TDA Progress Report 42-98

August 15, 1989

On the Decoder Error Probability of Linear Codes

K.-M. Cheung
Communications Systems Research Section

In this article, by using coding and combinatorial techniques, an approximate formula for the weight distribution of decodable words of most linear block codes is evaluated. This formula is then used to give an approximate expression for the decoder error probability $P_E(u)$ *of linear block codes, given that an error pattern of weight* u *has occurred. It is shown that* $P_E(u)$ *approaches the constant* Q *as* u *gets large, where* Q *is the probability that a completely random error pattern will cause decoder error.*

I. Introduction

Coding is used in a digital communication system to detect and correct errors introduced in the data stream by channel noise. An important parameter to evaluate the performance of a code is its decoder error probability. Let C be a linear (n, k, d) code over $GF(q)$, and $C^{\perp}$ be its $(n, n-k, d^{\perp})$ dual code. Let t be the number of errors the code is designed to correct. Let G be the generator matrix of C. Let A_u denote the number of codewords of weight u, and D_u denote the number of decodable words of weight u. Decodable words are defined as all words lying within distance t from a codeword. If the decoder is assumed to be a bounded distance decoder, then the weight distribution for the decodable words can be used to find the decoder error probability of the code.

When a codeword $\underline{c} \in C$ is transmitted over a communication channel, channel noise may corrupt the transmitted signals. As a result, the receiver gets a corrupted version of the transmitted codeword $\underline{c} + \underline{e}$, where $\underline{e}$ is an error pattern of some weight u. If $u \leq t$, then a bounded distance decoder on the receiver's end detects and corrects the error $\underline{e}$, and recovers $\underline{c}$. If $u > t$, the decoder fails, and it either

(1) Detects the presence of the error pattern $\underline{e}$, but is unable to correct it, or

(2) Misinterprets the received pattern $\underline{c}+\underline{e}$ for some other codeword $\underline{c}'$ if the received pattern falls into the radius t of the Hamming sphere around $\underline{c}'$.

Case (2) is, in most cases, more serious than case (1). This can occur when an error pattern $\underline{e}$ is of weight $u \geq d - t$. As pointed out in [1] and [2], if all error patterns of weight u are equally probable, the decoder error probabil-

ity given that an error pattern of weight u occurs, denoted by $P_E(u)$, is given by the following expression:

$$P_E(u) = \frac{D_u}{\binom{n}{u}(q-1)^u} \qquad d-t \le u \le n \tag{1}$$

In this article, by using combinatorial and coding techniques, an approximate formula for the weight distribution of decodable words for most linear block codes is evaluated. This formula together with Eq. (1) gives an approximate formula for the decoder error probability $P_E(u)$ for most linear block codes. It is also shown that

$$P_E(u) \overset{\text{large u}}{\longrightarrow} Q$$

where Q is the probability that a completely random error pattern will cause decoder error. That is,

$$Q = \frac{(q^k-1)V_n(t)}{q^n} \approx q^{-r}V_n(t) \tag{2}$$

where $r = n-k$ is the code's redundancy and $V_n(t) = \sum_{i=0}^{t}\binom{n}{i}(q-1)^i$ is the volume of a Hamming sphere of radius t.

II. Mathematical Preliminaries

In this section, combinatorial and coding techniques required to derive the results in later sections are introduced. These techniques are similar to those used in [6] to obtain the weight distribution of linear block codes.

A. Principle of Inclusion and Exclusion

Let χ be a set of N objects, and $P(1)$, $P(2), \ldots, P(u)$ be a set of u properties. Let $N(i_1, i_2, \ldots, i_r)$ be the number of objects with properties $P(i_1)$, $P(i_2), \ldots, P(i_r)$. The number of objects $N(\emptyset)$ with none of the properties is given by [3]:

$$N(\emptyset) = N - \sum_i N(i) + \sum_{i_1<i_2} N(i_1,i_2) + \ldots + (-1)^r \times \sum_{i_1<i_2\ldots<i_r} N(i_1,i_2,\ldots,i_r) + \ldots + (-1)^u N(1,2,3,\ldots,u) \tag{3}$$

There are $u+1$ terms on the right-hand side of Eq. (3), with the 0th term representing the total number of objects in χ. If all terms beyond the rth term on the right-hand side of Eq. (3) are ignored, then the resulting truncated sum is an upper bound when r is even, or a lower bound if r is odd. Thus the maximum error magnitude introduced in the inclusion and exclusion formula by ignoring all terms beyond the rth term does not exceed the magnitude of the $(r+1)$th term. This fact will be used later to upper bound the magnitude of the errors of the approximate weight distribution formula.

