

TABLE I  
COMPARISON BETWEEN UPPER ESTIMATES  
OF  $A_i$

$k$	$d'$	RES 1	RES 2	$H_1$	$H_2$
199	64	.408e-23	.169e-23	163	207
191	42	.408e-17	.899e-18	93	199*
139	26	.582e-3	.541e-3	93	191*

TABLE II  
COMPARISON BETWEEN UPPER ESTIMATES OF  $\max P_{ue}$

$k$	$d'$	$P_{ue}$ 1	$P_{ue}$ 2	$P_{ue}$ 3	$H_1$	$H_2$
199	64	.289e-14	.769e-15	.747e-15	163	207
191	42	.226e-15	.395e-16	.390e-16	93	199*
139	26	.489e-24	.121e-26	.114e-26	93	191*

TABLE III  
DIFFERENT UPPER ESTIMATES OF  $\max P_{ue}$  FOR BCH[255, 139, 31]

$\max P_{ue}$ computed via Method A	.489e-24
$\max P_{ue}$ computed via Method C	.121e-26
$\max P_{ue}$ computed via Method D	.114e-26
$\max P_{ue}$ computed via mixed approach, $s = 44$	.122e-26
$\max P_{ue}$ computed via mixed approach, $s = 38$	.254e-25

In Table I the first column contains the dimension of  $H$ , the second one contains the distance of the dual code of  $H$  (or a lower estimate), the following two columns contain the RES calculated as above, the last two columns<sup>1</sup> contain, respectively, the dimensions of  $H_1$  and  $H_2$ . Table II is similar to Table I except for the middle columns, which contain  $\max P_{ue}$  calculated as above.

If the computation of RES 2 (hence of  $P_{ue}$  2) is quite costly (as for the BCH[255, 139, 31]) and the researcher is only interested in the  $P_{ue}$ , we suggest a mixed approach between Method A and Method B (or Method C), in three steps. First, one calculates all  $\alpha_i$  by Method A. Second, one calculates only some  $\alpha_i$  by Method B (or Method C), more precisely, the  $\alpha_i$  with lower  $i$  (say  $i \leq s$ ). Finally, one calculates the estimate of the  $P_{ue}$  taking for  $i \leq s$  the  $\alpha_i$  obtained by the second step and, for the remaining values, the  $\alpha_i$  obtained by the first one.

A few results of the mixed approach for the  $\max P_{ue}$  of BCH[255, 139, 31] are shown in Table III.

ACKNOWLEDGMENT

The authors wish to thank T. Berger, P. Charpin, T. Fujiwara, and C. Traverso for their helpful suggestions.

REFERENCES

[1] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.  
 [2] Y. Desaki, T. Fujiwara, and T. Kasami, "The weight distributions of extended binary primitive BCH codes of length 128," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1364–1371, July 1997.  
 [3] T. Fujiwara and T. Kasami, "The weight distributions of (256,  $k$ ) extended binary primitive BCH codes with  $k \leq 63$  and  $k \geq 207$ ," IEICE, Tech. Rep. IT97-46 (1997-09), pp. 29–33.  
 [4] O. Keren and S. Litsyn, "More on the distance distribution of BCH codes," *IEEE Trans. Inform. Theory*, vol. 45, pp. 251–255, Jan. 1999.  
 [5] T. Kasami, T. Fujiwara, and S. Lin, "An approximation to the weight distribution of binary linear codes," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 769–780, Nov. 1985.

[6] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1977.  
 [7] W. W. Peterson and E. J. Weldon Jr., *Error Correcting Codes*. Cambridge, MA: MIT Press, 1972.  
 [8] T. Schaub, "A Linear Complexity Approach to Cyclic Codes," Dissertation ETH no. 8730 in Technical Sciences, Swiss Federal Inst. Technol., Zürich, Switzerland, 1988.  
 [9] D. Augot and F. Levy-dit-Vehel, "Bounds on the minimum distance of the duals of BCH codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1257–1260, July 1996.  
 [10] T. Kasami, N. Tokura, and S. Azumi, "On the weight enumeration of weights less than 2.5  $d$  of Reed–Muller codes," *Inform. Contr.*, vol. 30, pp. 380–395, Apr. 1976.  
 [11] F. Levy-dit-Vehel, "Bounds on the minimum distance of duals of extended BCH codes over  $F_p$ ," *Applicable Algebra in Eng. Commun. Comput. AAECC*, vol. 6, pp. 175–190, 1995.

