Codes for Multiple Localized Burst Error Correction

Adrian Mardjuadi and Jos H. Weber, Member, IEEE

Abstract—A new construction method for codes correcting multiple localized burst errors is proposed. The codes obtained by this method improve upon codes presented by Larsson in size, while keeping the encoding and decoding complexity low. Like the Larsson codes, the proposed codes are asymptotically optimal when the number of bursts to be corrected is fixed and the correctable burst length grows linearly with the codeword length. Unlike the Larsson codes, the proposed codes are also asymptotically optimal when the number of bursts to be corrected and the correctable burst length are both fixed.

Index Terms—Burst errors, decoding, encoding, error correction, localized errors.

I. INTRODUCTION

In the concept of channels with localized errors, it is assumed that the encoder knows the possible error locations but not the error values, while the decoder has no a priori information about either error locations or error values. This concept has been introduced by Bassalygo, Gelfand, and Pinsker [2]. Important results in this fast emerging research field have been presented by the same authors [3], by Ahlswede, Bassalygo, and Pinsker [1], and by Larsson [4]. Possible applications of codes designed for correction of localized errors are in the areas of storage media and mobile communications [4].

In this correspondence, a construction method for codes correcting (multiple) localized burst errors is proposed. The codes obtained by this method improve upon the best known codes in size. The correspondence is organized as follows. First, notations and definitions are given in Section II. Then, the construction method is presented in Section III. Finally, the proposed method is analyzed in Section IV.

II. PROBLEM DEFINITION

We will mainly use the same notations as in [4]. The problem consists of transmitting messages from a message set $\mathcal{M}$ of size $M$ over a $q$-ary channel. For this purpose, we consider block codes of length $n$ over the alphabet $\mathcal{Q} = \{0, 1, \ldots, q - 1\}$. The set of positions in a codeword is denoted by $\mathcal{N} = \{1, 2, \ldots, n\}$. The set $\mathcal{E}$ is a subset of $\mathcal{N}$ consisting of positions where errors may occur during transmission over the channel and it is known a priori to the encoder but not to the decoder. It is assumed that no errors will occur outside the set $\mathcal{E}$. The set $\mathcal{E}$ is called the configuration of unreliable positions. In the case of multiple localized burst errors, which is the topic of this correspondence, the set $\mathcal{E}$ consists of the union of $m$ sets of $b$ consecutive positions.

Denote by $e_i(\mathcal{E})$ a possible error vector for a certain set $\mathcal{E}$ and by $e_i(\mathcal{E})$ the $i$th component of $e(\mathcal{E})$. The encoder does not know the error values but it knows that there will be no errors in the positions outside $\mathcal{E}$. In other words, the encoder knows that $e_i(\mathcal{E}) = 0$ for all $i \in N \setminus \mathcal{E}$. The set of unreliable positions may change from one codeword to another.

Manuscript received June 20, 1996; revised March 9, 1998. The material in this correspondence was presented in part at the IEEE Fourth Symposium on Communications and Vehicular Technology in the Benelux, Gent, Belgium, October 7–8, 1996.

The authors are with the Faculty of Information Technology and Systems, Delft University of Technology, 2600 GA Delft, The Netherlands.

Publisher Item Identifier S 0018-9448/98$10.00 © 1998 IEEE

The encoding function $f$ is a mapping

$$f: \mathcal{M} \times \mathcal{E} \to \mathcal{Q}^n$$

i.e., the output from the encoder, $f(m, \mathcal{E})$, depends both on the message $m \in \mathcal{M}$ and on the set $\mathcal{E}$ of unreliable positions. The output of the channel will be the componentwise addition modulo $q$ of $f(m, \mathcal{E})$ and $e(\mathcal{E})$, denoted $f(m, \mathcal{E}) \oplus e(\mathcal{E})$. The decoding function $g$ is a mapping

$$g: \mathcal{Q}^n \to \mathcal{M} \cup \{\ast\}$$

where $\ast$ is a special symbol used to indicate that the decoder could not make a proper decision. Since the codes under consideration should correct up to $m$ localized bursts of length at most $b$ each, it must hold that

$$g(f(m, \mathcal{E}) \oplus e(\mathcal{E})) = m$$

for every message $m$, every configuration $\mathcal{E}$ which consists of the union of $m$ sets of $b$ consecutive positions, and any error vector $e(\mathcal{E})$ with zeros in the positions outside $\mathcal{E}$.

