A General Construction of Integer Codes Correcting Specific Errors in Binary Communication Channels

Hristo Kostadinov * and Nikolai Manev

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl.8, 1113 Sofia, Bulgaria; nlmanev@math.bas.bg
* Correspondence: hristo@math.bas.bg

Abstract: Integer codes have been successfully applied to various areas of communication and computer technology. They demonstrate good performance in correcting specific kinds of errors. In many cases, the used integer codes are constructed by computer search. This paper presents an algebraic construction of integer codes over the ring of integers modulo \( \mathbb{Z}_{A} = 2^n + 1 \) capable of correcting at least up to two bit errors in a single \( b \)-byte. Moreover, the codes can correct some configurations of three or more erroneous bits, but not all possible ones. The construction is based on the use of cyclotomic cosets of 2 modulo \( A \).

Keywords: integer codes; cyclotomic cosets; binary errors

MSC: 94B60

1. Introduction

In many binary communication channels, the majority of errors, besides uniformly distributed single-bit errors, consist of error configurations of several bits that are closely placed in a subsequence of \( b \) consecutive bits. We call such a subsequence \( b \)-bits byte (\( b \)-byte in short). We avoid the term \( b \)-burst, since its classical definition requires the first and the last bits to be 1s, while the error configurations considered in the paper have a more general form.

Integer codes are codes over a ring of integers modulo a given positive integer, and have been studied—though only from theoretical point of view—since the beginning of coding theory. Such codes for correcting asymmetric errors are proposed in [1,2]. The name “integer codes” appears for the first time in [3], where such codes are applied to magnetic recording and frame synchronization. Since that time, integer codes have been successfully used in many applications. Error control systems for flash memory [4–7] and coded modulation [8–10] are two examples for such applications. In [11], integer codes are applied to cryptographic construction. Integer codes over \( \mathbb{Z}_A \) with \( A = 2^b - 1 \) that correct configurations of error bits in a single \( b \)-byte are found via computer search in [12–16]. The specific use of integer codes is presented in [17].

Herein, we propose an algebraic construction of codes over the ring \( \mathbb{Z}_A \) of integers modulo \( A = 2^b + 1 \). The obtained codes can correct at least up to two bit errors in a \( b \)-byte. The codes can simultaneously correct some configurations of three or more erroneous bits, but not all possible ones.

The next section contains a summary of existing results and necessary notations and definitions that are required for reading the paper. The proposed construction based on the cyclotomic cosets of 2 modulo \( A = 2^b + 1 \) is described in Section 3. Section 4 discusses applications of constructed codes to error control in binary channels.

2. Preliminaries

Let us recall the basic definitions concerning integer codes.
Definition 1. Let $\mathbb{Z}_A$ be the ring of integers modulo $A$. An integer code of length $n$ with parity-check matrix $\mathbf{H} \in \mathbb{Z}_A^{m \times n}$, is referred to as a subset of $\mathbb{Z}_A^n$, defined by

$$C(\mathbf{H}) = \{ \mathbf{c} \in \mathbb{Z}_A^n \mid \mathbf{cH}^T = \mathbf{0} \}$$

where all zero vectors $\mathbf{0} \in \mathbb{Z}_A^m$. (We will write only $C$ if there is no possibility for ambiguity.)

$C$ is said to be $[n, n - m]_A$ integer code. Let us note that in this paper, the notation only means that $\mathbf{H}$ has $m$ rows, $n$ columns, and entries from $\mathbb{Z}_A$, not that $C$ is a linear code. For the not-prime $A$, the code $C$ is a submodule, not a linear subspace. However in our case, the difference does not concern correcting properties of considered codes.

The idea behind use of integer codes is that they demonstrate better performance in correcting specific types of errors. The parity-check matrix $\mathbf{H}$ can be chosen, thinking about dominant errors.