Separation of Random Number Generation and Resolvability

Karthik Visweswariah, *Member, IEEE*,  
 Sanjeev R. Kulkarni, *Senior Member, IEEE*, and  
 Sergio Verdú, *Fellow, IEEE*

**Abstract**—We consider the problem of determining when a given source can be used to approximate the output due to any input to a given channel. We provide achievability and converse results for a general source and channel. For the special case of a full-rank discrete memoryless channel we give a stronger converse result than we can give for a general channel.

**Index Terms**—Approximation theory, channel output statistics, random number generation, resolvability, source–channel separation.

I. INTRODUCTION

The classical separation theorem essentially states that source and channel coding can be done separately without losing optimality. In [5], a source–channel separation theorem was shown for sources and channels more general than those in the classical separation theorem. Here we investigate an analogous separation theorem in the case of resolvability and random number generation. Random number generation involves finding a deterministic transformation to generate a sequence of equiprobable bits from a given source of randomness. Resolvability is a property of a channel which gives the amount of randomness required to simulate, at the output of the channel, any distribution that can be achieved by a random input to the channel. We consider the two problems together, i.e., when can a given source of randomness be used to simulate the output of a given channel due to any input, to arbitrary accuracy. In Section II, we give some notation and background to state

Manuscript received May 5, 1999; revised April 6, 2000. This work was supported in part by the National Science Foundation under Grants NYI Award IRI-9457645 and NCR 9523805. The material in this correspondence was presented in part at the Allerton Conference on Communication, Control and Computing, Monticello, IL, September 1998.

K. Visweswariah is with IBM T. J. Watson Research Center, Yorktown Heights NY 10598 USA (e-mail: kv1@watson.ibm.com).

S. R. Kulkarni and S. Verdú are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: kulkarni@ee.princeton.edu; verdu@ee.princeton.edu).

Communicated by I. Csiszár, Associate Editor for Shannon Theory. Publisher Item Identifier S 0018-9448(00)07014-0.

<sup>1</sup>\*\* stands for "use of estimated values of weight distributions."

the problem precisely. Sections III and IV give achievability and converse results for a source to approximate the output of a given channel due to any input. We have not been able to show results fully analogous to [5] and there is a gap between the achievability and converse results that we have not been able to bridge. Section V deals with the special case when the channel is discrete-memoryless and of full rank. For this special case we will be able to show a converse which is stronger than the converse in Section IV for a general channel.

## II. PRELIMINARIES AND PROBLEM FORMULATION

In this section we give, some basic definitions, a precise statement of the problem to be considered, and results which are already known for the problem of separation of resolvability and random number generation.

The following definition is as in [3] and we repeat it here for the sake of completeness.

*Definition 1:* A channel  $\mathbf{W}$  with input alphabet  $A$  and output alphabet  $B$  is a sequence of conditional distributions

$$\mathbf{W} = \{W^n(y^n|x^n) : (x^n, y^n) \in A^n \times B^n\}_{n=1}^\infty.$$

*Definition 2:* A source of randomness  $\mathbf{X}$  is a sequence of finite-dimensional distributions  $\{P_{X^n}\}_{n=1}^\infty$  with  $X^n$  taking values in  $A^n$ .

Note that there are no consistency requirements on the sequence of finite-dimensional distributions, this is the difference between Definition 2 and the standard definition of a random process. Throughout we assume that the source and channel alphabets are finite.

We now give some notation that is used throughout this correspondence. Let the output distribution, when the input is distributed according to  $Q^n$ , be denoted by  $Q^n W^n$ . Thus

$$Q^n W^n(y^n) = \sum_{x^n \in A^n} W^n(y^n|x^n) Q^n(x^n).$$

Also let

$$i_{X^n W^n}(a^n, b^n) \triangleq \log \frac{W^n(b^n|a^n)}{P_{Y^n}(b^n)}$$

where  $P_{Y^n}$  is the output distribution when  $P_{X^n}$  is input to the channel  $W^n$ , and

$$h_{Z^n}(z^n) \triangleq \log \frac{1}{P_{Z^n}(z^n)}.$$

We will use  $l_1$  distance to measure the difference between two distributions on the same alphabet. We will denote the distance between  $P$  and  $Q$  by  $d(P, Q)$ . We note that

$$d(P, Q) = 2 \sup_{E \subseteq A} |P(E) - Q(E)|$$

where  $A$  is the alphabet on which the two distributions are defined. If  $X$  and  $Y$  are two random variables on the same alphabet we sometimes write  $d(X, Y)$  for the  $l_1$  distance between the distributions of  $X$  and  $Y$ .

