On the Symmetry of Good Nonlinear Codes

ROBERT J. McELIECE, MEMBER, IEEE

***Abstract*—It is shown that there are arbitrarily long "good" (in the sense of Gilbert) binary block codes that are preserved under very large permutation groups. This result contrasts sharply with the properties of linear codes: it is conjectured that long cyclic codes are bad, and known that long affine-invariant codes are bad.**

Introduction

IT IS the object of this paper to show that there exist arbitrarily long binary codes that have large permutation groups and also satisfy the Gilbert bound. Kasami's result[1] that long linear codes of length $n = 2^m$, which admit the affine group of order $n(n-1)$, cannot satisfy the Gilbert bound, makes it extremely unlikely that our result can be extended to linear codes. What our result does show, however, is that it cannot be merely the presence of a moderate-sized permutation group that forces a code to be bad, just as linearity does not necessarily degrade performance. The precise statement of our result is given in Theorem 1.

Theorem 1

For each $n \geq 1$, suppose P_n is a group of permutations on $\{1, 2, \cdots, n\}$, of order p_n, and that each nonidentity permutation in P_n has $\leq f_n$ fixed points.
We assume that

$$\log p_n = o(n),$$

$$f_n = o(n).$$

Then for any $0 < D < 1/2$, there exists a sequence of codes C_n (C_n has block length n) with rates R_n, so that

C_n admits P_n as a permutation group;
$d_{\min}(C_n) \geq Dn$;
$\liminf_{n\to\infty} R_n \geq 1 - H_2(D)$,

where $H_2(\cdot)$ is the binary entropy function.

For example, Theorem 1 permits the following choices of groups P_n.

P_n = cyclic group of order n.

$$P_n = \begin{cases} \text{affine group of order } n(n-1) \text{ if } n = 2^m \text{ for some } m. \\ \{1\} \text{ if } n \neq 2^m. \end{cases}$$

$$P_n = \begin{cases} \text{projective unimodular group PSL } (2, 2^m) \text{ if } n = 2^m \text{ for some } m. \\ \{1\} \text{ if } n \neq 2^m. \end{cases}$$

The proof of Theorem 1 is based on the usual proof of the Gilbert bound, and also on Theorem 2.

Theorem 2

Let

$$\pi = \begin{pmatrix} 1, & 2, & \cdots, & n \\ \pi(1), & \pi(2), & \cdots, & \pi(n) \end{pmatrix}$$

be a permutation without fixed points: $\pi(i) \neq i$ for all i. For each binary vector $\vec{x} = (x_1, x_2, \cdots, x_n)$ let $\pi\vec{x} = (x_{\pi(1)}, x_{\pi(2)}, \cdots, x_{\pi(n)})$, and define $N_\pi(\alpha, n)$ = number of vectors $\vec{x}$ for which $d(\vec{x}, \pi\vec{x}) \leq \alpha n$ (Hamming distance). Then if $\alpha \leq 1/2$,

$$N_\pi(\alpha, n) \leqq 2^{(n/2)(1+H_2(\alpha))}.$$

In Section II, we show how Theorem 2 may be used to prove Theorem 1. Section III is devoted to the proof of Theorem 2.

II. Theorem 2 Implies Theorem 1

Lemma 1

If P is a group of permutations on $\{1, 2, \cdots, n\}$, and if each nonidentity permutation of P has $\leq f$ fixed points, let $N_P(\alpha, n)$ = number of vectors $\vec{x}$ with $d(\vec{x}, \pi\vec{x}) \leq \alpha n$ for some $1 \neq \pi \in P$. Then if $\alpha \leq 1/2$, $f \leq n(1-2\alpha)$,

$$N_P(\alpha, n) \leq |P| \, 2^E,$$

where

$$E = f + \tfrac{1}{2}(n-f)\left[1 + H_2\left(\frac{\alpha n}{n-f}\right)\right].$$

Proof: Obviously

$$N_P(\alpha, n) \leq (|P| - 1) \max_{1 \neq \pi} N_\pi(\alpha, n).$$

Next, observe that if a permutation π has f_π fixed points, π acts as a permutation π^* without fixed points on $n - f_\pi$ points, so that $d(\vec{x}, \pi\vec{x}) = d(\vec{x}^*, \pi^*\vec{x}^*)$, where $\vec{x}^*$ is a vector of length $n - f_\pi$ obtained from $\vec{x}$ by deleting those co-

Manuscript received July 23, 1969; revised November 3, 1969. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, sponsored by the National Aeronautics and Space Administration under Contract NAS 7-100.