Definition 2. Let $k_i$ and $l_i$ be positive integers, $i = 1, \ldots, s$, $j = 1, \ldots, m$. The code $C(\mathbf{H})$ is said to be a single $(\pm k_1, \pm k_2, \ldots, \pm k_s, l_1, l_2, \ldots, l_m)$ that is error correctable if it can correct any error vector with only one nonzero entry with value $l_i$ or $\pm k_i$.

The existence of codes with the above defined property implies a lower bound for the size $A$ of the alphabet:

$$A \geq (2s + m)n + 1. \quad (1)$$

Indeed, the $(2s + m)n$ possible single $(\pm k_1, \pm k_2, \ldots, \pm k_s, l_1, l_2, \ldots, l_m)$ error vectors $\mathbf{e}$ must have pairwise disjoint syndromes $\mathbf{eH}^T$.

The sequence of bits in a binary communication channel can be considered as a sequence of blocks of bits with the same length $b$. (We call such blocks $b$-bytes, but $b$ is not necessarily equal to 8). The advantage of this consideration is that the blocks can be interpreted as nonnegative integers less than $2^b$, that is, elements of $\mathbb{Z}_A$ with $A \geq 2^b$. Hence, the binary input sequence can be transformed into a sequence of elements of $\mathbb{Z}_A$, and then it can be encoded using integer codes over $\mathbb{Z}_A$ with $A \geq 2^b$.

It seems the natural choice is $A = 2^b$, but it is not the right choice in any case. Indeed, the situation is more complex. Note that changing one bit in a $b$-byte is equivalent to adding or subtracting $2^i$, $i = 0, 1, \ldots, b - 1$ to the integer corresponding to the $b$-byte. However, $2^i$ divides $2^b$, which generates complications. If $A > 2^b$, there are elements of $A$ that cannot be transmitted over the channel; thus, they must not appear as coordinates of the codewords. Hence, the encoding procedure has to comply with this requirement.

Let $B \in \{0, 1, \ldots, 2^b - 1\}$ be the integer corresponding to a given $b$-byte sent through the communication channel. At the other end of the channel, the received $b$-byte may differ in several bits from the original. This difference results in adding a nonzero integer $E$, $-2^b < E < 2^b$, to $B$, that is, the integer value of the received $b$-byte is $V = B + E$.

Here is the correspondence between single-bit errors and $E$ in a $b$-byte.

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E1</strong></td>
<td>The $(i + 1)$-th bit $1 \rightarrow 0$: $E = -2^i, i = 0, 1, \ldots, b - 1$.</td>
</tr>
<tr>
<td><strong>E2</strong></td>
<td>The $(i + 1)$-th bit $0 \rightarrow 1$: $E = 2^i, i = 0, 1, \ldots, b - 1$.</td>
</tr>
</tbody>
</table>

All other error configurations in a single $b$-byte correspond to combinations of $E1$ and $E2$; thus, $E$ is a sum of powers (positive or negative) of 2. For example, in the case of transposition of two adjacent different bits (The existence of such errors has been reported in the literature), $E = 2^i - 2^{i-1} = 2^{i-1}$ or $E = -2^i + 2^{i-1} = -2^{i-1}$. Note that the same values correspond to the change of one bit in the $b$-byte. It may seem amusing, but the decoding procedure of integer codes corrects the $b$-byte. The explanation is in the difference between the corresponding values of $B$ (in $V = B + E$). We also discuss this fact in Section 4.

3. Construction of Integer Code

In this section, we propose a construction of a class of integer codes that is based on cyclotomic cosets.
Recall a cyclotomic coset of $q$ modulo $A$, $(A,q) = 1$, with leader $s$ is the set
\[ C_s = \{ sq^i \pmod{A} \mid i = 0,1,\ldots,m - 1 \}, \]
where $m$ is the minimal positive integer, such that $sq^m \equiv s \pmod{A}$. The two cyclotomic cosets are either disjoint, $C_s \cap C_t = \emptyset$; or they coincide, $C_s = C_t$.

Hereafter, in this paper, $A = 2^b + 1$.

