REFERENCES

[1] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.
[2] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.
[3] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.
[4] T. Berger and Z. Zhang, "Minimum breakdown degradation in binary source encoding," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 807–814, Nov. 1983.
[5] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 502–521, July 1987.
[6] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 721–726, Nov. 1985.
[7] H. S. Witsenhausen and A. D. Wyner, "Source coding for multiple descriptions II: A binary source," *Bell Syst. Tech. J.*, vol. 60, pp. 2281–2292, Dec. 1981.
[8] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, pp. 1417–1426, Oct. 1980.
[9] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.
[10] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 40, pp. 245–250, Jan. 1994.
[11] V. Vaishampayan, "Vector quantizer design for diversity systems," in *Proc. 25th Annu. Conf. on Information Sciences and Systems* (Baltimore, MD, Johns Hopkins Univ., Mar. 20–22, 1991), pp. 564–569.
[12] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 48–58, Jan. 1995.
[13] J.-C. Batllo, "Multiple description transform codes with an application to packetized speech," Master's thesis, Elec. Eng. Dep., Texas A&M Univ., College Station, TX, May 1994.
[14] V. Vaishampayan and A. A. Siddiqui, "Speech predictor design for diversity communication systems," in *Proc. 1995 IEEE Speech Coding Workshop*, Sept. 1995, pp. 20–22.
[15] S.-M. Yang and V. A. Vaishampayan, "Low delay communication for Rayleigh fading channels: An application of the multiple description quantizer," *IEEE Trans. Commun.*, vol. 43, pp. 2771–2783, Nov. 1995.
[16] K. Kintzley, "An application of multiple description scalar quantizers to speech coding on correlated fading channels," Master's thesis, Elec. Eng. Dep., Texas A&M Univ., College Station, TX, Aug. 1995.
[17] V. A. Vaishampayan, "Application of multiple description codes to image and video transmission over lossy networks," in *Proc. 7th Int. Workshop on Packet Video* (Brisbane, Australia, Mar. 1996), pp. 55–60.
[18] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.
[19] V. R. Algazi, "Useful approximations to optimal quantization," *IEEE Trans. Commun. Technol.*, vol. COM-14, pp. 297–301, June 1966.
[20] P. F. Panter and W. Dite, "Quantization in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, vol. 39, pp. 44–48, Jan. 1951.
[21] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, 1968.
[22] J. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with rth power distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 239–247, Mar. 1982.
[23] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.
[24] R. M. Gray and A. H. Gray, "Asymptotically optimal quantziers," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 143–144, Jan. 1977.
[25] H. L. Royden, *Real Analysis*. New York: Macmillan, 1968.

Identification Plus Transmission Over Channels with Perfect Feedback

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

***Abstract*—We determine the region of all identification and transmission rate-pairs achievable over a discrete memoryless channel with perfect and instantaneous feedback, for both randomized and deterministic encoding. As a by-product, we also have a new proof of Kemperman's strong converse to Shannon's coding theorem for DMC's with feedback.**

***Index Terms*—Channels with feedback, identification plus transmission coding, Kemperman's strong converse.**

I. INTRODUCTION

The following problem, called *identification*, was introduced by Ahlswede and Dueck [2]: suppose there are N events, any one of which may occur. The actual outcome is known to an agent (the *transmitter*) at the transmitting end of a discrete memoryless channel (DMC) with transition probability matrix W. The output of this channel can be observed by N agents (the *receivers*) who are interested in the outcome. However, receiver a $(1 \leq a \leq N)$ only wishes to know whether or not event a occurred, and not *which* event actually occurred. To this end, the transmitter must send a codeword bearing information about the outcome across the channel. Based on the channel output, each receiver must decide whether or not the event of interest to it occurred. It is required for each a that if event a is the actual outcome then i) receiver a should decide with high probability that event a occurred; and ii) for any $a' \neq a$, receiver a' should decide with high probability that event a' did not occur. The question of interest, then, is how fast N can grow with n, the number of uses of the channel permitted to the transmitter.

The surprising result of [2] is that N can actually grow as $\exp\{\exp[nR - o(n)]\}$ (*doubly* exponential in n), provided that the transmitter can use randomized encoding. Further, the identification capacity of the channel, defined as the maximum achievable (second-order) rate R in this situation, equals its Shannon capacity $C = \max_P I(P; W)$—the maximum mutual information between channel input and output, over all input distributions P. (Throughout this correspondence, all logarithms and exponentials will be to the base e.) Randomization is crucial here, in the sense that it is impossible to achieve any positive second-order rate R with deterministic encoding, i.e., the deterministic identification capacity is zero.

In [5], Han and Verdú studied a variant of the identification problem, called *identification plus transmission*. The situation here is as before, but now the transmitter also receives as input one of M equiprobable messages. When event a occurs, the transmitter is required to convey this message to receiver a, by sending a codeword

Manuscript received January 31, 1997; revised June 23, 1997. This work was supported by the NSF under Grants IRI 9005849, IRI 9310670, and NCR 9422513, and by the AT&T Foundation. The material in this correspondence was presented in part at the 34th Annual Allerton Conference on Communications and Control, Allerton, IL, October 2–4, 1996.

S. Venkatesan was with the Department of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA.

V. Anantharam is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA.

Publisher Item Identifier S 0018-9448(98)00071-6.

across the channel—thus the index a denotes the "address" to which the message must be sent. It is now required that if event a is the actual outcome then i) receiver a, the intended recipient of the message, should decide that event a occurred *and* decode the message correctly with high probability; and ii) for any $a' \neq a$, receiver a' should recognize with high probability that the message is not intended for it, i.e., that event a' did not occur. These probabilities are assumed to be averaged over M possible messages. The question of interest, then, is how fast M and N can simultaneously grow with the number of channel uses permitted to the transmitter.

The straightforward "time-sharing" solution to this problem would be to encode the message in $(1-\alpha)n$ symbols using a classical data transmission code, and encode the "address" in a header of αn symbols using an identification code (n is the number of channel uses). This scheme would permit

$$M = \exp[n(1-\alpha)C - o(n)]$$

and

$$N = \exp\{\exp[n\alpha C - o(n)]\}$$

and would require randomization for the address encoding. However, it is possible to do better. In [5], it was shown that in fact

$$M = \exp[nC - o(n)]$$

and

$$N = \exp\{\exp[nC - o(n)]\}$$

are *simultaneously* achievable, if the address and message are *jointly* encoded using an "identification plus transmission" code, instead of separately as above; moreover, the joint encoding does not require randomization (essentially because the message itself provides enough randomness for the address encoding).