We now define precisely what we mean by a source  $\mathbf{Z}$  being an approximating source for a channel  $\mathbf{W}$ .

*Definition 3:* For any  $\epsilon > 0$ , the source  $\mathbf{Z}$  with alphabet  $F$  is called an  $\epsilon$ -approximating source for the channel  $\mathbf{W}$  if for any arbitrary input source  $\tilde{\mathbf{X}}$ , there exists a sequence of deterministic mappings  $\{\phi_n: F^n \rightarrow A^n\}$  such that for sufficiently large  $n$

$$d(Y^n, \tilde{Y}^n) < \epsilon$$

where  $Y^n$  and  $\tilde{Y}^n$  are the outputs of the channel due to  $\phi(Z^n)$  and  $\tilde{X}^n$ , respectively.

Using this definition of an  $\epsilon$ -approximating source we can now define  $\mathbf{Z}$  to be an *approximating source* for  $\mathbf{W}$  if it is  $\epsilon$ -approximating for  $\mathbf{W}$  for all  $\epsilon > 0$ .

The resolvability of a channel is the minimum number of random bits required per input sample to approximate arbitrarily well the output of the channel due to any input process (see [3] for a formal definition of resolvability). The problem of resolvability of a channel was first considered by Han and Verdú ([3]) where it was shown that the resolvability of a channel is  $\sup_{\mathbf{X}} \bar{I}(\mathbf{X}, \mathbf{Y})$  (For a definition of  $\bar{I}(\mathbf{X}, \mathbf{Y})$  see [3]). The problem we consider here is that of finding necessary and sufficient conditions for a given source  $\mathbf{Z}$  to be an approximating source for a channel  $\mathbf{W}$ . To use a given source to approximate the output of a channel due to another source we could use the source to generate random bits at the best possible rate and then use a deterministic transformation of the random bits at the input of the channel. The problem of random bit generation was considered by Vembu and Verdú [6], where fundamental limits on the rate at which random bits can be generated from a given source were given. Using the results of [3] and [6] and the two-step process outlined above we can easily show the following sufficient condition.

*Theorem 1:* If  $\underline{H}(\mathbf{Z}) > S$  then  $\mathbf{Z}$  is an approximating source for a channel  $\mathbf{W}$ , where  $S$  is the resolvability of the channel  $\mathbf{W}$ .

The converse below can also be derived using the results in [3]. A source  $\mathbf{Z}$  can be approximated using  $\bar{H}(\mathbf{Z})$  random bits per sample [3, Theorem 3]. This also means any process derived from  $\mathbf{Z}$  by a deterministic transformation can be approximated using  $\bar{H}(\mathbf{Z})$  random bits per sample. Thus if a source  $\mathbf{Z}$  is an approximating source for a channel then we can approximate any output of the channel with  $\bar{H}(\mathbf{Z})$  random bits per sample and so the resolvability  $S$  of the channel must be smaller than  $\bar{H}(\mathbf{Z})$ . Thus we have the following theorem.

*Theorem 2:* If  $\bar{H}(\mathbf{Z}) < S$  then  $\mathbf{Z}$  is not an approximating source for a channel  $\mathbf{W}$ .

In the next two sections we will give an achievable and converse results in the spirit of [5].

## III. ACHIEVABILITY

In this section we give a sufficient condition for a source  $\mathbf{Z}$  to be an approximating source for a channel  $\mathbf{W}$  which implies the sufficient condition given by Theorem 1. The sufficient condition is analogous to a sufficient condition derived in [5] for the source-channel separation problem. We will first define the notion of a source strictly dominating a channel.

