TABLE I
COVERINGS FOR WOLFMANN'S DECODING SET

Permutations	Coverings											
π_1	12	13	14	15	16	17	18	19	20	21	22	23
π_2	6	7	8	9	10	11	18	19	20	21	22	23
π_3	3	4	5	9	10	11	15	16	17	21	22	23
π_4	6	7	8	9	10	11	12	13	14	15	16	17
π_5	3	4	5	6	7	8	15	16	17	18	19	20
π_6	3	4	5	6	7	8	12	13	14	21	22	23
π_7	3	4	5	9	10	11	12	13	14	18	19	20
π_8	0	1	2	3	4	5	6	7	8	9	10	11
π_9	0	1	2	3	4	5	12	13	14	15	16	17
π_{10}	0	1	2	6	7	8	12	13	14	18	19	20
π_{11}	0	1	2	3	4	5	18	19	20	21	22	23
π_{12}	0	1	2	9	10	11	12	13	14	21	22	23
π_{13}	0	1	2	9	10	11	15	16	17	18	19	20
π_{14}	0	1	2	6	7	8	15	16	17	21	22	23

TABLE II
BLOCK COVERINGS FOR WOLFMANN'S DECODING SET

Permutations	Blocks			
π_1	4	5	6	7
π_2	2	3	6	7
π_3	1	3	5	7
π_4	2	3	4	5
π_5	1	2	5	6
π_6	1	2	4	7
π_7	1	3	4	6
π_8	0	1	2	3
π_9	0	1	4	5
π_{10}	0	2	4	6
π_{11}	0	1	6	7
π_{12}	0	3	4	7
π_{13}	0	3	5	6
π_{14}	0	2	5	7

ing no errors. Clearly, this should be the first permutation applied by our decoding procedure. Next, we see that for a received word r containing a single error, either π_1 or π_8 will move the error out of the first 12 positions. If these are the first two permutations applied by our decoding scheme, then the average number of permutations required to decode r in the single error case is 1.5. Now, consider the situation in which two errors appear in r. Since $N(2, 12, 24) \geq 6$, we conclude that at least six permutations are required to guarantee the correction of r. Suppose such a set of six permutations exists within our decoding set, and that these are applied first. Then, on the average, 3.5 permutations are required to decode a received word containing two errors. Finally, when three errors occur in r, we expect an average of 7.5 permutations.

Suppose a codeword is transmitted over a binary symmetric channel with symbol error rate p. Then an estimate for the number of permutations required to decode a received word with three or fewer errors is given by $\mu_{\text{permutations}} = (1-p)^{24} + 1.5\binom{24}{1}p(1-p)^{23} + 3.5\binom{24}{2}p^2(1-p)^{22} + 7.5\binom{24}{3}p^3(1-p)^{21}$. Even for p as large as .05, we see that $\mu_{\text{permutations}} \approx 2.27$. Of course, this analysis depends upon our finding six permutations among our decoding set which will correct all double errors.

Now, consider a 24-length vector as consisting of eight blocks of three coordinate positions each. That is, for a received vector $r = (r_0, \cdots, r_{23})$, we have block $i = (r_{3i}, r_{3i+1}, r_{3i+2})$. Here, $i = 0, \cdots, 7$. In Table II, we present the coverings for Wolfmann's permutation decoding set in terms of the blocks of the received word r. Consider the case in which r contains two errors. Clearly, a permutation from our decoding set will suffice to correct r if the block(s) in which the two errors occur are covered by that permutation. However, we want six permutations which cover all pairs of blocks, to ensure that all double errors can be corrected. It can be easily verified that the set consisting of permutations $\{\pi_1, \pi_2, \pi_4, \pi_8, \pi_9, \pi_{11}\}$ will cover all pairs of blocks, and thus can decode any received vector containing two errors. So, our prior analysis can be implemented by sequencing the permutations of our decoding set as $\{\pi_1, \pi_8, \pi_2, \pi_4, \pi_9, \pi_{11}, \pi_3, \pi_5, \pi_6, \pi_7, \pi_{10}, \pi_{12}, \pi_{13}, \pi_{14}\}$.