The author is with Jet Propulsion Laboratory, Pasadena, Calif. 91103.

[1] T. Kasami, "An upper bound on k/n for affine-invariant codes with fixed d/n," *IEEE Trans. Information Theory* (Correspondence), vol. IT-15, pp. 174–176, January 1969.

ordinates fixed by π. Thus

$$N_\pi(\alpha, n) = 2^{f_\pi} N_{\pi^*}\left(\frac{\alpha n}{n - f_\pi}, n - f_\pi\right),$$

and so by Theorem 2, $N_\pi(\alpha, n) \leq 2^{E(f_\pi)}$, where

$$E(f_\pi) = f_\pi + \frac{n - f_\pi}{2}\left[1 + H_2\left(\frac{\alpha n}{n - f_\pi}\right)\right].$$

Finally, it is easy to see that $E(f)$ is an increasing function of f in the range $0 \leq f \leq n(1 - 2\alpha)$, so that if $f_\pi \leq f$ for all $\pi \in P$, $N_\pi(\alpha, n) \leq 2^{E(f)}$. Q.E.D.

Using Lemma 1, we now can prove Theorem 1.

For each n, let C_n be a code of largest possible rate R_n subject to the two restraints of having minimum distance $\geq Dn$ and being invariant under P_n. (The code consisting of $00 \cdots 0$ and $11 \cdots 1$ satisfies these hypotheses, so there is no doubt about the existence of such codes.) Our goal is to show that $\liminf R_n \geq 1 - H(D)$.

We enclose each word $\vec{x} \in C_n$ within a sphere of radius $[Dn]$. The total volume occupied by these spheres is, at most,

$$2^{nR_n} \sum_{k \leq Dn} \binom{n}{k} \leq 2^{n(R_n + H(D))},$$

and so there are at least $2^n - 2^{n(R_n+H(D))}$ words at a distance $\geq Dn$ from C_n. On the other hand, from Lemma 1, for sufficiently large n the number of vectors $\vec{x}$ with $d(\vec{x}, \pi\vec{x}) < Dn$ for some $\pi \in P$ is no more than $2^{nE(D,n)}$ where

$$E(D, n) = \phi_n + \frac{1 - \phi_n}{2}\left[1 + H\left(\frac{D}{1 - \phi_n}\right)\right] + \frac{1}{n}\log_2 p_n$$

with

$$\phi_n = \frac{1}{n} f_n.$$

Since R_n is assumed to be as large as possible, each $\vec{x}$ that is at distance $\geq Dn$ from C_n must have $d(\vec{x}, \pi\vec{x}) < Dn$ for some π, and so, in particular,

$$2^n - 2^{n(R_n+H(D))} \leq 2^{nE(D,n)}.$$

Hence,

$$2^{n(R_n+H(D))} \geq 2^n(1 - 2^{-n(1-E)})$$

$$n(R_n + H(D)) \geq n + \log_2(1 - 2^{-n(1-E)})$$

$$\geq n - \frac{1}{\ln 2} 2^{-n(1-E)}$$

$$\left(\text{since } \log_2(1 - y) \geq -\frac{1}{\ln 2} y\right).$$

Therefore,

$$R_n + H(D) \geq 1 - \frac{1}{\ln 2} 2^{-n(1-E)}.$$

But our hypotheses are $\phi_n \to 0$, $1/n \log p_n \to 0$, so that $E(D, n) \to \frac{1}{2}(1 + H(D)) < 1$ for all $D < \frac{1}{2}$. Hence $2^{-n(1-E(\alpha,n))} \to 0$ and so $\liminf R_n \geq 1 - H(D)$, as asserted.