There are two arguments for our choice:
- $A = 2^b + 1$ is coprime with 2; thus, we can use cyclotomic cosets of 2 modulo $A$.
- $A = 2^b + 1$ is the smallest integer greater than $2^k$. Hence, we have to avoid the use of only $2^k$ as a coordinate of a codeword.

Here are several simple but useful properties of cyclotomic cosets of 2 modulo $2^b + 1$. Their proofs are straightforward.

\begin{enumerate}
\item \textbf{P1} If $b$ has odd divisor $d$, then $2^d + 1$ divides $2^b + 1$. In partial, if $b$ is odd, $3|2^b + 1$.
\item \textbf{P2} If $A = 2^b + 1 = (2^k + 1)B$, then $C_B$ has length $2k$. If there is a cyclotomic coset of length $2k$, then $2^b + 1$ divides $A$.
\end{enumerate}

\textbf{Lemma 1.} If $A = 2^b + 1$, then the cyclotomic cosets $C_s$ of 2 modulo $A$ have even length. In partial for $s$ and $A$ coprime, the length $|C_s| = 2b$ and
\[ C_s = \{ s, 2s, \ldots, 2^{b-1}s, -s, -2s, \ldots, -2^{b-1}s \}. \]

\textbf{Proof.} $2^b \equiv -1 \pmod{A}$ gives $2^{2b} \equiv 1 \pmod{A}$. Hence, the order $m$ of 2 modulo $A$ is a divisor of $2b$. The assumption $m \mid b$ leads to $2^b \equiv 1 \pmod{A}$, which is a contradiction. Thus, $m = 2k$, where $k|b$, that is, $|C_s|$ is an even number. Let $d = \gcd(s,A)$, $s = ds_1$, $A = dA_1$. Let $l = |C_s|$, i.e., $l$ be the smallest positive integer such that $s(2^l - 1) \equiv 0 \pmod{A}$. If $d = 1$, then $A \mid (2^l - 1)$; thus (see above), $l = 2b$. Otherwise, we have $(2^l - 1) \equiv 0 \pmod{A_1}$, where $l$ is the minimal positive integer with this property. Since $2^{2b} \equiv 1 \pmod{A_1}$ we have $l|2b$. Hence, $l = 2k$, $k|b$. The structure of $C_s$ follows from $2^b \equiv -1 \pmod{A}$. \hfill \Box

Let $S$ be the set of the leaders of all cyclotomic cosets with length $2b$.

\textbf{Theorem 1.} Any integer code over $\mathbb{Z}_A$, $A = 2^b + 1$, having parity check matrix $H = (h_1, h_2, \ldots, h_n)$, where $h_j \in S$, can correct single $\pm 2^i - \text{error}$, $i = 0,1,\ldots,b - 1$.

\textbf{Proof.} Syndromes $\pm 2^ih_j$ belong to the coset $C_{h_j}$ and since cyclotomic cosets are disjoint, all syndromes corresponding to single $\pm 2^i - \text{errors}$ are pairwise different. \hfill \Box

We have to note that if $s \in \mathbb{Z}_A$ is the corresponding value to a given bit errors configuration in a $b$-byte then its cyclic shifts corresponds to $s2^i \in C_i$. This simple observation leads us to the following

\textbf{Theorem 2.} Let $\mathcal{E}$ be a set of cyclotomic cosets containing $C_1$. If there exist $h_2,\ldots,h_n \in \mathbb{Z}_A$ such that $\mathcal{E}, h_2\mathcal{E}, \ldots, h_n\mathcal{E}$ are disjoint sets of different cosets, then the code with check matrix $H = (h_1, h_2, \ldots, h_n)$ can correct error values belonging to $\mathcal{E}$, and of course, bit error configurations corresponding to them.