ACKNOWLEDGMENT

The author would like to thank Prof. V. Ramamurthi for the assistance in the editing of this work.

REFERENCES

[1] M. Blaum and J. Bruck, "Decoding the Golay code with Venn diagrams," *IEEE Trans. Inform. Theory*, vol. 36, pp. 906–910, July 1990.
[2] D. M. Gordon, "Minimal permutation sets for decoding the binary Golay codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 541–543, May 1982.
[3] F. J. MacWilliams, "Permutation decoding of systematic codes," *Bell Syst. Tech. J.*, vol. 43, pp. 485–505, 1964.
[4] J. Wolfmann, "A permutation decoding of the (24, 12, 8) Golay code," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 748–750, Sept. 1983.

A Recurrence Theorem for Dependent Processes with Applications to Data Compression

Andrew Nobel and Aaron D. Wyner, *Fellow, IEEE*

***Abstract*—In an earlier work, Wyner and Ziv proved theorems on recurrence times for strings in a random sequence, and applied these theorems to data compression and the Lempel–Ziv algorithm. It is shown that one of these theorems holds under an essentially weaker hypothesis. The new proof is considerably simpler than the original.**

I. INTRODUCTION

Let $\{X_i\}_{i=-\infty}^{\infty}$ be a stationary ergodic sequence of random variables X_i taking values in a finite set A. Let $\mathbb{P}$ be the distribution of the random vector $X = (\cdots, X_{-1}, X_0, X_1, \cdots)$. We will denote a particular realization of the process $\{X_i\}_{i=-\infty}^{\infty}$ by a vector $x \in \prod_{i=-\infty}^{\infty} A$; also, for $i < j$, x_i^j will denote the partial sequence $(x_i, x_{i+1}, \ldots, x_j)$. For every integer $l > 0$ and every sequence $x \in \prod_{i=-\infty}^{\infty} A$, we define the recurrence time $\hat{N}_l$ as follows:

$$\hat{N}_l(x) \text{ is the least integer } N \geq l \text{ for which } x_{-l+1}^0 = x_{N-l+1}^N.$$

If we think of the l-vector x_{-l+1}^0 as a "template," we must shift in $\hat{N}_l(x)$ places to the right before we find a match in x_1^∞. Of course, $\hat{N}_l(X)$ is a random variable.

Manuscript received April 2, 1991; revised January 18, 1992.

A. Nobel is with the Information Systems Laboratory, Durand Building, Stanford University, Stanford, CA 94305.
A. D. Wyner is with AT & T Bell Laboratories, Room 2C-365, 600 Mountain Avenue, Murray Hill, NJ 07974.

IEEE Log Number 9108034.

In [4], it was shown that the behavior of the recurrence time $\hat{N}_l(\boldsymbol{x})$ is related to the entropy rate H of the process $\{X_i\}$.

Theorem A: If the process $\{X_i\}$ has entropy rate H, then

$$(1/l)\log \hat{N}_l(\boldsymbol{X}) \to H \text{ in probability } \mathbb{P}, \qquad \text{as } l \to \infty.$$

Theorem A can be used to analyze the performance of the ordinary, *infinite-memory* version of the Lempel–Ziv data compression algorithm (cf. [4]). In this correspondence, we consider a related result from [4] that can be used to analyze the performance of a *finite data-base* version of the Lempel–Ziv algorithm (cf. [5]). Specifically, let $\hat{\mathbb{P}} = \mathbb{P}^0_{-\infty} \times \mathbb{P}^\infty_1$ be the distribution on $\{X_i\}_{i=-\infty}^{\infty}$ under which the collections $\{\cdots, X_{-1}, X_0\}$ and $\{X_1, X_2, \cdots\}$ are independent, but are distributed individually according to $\mathbb{P}$.

