Binary constant-weight codes correcting localized errors and defects

Ahlswede, Rudolf; Bassalygo, L. A.; Pinsker, M. S.

Problems of Information Transmission, Vol. 30, No. 2, 1994

BINARY CONSTANT-WEIGHT CODES CORRECTING LOCALIZED ERRORS AND DEFECTS

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

UDC 621.391.1:519.2

We establish here the asymptotically optimal transmission rate of binary constant weight codes correcting localized errors and defects.

This work shows how the methods of [1] and of [2] can be combined to simultaneously extend the results of [1-3]. We assume that the readeris familiar with the definitions of codes correcting localized errors and of codes correcting defects (see, e.g., [1-4]).

We suppose that t errors can occur and l defects occur during the transmission of a binary code word of length n and weight w, and that the encoder knows the subset of positions $E(E \subseteq \{1,2,\cdots,n\},\ |E|=t)$, where errors are possible, and the subset of positions $Z_0,\ Z_1(Z_0, Z_1 \subseteq \{1,2,\cdots,n\},\ Z=Z_0\cup Z_1,\ Z_0\cap Z_1 = \emptyset,\ |Z|=l)$, where defects are equal to 0 and 1, respectively. Furthermore we assume that $E\cap Z=\emptyset$.

As usual in such problems the decoder knows the numbers t and l (the sizes of the subsets E and Z). But here, in contrast to the standard case of defects, we suppose that the decoder knows a little more: he also knows the number of positions where the defects are equal to 1, that is, the size of the subset $Z_1(|Z_1|=s,\ |Z_0|=l-s)$.

Let $w=\omega n$, $t=\tau n$, $l=\lambda n$, and let $R'(\omega,\tau,\lambda)$ be the asymptotically optimal transmission rate of a binary constant-weight-ω code of length n correcting t localized errors and l defects, when the decoder also knows the number s of positions where the defects are equal to 1. We state our result.

Theorem. *If $\lambda+2\tau>1$, then $R'(\omega,\tau,\lambda)=0$, and otherwise*

$$R'(\omega,\tau,\lambda)=\begin{cases}(1-\lambda)\left[h\left(\frac{\omega+\tau}{1-\lambda}\right)-h\left(\frac{\tau}{1-\lambda}\right)\right], & if\ 0\le\omega\le\frac{1-\lambda-2\tau}{2}\\ (1-\lambda)\left[1-h\left(\frac{\tau}{1-\lambda}\right)\right], & if\ \frac{1-\lambda-2\tau}{2}\le\omega\le\frac{1+\lambda+2\tau}{2}\\ (1-\lambda)\left[h\left(\frac{\omega-\lambda-\tau}{1-\lambda}\right)-h\left(\frac{\tau}{1-\lambda}\right)\right], & if\ \frac{1+\lambda+2\tau}{2}\le\omega\le 1,\end{cases}\qquad(1)$$

where $h(\gamma)=-\gamma\log_2\gamma-(1-\gamma)\log_2(1-\gamma)$.

PROOF. It is easy to see that $R'(\omega,\tau,\lambda)=0$ if $\lambda+2\tau>1$ and that it does not exceed the right side of (1) (it is sufficient to consider the case where all defects concentrate in the first l positions and to use the lemma from [2] for the remaining $n-l$ positions).

Now we estimate $R'(\omega,\tau,\lambda)$ from below and we begin with some notations and definitions. Let $\mathcal{M}=\{m\}$ be the set of messages ($|\mathcal{M}|=M$), let $\mathcal{E}_t=\{E\}$ be all possible subsets of positions where t errors can occur ($|\mathcal{E}_t|=C_n^t$), and let $\mathcal{Z}_{l,s}=\{\overline{Z}=(Z_0,Z_1)\}$ be all possible pairs of subsets of positions where l defects with s symbols 1 occur ($|\mathcal{Z}_{l,s}|=C_n^lC_n^s$). Since the decoder knows the number s of positions where the defects are equal to 1, it is sufficient to construct a code correcting t localized errors and l defects with s unit symbols.

Since a code word $x(m,E,\overline{Z})$ depends on the message $m\in\mathcal{M}$, the subset $E\in\mathcal{E}_t$, and the pair $\overline{Z}\in\mathcal{Z}_{l,s}$, the list of code words $X_m=\bigcup_{E\in\mathcal{E}_t,\overline{Z}\in\mathcal{Z}_{l,s}}\{x(m,E,\overline{Z})\}$ corresponds to the message m and the code X represents a collection of M lists X_m ($X=\{X_m : m\in\mathcal{M}\}$).

