V. Conclusion

The stationary processes with maximum entropy rate in the class of processes whose pth-order marginal distribution satisfies some constraint were shown to be Markov processes. In particular, the $(p-1)$th-order stationary Markov process with pth-order marginal density $g(z_1, \cdots, z_p)$, with respect to some measure, was shown to possess maximum entropy rate in the class of stationary processes with pth-order marginal density equal to $g(z_1, \cdots, z_p)$. This result forms the basis for an extension of the usual Gaussian maximum entropy method of Burg to nonparametric settings.

A framework for the practical implementation of the proposed nonparametric maximum entropy method was also presented. Specifically, the pth-order marginal density should be estimated from the observed data, taking care that the estimate is compatible with the hypothesis of stationarity. This can be achieved by estimating the conditional density $g(x_p|x_{p-1}, \cdots, x_1)$ from the data, and setting $g(x_1, \cdots, x_{p-1})$ to be the corresponding stationary marginal. Then, for the purposes of extrapolation of distributions or prediction, the maximum entropy principle can be invoked, implying that the distribution of the data can be approximated by that of a $(p-1)$th-order stationary Markov process with the estimated pth-order marginal density. Notably, unless the estimated pth-order density is multivariate Gaussian, the nonparametric maximum entropy method would point to a *nonlinear* Markov model.

Acknowledgment

The author is grateful to T. Cover, D. Gatzouras, S. Lalley, and H. Rubin, for many helpful discussions.

References

[1] J. P. Burg, "Maximum entropy spectral analysis," *Proc. 37th Ann. Int. Meeting Soc. Explor. Geophys.*, Oklahoma City, OK, 1967.
[2] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley, 1991.
[3] D. G. Childers, *Modern Spectrum Analysis*. New York: IEEE Press, 1978.
[4] B. S. Choi and T. M. Cover, "An information-theoretic proof of Burg's maximum entropy spectrum," *Proc. IEEE*, vol. 72, no. 8, pp. 1094–1095, 1984.
[5] L. Devroye and L. Györfi, *Nonparametric Density Estimation: The L_1 View*. New York: John Wiley, 1985.
[6] J. L. Doob, *Stochastic Processes*. New York: John Wiley, 1953.
[7] L. Györfi, W. Härdle, P. Sarda, and P. Vieu, *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statistics No. 60. New York: Springer-Verlag, 1989.
[8] I. A. Ibragimov and Y. V. Linnik, *Independent and stationary sequences of random variables*. Groningen: Wolters-Noordhoff, 1971.
[9] L. Kavalieris, "The estimation of the order of an autoregression using recursive residual and cross-validation," *J. Time Ser. Anal.*, vol. 10, no. 3, pp. 271–282, 1989.
[10] O. E. Lanford, III and D. Ruelle, "Observables at infinity and states with short range correlation in statistical mechanics," *Commun. Math. Phys.*, vol. 13, pp. 194–215, 1969.
[11] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.
[12] W. Parry and S. Tuncel, *Classification Problems in Ergodic Theory*. Cambridge: Cambridge Univ. Press, 1982.
[13] M. Rosenblatt, *Markov Processes. Structure and Asymptotic Behavior*. New York: Springer-Verlag, 1971.
[14] S. M. Ross, *Stochastic Processes*. New York: John Wiley, 1983.
[15] G. Roussas, "Nonparametric estimation of the transition distribution function of a Markov process," *Ann. Math. Statist.*, vol. 40, pp. 1386–1400, 1969.
[16] ——, "Estimation of transition distribution function and its quantiles in Markov processes: Strong consistency and asymptotic normality," in *Nonparametric Functional Estimation and Related Topics*, G. Roussas, Ed. Amsterdam, The Netherlands: Kluwer Academic, 1991, pp. 443–462.
[17] L. Ruschendorf, "Consistency of estimators for multivariate density functions and for the mode," *Sankhya, Ser. A*, vol. 39, pp. 243–250, 1977.
[18] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: Univ. of Illinois Press, 1963.
[19] F. Spitzer, "A variational characterization of finite Markov chains," *Ann. Math. Statist.*, vol. 43, pp. 303–307, 1972.
[20] T. Subba Rao and M. M. Gabr, "A test for linearity of stationary time series," *J. Time Ser. Anal.*, vol. 1, pp. 145–158, 1980.
[21] C. Z. Wei, "On predictive least squares principles," *Ann. Statist.*, vol. 20, no. 1, pp. 1–42, 1992.
[22] S. Yakowitz, "Nonparametric estimation of Markov transition functions," *Ann. Statist.*, vol. 7, no. 3, pp. 671–679, 1979.
[23] ——, "Nonparametric density estimation, prediction, and regression for Markov sequences," *J. Amer. Statist. Assoc.*, vol. 80, pp. 215–221, 1985.
[24] ——, "Nonparametric density and regression estimation for Markov sequences without mixing assumptions," *J. Multivar. Analysis*, vol. 30, pp. 124–136, 1989.