Theorem B: If the process $\{X_i\}$ has entropy rate H, then

$$(1/l)\log \hat{N}_l(\boldsymbol{X}) \to H \text{ in probability } \hat{\mathbb{P}}, \qquad \text{as } n \to \infty, \quad (1)$$

provided that one of the following conditions is satisfied:

1) $\hat{\mathbb{P}} \ll \mathbb{P}$, i.e., the measure $\hat{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$;
2) there exists a $k \geq 0$ such that $X^0_{-\infty}, X^k_1, X^\infty_{k+1}$ is a Markov chain under $\mathbb{P}$.

In the next section, we show that Theorem B holds if $\mathbb{P}$ satisfies a certain "mixing condition." This condition is essentially weaker than condition 2), and our proof of sufficiency is considerably simpler than the proof given in [4]. In the final section we present a negative result that shows one cannot always "dither" a process in order to make it satisfy condition 1).

II. A New Proof of Sufficiency

Given any stationary process $\{X_i\}_{i=-\infty}^{\infty}$, with probability measure $\mathbb{P}$, we define the mixing coefficients $\alpha(k)$, $k = 1, 2, \cdots$, as follows:

$$\alpha(k) = \sup |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$
$$A \in \sigma(X^0_{-\infty})$$
$$B \in \sigma(X^\infty_k),$$

where $\sigma(X^0_{-\infty})$ denotes the least σ-field with respect to which the collection $\{\cdots, X_{-1}, X_0\}$ is measurable, and $\sigma(X^\infty_k)$ is defined similarly. Thus, $\alpha(k)$ is a measure of the amount of dependence between the past and the future of the process $\{X_i\}^\infty_{-\infty}$. If $\alpha(k) \to 0$ as $k \to \infty$ the process is said to be *α-mixing*. In particular, any α-mixing process is strongly mixing, and hence, ergodic (cf. [2]).

Theorem 1: Let $\{X_i\}_{i=-\infty}^{\infty}$ be a stationary process having entropy rate $H > 0$, and suppose that for some $\delta > 0$, $\alpha(k) \leq 1/k^{1+\delta}$. Then, $(1/l)\log \hat{N}_l(\boldsymbol{X}) \to H$ in probability $\hat{\mathbb{P}}$, as $l \to \infty$.

Proof: Fix $\epsilon > 0$. We begin by showing that

$$\hat{\mathbb{P}}\{(1/l)\log \hat{N}_l(\boldsymbol{X}) < H - \epsilon\} \to 0, \qquad \text{as } l \to \infty, \quad (2)$$

using an argument from [4]. First note that for every $\boldsymbol{a} \in A^l$, and every positive integer n,

$$\hat{\mathbb{P}}\{\hat{N}_l(\boldsymbol{X}) < n | X^0_{-l+1} = \boldsymbol{a}\}$$
$$\leq \sum_{k=l}^{n-1} \hat{\mathbb{P}}\{X^k_{k-l+1} = \boldsymbol{a} | X^0_{-l+1} = \boldsymbol{a}\}$$
$$= (n-l)\mathbb{P}(\boldsymbol{a}). \quad (3)$$

Now, let $\eta = \epsilon/2$. It is easy to see that

$$\hat{\mathbb{P}}\{\hat{N}_l(\boldsymbol{X}) < n\} \leq \hat{\mathbb{P}}\{\hat{N}_l(\boldsymbol{X}) < n | X^0_{-l+1} \in T(l,\eta)\}$$
$$+ \mathbb{P}\{X^0_{-l+1} \notin T(l,\eta)\}, \quad (4)$$

where $T(l,\eta) \triangleq \{\boldsymbol{a} \in A^l: |(-1/l)\log \mathbb{P}(\boldsymbol{a}) - H| < \eta\}$ is the set of η-typical sequences of length l. Letting $n = \lceil 2^{l(H-\epsilon)} \rceil$, and combining (3) and (4) gives

$$\hat{\mathbb{P}}\{(1/l)\log \hat{N}_l(\boldsymbol{x}) < H - \epsilon\} \leq (\lceil 2^{l(H-\epsilon)} \rceil - l) \cdot 2^{-l(H-\eta)}$$
$$+ \mathbb{P}(T(l,\eta)^c)$$
$$\leq 2^{-l\epsilon/2} + \mathbb{P}(T(l,\eta)^c),$$

which tends to zero as $l \to \infty$, by the AEP. Thus, (2) is established.