*Definition 4:* A source  $\mathbf{Z}$  is said to strictly dominate a channel  $\mathbf{W}$  if for every channel input process  $\mathbf{X}$  there exists a  $\delta > 0$  such that

$$\lim_{n \rightarrow \infty} \inf_{c_n \in \mathcal{R}} \left\{ P \left[ \frac{1}{n} h_{Z^n}(Z^n) \leq c_n + \delta \right] + P \left[ \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \geq c_n \right] \right\} = 0.$$

We note that  $\underline{H}(\mathbf{Z}) > S$  implies that  $\mathbf{Z}$  strictly dominates  $\mathbf{W}$  since if we take  $c_n = (\underline{H}(\mathbf{Z}) + S)/2$  the required condition is satisfied. The converse, however, is not true.

*Theorem 3:* If a source  $\mathbf{Z}$  strictly dominates a channel  $\mathbf{W}$  then  $\mathbf{Z}$  is an approximating source for the channel  $\mathbf{W}$ .

*Proof:* If the source  $\mathbf{Z}$  strictly dominates the channel  $\mathbf{W}$ , then for each input process  $\mathbf{X}$  there exists a sequence  $\{c_n\}_{n=1}^{\infty}$  and a  $\delta > 0$  such that

$$P \left[ \frac{1}{n} h_{Z^n}(Z^n) \leq c_n + \delta \right] \leq \tau_n \quad (1)$$

and

$$P \left[ \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \geq c_n \right] \leq \tau_n \quad (2)$$

where  $\tau_n \rightarrow 0$  as  $n \rightarrow \infty$ . Equation (1) implies that we can approximate any distribution with type less than  $\exp n (c_n + \frac{2}{3} \delta)$  using the distribution  $Z^n$ . What we mean by type here is the following: A distribution is of type  $k$  if all probabilities that it assigns are integral multiples of  $1/k$ . To show this we use the procedure in the Aggregation Lemma [6].

Define

$$S^{(n)} = \{z^n \in F^n : P_{Z^n}(z^n) \leq \exp^{-n(c_n + \delta)}\}.$$

Consider any  $M$ -type distribution  $P_M$  on  $\{1, 2, \dots, M\}$ . We will place elements of  $S^{(n)}$  in bin  $B_n(i)$  until we have

$$P_{Z^n}(B_n(i)) > P_M(i) - \exp^{-n(c_n + \delta)}.$$

We stop either when we complete this process for all  $i = 1, 2, \dots, M$  or when we run out of sequences in the set  $S^{(n)}$ . All remaining sequences in  $F^n$  are placed in  $B_n(1)$ . At the end of this process we have

$$\begin{aligned} & \sum_{i=1}^M |P_{Z^n}(B_n(i)) - P_M(i)| \\ & \leq \max \left( 2M \exp^{-n(c_n + \delta)}, 2\tau_n + M \exp^{-n(c_n + \delta)} \right). \end{aligned}$$

For any  $M \leq \exp n (c_n + \frac{2}{3} \delta)$  the right-hand side of the last equation goes to zero, since  $\tau_n \rightarrow 0$ .

We will now state and use a lemma the proof of which readily follows from the argument used in [3, Proof of Theorem 4].

*Lemma 1:* If

$$P \left[ \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \geq c_n \right] \leq \tau_n$$

where  $\tau_n \rightarrow 0$  as  $n \rightarrow \infty$  then for any  $\gamma > 0$  there exists a process  $\tilde{\mathbf{X}}$  (producing output  $\tilde{\mathbf{Y}}$ ) such that

$$\lim_{n \rightarrow \infty} d(Y^n, \tilde{Y}^n) = 0$$

and  $\tilde{X}^n$  is an  $M$ -type distribution with  $M \leq \exp n(c_n + \gamma)$

Equation (2) and Lemma 1 imply that there exists a process  $\tilde{\mathbf{X}}^n$  with type smaller than  $\exp n (c_n + \frac{2}{3} \delta)$  which approximates the output due to  $\mathbf{X}$ . We can approximate arbitrarily closely any distribution with type less than or equal to  $\exp n (c_n + \frac{2}{3} \delta)$  (and hence  $\tilde{X}^n$ ) using  $Z^n$ . This along with the fact that  $d(PW, QW) \leq d(P, Q)$  for any channel  $W$  and distributions  $P, Q$  imply that  $\mathbf{Z}$  can be used to approximate the output of the channel  $\mathbf{W}$  when the input process is  $\mathbf{X}$ . Since we can find a sequence  $\{c_n\}_{n=1}^{\infty}$  and a  $\delta > 0$  which satisfy (1) and (2) for any input process  $\mathbf{X}$ , the source  $\mathbf{Z}$  can approximate the output due to any process  $\mathbf{X}$ .  $\square$