Nonbinary Codes Correcting Localized Errors

R. Ahlswede, L. A. Bassalygo, and M. S. Pinsker

***Abstract*—A recursive construction yields for all alphabets asymptotically dense codes correcting a constant number of localized errors.**

***Index Terms*—Localized errors, nonbinary alphabets, asymptotic dense packing.**

I. Introduction and Main Result

The authors of the paper [1], where codes correcting localized errors were introduced, noticed that an extension of results to nonbinary alphabets causes difficulties. In particular this is the case with the asymptotic Hamming bound, when the number of errors $t = \tau n$ increases linearly with the length n of the code. Up to now we proved it only in a small fixed interval $[0, \tau_0]$, although we believe that Hamming's bound is the true bound in the interval $[0, \frac{1}{2}]$ (if $\tau \geq \frac{1}{2}$, then the rate of transmission is equal to 0 because it is impossible to transmit even two messages, when $t \geq \frac{n}{2}$). However, our joint efforts bore fruit for other, rather accurate, asymptotic estimates on the code size, when the number of errors t is constant. The binary case was studied in [2], but the method used there does not give the tight answer for nonbinary cases. Here, we present another method.

During the transmission of q-ary words of length n over the channel at most t errors occur, and the encoder knows the set E of t positions, where these errors are possible. The decoder does not know anything about these positions. Let $\mathcal{E}_t = \{E \mid E \subseteq \{1, 2, \cdots, n\}, |E| = t\}$

Manuscript received July 9, 1991; revised December 17, 1992. This work was presented at the Third International Workshop on Algebraic and Combinatorial Coding Theory, Tyrnovo, Bulgaria, June 22–28, 1992.

R. Ahlswede is with the Universität Bielefeld, Postfach 8640, D-4800 Bielefeld 1, Germany.

L. A. Bassalygo and M. S. Pinsker are with the Institute for Problems of Information Transmission, 19 Ermolova Str., 125047 Moscow, Russia.

IEEE Log Number 9209662.

be the set of all subsets from $\{1, 2, \cdots, n\}$ of size t and let $\mathcal{M}$ be a set of messages ($|\mathcal{M}| = M$). A code word $x(m, E)$ depends not only on the message $m \in \mathcal{M}$ but also on the configuration of possible errors E. So there exists the natural correspondence between the message $m \in \mathcal{M}$ and the list of code words $\bigcup_{E\in\mathcal{E}_t}\{x(m, E)\}$, which we use for the transmission of this message. Thus the code X for the set of messages $\mathcal{M}$ represents a collection of M lists $\{\bigcup_{E\in\mathcal{E}_t}\{x(m, E)\}, m \in \mathcal{M}\}$. Since we can use the same word for different configurations, the size of a list can be essentially smaller than the size of the set $\mathcal{E}_t(|\mathcal{E}_t| = \binom{n}{t})$.

Let us define the cylinder $\mathcal{C}(\mathfrak{a}, A)$ with the base $\mathfrak{a} = (a_1, \cdots, a_n)$ and the support A ($A \subseteq \{1, 2, \cdots, n\}$) as the set of words $(y_1, \cdots, y_n)$ with $y_i = a_i$, if $i \notin A$. It is clear that the size of the cylinder $\mathcal{C}(\mathfrak{a}, A)$ is equal to $q^{|A|}$ and the number of different cylinders with the same support A is equal to $q^{n-|A|}$.