Now let B_l be the event that there is a match of X^0_{-l+1} in the interval $[2^{l(H-\epsilon)}, 2^{l(H+\epsilon)}]$; that is,

$$B_l = \bigcup_{\boldsymbol{a} \in A^l} B_l(\boldsymbol{a}),$$

where

$$B_l(\boldsymbol{a}) \triangleq \{X^0_{-l+1} = \boldsymbol{a} \quad \text{and} \quad \exists N \in [2^{l(H-\epsilon)}, 2^{l(H+\epsilon)}]$$
$$\text{s.t. } X^N_{N-l+1} = \boldsymbol{a}\}.$$

By virtue of (2), it is enough to show that

$$\hat{\mathbb{P}}(B_l) \to 1, \qquad \text{as } l \to \infty. \quad (5)$$

Basically, (2) tells us that we will not find a match too soon, while (5) indicates that at least one match will occur in the desired interval.

To establish (5), we begin by noting that the events $B_l(\boldsymbol{a})$, $\boldsymbol{a} \in A^l$, are disjoint, so that

$$|\mathbb{P}(B_l) - \hat{\mathbb{P}}(B_l)| = \left| \sum_{\boldsymbol{a} \in A^l} \mathbb{P}(B_l(\boldsymbol{a})) - \sum_{\boldsymbol{a} \in A^l} \hat{\mathbb{P}}(B_l(\boldsymbol{a})) \right|$$
$$\leq \sum_{\boldsymbol{a} \in A^l} |\mathbb{P}(B_l(\boldsymbol{a})) - \hat{\mathbb{P}}(B_l(\boldsymbol{a}))|$$
$$= \sum_{\boldsymbol{a} \in T(l,\epsilon)} |\mathbb{P}(B_l(\boldsymbol{a})) - \hat{\mathbb{P}}(B_l(\boldsymbol{a}))|$$
$$+ \sum_{\boldsymbol{a} \notin T(l,\epsilon)} |\mathbb{P}(B_l(\boldsymbol{a})) - \hat{\mathbb{P}}(B_l(\boldsymbol{a}))|. \quad (6)$$

Theorem A guarantees that $\mathbb{P}(B_l) \to 1$ as $l \to \infty$, so (5) will follow if we can show that each term in (6) tends to zero as $l \to \infty$. Consider the second term in (6). We have

$$\sum_{\boldsymbol{a} \notin T(l,\epsilon)} |\mathbb{P}(B_l(\boldsymbol{a})) - \hat{\mathbb{P}}(B_l(\boldsymbol{a}))|$$
$$\leq \sum_{\boldsymbol{a} \notin T(l,\epsilon)} [\mathbb{P}(B_l(\boldsymbol{a})) + \hat{\mathbb{P}}(B_l(\boldsymbol{a}))]$$
$$\leq \sum_{\boldsymbol{a} \notin T(l,\epsilon)} 2\mathbb{P}(\boldsymbol{a})$$
$$= 2\mathbb{P}(T(l,\epsilon)^c), \quad (7)$$

which tends to zero as $l \to \infty$ by the AEP.

Now consider the first term in (6). Note that each event $B_l(\boldsymbol{a})$ is of the form $E \cap F$ with $E \in \sigma(X^0_{-\infty})$ and $F \in \sigma(X^\infty_{\lceil 2^{l(H-\epsilon)} \rceil - l+1}) \subseteq \sigma(X^\infty_{\lceil 2^{l(H-2\epsilon)} \rceil})$, for large l. By definition of

the coefficients $\alpha(k)$,

$$|\mathbb{P}(B_l(\boldsymbol{a})) - \hat{\mathbb{P}}(B_l(\boldsymbol{a}))| \le \alpha(2^{l(H-2\epsilon)}),$$