#### IV. CONVERSE

In this section we will give a converse result stating a condition under which the source will not be an approximating source for a channel in a certain sense. We would like to have a result completely analogous to the necessary condition in [5] for the source-channel separation problem, which would have been: If  $\mathbf{Z}$  is an approximating source

for a channel  $\mathbf{W}$  then the source dominates the channel. The notion of a source dominating a channel would be defined (analogous to [5]) as follows.

*Definition 5:* A source  $\mathbf{Z}$  dominates the channel  $\mathbf{W}$  if for every process  $\mathbf{X}$ , for every  $\delta > 0$ , and for every sequence of nonnegative numbers  $\{c_n\}_{n=1}^{\infty}$

$$\lim_{n \rightarrow \infty} P \left[ \frac{1}{n} h_{Z^n}(Z^n) \leq c_n - \delta \right] P \left[ \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \geq c_n \right] = 0.$$

It can be verified that if a source strictly dominates a channel then it dominates the channel. Also note that  $\bar{H}(\mathbf{Z}) < S$  implies that the source does not dominate the channel.

We have not been able to show this statement that we set out to prove but we give a weaker statement that neither implies nor is implied by Theorem 2. The result is stated in contrapositive form and is weaker than the statement that we would like to make in two ways. First it involves the notion of a source being a strong approximating source for a channel and secondly, the necessary condition that we show for a source to be strongly approximating for a channel is weaker than the notion of a source dominating a channel.

We now give the stronger definition of an approximating source, following which we can state a necessary condition for a given source to be a strong approximating source for a given channel.

*Definition 6:* For any  $\epsilon > 0$ , the source  $\mathbf{Z}$  with alphabet  $F$  is called a strong  $\epsilon$ -approximating source for the channel  $\mathbf{W}$  if for any arbitrary input source  $\tilde{\mathbf{X}}$ , there exists a sequence of deterministic mappings  $\{\phi_n : F^n \mapsto A^n\}$  such that  $P_{\phi_n(Z^n)} \ll P_{\tilde{X}^n}$  and such that for sufficiently large  $n$

$$d(Y^n, \tilde{Y}^n) < \epsilon$$

where  $Y^n$  and  $\tilde{Y}^n$  are the outputs of the channel due to  $\phi(Z^n)$  and  $\tilde{X}^n$ , respectively.

We call  $\mathbf{Z}$  a strong approximating source for  $\mathbf{W}$  if it is a strong  $\epsilon$ -approximating source for  $\mathbf{W}$  for all  $\epsilon > 0$ .

Note that the only difference between Definitions 3 and 6 is that the latter places an extra constraint that the deterministic transformation can only map to those sequences which have nonzero probability under the source whose output we are trying to approximate.

The following theorem states precisely a necessary condition for the source to be a strong approximating source for a channel. We note that in (4) below if we did not have  $\beta_n$  going to 1 but just being strictly bigger than 0 then this condition would be the negation of a source dominating a channel.

*Theorem 4:* Suppose that for some process  $\mathbf{X}$ , for some  $\alpha, \delta > 0$ , and for some sequence of nonnegative numbers  $\{c_n\}_{n=1}^{\infty}$

$$P \left[ \frac{1}{n} h_{Z^n}(Z^n) \leq c_n - \delta \right] > \alpha \quad (3)$$

and

$$P \left[ \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \geq c_n \right] > \beta_n \quad (4)$$

for all  $n \in I$  where  $I$  is an infinite set of integers and  $\beta_n \rightarrow 1$  as  $n \rightarrow \infty$ . Then the source  $\mathbf{Z}$  is not a strong approximating source for the channel  $\mathbf{W}$ .

*Proof:* Equation (4) along with Feinstein's lemma [2] implies that there exists a code of length  $n$  with  $\exp n (c_n - \frac{\delta}{10})$  codewords and with maximal probability of error less than  $\epsilon_n$  for  $n \in I$  where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Consider a good code with  $M = \exp n (c_n - \frac{\delta}{10})$  codewords. Let the codewords be  $\{b_i\}_{i=1}^M$  and their corresponding de-

coding sets be  $\{D_i\}_{i=1}^M$ . Consider now the process  $\tilde{X}^n$  which places mass  $\frac{1}{M}$  on each of the  $M$  codewords,  $\{b_i\}_{i=1}^M$ . We will show that the output due to this process cannot be approximated well with our source.