III. Proof of Theorem 2

Let us write the permutation π in cycle form: $\pi = C_1 C_2 \cdots C_r$, with $|C_i| = n_i$. For example,

$$\begin{pmatrix}1234567\\4561732\end{pmatrix} = (14)(257)(36),$$

with $n_1 = 2$, $n_2 = 3$, $n_3 = 2$. Now let d_m = number of vectors $\vec{x}$ for which $d(\vec{x}, \pi\vec{x}) = m$.

Lemma 2

$$D(z) = \sum_{m=0}^{n} d_m z^m = (1 + z)^n \prod_{i=1}^{r}\left(1 + \left(\frac{1 - z}{1 + z}\right)^{n_i}\right).$$

Proof: Since $d(\vec{x}, \pi\vec{x}) = w(\vec{x} + \pi\vec{x})$, where w is the Hamming weight, what is needed is an analysis of the weight spectrum of the range of the linear operator $T_\pi : \vec{x} \to \vec{x} + \pi\vec{x}$. Since the range of T_π is the set of vectors that are orthogonal (inner product: $\vec{x} \cdot \vec{z} = \sum x_i z_i \pmod 2$) to the null space of T_π, let us first identify the null space. If $\vec{x} = \pi\vec{x}$, then $x_1 = x_{\pi(1)} = x_{\pi^2(1)} = \cdots$, etc., so that $T_\pi\vec{x} = 0$ if and only if $\vec{x}$ is constant on the cycles of π; that is, $i, j \in C_k$ implies $x_i = x_j$. Thus, a vector $\vec{x}$ can be orthogonal to every vector in the null space of T_π if and only if $\vec{x}$ has even weight on each cycle; that is, $\sum \{x_i \mid i \in C_j\} \equiv 0 \pmod 2$ for each j. Thus, the number of vectors of weight m in the range of T_π is the number of vectors of weight m that have even weight on each cycle; and this number is clearly

$$d'_m = \sum_{\substack{m_1 + m_2 + \cdots + m_r = m \\ m_i \text{ all even}}} \binom{n_1}{m_1}\binom{n_2}{m_2} \cdots \binom{n_r}{m_r}. \tag{1}$$

If

$$D_i(z) = \sum_{\text{even } m} \binom{n_i}{m} z^m = \tfrac{1}{2}((1 + z)^{n_i} + (1 - z)^{n_i}),$$

from (1) we see that d'_m is the coefficient of z^m in the product $D_1(z)D_2(z) \cdots D_r(z)$. Furthermore, since there are 2^r vectors $\vec{x}$ such that $\vec{x} + \pi\vec{x} = 0$, each vector of the form $\vec{x} + \pi\vec{x}$ has 2^r preimages; that is, $d_m = 2^r d'_m$. Hence,

$$D(z) = 2^r \frac{1}{2^r} \prod_{i=1}^{r} ((1 + z)^{n_i} + (1 - z)^{n_i})$$

$$= (1 + z)^n \prod_{i=1}^{r}\left(1 + \left(\frac{1 - z}{1 + z}\right)^{n_i}\right). \quad \text{Q.E.D.}$$

Lemma 3 (The Chernoff Bound)

If

$$D(z) = \sum_{m=0}^{n} d_m z^m,$$

and if $d_m \geq 0$ for all m, then for every $s > 0$, $0 < \alpha < 1$,

$$\sum_{m \leq \alpha n} d_m \leq 2^{s\alpha n} D(2^{-s}).$$

Proof:

$$2^{s\alpha n} D(2^{-s}) = \sum_m d_m 2^{s(\alpha n - m)} \geq \sum_{m \leq \alpha n} d_m,$$

since

$$2^{s(\alpha n - m)} \geq 1 \qquad m \leq \alpha n.$$

Q.E.D.