As a result of the transmission of the codeword $x(m, E)$ every word of $\mathcal{C}(x(m, E), E)$ can appear as output of the channel. The code X corrects t localized errors, if the decoder can correctly recover every message $m \in \mathcal{M}$. The following condition is necessary and sufficient for it:

$$\mathcal{C}(x(m, E), E) \cap \mathcal{C}(x(m', E'), E') = \varnothing \quad \text{for all } E, E' \in \mathcal{E}_t,\ m, m' \in \mathcal{M},\ m \neq m'.$$

The maximal number of messages, which we can transmit by a code correcting t localized errors, is denoted by $L_q(n, t)$.

Proposition 1:

$$L_q(n, t) \leq \frac{q^n}{S_t},$$

where $S_t = \sum_{i=0}^{t}(q-1)^i\binom{n}{i}$ is the size of a sphere of radius t in the Hamming n space. A proof of this bound in the q-ary case can be given as for the binary case in [1] or [3]. The key inequality there has the following generalization.

Lemma 1: Let $\mathcal{C}(\mathfrak{a}_1, A_1), \cdots, \mathcal{C}(\mathfrak{a}_T, A_T)$ be cylinders with pairwise different supports $A_i \neq A_j$, $i \neq j$. Then, for the size of the union of the cylinders,

$$\left|\bigcup_{i=1}^{T}\mathcal{C}(\mathfrak{a}_i, A_i)\right| \geq \sum_{i=1}^{T}(q-1)^{|A_i|}.$$

The proof of this lemma can be made by an induction on n as in [1] or [3] (for other proofs see [4]).

It yields immediately

$$\left|\bigcup_{E\in\mathcal{E}_t}\mathcal{C}(x(m, E), E)\right| \geq \binom{n}{t}(q-1)^t.$$

However, it can be used more efficiently. Choose any monotone map $f: \bigcup_{i=1}^{t}\mathcal{E}_i \to \mathcal{E}_t$ with the property $F \subset f(F)$ for all $F \in \bigcup_{i=1}^{t}\mathcal{E}_i$. Then, we can write

$$\mathcal{C}(x(m, E), E) = \bigcup_{F\subset E}\mathcal{C}(x(m, f(F)), F).$$

Thus, by Lemma 1,

$$\left|\bigcup_{E\in\mathcal{E}_t}\mathcal{C}(x(m, E), E)\right| = \left|\bigcup_{F\in\bigcup_{i=0}^{t}\mathcal{E}_i}\mathcal{C}(x(m, f(F)), F)\right| \geq \sum_{i=0}^{t}\binom{n}{i}(q-1)^i \quad (1)$$

and Proposition 1 follows.

The following lower bound can be easily deduced by the standard greedy algorithm (maximal coding).

Proposition 2:

$$L_q(n, t) \geq \frac{q^n}{q^{2t}\binom{n}{t}}.$$

Already Propositions 1 and 2 imply the asymptotic equivalence within a constant

$$L_q(n, t) \asymp \frac{q^n}{n^t}, \qquad \text{when } t \text{ is fixed and } n \to \infty.$$

We draw attention to the fact that this equivalence is known for nonbinary error-correcting codes except $t = 1$ only for $t = 2$ and $q = 3, 4$ [5].

The purpose of our work is to find the precise constant in the equivalence. We construct asymptotically dense codes.

Theorem: For every constant t:

$$L_q(n, t) = \frac{q^n}{S_t}(1 + o(1)) = \frac{t!}{(q-1)^t}\cdot\frac{q^n}{n^t}(1 + o(1)),$$

where $o(1) \to 0$ as $n \to \infty$ ($o(1)$ depends certainly on t and q).

II. Proof of Theorem: The Induction Step

We proceed by induction on t. The induction beginning $t = 1$ is established at the end of the proof in Section IV. The induction step to t goes through the following scheme.