A trivial upper bound on the number of messages $M$ reads

$$M \leq q^{n-mb}$$

since there are $mb$ positions in a codeword where errors may occur. For the case $b = 1$, it has been shown [1] that the number of messages cannot exceed the Hamming bound, i.e.,

$$M \leq \frac{q^n}{\sum_{i=0}^{m} \binom{n}{i} q^i}$$

If $b > 1$ and $n$ is a multiple of $b$, then the bound from (5) can be used to improve upon the trivial bound from (4), by observing that a $q$-ary code of length $n$ correcting $m$ localized bursts of length $b$ generates a $q^b$-ary code of length $n/b$ of the same size correcting $m$ localized errors. This gives

$$M \leq \frac{(q^b)^{mb}}{\sum_{i=0}^{m} \binom{n/b}{i} q^{i b}}$$

Note that the upper bound from (6) is indeed better than the bound from (4), since

$$\sum_{i=0}^{m} \binom{n/b}{i} (q^b - 1)^i \geq \sum_{i=0}^{m} \binom{m}{i} q^i (q^b - 1)^i = (q^b - 1 + 1)^m = q^{mb}.$$
III. CONSTRUCTION METHOD

In [4] Larsson has presented a construction method for $q$-ary codes of length $n$ correcting $m$ localized bursts of length $b$. The main idea in this method is to use $2m + 1$ subblocks, each of length $r$, separated as much as possible in a code block of length $n$. In each of the first $m + 1$ error-free subblocks, the encoder transmits a message corresponding to the current set $E$ in order to inform the decoder on the unreliable positions. It can easily be shown that the number of possible configurations $E$ is

$$T(n, m, b) = \binom{n - mb + m}{m}$$  \hspace{1cm} (9)

so $r$ should be chosen

$$r = \lfloor \log_q T(n, m, b) \rfloor = \lfloor \log_q \binom{n - mb + m}{m} \rfloor$$ \hspace{1cm} (10)

By reading the contents of the $2m + 1$ subblocks, the receiver obtains the locations of the $m$ possible bursts through a majority decision. Since the number of positions remaining for the data is $n - mb - (m + 1)r$, the size of Larsson’s code reads

$$M = q^{n - mb - (m + 1)r}.$$ \hspace{1cm} (11)

In order to guarantee the existence of at least $m + 1$ error-free subblocks, the burst length $b$ must satisfy

$$b \leq \lfloor (n - (2m + 1)r)/2m \rfloor + 1.$$ \hspace{1cm} (12)

The main disadvantage of Larsson’s method is that the location information on the bursts is transmitted $m + 1$ times. In the new method proposed here, this information is sent only once. However, to guide the decoder on its way to retrieve this information, a price, such as a smaller allowable burst length $b$, must be paid as well. More precisely, the new method demands that the parameters $n, m, b$, and $q$ satisfy

$$b \leq \lfloor n/(m + 1)^2 \rfloor$$ \hspace{1cm} (13)

and

$$\log_q \binom{n - mb + m}{m} - (m + 1) \log_q (q - 1) \leq \lfloor n/(m + 1) \rfloor - m - b.$$ \hspace{1cm} (14)