for all l greater than some integer l_0. It follows that, for $l \ge l_0$,

$$\begin{aligned}\sum_{\boldsymbol{a}\in T(l,\epsilon)} |\mathbb{P}(B_l(\boldsymbol{a})) - \hat{\mathbb{P}}(B_l(\boldsymbol{a}))| &\le |T(l,\epsilon)|\alpha(2^{l(H-2\epsilon)}) \\ &\le 2^{l(H+\epsilon)} \cdot 2^{-l(1+\delta)(H-2\epsilon)} \\ &= 2^{l(-\delta H + \epsilon(2\delta+3))}. \qquad (8)\end{aligned}$$

Now $-\delta H + \epsilon(2\delta + 3)$ is negative for ϵ less than some number $\epsilon_0 > 0$, not depending on l, so for $\epsilon < \epsilon_0$ the second term in (6) tends to zero as $l \to \infty$. □

Remark: One can establish the conclusion of Theorem 1 easily, using a slightly stronger mixing condition known as absolute regularity. For $k \ge 1$ define

$$\beta(k) = \sup_{A\in\sigma(X_{-\infty}^0, X_k^\infty)} |\mathbb{P}(A) - \hat{\mathbb{P}}(A)|.$$

If $\beta(k) \to 0$ as $k \to \infty$ the process $\{X_i\}$ is said to be absolutely regular. In conjunction with equation (2), an easy argument shows that $(1/l)\log \hat{N}_l(X) \to H$ in probability $\hat{\mathbb{P}}$ if $\{X_I\}$ is absolutely regular.

We next examine the relationship of our condition and condition 2) of Theorem B. Suppose that $\{X_n\}$ satisfies condition 2), i.e., $X_{-\infty}^0, X_1^k, X_{k+1}^\infty$ is a Markov chain under $\mathbb{P}$. For $-\infty < n < \infty$, define $S_n = X_{n-k+1}^n$. The transformation $\{X_n\} \to \{S_n\}$ is one-to-one. Further, $\{S_n\}$ is a state sequence of a finite-state Markov chain (with state space A^k) with transition matrix say P. Since the $\{X_n\}$ are stationary and ergodic, all of the states in the Markov chain are recurrent, and the chain is irreducible. Denote the eigenvalues of P by λ_i, $1 \le i \le |A|^k$, ordered so that $|\lambda_1| \ge |\lambda_2| \ge \cdots$. Then, $\lambda_1 = 1$. Further, $|\lambda_2| < 1$, if and only if the Markov chain is aperiodic. We now state the following proposition.

Proposition: If condition 2) of Theorem B holds, and the corresponding Markov chain is aperiodic, then $\alpha(j) \le c|\lambda_2|^j = o(j^{-(1+\delta)})$.

Proof: Let $A^- \in \sigma(X_{-\infty}^0)$, $A^+ \in \sigma(X_j^\infty)$. Then,

$$\begin{aligned}&\mathbb{P}(A^- \cap A^+) \\ &\quad = \sum_{s_1, s_{j-1}} \mathbb{P}(A^-, A^+, S_1 = s_1, S_{j-1} = s_{j-1}) \\ &\quad \overset{(a)}{=} \sum_{s_1, s_{j-1}} \mathbb{P}(A^-|s_1)\mathbb{P}(A^+|s_{j-1})\mathbb{P}(s_1, s_{j-1}) \\ &\quad = \sum_{s_1, s_{j-1}} \mathbb{P}(A^-|s_1)\mathbb{P}(s_1)\mathbb{P}(A^+|s_{j-1})\mathbb{P}(s_{j-1}|s_1). \qquad (9)\end{aligned}$$