We divide the interval of transmission of length n into $b + 1$ successive intervals: the first interval shall be of length r and all the other b intervals shall be of length $k = \frac{n-r}{b}$ (more precisely, when $n - r$ isn't divisible by b, these intervals have the length $\lfloor\frac{n-r}{b}\rfloor$ or $\lceil\frac{n-r}{b}\rceil$). Now, for fixed E let us denote by $t_0, t_1, \cdots, t_b$ the number of possible errors in the ith interval ($i = 0, 1, \cdots, b$) ($t_0 + t_1 + \cdots + t_b = t$).

At first we guarantee that the decoder always knows whether $t_0 = 0$ or $t_0 > 0$. This we achieve by using in the first r positions always at least $t + 1$ 1's, if $t_0 = 0$, and always r 0's, if $t_0 > 0$.

Now we distinguish three cases.

Case 1: $t_0 > 0$. We use a code correcting $t - 1$ localized errors on the last $n - r$ positions.

Case 2: $t_0 = 0$ and $t_i < t$ for all i, $i = 1, \cdots, b$. In the first r positions, we transmit the values of $t_1, \cdots, t_b$, and we use for this transmission words of weight at least $t + 1$ with the additional restriction that in the first r_1 of the r positions there is at least one 0. Since the number of solutions of the equality $t_1 + \cdots + t_b = t$ with $t_i < t$, $i = 1, \cdots, b$, is equal to $\binom{b+t-1}{t} - b$, we can do this, if the numbers satisfy

$$\binom{b+t-1}{t} - b \leq q^r - \sum_{i=0}^{t}\binom{r}{i}(q-1)^i - (q-1)^{r_1}q^{r-r_1}. \quad (2)$$

Having transmitted the values of t_i, $i = 1, \cdots, b$, we use in the ith interval ($i = 1, \cdots, b$) a code correcting t_i localized errors.

Case 3: $t_0 = 0$ and $t_j = t$, $t_i = 0$, $i \neq j$. We transmit in the first r_1 from r positions ones only in order to distinguish this and the previous case, in the next $\lceil\log_q b\rceil$ positions we transmit the number j ($j = 1, 2, \cdots, b$), and in the next $r - r_1 - \lceil\log_q b\rceil$ positions and in all intervals, except for the jth, we transmit all possible sequences, and, finally, we use in the jth interval a code correcting t localized errors with parameters guaranteed by Proposition 2.

After this description of the code count how many messages can be transmitted in each one of the three cases. Of course this counting makes use of the induction hypothesis for $t' < t$.

Case 1: We can transmit at least

$$L_q(n-r, t-1) = \frac{q^{n-r}}{(q-1)^{t-1}\binom{n-r}{t-1}}(1+o(1)) \geq \frac{q^{n-r}}{(q-1)^t\binom{n}{t}} \cdot \frac{(q-1)n}{t}[1+o(1)] \quad (3)$$

messages, where $o(1) \to 0$ as $n-r \to \infty$. The inequality is based on the relations $\binom{n-r}{t-1} \leq \binom{n-1}{t-1} = \binom{n}{t}\frac{t}{n}$.

Case 2: We can transmit at least

$$\prod_{i=1}^{b} \cdot L_q(k, t_i) \geq \frac{q^{n-r}}{(q-1)^t k^t}(1+o(1)) \quad (4)$$

messages, where $o(1) \to 0$ as $k \to \infty$.

Case 3: We can transmit at least

$$L_q(k, t) \cdot q^{n-k-r_1-\lceil \log_q b \rceil} \geq \frac{q^{n-r_1-\lceil \log_2 b \rceil - 2t}}{\binom{k}{t}} \geq \frac{q^{n-r}}{\binom{k}{t}} \geq \frac{q^{n-r}}{k^t} \quad (5)$$

messages, provided that

$$r - r_1 \geq \lceil \log_q b \rceil + 2t. \quad (6)$$

Therefore, if (2) and (6) hold, then

$$L_q(n, t) \geq \min\left\{\frac{(q-1)n}{tq^r}, \frac{\binom{n}{t}}{k^t q^r}\right\} \cdot \frac{q^n}{S_t}(1+o(1)),$$

where $o(1) \to 0$ as $k \to \infty$ and $n-r \to \infty$. (7)

Under the conditions

$$q^r \leq \frac{(q-1)n}{t} \quad (8)$$

and

$$q^r \leq \frac{\binom{n}{t}}{k^t}[1+o(1)], \qquad \text{where } o(1) \to 0 \text{ as } b \to \infty. \quad (9)$$

we have from (7)

$$L_q(n, t) \geq \frac{q^n}{S_t}(1+o(1)), \quad (10)$$

where $o(1) \to 0$, if $k \to \infty$, $b \to \infty$.