Consider the set

$$S^n = \left\{ z^n : \frac{1}{n} h_{Z^n}(z^n) \leq c_n - \delta \right\}.$$

By (3),  $P_{Z^n}(S^n) > \alpha$  for all  $n \in I$ . Consider any deterministic mapping  $\phi_n$  (from  $F^n$  to  $A^n$ ) and the  $N$  codewords to which the sequences in  $S^n$  get mapped. Assume the codewords are numbered so that these codewords are  $\{b_i\}_{i=1}^N$ . Since  $|S^n| \leq \exp n(c_n - \delta)$ , we have  $N \leq \exp n(c_n - \delta)$ .

Let

$$B = \bigcup_{i=1}^N D_i$$

and  $\tilde{Y}^n$  be the output due to  $\tilde{X}^n$ .

$$\begin{aligned} P_{\tilde{Y}^n}(B) &= \frac{1}{M} \sum_{i=1}^N P(B|b_i) + \frac{1}{M} \sum_{i=N+1}^M P(B|b_i) \\ &\leq \frac{N}{M} + \frac{\epsilon_n}{M} (M - N) \\ &\leq 2\epsilon_n \end{aligned}$$

where the last inequality holds for sufficiently large  $n \in I$ . Let  $\mathbf{X}$  be  $\phi(\mathbf{Z})$  and let  $\mathbf{Y}$  be the output process with  $\mathbf{X}$  as the input to the channel

$$\begin{aligned} P_{Y^n}(B) &= \sum_{i=1}^M P(B|b_i) P_{X^n}(b_i) \\ &\geq \sum_{i=1}^N P(D_i|b_i) P_{X^n}(b_i) \\ &\geq (1 - \epsilon_n) P_{Z^n}(S^n) \\ &\geq (1 - \epsilon_n) \alpha. \end{aligned}$$

Thus we have

$$P_{\tilde{Y}^n}(B) - P_{Y^n}(B) \geq (1 - \epsilon_n) \alpha - 2\epsilon_n$$

for all sufficiently large  $n \in I$ . We note that the right-hand side of the equation above does not go to zero as  $n$  increases and so  $\mathbf{Z}$  cannot be an approximating source in the strong sense for the channel  $\mathbf{W}$ .  $\square$

## V. A SPECIAL CASE

We will now look at the special case of a full-rank, discrete, memoryless channel (FRDMC). A channel  $\mathbf{W}$  is of full rank if the transition vectors  $\{W(\cdot|a)\}_{a \in A}$  are linearly independent. For the FRDMC we will be able to show a result analogous to [5], that we would like to show for general channels.

*Theorem 5:* For a FRDMC, if the source  $\mathbf{Z}$  is an approximating source for the channel  $\mathbf{W}$  then the source  $\mathbf{Z}$  dominates the channel  $\mathbf{W}$ .

*Proof:* We will show that if the source  $\mathbf{Z}$  does not dominate the FRDMC  $\mathbf{W}$  then  $\mathbf{Z}$  is not an approximating source for the channel  $\mathbf{W}$ .

At the outset we mention why we need the assumption of full rank. We show that if source  $\mathbf{Z}$  does not dominate the channel  $\mathbf{W}$ , the source cannot approximate the output due to a particular independent and identically distributed (i.i.d.) input  $X$ . The reason we need the full rank assumption is that to approximate the output of an i.i.d. input to an FRDMC it is necessary to place mass on the typical sequences of  $X$ . This is not true if the channel is not full-rank.