In this correspondence, we study an analogous identification plus transmission problem when the transmitter has *perfect and instantaneous feedback* from the output of the DMC connecting it to the receivers. It is well known that feedback does not increase the data transmission capacity of a DMC. In marked contrast, feedback can have a dramatic effect on its identification capacity. In [1], the sequel to [2], it was shown that feedback increases the identification capacity of a DMC W with positive Shannon capacity to $\max_P H(PW)$—the maximum unconditional output entropy, over all input distributions P—when randomized encoding is allowed. It was also shown that feedback increases the deterministic identification capacity from zero to $\max_x H[W(\cdot|x)]$—the maximum, over all input symbols x, of the conditional output entropy when x is transmitted.

Here, we determine the region of all rate-pairs (R_1, R_2) such that the transmitter can reliably send one of $M = \exp[nR_1 - o(n)]$ messages to one of $N = \exp\{\exp[nR_2 - o(n)]\}$ receivers across the DMC W, in the presence of feedback. As in [1], we consider both the case where randomized encoding is allowed, and the case where the encoding must be deterministic. The problem is formulated more precisely in Section II, and the results are stated in Theorem 2.1. The identification theorems of [1] can be viewed as special cases of these results, obtained by setting the transmission rate requirement R_1 to zero. As in [1], the converses proved here are "strong." As a by-product of these converses, we also have a new proof of the strong converse to Shannon's coding theorem for DMC's with feedback, a result first proved by Kemperman [6].

II. Statement of Problem and Results

The DMC connecting the transmitter and the receivers is assumed to have finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and transition probability function $W = \{W(y|x)\colon x \in \mathcal{X}, y \in \mathcal{Y}\}$. In the presence of feedback, the transmitter can adapt its channel input at each step of communication to the outputs from all previous steps. Accordingly, we define an *n-step feedback function* as a vector $f = (f_1, \cdots, f_n)$, where $f_k\colon \mathcal{Y}^{k-1} \to \mathcal{X}$. When the transmitter uses this feedback function for communication, it sends the symbol $X_k = f_k(Y^{k-1})$ in step k, $1 \leq k \leq n$, where Y^{k-1} is the sequence of channel outputs in the first $k-1$ steps. The corresponding probability that $Y^n = y^n$, denoted $Q_f(y^n)$, equals $W^n(y^n \mid f(y^n))$, where $f(y^n)$ is the $x^n \in \mathcal{X}^n$ given by $x_k = f_k(y^{k-1}), 1 \leq k \leq n$. Clearly, the probability that $(X^n, Y^n) = (x^n, y^n)$, denoted $W_f(x^n, y^n)$, equals $Q_f(y^n)$ if $x^n = f(y^n)$, and 0 otherwise.

An *n-step feedback strategy* F for the channel W is defined as a probability distribution (p.d.) on $\mathcal{F}_n$, the set of all n-step feedback functions. To communicate according to the strategy F, the transmitter randomly chooses $f \in \mathcal{F}_n$ with distribution F, and then uses it as described above, to decide its channel input at each step. The probability that $(X^n, Y^n) = (x^n, y^n)$ under the strategy F, denoted $W_F(x^n, y^n)$, is then

$$\sum_{f \in \mathcal{F}_n} F(f) W_f(x^n, y^n)$$

and the corresponding marginal probability that $Y^n = y^n$, denoted $Q_F(y^n)$, is

$$\sum_{f \in \mathcal{F}_n} F(f) Q_f(y^n).$$

A general strategy, as defined above, is allowed to use randomization. The strategy F is called *deterministic* if $F(f) = 1$ for some $f \in \mathcal{F}_n$. Clearly, such a strategy does not require any randomization.

We will now define the identification plus transmission codes to be studied here. In the definition below, and in the rest of the paper, $[J] \stackrel{\text{def}}{=} \{1, 2, \cdots, J\}$ for any positive integer J.

Definition 2.1: An (n, N, M, λ, μ) identification plus transmission (IT) code is a collection

$$\{(F_{a,m}, \mathcal{D}_{a,m})\colon (a, m) \in [N] \times [M]\}$$

where $F_{a,m}$ is an n-step strategy, $\mathcal{D}_{a,m} \subseteq \mathcal{Y}^n$, and for each $a \in [N]$:

1) $\mathcal{D}_{a,m} \cap \mathcal{D}_{a,m'}$ is empty if $m \neq m'$;
2) $(1/M)\sum_m Q_{F_{a,m}}(\mathcal{D}_{a,m}) > 1 - \lambda$;
3) $(1/M)\sum_m Q_{F_{a,m}}(\mathcal{D}_{a'}) < \mu$ for all $a' \neq a$, where $\mathcal{D}_{a'} = \bigcup_m \mathcal{D}_{a',m}$.

The code is called deterministic if all the strategies $F_{a,m}$ are deterministic.

The interpretation of the above code is as follows: if the transmitter wishes to send message m to receiver a, it communicates according to the strategy $F_{a,m}$. After n steps, receiver a decides that it is indeed the intended recipient if the received sequence Y^n falls in $\mathcal{D}_a = \bigcup_m \mathcal{D}_{a,m}$; in this case, because of Condition 1) in the definition, there is a unique m such that $Y^n \in \mathcal{D}_{a,m}$, and it takes this m as the transmitted message. Otherwise, i.e., if $Y^n \notin \mathcal{D}_a$, receiver a decides that the message is not intended for it. Condition 2) guarantees that the intended recipient decodes the transmitted message correctly with probability greater than $1-\lambda$, while Condition 3) guarantees that any other receiver wrongly decides it is the recipient with probability less than μ. Note that these probabilities are averaged over the M possible messages.

The rate-pair (R_1, R_2) will be called *(λ, μ)-achievable* if there exists a sequence of $(n, N_n, M_n, \lambda, \mu)$ IT codes such that

$$\liminf n^{-1} \log M_n = R_1$$

and

$$\liminf n^{-1} \log \log N_n = R_2.$$

If this sequence can be chosen to be deterministic, then (R_1, R_2) will be called *deterministically* (λ, μ)*-achievable*.

Theorem 2.1. Main Theorem: Assume that the discrete memoryless channel W has positive Shannon capacity $C = \max_P I(P; W)$. Then, for any (λ, μ) satisfying $\lambda > 0$, $\mu > 0$, and $\lambda + \mu < 1$, the rate-pair (R_1, R_2) is (λ, μ)-achievable if and only if

$$R_1 \le C \quad \text{and} \quad R_2 \le \max_{P:\, I(P;W) \ge R_1} H(PW)$$

and is deterministically (λ, μ)-achievable if and only if

$$R_1 \le C \quad \text{and} \quad R_2 \le R_1 + \max_{P:\, I(P;W) \ge R_1} H(W|P).$$

The assumptions $\lambda > 0$ and $\mu > 0$ are, of course, reasonable. We also need the assumption $\lambda + \mu < 1$ in order to get meaningful results; if $\lambda + \mu > 1$, it can be shown that arbitrarily high identification rates R_2 are achievable. The assumption that $C > 0$ precludes trivialities of the opposite kind; if $C = 0$, it can be shown that there does not exist any (n, N, M, λ, μ) IT code with $N > 1$ or $M > (1 - \lambda)^{-1}$ (assuming $\lambda > 0$, $\mu > 0$, and $\lambda + \mu < 1$).