Finally, we have to choose the parameters r, r_1, and b in a suitable way:

$$r = \left\lfloor \log_q \frac{(q-1)n}{t} \right\rfloor$$
$$r_1 = r - \lceil \log_q b \rceil - 2t$$
$$b = \text{maximal number such that (2) holds.} \quad (11)$$

Clearly with these parameters (2), (6), and (8) hold by definition and it remains only to verify (9). But the definition of b, (11), and (2) imply

$$q^r \leq \binom{b}{t}(1+o(1)), \qquad \text{where } o(1) \to 0 \text{ as } b \to \infty.$$

Since

$$\binom{b}{t} \leq \frac{\binom{n}{t}}{k^t},$$

(9) holds.

III. An Auxiliary Result for Perfect Codes with One Localized Error

Recently Kabatyansky [6] constructed codes for one localized error of length $D = \frac{d^4-1}{q-1}$ $(d = 1, 2, \cdots)$, which are perfect and thus,

$$L_q(D, 1) = \frac{q^D}{(q-1)D+1} = \frac{q^D}{S_1}(1+o(1)). \quad (12)$$

From his construction, it follows also that

$$L_q(D-1, 1) \geq \frac{q^{D-1}}{(q-1)D+1}. \quad (13)$$

Subsequently we established in [7] the following recurrence relation of independent interest. We repeat the proof, because it is brief.

Lemma 2: If $N \leq (q-1)T + 1$, then

$$L_q(N+T, 1) \geq q^{N-1} L_q(T, 1).$$

Proof: On the first N positions, we always transmit a parity check (mod q) code of the size q^{N-1}. The last T positions we reserve for the code, which allows us to transmit $L_q(T, 1)$ messages and to correct a single localized error. The method of transmission depends on the position of the localized error in the following way.

1) The error is on the last T positions. We use our code as a code, correcting a single localized error.
2) The error is in the first N positions. We use our code for the transmission of both, the message and the number of the position, where the error can occur. According to (1) at least $(q-1)T + 1$ different words in the output of the channel of length T corresponds to each of $L_q(T, 1)$ messages and therefore we can make the successful transmission if

$$N \leq (q-1)T + 1.$$

The proof is complete. □

Now, we derive (12) and (13) and thus, make our paper independent of work that has not appeared yet. Clearly, for $d = 1$ we have $D = 1$ and (12) holds, because obviously $L_q(1, 1) = 1$. We proceed now inductively from d to $d+1$. We set $D' = \frac{q^{d+1}-1}{q-1}$ and observe that for $N = (q-1)D + 1$ we have $D' = q^d + q^{d-1} + \cdots + 1 = qD + 1 = D + N$. By Lemma 2,

$$L_q(D', 1) \geq q^{(q-1)D} L_q(D, 1)$$

and by the induction hypothesis

$$L_q(D', 1) \geq q^{(q-1)D} \frac{q^D}{(q-1)D+1} = \frac{q^{qD}}{(q-1)D+1} = \frac{q^{d'}}{q(q-1)D+q} = \frac{q^{D'}}{(q-1)D'+1}.$$

Finally, we verify (13). With the choices $T = \frac{q^{d-1}-1}{q-1}$ and $N = (q-1)T$, we have by Lemma 2 and (12),

$$L_q(D-1, 1) = L_q(N+T, 1) \geq q^{(q-1)T-1}\frac{q^T}{(q-1)T+1} = \frac{q^{D-1}}{(q-1)T+1} \geq \frac{q^{D-1}}{(q-1)D+1} \text{ (because } D \geq T).$$

Remark: When q is a prime power then the maximal size of a code of length n, correcting a *single* (not necessarily localized) error, already equals $\frac{q^n}{S_1}(1+o(1))$, where $o(1) \to o$ as $n \to \infty$ ([8]). Such a strong result is not available for general q. However, next we establish the asymptotic Hamming bound for general q in case of a localized error $(t = 1)$.