Step (a) follows from the Markov condition 2). Now it is a property of Markov chains (see, e.g., [3, Section XVI.1, (1.9)] that

$$\mathbb{P}(s_{j-1}|s_1) = \mathbb{P}(s_{j-1})(1 + \epsilon_j),$$

where $|\epsilon_j| \le c|\lambda_2|^j$. Thus, (9) becomes

$$\begin{aligned}\mathbb{P}(A^- \cap A^+) &= \left(\sum_{s_1} \mathbb{P}(A^-|s_1)\mathbb{P}(s_1)\right) \\ &\quad \cdot \left(\sum_{s_{j-1}} \mathbb{P}(A^+|s_{j-1})\mathbb{P}(s_{j-1})\right)(1 + \epsilon_j) \\ &= \mathbb{P}(A^-)\mathbb{P}(A^+)(1 + \epsilon_j).\end{aligned}$$

Thus,

$$|\mathbb{P}(A^- \cap A^+) - \mathbb{P}(A^-)\mathbb{P}(A^+)| \le |\epsilon_j| \le c|\lambda_2|^j,$$

establishing the proposition. □

III. Dithering and a Counter-Example

Condition 1) of Theorem B is somewhat restrictive: for many processes with memory, $\hat{\mathbb{P}}$ is not absolutely continuous with respect to $\mathbb{P}$. In many cases, however, one can modify the process $\{X_i\}_{i=-\infty}^\infty$ slightly in order to obtain a new process $\{Z_i\}_{i=-\infty}^\infty$ for which condition 1) holds. The procedure by which one modifies the original process is known as "dithering;" it is illustrated in the following example.

Example 1: Let $\mathbb{P}$ define a binary Markov process $\{X_i\}_{i=-\infty}^\infty$ for which a sequence of two consecutive ones is not allowed. Then, $\mathbb{P}\{X_0^1 = (1,1)\} = 0$ and $\hat{\mathbb{P}}\{X_0^1 = (1,1)\} > 0$, so that condition 1) is not satisfied. Now let $\{Y_i\}_{i=-\infty}^\infty$ be an i.i.d. sequence of binary random variables with $\Pr\{Y_i = 1\} = \epsilon$, ϵ small. Define the *dithered* process $\{Z_i = X_i \oplus Y_i\}_{i=-\infty}^\infty$, where $\oplus$ denotes modulo-2 addition. It is easy to see that the distribution $\mathbb{Q}$ of $\{Z_i\}$ satisfies $\hat{\mathbb{Q}} \ll \mathbb{Q}$. Moreover, X_k and Z_k agree with probability $(1 - \epsilon)$, so we have not changed the original process very much.

In a similar fashion, one can dither any sequence of random variables taking values in a finite set A. Will dithering always yield a process $\{Z_i\}$ with a distribution $\mathbb{Q}$ satisfying $\hat{\mathbb{Q}} \ll \mathbb{Q}$? The answer, unfortunately, is no. As the previous example indicates, dithering will eliminate problems caused by constraints arising from short-range memory in the process. This is a consequence of the fact that in the dithered process every finite-length sequence has positive probability. However, for long-range dependencies, those that persist along an entire sample sequence, dithering may not be effective. This is illustrated in the following (counter) example.

Example 2: Let $\{U_i\}_{i=-\infty}^\infty$, θ be independent identically distributed Bernoulli 1/2 random variables. Replicate each of the random variables U_i to get a process $\{V_j\}_{j=-\infty}^\infty$ with $V_{2i} = V_{2i+1} = U_i$. Finally, define $\{X_i = V_{i+\theta}\}_{i=-\infty}^\infty$. Every realization of the process $\{X_i\}$ is a sequence of pairs 00 or 11 with (random) phase θ: if $\theta = 0$, pairs begin at even-numbered time instants; if $\theta = 1$, they begin at odd-numbered time instants. It can be shown that the process $\{X_i\}$ is stationary and ergodic (cf. [1]).