B. Facts on Coding Theory

A linear (n,k,d) code over $GF(q)$ can be generated by a $k \times n$ generator matrix G, not necessarily unique and such that $\text{rank}(G) = k$. Let l be the maximum number such that no l or fewer columns of G add to zero. Then

$$l \le k \tag{4}$$

Equality in Eq. (4) is achieved in the case of *maximum distance separable* (MDS) codes. Since G is the parity-check matrix of $C^\perp$, $l = d^\perp - 1$. Let col_{i_1}, $\text{col}_{i_2}, \ldots, \text{col}_{i_j}$ be any j particular columns of G, $j \le l \le k$. It is obvious that there exists a $k \times n$ generator matrix G' of C and a $k \times k$ nonsingular matrix K such that

$$G' = KG \tag{5}$$

and col_{i_1}, $\text{col}_{i_2}, \ldots, \text{col}_{i_j}$ of G' form a $k \times j$ submatrix of the form $\begin{pmatrix} \text{I} \\ \ldots \\ 0 \end{pmatrix}$. This fact guarantees that given any pattern of j symbols on the i_1th, i_2th, ..., i_jth coordinates, the number of codewords with the j-symbol pattern on the i_1th, i_2th, ..., i_jth coordinates equals q^{k-j} for $j \le l$. This fact is important in the next section to evaluate the cardinalities of some sets of decodable words.

III. Derivation of Formulae

Let D be the set of decodable words of C. Let $\underline{d}$ be a decodable word with Hamming weight u, $u \ge n-l$. Let the coordinates of $\underline{d}$ be indexed by $\{0,1,\ldots,n-1\}$. Then $\underline{d}$ has v zeros ($v \le l$), where $v = n-u$. Let V be a set of v coordinates, $|V| = v$. Let $\{i_1, i_2, \ldots, i_j\} \subseteq \{0,1,\ldots,n-1\} - V$ be a set of j coordinates. Define $S(i_1, i_2, \ldots, i_j) = \{\underline{d}:$

$\underline{d} \in D$ and $\underline{d}$ has zeros in $V \cup \{i_1, i_2, \ldots, i_j\}\}$. A decodable word $\underline{d} \in S(i_1, i_2, \ldots, i_j)$ always has at least $v + j$ zeros. For $0 \leq j \leq l - v$, the number of zeros in the decodable words of $S(i_1, i_2, \ldots, i_j)$ is less than or equal to l. Now, since all words lying within the Hamming spheres (with volume $V_n(t)$) that surround codewords are decodable words, there are $V_n(t)$ disjoint cosets that contain decodable words. Each coset can be constructed by adding a coset leader $\underline{a}$ (Hamming weight of $\underline{a} \leq t$) to each codeword in C. Thus from the discussion in Section II.B, for each of the $V_n(t)$ different coset leaders (each corresponding to a coset), there are q^{k-v-j} codewords in C which, when added to the coset leader, give decodable words with zeros in the $v + j$ coordinates. The number of decodable words in $S(i_1, i_2, \ldots, i_j)$ is then given by

$$|S(i_1, i_2, \ldots, i_j)| = q^{k-v-j} V_n(t) \quad \text{for } 0 \leq j \leq l - v \tag{6}$$