Note that the statement "(R_1, R_2) is deterministically achievable" does not imply "(R_1', R_2) is deterministically achievable if $R_1' < R_1$!" The reason is that, with deterministic IT codes, the transmission rate determines the amount of randomization available for identification coding. There is no such pathology in the result for general IT codes.

We will prove the achievability parts of Theorem 2.1 in Section III, and the converse parts in Section IV. The achievability results can actually be proved with Condition 2) in Definition 2.1 replaced by the stronger condition "$Q_{F_{a,m}}(\mathcal{D}_{a,m}) > 1 - \lambda$ for all m." However, it is *not* possible similarly to replace Condition 3) by "$Q_{F_{a,m}}(\mathcal{D}_{a'}) < \mu$ for all m and all $a' \ne a$" without affecting the results. Averaging over messages is essential in controlling the probability of a receiver wrongly deciding that it is the intended recipient.

III. Proofs of the Achievability Parts

The achievability proofs are based on those of [1]. We will first consider the general case in Section III-A. With very minor changes, the same arguments will work in the deterministic case as well (see Section III-B). We will need the following two lemmas.

Lemma 3.1: Let P be an n-type on $\mathcal{X}$ for some $n \ge 1$ (i.e., P is a p.d. on $\mathcal{X}$ such that $nP(x)$ is an integer for all $x \in \mathcal{X}$), and let $d = |\mathcal{X}|\,|\mathcal{Y}|$.

1) If $R' \ge 0$ and $J \le \exp(nR')$, then there exist sequences $\boldsymbol{c}_1, \cdots, \boldsymbol{c}_J$ in $\mathcal{X}^n$, all of type P, and a partition of $\mathcal{Y}^n$ into subsets $\mathcal{C}_1, \cdots, \mathcal{C}_J$, such that

$$1 - W^n(\mathcal{C}_j|\boldsymbol{c}_j) \le (n+1)^{4d} \exp[-nE'(R', P)]$$

for all $j \in [J]$. Here, $E'(R', P)$ is a continuous function of (R', P) which is positive if $R' < I(P; W)$ and zero otherwise.

2) If $R'' \ge 0$ and $K \le \exp(nR'')$, then, for any $\boldsymbol{c} \in \mathcal{X}^n$ of type P and any $\mathcal{C} \subseteq \mathcal{Y}^n$, there exists a partition of $\mathcal{C}$ into subsets $\mathcal{C}_e, \mathcal{C}_1, \cdots, \mathcal{C}_K$, such that

$$W^n(\mathcal{C}_k|\boldsymbol{c}) = (1/K)[W^n(\mathcal{C}|\boldsymbol{c}) - W^n(\mathcal{C}_e|\boldsymbol{c})]$$

for all $k \in [K]$, and

$$W^n(\mathcal{C}_e|\boldsymbol{c}) \le (n+1)^d \exp[-nE''(R'', P)].$$

Here, $E''(R'', P)$ is a continuous function of (R'', P) which is positive if $R'' < H(W|P)$ and zero otherwise.

Proof: See the Appendix. □

The first part of Lemma 3.1 is a standard result in channel coding. It implies the existence of an "equitype" transmission code for the DMC W (*without* feedback) of blocklength n with about $\exp[nI(P; W)]$ codewords—P is the common type of the codewords—whose maximal error probability over W decays exponentially with n. By the second part of Lemma 3.1, the decoding set corresponding to each codeword in such a code can be partitioned into about $\exp[nH(W|P)]$ sets—all of which have *exactly* the same probability—and a remaining "error" set whose probability decays exponentially with n.

Lemma 3.2: Let $0 < \epsilon < 1$, and let $J \ge 2/\epsilon$, $S \ge 1$, and $N < \exp(\epsilon^2 S/2)$ be integers. Then, there exists an $N \times S$ array (N rows and S columns) with entries from $[J]$, any two rows of which are at a Hamming distance greater than $(1 - \epsilon)S$ from each other.

Proof: See the Appendix. □

Lemma 3.2 is based on the arguments in [1, Sec. III], though it is stated there differently. It is the essence of the "$\sqrt{n}$ trick" of [1], which can be used to prove all known achievability results in identification theory. (In the absence of feedback, the original "Gilbert bound" approach of [2] may be simpler.) In its present form, the name "array trick" may be more appropriate.

A. The General Case

Fix any $\delta > 0$ and $\epsilon > 0$. It suffices to prove that if

$$R_1 < C - \delta \quad \text{and} \quad R_2 < \max_{P:\, I(P;W) \ge R_1 + \delta} H(PW) - \delta$$

then (R_1, R_2) is $(4\epsilon, 4\epsilon)$-achievable. So let $M_n = \lfloor \exp(nR_1) \rfloor$ and $N_n = \lfloor \exp[\exp(nR_2)] \rfloor$. We will show that for all large n there exists an $(n + t, N_n, M_n, 4\epsilon, 4\epsilon)$ IT code, where t does not depend on n.

Let P^* be a p.d. on $\mathcal{X}$ that maximizes $H(PW)$ subject to $I(P; W) \ge R_1 + \delta$. Let

$$R' = I(P^*; W) - \delta/2$$

and

$$R'' = H(W|P^*) - \delta/2.$$

Pick any sequence $\{P_n\}$, with P_n an n-type on $\mathcal{X}$, such that $P_n \to P^*$ as $n \to \infty$.

Let $L_n = \lfloor \exp[n(R' - R_1)] \rfloor$. Then, $M_n L_n \le \exp(nR')$, and the first part of Lemma 3.1 guarantees the existence of sequences $\boldsymbol{c}_{ml} \in \mathcal{X}^n$, all of type P_n, and sets $\mathcal{C}_{ml}$ partitioning $\mathcal{Y}^n$ and satisfying $1 - W^n(\mathcal{C}_{ml}|\boldsymbol{c}_{ml}) \le \alpha_n$ for all $(m, l) \in [M_n] \times [L_n]$. Here

$$\alpha_n = \exp[-nE'(R', P_n) + o(n)] < \epsilon$$

for all large n, because

$$E'(R', P_n) \to E'(R', P^*) > 0.$$

Let $K_n = \lfloor \exp(nR'') \rfloor$. Then, by the second part of Lemma 3.1, each set $\mathcal{C}_{ml}$ can be partitioned further into subsets $\mathcal{C}_{mlk}$, $k \in \{e\} \cup [K_n]$, such that $W^n(\mathcal{C}_{mlk}|\boldsymbol{c}_{ml})$ is the same for all $k \in [K_n]$, and $W^n(\mathcal{C}_{mle}|\boldsymbol{c}_{ml}) \le \beta_n$. Here