We say that a word $x=(x_1,\cdots,x_n)$ of weight r ($w+s\ge r\ge w$) is good for $E\in\mathcal{E}_t$, $\overline{Z}\in\mathcal{Z}_{l,s}$ if it satisfies the following conditions:

(a) $x_i=0$ when $i\in E\cup Z_0$

Translated from *Problemy Peredachi Informatsii*, Vol. 30, No. 2, pp. 10-13, April-June, 1994. Original article submitted May 25, 1993.

(b) w symbols 1 of the word x are outside of Z_1
(thus $r-w$ symbols 1 of the word x are inside of Z_1). (2)

We denote by $c(x, E, \overline{Z}) = (c_1, \cdots, c_n)$ the following word:

$$c_i = \begin{cases} j, & i \in Z_j,\ j = 0, 1 \\ 1, & i \in E \\ x_i, & i \notin Z \cup E. \end{cases}$$

Finally we say that a word $a = (a_1, \cdots, a_n)$ covers a word $b = (b_1, \cdots, b_n)$ if $a_i \geq b_i$, $i = 1, 2, \cdots, n$.

Now we shall prove the lemma below, which presents a nonstandard part of the proof of our theorem (for the first analogue see [4]). But before we do that, we would like to give a small heuristic explanation: for problems such as ours, the nearest code word decoding rule that usually accompanies a random choice of a code does not permit us to achieve the optimal transmission rate. However, if we can make the random choice of a code so that the number of possible concurrences of the transmitted messages during the decoding is not large, all of them are known to the encoder, and the number of code words fitting a triple $(m, E, \overline{Z})$ are large enough, then we can organize a correct transmission of messages using a special encoding and decoding.

Lemma. *Let p and q be natural numbers, let the size M of the set of messages satisfy $M \leq 2^n / C_n^t$, and let the constant-weight-r code $X = \{X_m : m \in \mathcal{M}\}$ satisfy the following conditions:*

(a) any word of weight $w + s + t$ covers at most q code words;

(b) any word belongs to at most p code lists;

(c) for any $m \in \mathcal{M}$, $E \in \mathcal{E}_t$, $\overline{Z} \in \mathcal{Z}_{l,s}$, there exists at least $n2^{pq}$ good words from X_m for $E, \overline{Z}$.

Then the code X corrects t localized errors and l defects with s symbols 1.

PROOF. First of all, the limitation $M \leq 2^n / C_n^t$ is quite natural since the Hamming bound is right for a code correcting localized errors [2]. Let

$$\mathcal{B} = \{(m, x) : m \in \mathcal{M},\ x \in X_m\},$$

$$\widetilde{\mathcal{B}}^2 = \{\{(m, x)\} \cup \{m', x')\} : (m, x) \in \mathcal{B},\ (m', x') \in \mathcal{B},\ m \neq m'\},$$

and let $\Phi = \{\varphi\}$ be the set of functions on $\widetilde{\mathcal{B}}^2$ taking in every element $\{(m, x)\} \cup \{(m', x')\} \in \widetilde{\mathcal{B}}^2$ the value m or m'. The number of these functions equals $2^{|\widetilde{\mathcal{B}}^2|}$. We shall call the function $\varphi \in \Phi$ good for the code X if it satisfies the following condition:

for any $m \in \mathcal{M}$, $E \in \mathcal{E}_t$, $\overline{Z} \in \mathcal{Z}_{l,s}$, there exists a good code word
$x \in X_m$ for $E, \overline{Z}$ with $\varphi\{\{(m, x)\} \cup \{(m', x')\}\} = m$
for all pairs (m', x'), where x' is covered by the word $c(x, E, \overline{Z})$.

It is not difficult to see that the code X corrects t localized errors and l defects with s symbols 1 if there exists a good function for it (this good function gives the natural encoding and decoding).

Since the number of bad functions does not exceed

$$M|\mathcal{E}_t||\mathcal{Z}_{l,s}|(2^{pq-1} - 1)^{n2^{pq}} \cdot 2^{|\widetilde{\mathcal{B}}^2| - n2^{pq}(pq-1)} \leq 6^n e^{-2n} 2^{|\widetilde{\mathcal{B}}^2|} < 2^{|\widetilde{\mathcal{B}}^2|},$$

a good function exists for the code X. The proof is complete.