IV. Proof of Theorem: The Induction Beginning

For the value $t = 1$, only the cases 1 and 3 occur. This simplifies matters, because there is now no need to inform the decoder whether case 2 or case 3 occured. On the other hand, since $\binom{n}{1}$ is linear in n, rather accurate estimates are needed and in particular the inequality in Proposition 2 is too crude for the present purposes. Instead, we use (12) and (13), consequences of Lemma 2.

For general n, we devide the interval of transmission of length n into $b+1$ successive blocks such that the 0th block has length $r \triangleq \lfloor \log_1(q-1)n \rfloor$ and all others have length D or $D-1$ (and so $\lfloor \frac{n-r}{D} \rfloor \le b \le \lceil \frac{n-r}{D-1} \rceil$).

Case 1: The one error can only be in the 0th block, where we send (as previously) 0's. In the other blocks, all sequences can be sent. Therefore, we can transmit q^{n-r} messages, where

$$q^{n-r} = q^{n-\lfloor \log_q(q-1)n \rfloor} \ge \frac{q^n}{(q-1)n+1} = \frac{q^n}{S_1}.$$

Case 3: Let the position of a possible error be in the jth block.

In the 0th block we use only sequences with at most $r-2$ 0's (to distinguish for the decoder this case from case 1) and we also encode that an error may occur in the jth block. We cannot waste even one position! This we achieve by partitioning the $q^r - r(q-1) - 1$ sequences with at most $r-2$ 0's into b sets $P_1, \cdots, P_b$ such that

$$\left| |P_i| - \frac{q^r - r(q-1) - 1}{b} \right| \le 1,$$

for $i = 1, 2, \cdots, b$.

Words in P_j inform the decoder that an error may occur in the jth block. There the sender uses a code meeting the bounds in (12) or (13) and in the remaining blocks all sequences can be used.

Therefore, we can transmit at least

$$\left\lfloor \frac{q^r - r(q-1) - 1}{b} \right\rfloor \frac{q^{n-r}}{(q-1)D+1} \ge \frac{q^n}{S_1}(1 + o(1))$$

messages, where $o(a) \to o$ as $D \to \infty$ and $b \to \infty$. The proof is complete. □

References

[1] L. A. Bassalygo, S. I. Gelfand, and M. S. Pinsker, "Coding for channels with localized errors," *Proc. 4th Soviet–Swedish Workshop in Inform. Theory,* Gotland, Sweden, 1989, pp. 95–99.

[2] ——, "Coding for channels with partially localized errors," *IEEE Inform. Theory,* vol. 37, no. 3, pp. 880–884, May 1991.

[3] ——, "Simple methods of deduction of lower bounds in coding theory," *Probl. Peredac. Inform.,* vol. 27, no. 4, pp. 3–8, 1991.

[4] R. Ahlswede, "Coding for channels with localized errors: The nonbinary cases," Preprint 89-020 of SFB 343, Universität Bielefeld.

[5] I. I. Dumer and V. A. Zinovyev, "New maximum codes over GF(4)," *Probl. Peredac. Inform.,* vol. 14, no. 3, pp. 24–34, 1978.

[6] G. A. Kabatyansky, "The construction of code correcting single localized error," *Proc. III Int. Workshop on Algebraic and Combinat. Coding Theory,* Tyrnovo, Bulgaria, June, 1992, pp. 22–28, in print.

[7] R. Ahlswede, L. A. Bassalygo, and M. S. Pinsker, "Asymptotically dense nonbinary codes correcting a constant number of localized errors," *Proc. III Int. Workshop on Albegraic and Combinat. Coding Theory,* Tyrnovo, Bulgaria, June 1992, pp. 22–28, in print.

[8] G. A. Kabtyansky and V. I. Panchenko, "Packings and coverings of Hamming spaces with unit spheres," *Probl. Peredac. Inform.,* vol. 24, no. 4, pp. 3–16, 1988.