If a source  $\mathbf{Z}$  does not dominate  $\mathbf{W}$  then there is a process  $\bar{\mathbf{X}}$  such that for some  $\alpha, \delta > 0$  and for some sequence of nonnegative numbers  $\{c_n\}_{n=1}^\infty$

$$P \left[ \frac{1}{n} h_{Z^n}(Z^n) \leq c_n - \delta \right] > \alpha \quad (5)$$

and

$$P \left[ \frac{1}{n} i_{\bar{X}^n W^n}(\bar{X}^n; \bar{Y}^n) \geq c_n \right] > \alpha \quad (6)$$

for all  $n \in I$  where  $I$  is an infinite set of integers. From [4, Corollary 2] we have that for an FRDMC  $S = C$ . But for a DMC,  $C = \sup_X I(X; Y)$ . Since  $S = \sup_X \bar{I}(\mathbf{X}; \mathbf{Y})$  we have for a FRDMC

$$\sup_{\mathbf{X}} \bar{I}(\mathbf{X}; \mathbf{Y}) = \sup_X I(X; Y).$$

Thus for an FRDMC if there is a process  $\bar{\mathbf{X}}$  such that (5) and (6) are satisfied then there exists an i.i.d. process  $\mathbf{X}$  such that

$$c_n \leq I(X; Y) + \frac{\delta}{10} \quad (7)$$

for sufficiently large  $n \in I$ . We will show that the source  $\mathbf{Z}$  cannot approximate the output due to process  $\mathbf{X}$ .

Define the set of sequences in  $A^n$  that are not  $\gamma$ -typical ( $\gamma > 0$ ) by

$$D_{X^n}(\gamma) = \left\{ x^n : \left| \frac{1}{n} N(a|x^n) - P_X(a) \right| > \gamma \text{ for some } a \in A \right\}$$

where  $N(a|x^n)$  denotes the number of times that  $a$  occurs in the sequence  $x^n$ .

Also define the set of sequences jointly  $\gamma$ -typical with  $x^n$  by

$$T_W^n(x^n, \gamma) = \left\{ y^n : \left| \frac{1}{n} N(a, b|x^n, y^n) - \frac{1}{n} N(a|x^n) W(b|a) \right| \leq \gamma \text{ for all } (a, b) \in A \times B \right\}$$

where  $N(a, b|x^n, y^n)$  denotes the number of times that  $(a, b)$  occurs in  $(x^n, y^n)$ .

Let  $W$  denote the matrix whose columns are the transition vectors  $\{W(\cdot|a)\}_{a \in A}$ . Let  $\mathbf{r}$  be an  $|A|$ -dimensional vector and let  $\mathbf{s} = W\mathbf{r}$ . Since  $W$  is of full rank if  $|r_i| > \gamma$  for some  $i \in \{1, 2, \dots, |A|\}$  then  $|s_j| > k\gamma$  for some  $j \in \{1, 2, \dots, |B|\}$  and some  $k > 0$  independent of  $\mathbf{r}$ . Thus we have that if  $x^n \in D_X^n(\gamma)$  then  $W\mathbf{N}(x^n)/n$  differs from  $W\mathbf{P}_X$  by  $k\gamma$  in at least one component, where  $\mathbf{N}(x^n)$ ,  $\mathbf{P}_X$  are column vectors consisting of  $N(a|x^n)$ ,  $P_X(a)$ , respectively, for  $a \in A$ . Now if  $y^n \in T_W^n(x^n, \gamma)$  then

$$\left| \frac{1}{n} N(b|y^n) - \sum_{a \in A} \frac{1}{n} N(a|x^n) W(b|a) \right| \leq |A|\gamma$$

for all  $b \in B$ . Thus we have that if

$$x^n \in D_X^n \left( \frac{\gamma}{k} \right) \quad \text{and} \quad y^n \in T_W^n \left( x^n, \frac{\gamma}{3|A|} \right)$$

then  $\mathbf{N}(y^n)/n$  differs from  $\mathbf{P}_Y$  in some component by at least  $\frac{2\gamma}{3}$ . Thus  $y^n \in D_Y^n \left( \frac{2\gamma}{3} \right)$ . We have shown that if  $x^n \in D_X^n \left( \frac{\gamma}{k} \right)$  then

$$T_W^n \left( x^n, \frac{\gamma}{3|A|} \right) \subseteq D_Y^n \left( \frac{2\gamma}{3} \right).$$