The method can be described as follows. The set of codeword positions $\mathcal{V}$ is partitioned into $m + 1$ subsets $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{m+1}$, each containing consecutive positions, such that

$$\lfloor n/(m + 1) \rfloor \leq |\mathcal{P}_i| \leq |\mathcal{P}_{i+1}| \leq \lfloor n/(m + 1) \rfloor$$ \hspace{1cm} (15)

for $1 \leq i \leq m$. Each subset $\mathcal{P}_i$ is partitioned into $m + 1$ subsubsets $\mathcal{L}_{i, 1}, \mathcal{L}_{i, 2}, \ldots, \mathcal{L}_{i, m+1}$, each containing consecutive positions, such that

$$\lfloor n/(m + 1)^2 \rfloor \leq |\mathcal{L}_{i,j}| \leq |\mathcal{L}_{i,j+1}| \leq \lfloor n/(m + 1)^2 \rfloor$$ \hspace{1cm} (16)

for $1 \leq j \leq m$. Let $I_{i,j}$ denote the value of the smallest element in $\mathcal{L}_{i,j}$ for $1 \leq i, j \leq m + 1$, and let

$$\mathcal{R}_i = \{I_{i,j}|1 \leq j \leq m + 1\}.$$ \hspace{1cm} (17)

for $1 \leq i \leq m + 1$. Let $\mathcal{S}_i$ denote the subset of $\mathcal{P}_i$, containing the

$$s = \max \left\{ 0, \left[ \log_q \binom{n - mb + m}{m} - (m + 1) \log_q (q - 1) \right] \right\}$$ \hspace{1cm} (18)
smallest elements in $\mathcal{P}_i \backslash \mathcal{R}_i$, for $1 \leq i \leq m + 1$. The set $\mathcal{E}$ of unreliable positions can be written as
\begin{equation}
\mathcal{E} = \bigcup_{i=1}^{m+1} \mathcal{E}_i
\end{equation}
where each $\mathcal{E}_i$ consists of $b$ consecutive positions and all $\mathcal{E}_i$ are mutually disjoint. Let $E_i$ denote the value of the largest element in $\mathcal{E}_i$ and let $\mathcal{V} = \{E_1, E_2, \ldots, E_m\}$. Next, define
\begin{equation}
z = \min \{i | \mathcal{P}_i \cap \mathcal{V} = \emptyset\}.
\end{equation}
Note that $z$ always exists, since there are $m + 1$ disjoint sets $\mathcal{P}_i$, while $\mathcal{V}$ contains only $m$ elements. Finally, define
\begin{equation}
y_i = \min \{j | \mathcal{E}_{i,j} \cap \mathcal{V} = \emptyset\}
\end{equation}
for $1 \leq i \leq m + 1$. Note that $y_i$ exists for all $i$, since for any $i$ there are $m + 1$ disjoint sets $\mathcal{E}_{i,j}$, while $\mathcal{V}$ contains only $m$ elements. Let
\begin{equation}
\mathcal{Y} = \{L_{i,y_i} | 1 \leq i \leq m + 1, i \neq z\}.
\end{equation}

Take into account that $\mathcal{N}, \mathcal{P}_i, \mathcal{L}_{i,j}, \mathcal{R}_i, s$, and $\mathcal{S}_i$ are fixed and are known a priori to both the encoder and the decoder. Table I gives an example for the case $b = 10$, $m = 2$, $n = 300$, and $q = 4$, i.e., in case of a quaternary code of length 300 correcting two localized bursts of length ten. On the other hand, $\mathcal{E}_i, E_i, \mathcal{V}, z, y_i$, and $\mathcal{Y}$ depend on the actual set $\mathcal{E}$, which implies that they may change from codeword to codeword and are known a priori only to the encoder. Continuing the example of Table I with the set $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$, where $\mathcal{E}_1 = \{31, 32, \ldots, 40\}$ and $\mathcal{E}_2 = \{195, 196, \ldots, 204\}$, we obtain $\mathcal{V} = \{40, 204\}, z = 2, y_1 = 1, y_2 = 1, y_3 = 2$, and $\mathcal{Y} = \{1, 234\}$.