\textbf{Proof.} The proof is straightforward. \hfill \Box

\section*{4. Application of the Constructed Codes}

In this section, we will apply the considered integer codes to correct up to two erroneous bits in a $b$-byte. As we have noted in the Introduction, there are binary channels
where these types of errors dominate. The experiments with simulations of several channels we have made also confirm this observation. We shall explain the idea behind these simulations by describing in short one such form of optical wireless communication (for an example, we will describe one of our experiments in this direction). We generated a very long random binary sequence and applied simulation procedure of the optical wireless channel given in [18] to determine the error sequence (the difference between the input and output sequences). Then, with a code of length \( n \) over \( \mathbb{Z}_{2^b+1} \) in the mind, we divided the error sequence into blocks of \( n \)-bytes. The next step is to count the number of 1s in each block and how they are distributed among \( b \)-bytes: what part of blocks contain only one 1, what part of blocks have two 1s in one byte, and so on. Table 1 illustrates the results that we obtained. With the exception of the cases of the very small signal-to-noise ratio (SNR \( \leq 4 \) decibels (dB)), the considered error configurations (columns two and three) are the majority. Very soon (SNR \( \geq 7 \)dB), the error configuration that consists of one erroneous bit per a codeword (see second column of Table 1) becomes the dominant error type.

Table 1. The case \( n = 3, b = 10 \).

<table>
<thead>
<tr>
<th>SNR in dB</th>
<th>1 Bit in a Codeword</th>
<th>2 Bits in 1 Symbol</th>
<th>2 Bits in 2 Symbols</th>
<th>3 or More Bits in a Codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00223</td>
<td>0.00067</td>
<td>0.00423</td>
<td>0.02573</td>
</tr>
<tr>
<td>2</td>
<td>0.00431</td>
<td>0.00167</td>
<td>0.00555</td>
<td>0.02057</td>
</tr>
<tr>
<td>3</td>
<td>0.00727</td>
<td>0.00319</td>
<td>0.00584</td>
<td>0.01419</td>
</tr>
<tr>
<td>4</td>
<td>0.01046</td>
<td>0.00453</td>
<td>0.00452</td>
<td>0.00797</td>
</tr>
<tr>
<td>5</td>
<td>0.01237</td>
<td>0.00461</td>
<td>0.00247</td>
<td>0.00341</td>
</tr>
<tr>
<td>6</td>
<td>0.01171</td>
<td>0.00372</td>
<td>0.00088</td>
<td>0.00125</td>
</tr>
<tr>
<td>7</td>
<td>0.00871</td>
<td>0.00141</td>
<td>0.00022</td>
<td>0.00021</td>
</tr>
<tr>
<td>8</td>
<td>0.00503</td>
<td>0.00042</td>
<td>( 3 \times 10^{-5} )</td>
<td>( 2 \times 10^{-5} )</td>
</tr>
<tr>
<td>9</td>
<td>0.00221</td>
<td>0.00008</td>
<td>( 2 \times 10^{-6} )</td>
<td>( 2 \times 10^{-6} )</td>
</tr>
<tr>
<td>10</td>
<td>0.00077</td>
<td>( 9 \times 10^{-6} )</td>
<td>( 6 \times 10^{-8} )</td>
<td>( 1 \times 10^{-7} )</td>
</tr>
</tbody>
</table>

Now, let us fix some notations. In this paper, the leftmost bit is the least significant bit and the rightmost bit is the most significant in a \( b \)-byte, that is, the latter represents the number \( 2^{b-1} \).