But

$$P_{Y^n} \left( D_Y^n \left( \frac{2\gamma}{3} \right) \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty$$

and

$$W^n \left( T_W^n \left( x^n, \frac{\gamma}{3|A|} \right) | x^n \right) \rightarrow 1, \quad \text{as } n \rightarrow \infty$$

at a rate independent of  $x^n$  (from [1, Lemma 2.12]). So if the input  $\phi^n(Z^n)$  approximates the output due to  $X^n$  then we must have

$$P_{Z^n} \left( \phi^n(Z^n) \in D_X^n \left( \frac{\gamma}{k} \right) \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty$$

since if this were not so

$$P_{\phi^n(Z^n)} W^n \left( D_Y^n \left( \frac{2\gamma}{3} \right) \right) - P_{Y^n} \left( D_Y^n \left( \frac{2\gamma}{3} \right) \right) > \beta$$

infinitely often, for some  $\beta > 0$ . This would imply that the input  $\phi^n(Z^n)$  does not approximate the output due to  $X^n$ .

Now by a slight modification of [1, Lemma 2.13]

$$\left| \frac{1}{n} \log |T_W^n(x^n, \gamma/3|A|)| - H(Y|X) \right| \leq \epsilon(\gamma)$$

for every  $x^n \in D_{X^n}^c \left( \frac{\gamma}{k} \right)$  where  $\epsilon(\gamma)$  is continuous in  $\gamma$ , independent of  $x^n$  and  $\epsilon(\gamma) \rightarrow 0$  as  $\gamma \rightarrow 0$ . Define

$$Q^n = \left\{ z^n : \frac{1}{n} h_{Z^n}(z^n) \leq c_n - \delta \right\}.$$

Clearly,  $|Q^n| \leq \exp(n(c_n - \delta))$ . Also define  $R^n$  as the image of  $Q^n$  under the mapping  $\phi^n$  and

$$S^n = \bigcup_{x^n \in R^n \cap D_{X^n}^c \left( \frac{\gamma}{k} \right)} T_W^n \left( x^n, \frac{\gamma}{3|A|} \right).$$

Then we have

$$|S^n| \leq \exp n(c_n - \delta) \exp n(H(Y|X) + \epsilon(\gamma))$$

and  $P_{\phi^n(Z^n)} W^n(S^n) \geq \frac{\alpha}{2}$  for sufficiently large  $n \in I$ . This is because

$$W^n \left( T_W^n \left( x^n, \frac{\gamma}{3|A|} \right) | x^n \right) \rightarrow 1$$

as  $n \rightarrow \infty$  at a rate independent of  $x^n$

$$P_{Z^n} \left[ \phi^n(Z^n) \in D_{X^n}^c \left( \frac{\gamma}{k} \right) \right] \rightarrow 1$$

as  $n \rightarrow \infty$  and

$$P_{Z^n}[\phi^n(Z^n) \in R^n] > \alpha$$

for all  $n \in I$ . Using (7) we can upper-bound  $|S^n|$  by

$$|S^n| \leq \exp n \left( H(Y) - \frac{9\delta}{10} + \epsilon(\gamma) \right) \leq \exp n \left( H(Y) - \frac{\delta}{2} \right)$$

for sufficiently large  $n \in I$ , where the second inequality holds if we choose  $\gamma > 0$  sufficiently small. We have  $P_{Y^n}(S^n) \rightarrow 0$  as  $n \rightarrow \infty$  (by the strong source coding theorem for i.i.d. sources). For any deterministic mapping  $\phi^n$  we can find a set  $S^n \subseteq B^n$  such that  $P_{\phi^n(Z^n)} W^n(S^n) \geq \frac{\alpha}{2}$  for sufficiently large  $n \in I$  and  $P_{Y^n}(S^n) \rightarrow$

0 as  $n \rightarrow \infty$ . Hence,  $Z$  cannot be an approximating source for the FRDMC  $W$ .  $\square$

We note that Theorem 5 implies that if  $H(Z) < S$  then the source  $Z$  is not an approximating source for a FRDMC channel with resolvability  $S$ .

REFERENCES

- [1] I. Csiszár and Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [2] A. Feinstein, "A new basic theorem in information theory," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 2-22, Jan. 1954.
- [3] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, pp. 752-772, May 1993.
- [4] —, "Spectrum invariancy under output approximation for full rank discrete memoryless channels" (in Russian), *Probl. Pered. Inform.*, no. 2, pp. 101-118, 1993.
- [5] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, pp. 44-54, Jan. 1995.
- [6] S. Vembu and S. Verdú, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1322-1332, Sept. 1995.