The main idea in the construction method is to utilize the subset $\mathcal{R}_i \cup \mathcal{S}_i$ of $\mathcal{P}_i$ for the burst location information, and to enable the decoder to determine $z$, by setting all positions from $\mathcal{Y}$ to zero and using only symbols from the subalphabet $\mathcal{Q} \backslash \{1, 2, \ldots, q - 1\}$ in all positions from $\mathcal{R}_z$. Note that the sets $\mathcal{E}, \mathcal{R}_i, \mathcal{S}_i$, and $\mathcal{Y}$ are mutually disjoint because of (13) and (14). Hence, the positions from $\mathcal{R}_z, \mathcal{S}_i$, and $\mathcal{Y}$ are guaranteed to be error-free. Further, $\mathcal{R}_z \cup \mathcal{S}_i$ can indeed be used to identify any set of unreliable positions $\mathcal{E}$, since $(q - 1)^{m+1} q^r \geq T(n, m, b)$ because of (18) and (9).

In the encoding procedure we compose a codeword
\begin{equation}
f(m, \mathcal{E}) = (c(1), c(2), \ldots, c(n))
\end{equation}
where all $c(i) \in \mathcal{Q}$, as follows:
1) From $\mathcal{E}$, determine first $\mathcal{V}$ and then $z$ and $\mathcal{Y}$.
2) Put location information on the bursts into the $c(i)$ with $i \in \mathcal{R}_z \cup \mathcal{S}_i$, in such a way that $c(i) \neq 0$ for all $i \in \mathcal{R}_z$.
3) Set $c(i) = 0$ for all $i \in \mathcal{Y}$ and fill the $c(i)$ with $i \in \mathcal{E}$ arbitrarily.
4) Put data corresponding to the message $m$ in the $c(i)$ with $i \in \mathcal{N} \backslash (\mathcal{E} \cup \mathcal{R}_z \cup \mathcal{S}_i \cup \mathcal{Y})$.
Note that the number of data positions is
\[ |\mathcal{V}| - |\mathcal{E}| - |\mathcal{R}| - |\mathcal{S}| - |\mathcal{Y}| = n - mb - (m+1) - s - m = n - m(b+2) - (s+1). \]
Hence, the size of the code reads
\[ M = q^{n-m(b+2)-(s+1)}. \]  
(23)

The decoding procedure, when receiving a word 
\[ f(m, \mathcal{E}) = e(\mathcal{E}) = (r(1), r(2), \ldots, r(n)) \]
reads as follows.

1) Calculate
\[ z = \min \{ i | r(L_{i,j}) \neq 0 \forall 1 \leq j \leq m+1 \}. \]  
(24)
If no \( i \) with \( r(L_{i,j}) \neq 0 \) for all \( 1 \leq j \leq m+1 \) is found, then an incorrectable error pattern is detected, and the decoding procedure is stopped (decoder output is +).

2) Read the burst location information \( \mathcal{E} \) from the \( r(i) \) with \( i \in \mathcal{R} \cup \mathcal{S} \).

3) Knowing \( \mathcal{E} \), compose the sets \( \mathcal{V} \) and \( \mathcal{Y} \), and read the data from the \( r(i) \) with \( i \in \mathcal{N} \setminus (\mathcal{E} \cup \mathcal{R} \cup \mathcal{S} \cup \mathcal{Y}) \).

IV. Analysis

The (code) rate \( R \) of a \( q \)-ary code of length \( n \) and size \( M \) is defined as
\[ R = \frac{\log_q M}{n}. \]  
(25)
In this section we analyze the rate of the Larsson codes and the proposed codes in comparison with upper bounds on the rate of codes for multiple localized burst error correction as well as upper bounds on the rate of conventional codes for multiple burst error correction.