Let \( a_1 \ldots a_i \ldots a_{i+k} \ldots a_b \) be an arbitrary \( b \)-byte. Below lists the possible error transformation of \( i \)-th and \( (i+k) \)-th bits and the corresponds to them the alteration \( E \) of the integer value of the byte.

\[
\begin{align*}
T1 \quad & \ldots 0 \ldots 0 \ldots \rightarrow \ldots 1 \ldots 1 \ldots \quad E = 2^{i-1} + 2^{k-1} = (1 + 2^k)2^{i-1} \\
T2 \quad & \ldots 1 \ldots 1 \ldots \rightarrow \ldots 0 \ldots 0 \ldots \quad E = -2^{i-1} - 2^{k-1} = -(1 + 2^k)2^{i-1} \\
T3 \quad & \ldots 1 \ldots 0 \ldots \rightarrow \ldots 0 \ldots 1 \ldots \quad E = -2^{i-1} + 2^{k-1} = (2^k - 1)2^{i-1} \\
T4 \quad & \ldots 0 \ldots 1 \ldots \rightarrow \ldots 1 \ldots 0 \ldots \quad E = 2^{i-1} - 2^{k-1} = -(2^k - 1)2^{i-1}
\end{align*}
\]

where \( k = 1, \ldots, b-1 \) and \( i = 1, \ldots, b-k \).

Obviously, the error values \( E \in C_{2^b+1} \) for T1 and T2, while \( E \in C_{2^{b-k}+1} \) for T3 and T4. However, since \( A = 2^b+1 \), we have

\[
2^k + 1 \equiv 2^k - 2^b = -2^k(2^{b-k} - 1) \pmod{A}
\]

as well as

\[
2^k - 1 \equiv 2^k + 2^b = 2^k(2^{b-k} + 1) \pmod{A}.
\]

Hence,

\[
C_{2^b+1} = C_{2^{b-k}+1}, \quad k = 1, \ldots, b-1
\]
If Theorem 3.

The blue curve gives BER in the case when no error correcting codes are used. The green performance, but this is the price we pay for the higher code rate.

Hence the set of error values is a proper subset of

\[ \mathcal{E} = \{ \cup C_{2^k \pm 1}, k = 1, \ldots, [b/2] \} \]

**Theorem 3.** If \( \mathcal{E} \) consists of \( t \) cyclotomic cosets, then

\[ 2btn \leq 2^b \]

is a necessary condition for the existence of an integer code that can correct the error configurations of types T1–T4 defined by \( \mathcal{E} \). In partial a necessary condition for the code to correct all the considered error configurations T1–T4 is

\[ 2b(b - 2)n \leq 2^b \]  

(2)

**Proof.** If \( \mathbf{H} = (h_1, h_2, \ldots, h_n) \) is a parity check matrix of an integer code capable to correct all error configurations T1–T4, the sets of cyclotomic cosets \( \{ h_i \mathcal{E} \} \) have to be disjoint, with lengths of \( 2b \) each. Hence, we obtain the following necessary condition (inequality):

\[ 2bnt \leq A - 1 = 2^b, \quad n \geq 2. \]

Since \( t \) is at most \( b - 2 \) we obtain (2).

**Example 1** \((b = 9, n = 2)\) (see Appendix A.2)). The minimal \( b \) that satisfies inequality (2) (for the minimal \( n = 2 \)) is \( b = 8 \), but a required code does not exist. Such a code exists for \( b = 9 \) and \( n = 2 \). In that case, \( \mathbf{H} = (1\ 2\ 3) \) and

\[ \mathcal{E} \subseteq C_1 \cup C_3 \cup C_5 \cup C_7 \cup C_9 \cup C_{15} \cup C_{17} \]

\[ 23\mathcal{E} \subseteq C_{23} \cup C_{39} \cup C_{53} \cup C_{11} \cup C_{45} \cup C_{21} \cup C_{25} \]

**Example 2** \((b = 10, n = 3)\) (see Appendix A.3)). In this case, a code that corrects all two-bit errors in a single \( b \)-byte per a codeword exists: the code with \( \mathbf{H} = (1\ 19\ 27) \)

\[ \mathcal{E} \subseteq C_1 \cup C_3 \cup C_5 \cup C_7 \cup C_9 \cup C_{15} \cup C_{17} \cup C_{31}, \quad C_{33} = C_{31} \]

\[ 19\mathcal{E} \subseteq C_{19} \cup C_{37} \cup C_{35} \cup C_{39} \cup C_{171} \cup C_{105} \cup C_{43} \cup C_{109} \]

\[ 27\mathcal{E} \subseteq C_{27} \cup C_{59} \cup C_{55} \cup C_{51} \cup C_{155} \cup C_{107} \cup C_{47} \]

Length \( n = 3 \) is the maximal possible value for the existence of such a code.