$$\beta_n = \exp[-nE''(R'', P_n) + o(n)] < \epsilon$$

for all large n, because

$$E''(R'', P_n) \to E''(R'', P^*) > 0.$$

Note that

$$\lim n^{-1} \log(M_n L_n K_n) = H(P^*W) - \delta > R_2$$

by assumption. Hence, for all large n, $\exp(nR_2) < (\epsilon^2/2) M_n L_n K_n$ and $N_n < \exp[(\epsilon^2/2) M_n L_n K_n]$. By Lemma 3.2, then, there exists an $N_n \times (M_n L_n K_n)$ array with entries from $[J]$, any two rows of which are at a Hamming distance greater than $(1 - \epsilon) M_n L_n K_n$, if

n is large. Here, we may take $J = \lceil 2/\epsilon \rceil$. We will denote this array by $\mathcal{A}$, and think of its rows as being indexed by the N_n receivers. Its columns will be indexed by triples $(m, l, k) \in [M_n] \times [L_n] \times [K_n]$. $\mathcal{A}(a; m, l, k)$ will denote the array element in row a and column (m, l, k).

Finally, pick an integer t large enough that there exist sequences $\tilde{\boldsymbol{c}}_1, \cdots, \tilde{\boldsymbol{c}}_J$ in $\mathcal{X}^t$ and a partition $\tilde{\mathcal{C}}_1, \cdots, \tilde{\mathcal{C}}_J$ of $\mathcal{Y}^t$, satisfying $1 - W^t(\tilde{\mathcal{C}}_j|\tilde{\boldsymbol{c}}_j) < \epsilon$ for all $j \in [J]$. This is possible because the channel has positive Shannon capacity.

The transmitter encodes any address–message pair $(a, m) \in [N_n] \times [M_n]$ in two stages. In the first stage, it picks a random $l \in [L_n]$ with a uniform distribution, and sends the sequence $\boldsymbol{c}_{ml}$ across the channel. There is then a unique triple

$$(\hat{m}, \hat{l}, k) \in [M_n] \times [L_n] \times (\{e\} \cup [K_n])$$

such that the corresponding channel output sequence lies in $\mathcal{C}_{\hat{m}\hat{l}k}$. This triple, which is known to the transmitter also because of feedback, identifies a column of the array $\mathcal{A}$ if $k \in [K_n]$ (i.e., if $k \neq e$).

If $k \in [K_n]$, then the transmitter sends the sequence $\tilde{\boldsymbol{c}}_j$ in the second stage, where $j = \mathcal{A}(a; m, l, k)$ is the integer in row a and column (m, l, k) of the array $\mathcal{A}$; correspondingly, there is a unique $\hat{\jmath} \in [J]$ such that the output sequence falls in $\tilde{\mathcal{C}}_{\hat{\jmath}}$. On the other hand, if $k = e$, the transmitter and receivers declare an error, and the transmitter sends a dummy sequence of length t, say $\tilde{\boldsymbol{c}}_e$, in the second stage. We have thus implicitly defined $(n + t)$-step strategies $F_{a,m}$ for each $(a, m) \in [N_n] \times [M_n]$.

At the end of $n + t$ steps, receiver a' simply checks if $\mathcal{A}(a'; \hat{m}, \hat{l}, k) = \hat{\jmath}$. If so, it assumes that it is indeed the intended recipient, and that the transmitted message is $\hat{m}$. Otherwise, i.e., if $\mathcal{A}(a'; \hat{m}, \hat{l}, k) \neq \hat{\jmath}$, it decides that the message is not intended for it. Formally, this means that the decoding region $\mathcal{D}_{a,m} \subset \mathcal{Y}^{n+t}$ equals $\bigcup_{(l,k,j)} \mathcal{C}_{mlk} \times \tilde{\mathcal{C}}_j$, the union extending over all those triples (l, k, j) for which $\mathcal{A}(a; m, l, k) = j$.

We will now bound the error probabilities of this IT code. Suppose the transmitter attempts to convey message m to receiver a. If $(\hat{m}, \hat{l}) = (m, l)$, $k \neq e$, and $\hat{\jmath} = j$, then receiver a will recognize that the message is intended for it, and decode it correctly as m. Therefore, by a union bound

$$1 - Q_{F_{a,m}}(\mathcal{D}_{a,m}) \leq \alpha_n + \beta_n + \epsilon < 3\epsilon$$

for all large n.

Now, consider any receiver $a' \neq a$. If $(\hat{m}, \hat{l}) = (m, l)$, $k \neq e$, $\hat{\jmath} = j$, *and* $\mathcal{A}(a'; m, l, k) \neq \mathcal{A}(a; m, l, k)$, then receiver a' will correctly recognize that it is not the intended recipient. As before, the probability that either $(\hat{m}, \hat{l}) \neq (m, l)$, or $k = e$, or $\hat{\jmath} \neq j$ is at most $\alpha_n + \beta_n + \epsilon$. Further, given that $(\hat{m}, \hat{l}) = (m, l)$, $k \neq e$, and $\hat{\jmath} = j$, the probability that $\mathcal{A}(a'; m, l, k) = \mathcal{A}(a; m, l, k)$ is equal to $E_m(a, a')/L_n K_n$, where $E_m(a, a')$ is the number of pairs (l, k) such that $\mathcal{A}(a'; m, l, k) = \mathcal{A}(a; m, l, k)$. This is because l is chosen with a uniform distribution over $[L_n]$, and, conditional on $k \neq e$, k has a uniform distribution on $[K_n]$ for all values of (m, l). Thus

$$Q_{F_{a,m}}(\mathcal{D}_{a'}) \leq \alpha_n + \beta_n + \epsilon + \frac{E_m(a, a')}{L_n K_n}.$$

But then

$$\begin{aligned} &\frac{1}{M_n} \sum_{m=1}^{M_n} Q_{F_{a,m}}(\mathcal{D}_{a'}) \\ &\quad \leq \alpha_n + \beta_n + \epsilon + \frac{1}{M_n} \sum_{m=1}^{M_n} \frac{E_m(a, a')}{L_n K_n} \\ &\quad < 4\epsilon, \qquad \text{for all large } n \end{aligned}$$

because $\sum_m E_m(a, a')$ is just the number of positions in which rows a and a' of the array $\mathcal{A}$ agree, and this number is smaller than $\epsilon(M_n L_n K_n)$ for all large n. This completes the proof.