We can now prove Theorem 2 without difficulty. From Lemmas 1 and 3,

$$N_\pi(\alpha, n) \leq 2^{s\alpha n}(1 + 2^{-s})^n \prod_{i=1}^{r} \left(1 + \left(\frac{1 - 2^{-s}}{1 + 2^{-s}}\right)^{n_i}\right)$$

all $s > 0$.

First notice that if $\theta \geq 0$, $m \geq 2$, $1 + \theta^m \leq (1 + \theta^2)^{m/2}$, and since we have assumed that $n_i \geq 2$ for all i,

$$N_\pi(\alpha, n) \leq 2^{\alpha s n}(1 + 2^{-s})^n \left(1 + \left(\frac{1 - 2^{-s}}{1 + 2^{-s}}\right)^2\right)^{n/2}$$

$$= 2^{\alpha s n} 2^{n/2}(1 + 2^{-2s})^{n/2}.$$

It is easy to verify that the minimum of this last expression occurs at

$$s = -\tfrac{1}{2} \log_2 \left(\frac{\alpha}{1 - \alpha}\right),$$

and is $2^{(n/2)(1 + H(\alpha))}$.

Q.E.D.

Sequential Decoding of Systematic and Nonsystematic Convolutional Codes With Arbitrary Decoder Bias

EDWARD A. BUCHER, MEMBER, IEEE

Abstract—**This paper presents several results involving Fano's sequential decoding algorithm for convolutional codes. An upper bound to the ath moment of decoder computation is obtained for arbitrary decoder bias B and $a \leq 1$. An upper bound on error probability with sequential decoding is derived for both systematic and nonsystematic convolutional codes. This error bound involves the exact value of the decoder bias B. It is shown that there is a trade-off between sequential decoder computation and error probability as the bias B is varied. It is also shown that for many values of B, sequential decoding of systematic convolutional codes gives an exponentially larger error probability than sequential decoding of nonsystematic convolutional codes when both codes are designed with exponentially equal *optimum* decoder error probabilities.**

I. Introduction

A RECENT paper [1] extends Viterbi's [9] upper and lower bounds to error probability for convolutional codes to include systematic as well as nonsystematic convolutional codes with optimum decoding. These results are of the form

$$\exp\{-N_e[E_L(R) - 0(N_e)]\} \leq P(E) \leq f(L) \exp\{-N_e E_u(R)\} \qquad (1)$$

where N_e, the effective constraint length, is the number of channel symbols directly affected by a given information symbol after the symbol's first appearance in the codeword. The function $0(N_e)$ approaches zero as N_e approaches infinity and $f(L)$ is a linear function of the information block length L.

This paper examines the performance of sequential decoding with systematic and nonsystematic convolutional codes. The Fano algorithm [2] sequential decoder provides a practical yet powerful method of decoding convolutional codes [3]. A combination of results by Jacobs and Berlekamp [4], Savage [5], Falconer [6], and Jelinek [7] shows that the number of sequential decoder computations required to decode an information symbol in an infinite constraint length convolutional code has a Pareto distribution. Thus,

$$P[\text{number of computations} > N] \approx N^{-\alpha} \qquad (2)$$

for large N. The Pareto exponent α is obtained by noting that the ath moment of computation is bounded for all positive $a < \alpha$ and unbounded for $a \geq \alpha$. Thus, α is the smallest positive a for which the ath moment of computation is unbounded.

We show that there is a trade-off between sequential decoder computation and error probability for convolutional codes. This trade-off involves the value of the decoder bias B. Setting B equal to the data rate R maximizes α but gives an upper bound on error probability in which the sequential decoding upper-bound error exponent $E_{Us}(R, B)$ is substantially smaller than $E_U(R)$ the optimum decoder upper-bound error exponent. On the other hand, making B somewhat larger than R decreases error probability until $E_{Us}(R, B) = E_U(R)$, but this increase in B implies a smaller α and more decoder computation. We also find that $E_{Us}(R, B)$ is often substantially

Manuscript received April 1, 1969; revised January 6, 1970. This work is part of the author's Ph.D. dissertation, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass.

The author is with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass. 02173.