**Example 3** \((b = 8, n = 4)\) (see Appendix A.1)). If we restrict ourselves to \( \mathcal{E} = C_1 \cup C_3 \cup C_5 \), that is, to the case when we correct only erroneous bits that are either adjacent or there is one correct bit between them, we can construct codes with larger rates.

When \( b = 8 \), i.e. \( A = 257 \), the code with \( \mathbf{H} = (1\ 7\ 9\ 19) \) (the rate is 3/4) corrects a single \( b \)-byte of the aforesaid configurations with up to 2 erroneous bits. Indeed,

\[ 7\mathcal{E} = C_7 \cup C_{21} \cup C_{23}; \quad 9\mathcal{E} = C_9 \cup C_{27} \cup C_{45}; \quad 19\mathcal{E} = C_{19} \cup C_{25} \cup C_{11}. \]

Figure 1 presents in graphical form probability of bit error (BER) versus signal/noise ratio per bit in decibels for the aforesaid simulation of noisy wireless optical channel [18]. The blue curve gives BER in the case when no error correcting codes are used. The green and red curves correspond to applying the codes given in Example 1 and Example 2, respectively. The reader can note that the code from Example 2 demonstrates slightly worse performance, but this is the price we pay for the higher code rate.
5. Discussion

In this paper, we propose a general construction of codes over $\mathbb{Z}_{2^b+1}$, which can be used for correcting two erroneous bits (as well as several specific configurations of erroneous bits) in a single $b$-byte in a codeword of $n \geq 2b$-bytes. Like any integer codes, the proposed codes have very simple encoding and decoding without involving arithmetic in finite fields. The proposed construction avoids the exhaustive computer search commonly used for construction similar codes. Determining the cyclotomic cosets of 2 modulo $2^b+1$ can be performed with the simplest mathematical packages (even manually).

Figure 1 (and other our experiments) show that the proposed codes demonstrate good performance. This performance is in some sense unexpected, since the application of integer codes to the considered binary channel is not so proper as their application to coded modulation and flash memory. For example, we performed a comparison between the code given in Example 2 and a BCH code with close length and rate. Our expectation was for the BCH code to demonstrate better performance since the considered errors are not strongly dominant. However, the results were similar. Hence, with their simple encoding and decoding procedures, integer codes are at least commensurable with other codes with similar parameters, but have complicated encoding and decoding.

Author Contributions: All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially supported by the National Science Fund of Bulgaria under Grant KP-06-N32/2-2019 and by Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014–2020) and co-financed by the European Union through the European structural and Investment funds.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the access to the e-infrastructure provided by the Grant No. D01-168/28.07.2022 “National Centre for High Performance and Distributed Computing” of the Ministry of Education and Science of Bulgaria. The authors would like to thank to the anonymous reviewers for their valuable remarks on the content of the paper.
Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. List of the Leaders of Cyclotomic Cosets of 2 Modulo A with Maximal Length 2b for Several Values of b

Appendix A.1. The Case b = 8, \( A = 257 \), Prime

\[ S = \{1, 3, 5, 7, 9, 11, 13, 15, 19, 21, 23, 25, 27, 37, 43, 45\}. \]

Appendix A.2. The Case b = 9, \( A = 513 = 3 \times 9 \times 19 \)

\[ S = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 35, 37, 39, 41, 43, 45, 51, 53, 55, 75, 77, 83, 85\}. \]

\[ |C_{171}| = 6; \quad |C_{171}| = 2. \]

Appendix A.3. The Case b = 10, \( A = 1025 = 25 \times 41 \)


\[ |C_{205}| = 4. \]

References


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.