The Maximal Error Capacity of Arbitrarily Varying Channels for Constant List Sizes

Rudolf Ahlswede

***Abstract*—The capacity of an arbitrarily varying channel (AVC) for list codes of arbitrarily small list rate under the maximal error criterion has previously been determined. Here, the following sharper result is proved: For an AVC $\mathcal{A}$, any rate R below the list code capacity $C_l(\mathcal{A})$ is achievable with the list size $L(\mathcal{A}, R) = \left\lfloor \frac{\log|\mathcal{Y}|}{C_l(\mathcal{A})-R} \right\rfloor + 1$, where $\mathcal{Y}$ is the ontpnt alphabet. For the average error criterion, the corresponding resnlt was conjectured by Pinsker and proved by Ahlswede and Cai.**

***Index Terms*—Arbitrarily varying channel, list codes, maximal error, balauced hypergraph packing.**

I. Known Results

An AVC is defined here by a sequence $\mathcal{A} = (\{P(\cdot|\cdot|s^n): s^n \in \mathcal{S}^n\})_{n=1}^{\infty}$ of sets of transmission probabilities, where for a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$ and a finite set $\mathcal{W} = \{w(\cdot|\cdot|s): s \in \mathcal{S}\}$ of stochastic $|\mathcal{X}| \times |\mathcal{Y}|$-matrices

$$P(y^n|x^n|s^n) = \prod_{t=1}^{n} w(y_t|x_t|s_t), \tag{1}$$

for all $x^n = (x_1, x_2, \cdots, x_n) \in \mathcal{X}^n = \prod_1^n \mathcal{X}$, for all $y^n \in \mathcal{Y}^n$, and for all $s^n \in \mathcal{S}^n$. Let L be a positive integer. An (n, N, L) list code is a system

$$\{(u_i, \mathcal{D}_i): 1 \le i \le N\},$$

where $u_i \in \mathcal{X}^n$, $\mathcal{D}_i \subset \mathcal{Y}^n$ and

$$\sum_{i=1}^{N} 1_{\mathcal{D}_i}(y^n) \le L, \qquad \text{for all } y^n \in \mathcal{Y}^n. \tag{2}$$

$1_{\mathcal{B}}$ denotes the indicator function of a set $\mathcal{B}$.

We speak of an (n, N, L, λ) code for $\mathcal{A}$, if in addition for all $s^n \in \mathcal{S}^n$

$$P(\mathcal{D}_i|u_i|s^n) \ge 1 - \lambda, \qquad \text{for } i = 1, 2, \cdots, N. \tag{3}$$

We call a number $C_l(\mathcal{A})$ the list code capacity of $\mathcal{A}$, if the two conditions hold

1) For any $\epsilon > 0, \delta > 0$ and $\lambda \in (0, 1)$ there is an $(n, \exp\{n(C_l(\mathcal{A}) - \delta)\}, \exp\{n\epsilon\}, \lambda)$ code for all large n.
2) For any $\delta > 0$ and $\lambda \in (0, 1)$ there is no ϵ, $0 < \epsilon < \delta$, such that $(n, \exp\{n(C_l(\mathcal{A}) + \delta)\}, \exp\{n\epsilon\}, \lambda))$ codes exist for all large n.

The so-called row-convex hull of $\mathcal{W}$ is defined as

$$\bar{\mathcal{W}} = \{w(\cdot \mid \cdot): w(\cdot \mid x) \in \operatorname{conv}\{w(\cdot|x|s): s \in \mathcal{S}\} \text{ for all } x \in \mathcal{X}\}. \tag{4}$$

Let $\mathcal{P}(\mathcal{X})$ stand for the set of probability distributions on $\mathcal{X}$. Denoting by $I(P, w)$ the mutual information for input distribution P and channel matrix w, we can introduce

$$\bar{C} = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{w \in \bar{\mathcal{W}}} I(P, w). \tag{5}$$

Manuscript received April 5, 1992; revised November 11, 1992.

The author is with The Universität Bielefeld, Postfach 8640, 4800 Bielefeld 1, Germany.

IEEE Log Number 9209659.