First, we elaborate upon the example started in the previous section, i.e., the case of quaternary codes of length 300 correcting two localized bursts of length ten \( (q=4, n=300, m=2, b=10) \). In Figs. 1–4 we give (upper bounds on) code rates, while fixing three of the code parameters \( q, n, m, \) and \( b \), and varying the fourth one. In each of the four figures, the following five (bounds on) code rates are indicated.

- Code rate of the Larsson code (“squares”), derived from (25) and (11)
\[ 1 - (mb + mr + r)/n \]  
(26)
where \( r \) is given by (10).

- Code rate of the proposed code (“circles”), derived from (25) and (23)
\[ 1 - (mb + 2m + s + 1)/n \]  
(27)
where \( s \) is given by (18).

- Trivial upper bound on the rate of codes for multiple localized burst error correction (“diamonds”), derived from (25) and (4)
\[ 1 - mb/n. \]  
(28)

- Hamming-type upper bound on the rate of codes for multiple localized burst error correction (“asterisks”), derived from (25) and (6)
\[ 1 - S/n. \]  
(29)
where
\[ S = \log_q \sum_{i=0}^{m} \left( \frac{n/b}{i} \right) (q^b - 1)^i. \]
Fig. 4. Code rate as a function of the number of bursts \( m \) if \( q = 4, n = 300, \) and \( b = 10 \); index: (26) \( \rightarrow \) “squares,” (27) \( \rightarrow \) “circles,” (28) \( \rightarrow \) “diamonds,” (29) \( \rightarrow \) “asterisks,” (30) \( \rightarrow \) “triangles.”

* Reiger-type upper bound on the rate of conventional codes for multiple burst error correction (“triangles”), derived from (25) and (8)

\[
1 - 2mb/n. \tag{30}
\]

The figures show that for the examples under consideration (big improvements in code rate are indeed possible for the localized burst error concept over the conventional burst error concept, that the proposed codes are considerably larger than the Larsson codes, and finally, that the proposed codes are quite close to the theoretical optimum.

Next, we consider the asymptotic rate when \( m \) is fixed and \( b \) grows linearly with \( n \), i.e., \( b/n \rightarrow \beta, \ n \rightarrow \infty \), where \( \beta \) is a constant satisfying \( 0 < \beta < 1/m \). Since \( r \) and \( s \), as given in (10) and (18), at most grow logarithmically with \( n \), it follows from (12) that the Larsson method requires \( 0 < \beta < 1/2m \) and from (13) that the proposed construction method requires \( 0 < \beta < 1/(m + 1)^2 \), and from (26) and (27) that the rates of both the Larsson codes and the proposed codes asymptotically equal the upper bound \( 1 - m\beta \) from (28). Hence, both the Larsson method and the proposed method are asymptotically optimal when \( b \) grows linearly with \( n \).

However, when considering the code size \( M \) in case both \( m \) and \( b \) are fixed and \( n \) tends to infinity, the situation is different. It follows from (6) that

\[
M \leq h_1(q, b, m) \frac{q^n}{n^m}(1 + o(1)) \tag{31}
\]

where \( o(1) \rightarrow 0 \) as \( n \rightarrow \infty \), and \( h_1(q, b, m) \) is a function of \( q, b, \) and \( m \) but not of \( n \). For the size of the Larsson codes we obtain from (11) and (10) that

\[
M = h_2(q, b, m) \frac{q^n}{n^m(m+1)}(1 + o(1)) \tag{32}
\]

where \( o(1) \rightarrow 0 \) as \( n \rightarrow \infty \), and \( h_2(q, b, m) \) is a function of \( q, b, \) and \( m \) but not of \( n \). Hence, when \( m \) and \( b \) are both fixed, the proposed codes are asymptotically optimal, while the Larsson codes are not.

**ACKNOWLEDGMENT**

The authors wish to thank the two referees and Khaled Abdel-Ghaffar for their valuable comments and suggestions.

**REFERENCES**