Here we would like to draw attention to the following fact: the weight of the code words in the lemma is equal to r, and not to w as in the theorem. But for transmission we use only good words, and they have just w symbols 1 outside the positions $E \cup Z$ by the definition (2). Since nothing changes at the output of the channel if we transmit such words of weight w in place of the corresponding good words of weight r, it remains only to show the existence of a code satisfying the conditions of the lemma, whose transmission rate is asymptotically equal to the right side of (1).

This is proved by the standard method of a random choice, and we shall begin with the case when $\omega \leq \frac{1-\lambda-2\tau}{2}$. We consider all possible ordered collections of M lists with N words (with repetitions) each of weight r ($w \le r \le w+s$). The number of such collections is equal to $(C_n^r)^{MN}$. We shall say that a collection $Y = \{Y_k : k = 1, \cdots, M\}$ is good if it satisfies the following conditions:

(a) any word of weight $w+s+t$ covers at most q words in the collection;
(b) any word belongs to at most p lists in the collection;
(c) for any $k (k = 1, \cdots, M)$, $E \in \mathcal{E}_t$, $\overline{Z} \in \mathcal{Z}_{l,s}$, there exists a word from Y_k which is good for $E, \overline{Z}$ (recall the definition (2) of a good word for $E, \overline{Z}$). (3)

We shall now estimate from above the number of bad collections. The number of collections that do not satisfy:

1. condition (3a) does not exceed

$$C_n^{w+s+t} C_{MN}^{q+1} (C_{w+s+t}^r)^{q+1} (C_n^r)^{MN-q-1};$$

2. condition (3b) does not exceed

$$C_n^r C_{MN}^{p+1} (C_n^r)^{MN-p-1};$$

3. condition (3c) does not exceed

$$M C_n^t C_n^l C_l^s (C_n^r - C_{n-t-l}^w C_s^{r-w})^N (C_n^r)^{MN-N}.$$

It is not difficult to see that a good collection exists when

$$q = n^{\frac{1}{3}},\ p = n^{\frac{1}{3}},\ r = \frac{ws}{w+t} + w,\ N = \frac{n^2 C_n^r}{C_{n-t-l}^w C_s^{r-w}},$$

and

$$M = \frac{C_{n-l}^{w+t}}{n^3 2^{n^{\frac{2}{3}}} C_{n-l}^t} = 2^{(1-\lambda)[h(\frac{w+r}{1-\lambda}) - h(\frac{r}{1-\lambda})]n + 0(n)}. \tag{4}$$

Now, using a good collection for M messages, we shall easily construct the code satisfying the conditions of the lemma for $M/np2^{pg}$ messages: it is sufficient to partition M messages into $M/np2^{pq}$ groups of $np2^{pq}$ messages each and, accordingly, to join different good words of corresponding $np2^{pq}$ lists in one code list. By (4) the theorem is proved when $\omega \le \frac{1-\lambda-2\tau}{2}$.

In the interval $\frac{1-\lambda-2\tau}{2} \le \omega \le \frac{1}{2}$ the right side of (1) does not depend on the weight w. Thus, using the code from the lemma for the weight $\frac{n-l-2t}{2}$ we can construct a code with the same transmission rate for any weight w, $\frac{n-l-2t}{2} \le w \le \frac{n}{2}$, by adding the corresponding number of symbols 1 in the positions from E (code words of the lemma have 0 on these positions).

When $w \ge \frac{n}{2}$, it is enough to construct a code for the weight $n-w$ and afterwards to invert its words. The proof of the theorem is complete.

REFERENCES

1. L. A. Bassalygo and M. S. Pinsker, "Binary constant weight codes correcting localized errors," *Probl. Peredachi Inf.*, 28, No. 4, 3–6 (1992).
2. L. A. Bassalygo, S. I. Gelfand and M. S. Pinsker. "Coding for channels with localized errors," *Proc. Fourth Soviet-Swedish Workshop in Information Theory*, Gotland, Sweden, 95–99 (1989).
3. A. V. Kuznetsov and B. S. Tsybakov. "Coding for memory with defected cells," *Probl. Peredachi Inf.*, 10, No. 2, 52–60 (1974).
4. L. A. Bassalygo and M. S. Pinsker. "Error correction in channels with defects," *Dokl. Akad. Nauk SSSR*, 234, No. 6, 1361–1364 (1978).