For $l - v + 1 \leq j \leq n - v - d + t$, the number of zeros in the decodable words of $S(i_1, i_2, \ldots, i_j)$ exceeds l, and apparently there is no simple expression for $|S(i_1, i_2, \ldots, i_j)|$. For $n - v - d + t + 1 \leq j \leq n - v - t$, the number of zeros in a decodable word is greater than or equal to $n - d + t + 1$, but less than or equal to $n - t$. Thus any decodable word in $S(i_1, i_2, \ldots, i_j)$ has weight less than or equal to $d - t - 1$. It is not hard to see that the elements of $S(i_1, i_2, \ldots, i_j)$ cannot be decoded into a codeword of weight other than $\underline{0}$. Therefore, $S(i_1, i_2, \ldots, i_j)$ contains all words of weight less than or equal to t in the coordinates $\{0, 1, \ldots n - 1\} - (V \cup \{i_1, i_2, \ldots, i_j\})$. Thus,

$$|S(i_1, i_2, \ldots, i_j)| = \sum_{i=0}^{t} \binom{u - j}{i} (q - 1)^i \quad \text{for } n - v - d + t + 1 \leq j \leq n - v - t \tag{7}$$

For $n - v - t + 1 \leq j \leq n - v$, since j is greater than or equal to $n - v - t + 1$, the number of zeros $v + j$ is greater than or equal to $n - t + 1$. Therefore, the number of nonzero components is less than t. Thus, all words with zeros on $V \cup \{i_1, i_2, \ldots, i_j\}$ are decodable and thus

$$|S(i_1, i_2, \ldots, i_j)| = q^{u-j} \quad \text{for } n - v - t + 1 \leq j \leq n - v \tag{8}$$

In the cases for $0 \leq j \leq l - v$, $n - v - d + t + 1 \leq j \leq n - v - t$, and $n - v - t + 1 \leq j \leq n - v$, the set $i_1, i_2, \ldots, i_j$ can be chosen arbitrarily from a set of $u = n - v$ coordinates. Thus for every choice of j, there are $\binom{u}{j}$ sets $S(i_1, i_2, \ldots, i_j)$. By the principle of inclusion and exclusion, the number of decodable words with exactly v zeros in V, which is denoted by D'_V, is:

$$\begin{aligned} D'_V &= |S(\emptyset)| + (-1) \sum_{i_1} |S(i_1)| + \ldots + (-1)^r \\ &\quad \times \sum_{i_1 < i_2 < \ldots < i_r} |S(i_1, i_2, \ldots, i_r)| \\ &\quad + \ldots + (-1)^{n-v} S(i_1, i_2, \ldots, i_{n-v}) \\ &= \sum_{j=0}^{l-v} (-1)^j \binom{u}{j} q^{k-v-j} V_n(t) + \sum_{j=l-v+1}^{n-v-d+t} (-1)^j \\ &\quad \times \sum_{i_1 < i_2 < \ldots < i_j} |S(i_1, i_2, \ldots, i_j)| \\ &\quad + \sum_{j=n-v-d+t+1}^{n-v-t} (-1)^j \binom{u}{j} \sum_{i=0}^{t} \binom{n - v - j}{i} (q - 1)^i \\ &\quad + \sum_{j=n-v-t+1}^{n-v} (-1)^j \binom{u}{j} q^{u-j} \end{aligned} \tag{9}$$

If all the terms beyond the $l - v - 1$ terms are ignored in the above inclusion and exclusion formula, Eq. (9) is reduced to

$$D'_V = \sum_{j=0}^{l-v-1} (-1)^j \binom{u}{j} q^{k-v-j} V_n(t) + E_1 \tag{10}$$

where

$$E_1 = (-1)^{l-v}\binom{u}{l-v}q^{k-l}V_n(t)$$

$$+\sum_{j=l-v+1}^{n-v}\sum_{i_1<\cdots<i_j}(-1)^j|S(i_1,\ldots,i_j)|$$

and $|E_1| \le \binom{u}{l-v}q^{k-l}V_n(t)$ (from the discussion in Section II.A). If $E_2 = \sum_{j=l-v}^{u}(-1)^j\binom{u}{j}q^{k-v-j}V_n(t)$ is added and subtracted from Eq. (10), one has

$$D'_V = \frac{(q-1)^u}{q^{n-k}}V_n(t) + E_1 + E_2 \qquad (11)$$

If $\binom{u}{l-v}q \ge \binom{u}{l-v+1}$, that is, if $u \ge \frac{q+1}{q}(n-l)-1$, E_2 is a sum of terms with alternate signs and descending magnitude. Then $|E_2| \le \binom{u}{l-v}q^{k-l}V_n(t)$. Thus