B. The Deterministic Case

It suffices to prove that if $R_1 < C - \delta$ and

$$R_2 < R_1 + \max_{P:\, I(P;W) \geq R_1 + \delta} H(W|P) - \delta$$

for some $\delta > 0$, then for any $\epsilon > 0$ and all large n there exists an $(n + t, N_n, M_n, 4\epsilon, 4\epsilon)$ deterministic IT code, with $M_n = \lfloor \exp(nR_1) \rfloor$ and $N_n = \lfloor \exp[\exp(nR_2)] \rfloor$ (t being a constant, as before). We have a proof of the existence of such codes if we simply change three sentences in the proof for the general case, starting from Section III-A. These are the first sentences of paragraphs 2, 3, and 5 of Section III-A. The first of these must be changed to "Let P^* be a p.d. on $\mathcal{X}$ that maximizes $H(W|P)$ subject to $I(P; W) \geq R_1 + \delta$;" the second to "Let $L_n = 1$;" and the third to "Note that

$$\lim n^{-1} \log(M_n L_n K_n) = R_1 + H(W|P^*) - \delta/2 > R_2$$

by assumption." But for these changes, the proof in the general case carries over word-for-word. The resulting sequence of IT codes is indeed deterministic because $L_n = 1$ here; an inspection of the previous proof shows that the transmitter needs randomization *only* to generate a random $l \in [L_n]$.

IV. Proofs of the Converse Parts

Consider any sequence of $(n, N_n, M_n, \lambda, \mu)$ IT codes $\{(F_{a,m}, \mathcal{D}_{a,m})\}$ achieving the rate-pair (R_1, R_2), with $\lambda > 0$, $\mu > 0$, and $\lambda + \mu < 1$. We will now outline the ideas for bounding R_1 and R_2.

To begin with, $\{(F_{1,m}, \mathcal{D}_{1,m})\colon m \in [M_n]\}$ is a sequence of (n, M_n) transmission codes with average error probability λ, for the DMC W with feedback. The encoding may involve randomization, but this does not help at all in a transmission code. Since $\lambda < 1$, Kemperman's strong converse to Shannon's theorem for DMC's with feedback [6] yields $\limsup n^{-1} \log M_n \leq C$, which implies that

$$R_1 = \liminf n^{-1} \log M_n \leq C.$$

However, it turns out that we can prove Kemperman's result here with very little additional effort, and will therefore not appeal to it directly. Our proof of this result is different from the original one.

The idea for bounding the identification rate is similar to that in [1]. Fix any $\gamma \in (\mu, 1 - \lambda)$. Suppose we could find subsets $\mathcal{D}_a^* \subseteq \mathcal{D}_a$ and a number K_n such that

$$\frac{1}{M_n} \sum_{m=1}^{M_n} Q_{F_{a,m}}(\mathcal{D}_a^*) > \gamma \quad \text{and} \quad |\mathcal{D}_a^*| \leq K_n, \qquad \text{for each } a \in [N_n].$$

Then, since

$$(1/M_n) \sum_m Q_{F_{a,m}}(\mathcal{D}_{a'}^*) < \mu < \gamma$$

for all $a' \neq a$, the sets $\mathcal{D}_a^*$ would have to be distinct. But this would imply that N_n is no bigger than the number of distinct subsets of $\mathcal{Y}^n$ whose size is at most K_n, which in turn is bounded above by $|\mathcal{Y}^n|^{K_n}$. Thus we would have

$$n^{-1} \log \log N_n \leq n^{-1} \log K_n + o(1).$$

We will prove that if n is large enough then there exist subsets $\mathcal{D}_a^* \subseteq \mathcal{D}_a$ satisfying the above conditions, with $n^{-1} \log K_n$ being

$$\max_{P:\, I(P;W) \geq R_1 - \delta} H(PW) + o(1)$$

in the general case, and

$$n^{-1}\log M_n + \max_{P:\, I(P;W)\geq R_1-\delta} H(W|P) + o(1)$$

in the deterministic case. Here, δ is an arbitrary positive number. The required bounds on the identification rate $R_2 = \liminf n^{-1}\log\log N_n$ in the two cases will then follow from the continuity in R of

$$\max_{I(P;W)\geq R} H(PW)$$

and

$$\max_{I(P;W)\geq R} H(W|P).$$

For any n-step strategy F, and

$$(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$$

let $P_k^F(x|x^{k-1}, y^{k-1})$ be the probability that $X_k = x$ given

$$(X^{k-1}, Y^{k-1}) = (x^{k-1}, y^{k-1})$$

under the strategy F. Let $P^F_{x^n,y^n}$ be the p.d. on $\mathcal{X}$ given by

$$P^F_{x^n,y^n}(x) = n^{-1}\sum_{k=1}^{n} P_k^F(x|x^{k-1}, y^{k-1}).$$

The "typical set" $\mathcal{E}(F)$ for the strategy F is then defined as the set of (x^n, y^n) such that $W_F(x^n, y^n) > 0$, and

$$|N(x, y|x^n, y^n) - nP^F_{x^n,y^n}(x)W(y|x)| \leq n^{3/4}\sqrt{W(y|x)}, \quad \text{for all } (x, y).$$

Here

$$N(x, y|x^n, y^n) = |\{k: (x_k, y_k) = (x, y)\}|.$$

Lemma 4.1: Let $d = |\mathcal{X}|\,|\mathcal{Y}|$. For any n-step strategy F

$$1 - W_F(\mathcal{E}(F)) \leq dn^{-1/2}.$$

If $(x^n, y^n) \in \mathcal{E}(F)$ and y^n has type Q, then

a) $|\log Q^n(y^n) + nH(P^F_{x^n,y^n}W)| \leq dn^{7/8}$;
b) $|\log W^n(y^n|x^n) + nH(W|P^F_{x^n,y^n})| \leq dn^{7/8}$; and
c) $|\log(W^n(y^n|x^n)/Q^n(y^n)) - nI(P^F_{x^n,y^n}; W)| \leq 2dn^{7/8}$.

Proof: See the Appendix. □

For any $\alpha \geq 0$, let

$$\mathcal{B}_\alpha(F) = \{y^n: \exists\, x^n \in \mathcal{E}_{y^n}(F) \text{ such that} I(P^F_{x^n,y^n}; W) > \alpha\}$$

where $\mathcal{E}_{y^n}(F) = \{x^n: (x^n, y^n) \in \mathcal{E}(F)\}$ is the "section" of $\mathcal{E}(F)$ at y^n.

Lemma 4.2: Let $F_1, \cdots, F_M$ be n-step strategies, and $\mathcal{D}_1, \cdots, \mathcal{D}_M$ pairwise-disjoint subsets of $\mathcal{Y}^n$. Let

$$\mathcal{B}^c_\alpha(F_m) = \mathcal{Y}^n - \mathcal{B}_\alpha(F_m).$$

Then

1) $(1/M)\sum_m Q_{F_m}(\mathcal{D}_m \cap \mathcal{B}^c_\alpha(F_m)) \leq (1/M)\exp\{n\alpha + o(n)\} + o(1)$

2) $\left|\bigcup_m \mathcal{D}_m \cap \mathcal{B}_\alpha(F_m)\right| \leq \exp\{n[\max_{P:\, I(P;W)\geq\alpha} H(PW)] + o(n)\}$

3) $\left|\bigcup_m \mathcal{D}_m \cap \mathcal{B}_\alpha(F_m)\right| \leq M\cdot\exp\{n[\max_{P:\, I(P;W)\geq\alpha} H(W|P)] + o(n)\}$

if $F_1, \cdots, F_M$ are deterministic.