Fix $0 < \delta < 1/16$ and let $\epsilon > 0$ be such that $(1 - \epsilon)^2 \ge 1 - \delta$. Let $\{Y_i\}_{i=-\infty}^\infty$ be a sequence of i.i.d. binary random variables with $\Pr\{Y_i = 1\} = \epsilon$. Define the dithered process $\{Z_i = X_i \oplus Y_i\}$, with distribution $\mathbb{Q}$, as before. We will exhibit a set A for which $\mathbb{Q}(A) = 0$ while $\hat{\mathbb{Q}}(A) > 0$. By the ergodic theorem, the random variables

$$S = \lim_{n\to\infty} 1/n \sum_{k=0}^{n-1} X_{2k}X_{2k+1} \quad \text{and}$$

$$T = \lim_{n\to\infty} 1/n \sum_{k=0}^{n-1} Z_{2k}Z_{2k+1}$$

are well defined with probability 1. Now $X_{2k}X_{2k+1}$ and $Z_{2k}Z_{2k+1}$ can differ only if $(Y_{2k}, Y_{2k+1}) \ne (0,0)$, so

$$\begin{aligned}&\left|\frac{1}{n}\sum_{k=0}^{n-1} X_{2k}X_{2k+1} - \frac{1}{n}\sum_{k=0}^{n-1} Z_{2k}Z_{2k+1}\right| \\ &\qquad \le 1 - \frac{1}{n}\sum_{k=0}^{n-1} I_{\{(0,0)\}}(Y_{2k}, Y_{2k+1}),\end{aligned}$$

where I_C denotes the indicator function of the event C. Letting $n \to \infty$ we see that with probability 1,

$$|S - T| \le 1 - \Pr\{(Y_0, Y_1) = (0,0)\}$$

$$= 1 - (1 - \epsilon)^2 \le \delta.$$

From the definition of the process $\{X_i\}$, it follows that

$$S = \begin{cases} 1/2, & w/p1, \quad \text{if } \theta = 0, \\ 1/4, & w/p1, \quad \text{if } \theta = 1, \end{cases}$$

so T is within δ of $1/2$ or $1/4$, depending on the phase of the process $\{X_i\}$.

Now define the counterpart of T for negative time, namely

$$T' = \lim_{n \to \infty} 1/n \sum_{k=-n}^{-1} Z_{2k} Z_{2k+1},$$

and let $R = (T + T')/2$. The analysis above of the random variable T applies also to T': under the distribution $\mathbb{Q}$ both T and T' are either within δ of $1/4$, or within δ of $1/2$, because $\{X_i\}$ maintains the same phase for positive and negative time. Since $\delta < 1/16$, we have

$$\mathbb{Q}\{R \in (3/8 - \delta, 3/8 + \delta)\} = 0. \tag{9}$$

Under $\hat{\mathbb{Q}}$, T and T' are independent, so there is some positive probability that $T \in (1/4 - \delta, 1/4 + \delta)$ and $T' \in (1/2 - \delta, 1/2 + \delta)$, in which case $R \in (3/8 - \delta, 3/8 + \delta)$. Hence,

$$\hat{\mathbb{Q}}\{R \in (3/8 - \delta, 3/8 + \delta)\} > 0. \tag{10}$$

It follows from (9) and (10) that $\hat{\mathbb{Q}}$ is not absolutely continuous with respect to $\mathbb{Q}$.

References

[1] P. Billingsley, *Ergodic Theory and Information.* New York: John Wiley, 1965.

[2] R. C. Bradley, "Basic properties of strong mixing conditions," in *Progress in Probability and Statistics*, E. Eberlein and M. S. Taqqu, Eds. Boston: Birkhäuser, 1986, pp. 165–192.

[3] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: John Wiley, 1957.

[4] A. Wyner and J. Ziv, "Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1250–1258, Nov. 1989.

[5] ——, "Fixed data base version of the Lempel–Ziv data compression algorithm," to appear in *IEEE Trans. Inform. Theory.*

Optimum Codes of Dimension 3 and 4 Over GF(4)

M. C. Bhandari and M. S. Garg

***Abstract*—$n_q(k, d)$, the length of a q-ary optimum code for given k and d, for $q = 4$ and $k = 3, 4$ is discussed. The problem is completely solved for $k = 3$ and exact value of $n_4(4, d)$ is determined for all but 52 values of d.**

***Index Terms*—Optimum codes, minimal length, linear codes, q-ary codes, codes over GF(4).**

Manuscript received July 12, 1990; revised January 17, 1992.

