. On identification via multiway channels with feedback. *IEEE Transactions on Information Theory*, Vol. 37(No. 5), September 1991.

[5] R.G. Gallager. *Information Theory and Reliable Communication.* John Wiley, 1968.

[6] J.H.B. Kemperman. Strong converses for a general memoryless channel with feedback. *Trans. 6th Prague Conf. Information Theory, Stat. Dec. Fct's and Rand. Proc.*, 1973.

[7] L. Lovász. Communication complexity: A survey. In B.H. Korte et al., editors, *Paths, Flows and VLSI layout.* Springer-Verlag, 1990.

[8] U.M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, Vol. 39(No. 3), May 1993.

[9] A. Orlitsky and A. El Gamal. Communication complexity. In Y. Abu-Mostafa, editor, *Complexity in Information Theory.* Springer-Verlag, 1988.