Proof: See the Appendix. □

We will now return to the sequence of $(n, N_n, M_n, \lambda, \mu)$ IT codes $\{(F_{a,m}, \mathcal{D}_{a,m})\}$ at the beginning of this section, and complete the proofs of the converses. First, note that if $\alpha = C$ then $\mathcal{B}_\alpha(F_{1,m})$ is empty for all m, so that $\mathcal{D}_{1,m} = \mathcal{D}_{1,m} \cap \mathcal{B}^c_\alpha(F_{1,m})$. Therefore,

$$\begin{aligned} 1 - \lambda &< \frac{1}{M_n}\sum_{m=1}^{M_n} Q_{F_{1,m}}(\mathcal{D}_{1,m}) \\ &\leq \frac{(n+1)^d}{M_n}\exp\{nC + 2\,dn^{7/8}\} + dn^{-1/2} \end{aligned}$$

where the second inequality is by Part 1) of Lemma 4.2, applied with $\alpha = C$. Since $\lambda < 1$, the above inequalities imply that $\limsup n^{-1}\log M_n \leq C$, whence $R_1 \leq C$. We have now proved Kemperman's strong converse for DMC's with feedback, and the required bound on the transmission rate for general IT codes as well as deterministic ones.

To bound the identification rate, we will choose the set $\mathcal{D}^*_a$ to be $\bigcup_m \mathcal{D}_{a,m} \cap \mathcal{B}_\alpha(F_{a,m})$ for each a, with $\alpha = R_1 - \delta$ (δ is an arbitrary positive number). This amounts to throwing away all sequences y^n in the decoding region $\mathcal{D}_{a,m}$ which are, roughly speaking, either "atypical" (i.e., $\mathcal{E}_{y^n}(F_{a,m})$ is empty) or of "low mutual information" (i.e., $I(P^{F_{a,m}}_{x^n,y^n}; W) \leq R_1 - \delta$ for all $x^n \in \mathcal{E}_{y^n}(F_{a,m})$). The intuition is that, if n is large, such sequences cannot contribute significantly to the probability of $\mathcal{D}_a$ because the IT code is required to transmit messages at rate R_1. More precisely, for any a

$$\begin{aligned} \frac{1}{M_n}\sum_{m=1}^{M_n} Q_{F_{a,m}}(\mathcal{D}^*_a) &\geq \frac{1}{M_n}\sum_{m=1}^{M_n} Q_{F_{a,m}}(\mathcal{D}_{a,m} \cap \mathcal{B}_\alpha(F_{a,m})) \\ &= \frac{1}{M_n}\sum_{m=1}^{M_n} Q_{F_{a,m}}(\mathcal{D}_{a,m}) \\ &\quad - \frac{1}{M_n}\sum_{m=1}^{M_n} Q_{F_{a,m}}(\mathcal{D}_{a,m} \cap \mathcal{B}^c_\alpha(F_{a,m})) \\ &> 1 - \lambda - \frac{\exp[n(R_1 - \delta) + o(n)]}{M_n} - o(1) \\ &> 1 - \lambda - \frac{1 - \lambda - \mu}{2}, \qquad \text{for all large } n \\ &> \gamma, \qquad \text{for all large } n \end{aligned}$$

as required earlier. In the second inequality above, we have once again used Part 1) of Lemma 4.2, this time with $\alpha = R_1 - \delta$.

The removal of the "inessential" sequences also trims down $\mathcal{D}_a$ to the right size: $|\mathcal{D}^*_a|$ is bounded above by the right-hand sides of the inequalities in Parts 2) and 3) of Lemma 4.2 in the general and deterministic cases, respectively, with α replaced by $R_1 - \delta$ and M by M_n. Thus the sets $\mathcal{D}^*_a$ have all the properties postulated earlier, and the converses are proved.

Appendix

Proof of Lemma 3.1: The first part is a standard result—see [4, Theorem 5.2, p. 165]. We will, therefore, only sketch the proof of the second part.

Let $\mathcal{W}_n(P)$ be the set of those DMC's V (with alphabets $\mathcal{X}$ and $\mathcal{Y}$) such that $nP(x)V(y|x)$ is an integer for all x, y. For any such V, define $\mathcal{T}_V(\boldsymbol{c})$ to be the set of y^n such that

$$N(x, y|\boldsymbol{c}, y^n) = nP(x)V(y|x)$$

for all x, y. Here, $N(x, y|\boldsymbol{c}, y^n)$ is the number of occurrences of the pair (x, y) in $(\boldsymbol{c}, y^n)$. For each $V \in \mathcal{W}_n(P)$, construct K pairwise-disjoint subsets $\mathcal{C}_1(V), \cdots, \mathcal{C}_K(V)$ of $\mathcal{C} \cap \mathcal{T}_V(\boldsymbol{c})$, each of size exactly $\lfloor|\mathcal{C} \cap \mathcal{T}_V(\boldsymbol{c})|/K\rfloor$ (the subsets are otherwise arbitrary). Let $\mathcal{C}_k = \bigcup_V \mathcal{C}_k(V)$, $k \in [K]$.

Then $W^n(\mathcal{C}_k|\boldsymbol{c})$ is the same for all $k \in [K]$, since the number of sequences in $\mathcal{C}_k$ of a given conditional type w.r.t. $\boldsymbol{c}$ is the same for all k. To upper-bound $W^n(\mathcal{C}_e|\boldsymbol{c})$, where

$$\mathcal{C}_e = \mathcal{C} - \bigcup_{k=1}^{K} \mathcal{C}_k$$

note that $\mathcal{C}_e$ contains exactly $|\mathcal{C} \cap \mathcal{T}_V(\boldsymbol{c})| \bmod K$ sequences from $\mathcal{T}_V(\boldsymbol{c})$. So

$$\begin{aligned}|\mathcal{C} \cap \mathcal{T}_V(\boldsymbol{c})| \bmod K &\le \min\{K, |\mathcal{T}_V(\boldsymbol{c})|\} \\ &\le \min\{\exp(nR''), \exp[nH(V|P)]\} \\ &= \exp\{n[H(V|P) - (H(V|P) - R'')^+]\}\end{aligned}$$

and

$$\begin{aligned}W^n(\mathcal{C}_e|\boldsymbol{c}) &= \sum_{V \in \mathcal{W}_n(P)} [|\mathcal{C} \cap \mathcal{T}_V(\boldsymbol{c})| \bmod K] \\ &\quad \cdot \exp\{-n[H(V|P) + D(V\|W|P)]\} \\ &\le (n+1)^d \exp[-nE''(R'', P)].\end{aligned}$$