The authors are with the Department of Mathematics, Indian Institute of Technology, Kanpur-208016, India.

IEEE Log Number 9108035.

I. Introduction and Preliminary Results

Let GF(q) denote the Galois field of q elements ($q = p^m$, for some prime p and some positive integer m). An $[n, k, d]$ code is a k-dimensional linear subspace of GF$(q)^n$ over GF(q) with minimum distance d. Let $n_q(k, d)$ be the minimum value of n for which an $[n, k, d]$ code exists. Many researchers have worked on determining $n(k, d) = n_2(k, d)$, and it is known for $k \le 7$ [1], [17]. Bounds on $n(k, d)$ for $k = 8$ and 9 are given in [2], [6], and [8]. In 1965, Solomon and Stiffler [15] generalized the result of Griesmer and gave the following lower bound, called the Griesmer bound for $n_q(k, d)$:

$$n_q(k, d) \ge \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil \equiv g_q(k, d), \tag{1}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. A $[g_q(k, d), k, d]$ code if exists is called a code meeting the Griesmer bound, while an $[n_q(k, d), k, d]$ code is called an optimum code.

It was shown independently by Hamada and Tamari [12] and Dodunekov [4] that the bound given by (1) is reached for $k = 2$. In recent years Hamada [9] in cooperation with Deza [10], Helleseth [11], and Tamari [12], [13] has characterized many values of d for which $n_q(k, d) = g_q(k, d)$, $k \ge 3$ by relating the problem to that of characterizing min. hypers in a finite projective geometry. Some basic general results on $n_q(k, d)$ are given by Dodunekov [4]. He has shown that codes of dimension 3 need not meet the Griesmer bound [5]. Intervals for d in which equality (inequality) holds are also given.

In Section II, we determine $n_4(3, d)$ for all d. In Section III, we discuss the known results and a certain relation between codes over GF(p^m) and codes over GF(p^s). In Section IV, we obtain further improvements by constructing new codes. We conclude by collecting a table of bounds on $n_q(4, d)$.

Let $s = [d/(q-1)q^{k-1}]$ and let $s(q-1)q^{k-1} - d = \sum_{i=o}^{p} a_i q^{u_i - 1}$, $k = u_0 > u_1 > u_2 \cdots > u_p \ge 1$, $0 \le a_0 \le q - 2$, $1 \le a_i \le q - 1$, for $i = 1, 2, \cdots, p$. Then, the following four theorems summarize results from [4] and [5].

Theorem 1[4]: For given d, k, and q, the optimum code meets the Griesmer bound if any one of the following holds:

$$a_0 = 0 \quad \text{and} \quad \sum_{i=1}^{p} u_i \le sk \tag{2}$$

$$a_0 = 0, u_{i+1} = u_i - 1, \text{ for } i = s, s+1, \cdots, p-1 \quad \text{and}$$

$$n_q(u_s + 1, d_1) = g_q(u_s + 1, d_1), \qquad \text{for } d_1 = q^{u_p - 1} + \sum_{i=s}^{p} (q - 1 - a_i) q^{u_i - 1} \tag{3}$$

$$a_0 > 0 \quad \text{and} \quad s(q-1) \ge \sum_{i=0}^{p} a_i \tag{4}$$

$$d > [(k-2)(q-1) - (q-2)]q^{k-1} - 2q, \; k \ge 3 \tag{5}$$

$$q = 2^m, \; k = 3 \quad \text{and} \quad d \le q + 2. \tag{6}$$

Theorem 2[5]: $n_q(3, q(q-2)) = 1 + g_q(3, q(q-2))$.

Theorem 3[4]: If $k \ge q \ge 3$ and $2q^i < d \le q^{i+1}$ for $0 \le i \le k - q$, then $n_q(k, d) > g_q(k, d)$.

0018-9448/92$03.00