$$D'_V = \frac{(q-1)^u}{q^{n-k}}V_n(t) + E \qquad (12)$$

where $E = E_1 + E_2$ and $|E| \le 2\binom{u}{l-v}q^{k-l}V_n(t)$. D'_V can thus be approximated by $\frac{(q-1)^u}{q^{n-k}}V_n(t)$, and the goodness of approximation depends on how small the ratio $R = E/[(q-1)^u q^{-(n-k)}V_n(t)]$ is. By using the upper bound on $|E|$, an upper bound on this ratio is given by

$$R \le \frac{2\binom{u}{n-l}q^{k-l}}{(q-1)^u} \qquad (13)$$

Since $v \le l$, there are $\binom{n}{v} = \binom{n}{u}$ ways to choose v zeros from $\{0,1,\ldots,n-1\}$. Then D_u can be approximated by the following expression:

$$D_u = \sum_{|V|=n-u} D'_V \approx q^{-(n-k)}\binom{n}{u}(q-1)^u V_n(t) \qquad (14)$$

for $u \ge \max\{n-l, \frac{q+1}{q}(n-l)-1\}$.

Strictly speaking, the derivation of Eq. (14) is valid only for $u \ge \max\{n-l, \frac{q+1}{q}(n-l)-1\}$. However, it is observed that in most cases Eq. (14) is also a close approximation to D_u for u considerably smaller than $n-l$ (as in the case of Reed-Solomon codes). The upper bound of R derived above has a denominator term $(q-1)^u$ and this indicates that this approximation formula is good for nonbinary linear codes, and is not useful for binary linear codes. The looseness of this approximation for binary linear codes is best illustrated by extended binary codes that have only even weights. In the case of binary primitive codes, Kasami et al. [4] generalized Sidel'nikov's approach [5] and showed that the weights of most binary primitive codes have approximate binomial distribution.

Cheung [6] later showed this is also true for nonbinary codes. It is conjectured in this article that the approximate Eq. (14) for the weight distribution of decodable words is also good for binary primitive codes. For nonbinary linear codes, the upper bound on R shows that the approximation in Eq. (14) is particularly good for codes with large alphabet sets. The upper bound on R for the (31,15,17) Reed-Solomon code over $GF(32)$ is given in Table 1. The weight distribution of decodable words and its approximation (using Eq. 14) of the (31,15,17) Reed-Solomon code are given in Table 2.

Given the approximate formula of D_u, an approximate decoder error probability $P_E(u)$ is obtained by substituting Eq. (14) into Eq. (1). It is observed that $P_E(u)$ approximates the constant $Q = q^{-r}V_n(t)$ as u gets large, where Q is the probability that a completely random error pattern will cause decoder error. An upper bound of R given by Eq. (13) shows that $P_E(u)$ approaches Q "nearly exponentially" (for nonbinary codes) as u increases.

IV. Conclusion

In this article, by using the inclusion and exclusion principle, an approximate formula for the weight distribution of decodable words of most linear block codes is derived. The decoder error probability $P_E(u)$, which is a function of D_u, is then shown to approach the constant Q as u gets large, where Q is the probability that a completely random error pattern will cause decoder error.

References

[1] R. J. McEliece and L. Swanson, "On the Decoder Error Probability for Reed-Solomon Codes," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 701–703, 1986.

[2] K. Cheung, "More on the Decoder Error Probability of Reed-Solomon Codes," to appear in *IEEE Trans. Inform. Theory.*

[3] R. Stanley, *Enumerative Combinatorics, Vol. I*, Monterey, California: Wadsworth and Brooks-Cole, 1986.

[4] T. Kasami, T. Fujiwara, and S. Lin, "An Approximation to the Weight Distribution of Binary Linear Codes," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 6, pp. 769–780, November 1985.

[5] V. M. Sidel'nikov, "Weight Spectrum of Binary Bose-Chaudhuri-Hocquenghem Codes," *Probl. Peredachi Inform.*, vol. 7, no. 1, pp. 14–22, January–March, 1971.

[6] K. Cheung, "The Weight Distribution and Randomness of Linear Codes," *TDA Progress Report 42-97*, vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 208–215, May 15, 1989.