Here

$$E''(R'', P) = \min_V \{D(V\|W|P) + [H(V|P) - R'']^+\}$$

(the minimum is over all DMC's V with alphabets $\mathcal{X}$ and $\mathcal{Y}$). The stated properties of $E''(R'', P)$ are easy to establish. □

Proof of Lemma 3.2: Let the first row of the array be arbitrary. Then, choose a random second row, by picking each element independently and equiprobably from $[J]$. Since $\epsilon > 1/J$, the probability that the second row matches the first at least in ϵS positions is, by a Chernoff bound, no greater than

$$\exp\{-S \cdot D(\epsilon\|1/J)\} \le \exp(-S\epsilon^2/2) < 1$$

(we have used the bound $D(p\|q) \ge 2(p-q)^2$ for all p, q in $[0, 1]$, proved in [3, Lemma 12.6.1, p. 300]). Hence, there exists a "good" $2 \times S$ array.

In general, if there exists a "good" $L \times S$ array for some $L \ge 2$, and we pick an $(L+1)$th row randomly as above, then the probability that this row matches any of the other L rows at least in ϵS positions is bounded by

$$L \cdot \exp\{-S \cdot D(\epsilon\|1/J)\} \le L \cdot \exp(-S\epsilon^2/2) < 1$$

if $L < N$. This proves the existence of a "good" $N \times S$ array. □

Proof of Lemma 4.1: Let (X^n, Y^n) be the random pair of input and output sequences when the strategy F is used. Fix a pair (x, y). For $1 \le k \le n$, let $A_k = 1$ if $(X_k, Y_k) = (x, y)$, and 0 otherwise. Then,

$$E[A_k|X^{k-1}, Y^{k-1}] = P_k^F(x|X^{k-1}, Y^{k-1})W(y|x).$$

Thus if

$$\tilde{A_k} = A_k - E[A_k|X^{k-1}, Y^{k-1}]$$

then

$$\sum_k \tilde{A_k} = N(x, y|X^n, Y^n) - nP^F_{X^n, Y^n}(x)W(y|x).$$

It can be verified easily that $\mathrm{Var}(\tilde{A_k}) \le W(y|x)$, and that the $\tilde{A_k}$'s are pairwise-uncorrelated. Hence, by Chebyshev's inequality,

$$\Pr[|\sum_k \tilde{A_k}| > n^{3/4}\sqrt{W(y|x)}] \le n^{-1/2}.$$

By a union bound over all (x, y), we then have $1 - W_F(\mathcal{E}(F)) \le dn^{-1/2}$.

Suppose $(x^n, y^n) \in \mathcal{E}(F)$ and y^n has type Q. Then, for any $y \in \mathcal{Y}$

$$\begin{aligned}|Q(y) - P^F_{x^n, y^n}W(y)| &= n^{-1}|\sum_x [N(x, y|x^n, y^n) \\ &\quad - nP^F_{x^n, y^n}(x)W(y|x)]| \le |\mathcal{X}|n^{-1/4}.\end{aligned}$$

Now, if P_1 and P_2 are probability distributions on a finite set $\mathcal{Z}$, and $|P_1(z) - P_2(z)| \le \beta$ for all $z \in \mathcal{Z}$, then $|H(P_1) - H(P_2)| \le |\mathcal{Z}|\sqrt{\beta}$ (this is a weaker, but more convenient, version of Lemma 2.7 on p. 33 of [4]). Therefore,

$$|H(Q) - H(P^F_{x^n, y^n}W)| \le |\mathcal{Y}|(|\mathcal{X}|n^{-1/4})^{1/2} \le dn^{-1/8}.$$

Since $\log Q^n(y^n) = -nH(Q)$, Part a) of the Lemma is proved.

For Part b), note that

$$\begin{aligned}|\log W^n(y^n|x^n) + nH(W|P^F_{x^n, y^n})| &= \Big|\sum_{x, y}[N(x, y|x^n, y^n) \\ &\quad - nP^F_{x^n, y^n}(x)W(y|x)] \log W(y|x)\Big| \\ &\le \sum_{x, y} n^{3/4}\sqrt{W(y|x)}|\log W(y|x)| \le dn^{7/8}\end{aligned}$$

since $n^{3/4} \le n^{7/8}$, and $|\sqrt{z}\log z| \le 1$ if $0 \le z \le 1$. Part c) is an obvious consequence of Parts a) and b). □

Proof of Lemma 4.2: Let F be any n-step strategy, $\mathcal{D} \subseteq \mathcal{Y}^n$, and $\mathcal{D}' = \mathcal{D} \cap \mathcal{B}^c_\alpha(F)$. Let $\mathcal{P}_n(\mathcal{Y})$ be the set of n-types on $\mathcal{Y}$, and $\mathcal{T}_Q$ the set of y^n with type Q. Then

$$\begin{aligned}Q_F(\mathcal{D}') &\le W_F([\mathcal{X}^n \times \mathcal{D}'] \cap \mathcal{E}(F)) + dn^{-1/2} \\ &= \sum_{Q \in \mathcal{P}_n(\mathcal{Y})} \sum_{y^n \in \mathcal{D}' \cap \mathcal{T}_Q} W_F(\mathcal{E}_{y^n}(F) \times \{y^n\}) + dn^{-1/2}\end{aligned}$$

where the inequality is because $1 - W_F(\mathcal{E}(F)) \le dn^{-1/2}$.

Now, for any $y^n \in \mathcal{D}' \cap \mathcal{T}_Q$

$$\begin{aligned}W_F(\mathcal{E}_{y^n}(F) \times \{y^n\}) &= \sum_{x^n \in \mathcal{E}_{y^n}(F)} P^F_{y^n}(x^n)W^n(y^n|x^n) \\ &\le Q^n(y^n)\exp\{n\alpha + 2dn^{7/8}\}.\end{aligned}$$

Here

$$P^F_{y^n}(x^n) = \prod_k P^F_k(x_k|x^{k-1}, y^{k-1}).$$

The inequality is by Part c) of Lemma 4.1, and the fact that if $y^n \in \mathcal{B}^c_\alpha(F)$ then $I(P^F_{x^n, y^n}; W) \le \alpha$ for all $x^n \in \mathcal{E}_{y^n}(F)$.

From the last two paragraphs, it follows that

$$Q_F(\mathcal{D}') \le \sum_{Q \in \mathcal{P}_n(\mathcal{Y})} Q^n(\mathcal{D})\exp\{n\alpha + 2dn^{7/8}\} + dn^{-1/2}.$$

Hence, if $F_1, \cdots, F_M$ are n-step strategies, and $\mathcal{D}_1, \cdots, \mathcal{D}_M$ are pairwise-disjoint

$$\begin{aligned}(1/M)\sum_m Q_{F_m}(\mathcal{D}_m \cap \mathcal{B}^c_\alpha(F_m)) &\\ \le (1/M)\exp\{n\alpha + 2dn^{7/8}\}|\mathcal{P}_n(\mathcal{Y})| + dn^{-1/2}.&\end{aligned}$$

Since $|\mathcal{P}_n(\mathcal{Y})| \le (n+1)^d$, Part 1) is proved.

If $y^n \in \mathcal{B}_\alpha(F) \cap \mathcal{T}_Q$, then there exists $x^n \in \mathcal{E}_{y^n}(F)$ satisfying $I(P^F_{x^n, y^n}; W) > \alpha$, and

$$\begin{aligned}-n^{-1}\log Q^n(y^n) &\le H(P^F_{x^n, y^n}W) + dn^{-1/8} \\ &\le \max_{P:\ I(P;W) \ge \alpha} H(PW) + dn^{-1/8}.\end{aligned}$$

Here the first inequality is by Part a) of Lemma 4.1. It follows that the number of sequences of type Q in $\bigcup_F \mathcal{B}_\alpha(F)$ (the union over all n-step strategies F) is at most

$$\exp\{n[\max_{P:\, I(P;W)\geq\alpha} H(PW)] + dn^{7/8}\}.$$

Since the number of types Q is at most $(n+1)^d$, we then have

$$\left|\bigcup_F \mathcal{B}_\alpha(F)\right| \leq (n+1)^d \exp\{n[\max_{P:\, I(P;W)\geq\alpha} H(PW)] + dn^{7/8}\}.$$

But this clearly implies Part 2).

Suppose F is a deterministic n-step strategy, say $F(f)=1$. Then, $(x^n, y^n) \in \mathcal{E}(F)$ implies that $x^n = f(y^n)$ (because of the condition $W_F(x^n, y^n) > 0$ in the definition of $\mathcal{E}(F)$). Thus $y^n \in \mathcal{B}_\alpha(F)$ implies $I(P^F_{f(y^n),\, y^n}; W) > \alpha$, and

$$\begin{aligned} -n^{-1}\log Q_F(y^n) &= -n^{-1}\log W^n(y^n|f(y^n)) \\ &\leq H(W|P^F_{f(y^n),\, y^n}) + dn^{-1/8} \\ &\leq \max_{P:\, I(P;W)\geq\alpha} H(W|P) + dn^{-1/8}. \end{aligned}$$

Here the first inequality is by Part b) of Lemma 4.1. It follows that

$$|\mathcal{B}_\alpha(F)| \leq \exp\{n[\max_{P:\, I(P;W)\geq\alpha} H(W|P)] + dn^{7/8}\}$$

for any deterministic n-step strategy F. This, together with the disjointness of the sets $\mathcal{D}_m$, implies Part 3). □

References

[1] R. Ahlswede and G. Dueck, "Identification in the presence of feedback—A discovery of new capacity formulas," *IEEE Trans. Inform. Theory*, vol. 35, pp. 30–36, Jan. 1989.

[2] ——, "Identification via channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 15–29, Jan. 1989.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

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

[5] T. S. Han and S. Verdú, "New results in the theory of identification via channels," *IEEE Trans. Inform. Theory*, vol. 38, pp. 14–25, Jan. 1992.

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

Goppa Codes and Trace Operator

P. Véron

***Abstract*—We study Goppa codes, $\Gamma(L, g)$, defined by the polynomial**

$$g(z) = a(z)\mathrm{Tr}_{\mathbb{F}_{p^{ms}}:\mathbb{F}_{p^s}}(b(z)).$$

It is shown that the dimension of these codes never reaches the general, well-known, bound for Goppa codes. New bounds are proposed depending on the value of m and p. Furthermore, we prove that when $p=2$ these codes have only even weights.

***Index Terms*—Goppa codes, parameters of Goppa codes, redundancy of Goppa codes, trace operator.**

I. Introduction

Binary Goppa codes defined by the polynomial $g(z) = z^{2^s} + z$ have been introduced in [6]. Their dimension have been studied in [8] and [9], where a new bound has been proposed. In this correspondence, we generalize these results by studying the Goppa Codes which are defined by the polynomial

$$g(z) = a(z)\mathrm{Tr}_{\mathbb{F}_{p^{ms}}:\mathbb{F}_{p^s}}(b(z))$$

where $a(z)$ and $b(z)$ are two arbitrary elements of $\mathbb{F}_{p^{ms}}[z]$. We first show that the usual bound cannot be reached by giving a general new bound. Then, we treat the peculiar cases $m=2$ and $p=2$. Moreoever, we give a general property over the components of the codewords which shows that their weight is even when $p=2$.

II. Generalities, Definitions

Definition 2.1: Let p be a prime number. Let s and m be two integers, $m>1$, $g(z)$ be a polynomial over $\mathbb{F}_{p^{ms}}$, and $L = \{\alpha_1, \cdots, \alpha_n\}$ be a subset of $\mathbb{F}_{p^{ms}}$, such that $g(\alpha_i) \neq 0, \forall i = 1, \cdots, n$. The Goppa code $\Gamma(L, g)$, of length n, over $\mathbb{F}_p$, is defined as the set of words $c = (c_1, \cdots, c_n)$, $c_i \in \mathbb{F}_p$, such that

$$\mathcal{R}_c(z) = \sum_{i=1}^{n} \frac{c_i}{z-\alpha_i} \equiv 0 \pmod{g(z)}.$$

Let us denote by r the degree of $g(z)$, then

Proposition 2.2: A parity check matrix of the code $\Gamma(L, g)$ is

$$H = \begin{pmatrix} g(\alpha_1)^{-1} & \cdots & g(\alpha_n)^{-1} \\ \alpha_1 g(\alpha_1)^{-1} & \cdots & \alpha_n g(\alpha_n)^{-1} \\ \vdots & \vdots & \vdots \\ \alpha_1^{r-1} g(\alpha_1)^{-1} & \cdots & \alpha_n^{r-1} g(\alpha_n)^{-1} \end{pmatrix}.$$

Remark: This matrix satisfies $c \in \Gamma(L, g) \Leftrightarrow Hc^t = 0$, but its rows are in $\mathbb{F}_{p^{ms}}$, so they cannot generate the dual of the code $\Gamma(L, g)$, which is defined over $\mathbb{F}_p$.

Manuscript received April 8, 1996; revised June 30, 1997. The material in this correspondence was presented in part at the 3rd International Conference on Finite Fields, Glasgow, Scotland, U.K., July 1995.

The author is with G.E.C.T., Université de Toulon et du Var, 83957 La Garde Cedex, France.

Publisher Item Identifier S 0018-9448(98)00092-3.