Encoding and Decoding of Balanced q-ary Sequences Using a Gray Code Prefix

Elie N. Mambou & Theo G. Swart

Department of Electrical and Electronic Engineering Science,
University of Johannesburg (UJ)
South Africa

2016 IEEE International Symposium on Information Theory
Universitat Pompeu Fabra Campus Ciutadella, Barcelona, Spain

July 10-15, 2016

Overview

Definition of balanced codeword

- Consider a q-ary information sequence $\mathbf{x} = x_0 x_1 x_2 \ldots x_{k-1}$, $x_i \in \{0, 1, \ldots, q-1\}$, of length k.
- Let the prefix that will be appended to $\mathbf{x}$ be of length r; and let the information and the prefix together be denoted by $\mathbf{c} = c_0 c_1 c_2 \ldots c_{k-1}$, $c_i \in \{0, 1, \ldots, q-1\}$, of length $n = k + r$.
- If $w(\mathbf{c})$ represents the weight of $\mathbf{c}$ then

$$w(\mathbf{c}) = \sum_{i=0}^{k-1} c_i.$$

- $\mathbf{c}$ is said to be balanced if

$$w(\mathbf{c}) = \frac{n(q-1)}{2} = \beta, \text{ where } \beta \text{ is an integer.}$$

Balancing of q-ary sequences

- It has been proven [1], that $\mathbf{x}$, can always be balanced by adding modulo q one sequence from a set of balancing sequences $\mathbf{b}(s,p) = b_1 b_2 \ldots b_k$ generated as follows:

$$b_i = \begin{cases} s, & i-1 \geq p, \\ s+1 \pmod{q}, & i-1 < p, \end{cases} \text{ where } \begin{cases} 0 \leq s \leq q-1, \\ 0 \leq p \leq k-1. \end{cases}$$

- Let z be the iterator through these balancing sequences, with $z = sk + p$, $0 \leq z \leq kq-1$. $\mathbf{b}(s,p)$ and $\mathbf{b}(z)$ refers to the same.
- Let $\mathbf{y}$ denote the sequence after a balancing sequence is added, $\mathbf{y} = \mathbf{x} \oplus_q \mathbf{b}(z)$. At least one $\mathbf{b}(z)$ will lead to a balanced output $\mathbf{y}$.

[1] T. G. Swart and J. H. Weber, "Efficient balancing of q-ary sequences with parallel decoding," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009.

Balancing of q-ary sequences (Cont'd)

Example (1)

For $q = 3$, $k = 3$, consider the sequence 202. The balancing value is $\beta = 3$.

z	$\mathbf{b}(z)$	$\mathbf{x} \oplus_q \mathbf{b}(z) = \mathbf{y}$	$w(\mathbf{y})$	Balanced?
0	000	$202 \oplus_3 000 = 202$	4	
1	100	$202 \oplus_3 100 = 002$	2	
2	110	$202 \oplus_3 110 = 012$	**3**	✓
3	111	$202 \oplus_3 111 = 010$	1	
4	211	$202 \oplus_3 211 = 110$	2	
5	221	$202 \oplus_3 221 = 120$	**3**	✓
6	222	$202 \oplus_3 222 = 121$	4	
7	022	$202 \oplus_3 022 = 221$	5	
8	002	$202 \oplus_3 002 = 201$	**3**	✓

q-ary Gray Codes

- Invented by Frank Gray [2]; originally used to solve problems in pulse code communication; and extended to several other fields.
- $\mathbf{d} = d_1 d_2 \dots d_{r'}$ denotes a sequence amongst the set of q-ary sequences of length r' listed in lexicographic order. They are mapped to Gray code sequences, $\mathbf{g} = g_1 g_2 \dots g_{r'}$. Any two adjacent sequences differ in only one symbol position, with weight difference of either -1 or $+1$.
- *4-ary Gray code of length 2*

z	**d**	**g**	z	**d**	**g**	z	**d**	**g**	z	**d**	**g**
0	00	00	4	10	13	8	20	20	12	30	33
1	01	01	5	11	12	9	21	21	13	31	32
2	02	02	6	12	11	10	22	22	14	32	31
3	03	03	7	13	10	11	23	23	15	33	30

[2] F. Gray, "Pulse code communication," *U. S. Patent 2632058*, 1953.

Encoding and Decoding of q-ary Gray codes [3]

Gray code encoding algorithm The parity of the sum S_i of the first $i-1$ digits of $\mathbf{g}$ determines the Gray code symbols, where $2 \leq i \leq r'$ and $g_1 = d_1$, then

$$S_i = \sum_{j=1}^{i-1} g_j, \quad \text{and} \quad g_i = \begin{cases} d_i, & \text{if } S_i \text{ is even,} \\ q-1-d_i, & \text{if } S_i \text{ is odd.} \end{cases}$$

Gray code decoding algorithm

$$S_i = \sum_{j=1}^{i-1} g_j, \quad \text{and} \quad d_i = \begin{cases} g_i, & \text{if } S_i \text{ is even,} \\ q-1-g_i, & \text{if } S_i \text{ is odd.} \end{cases}$$

[3] D.-J. Guan, "Generalized Gray codes with applications," in *Proc. National Science Council, Republic of China, Part A*, 1998.

Problem Statement

- The prefix plays an important role as it helps to decode the source information at the receiver end.
- Various previous schemes for balancing sequences were based on the assumption that a prefix can be sent [4]; without actually implementing the prefix.
- We propose an efficient, fast and easy algorithm to encode and decode balanced sequences with prefixes based on Gray codes.

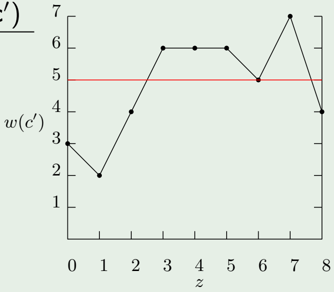
[4] T. G. Swart and J. H. Weber, "Efficient balancing of q-ary sequences with parallel decoding," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009.

Encoding

Example (2)

For $q = 3$, $k = 3$, consider the sequence 201; the condition $k = q^t$ with $t \in N$ must hold. Length of Gray code $r' = \log_q k + 1 = 2$ so $n = 5$. Graph goes through β.

z	$\mathbf{x} \oplus_q \mathbf{b}(z) = \mathbf{y}$	$\mathbf{c}' = [\mathbf{g}\|\mathbf{y}]$	$w(\mathbf{c}')$
0	$201 \oplus_3 000 = 201$	<u>00</u>201	3
1	$201 \oplus_3 100 = 001$	<u>01</u>001	2
2	$201 \oplus_3 110 = 011$	<u>02</u>011	4
3	$201 \oplus_3 111 = 012$	<u>12</u>012	6
4	$201 \oplus_3 211 = 112$	<u>11</u>112	6
5	$201 \oplus_3 221 = 122$	<u>10</u>122	6
6	$201 \oplus_3 222 = 120$	<u>20</u>120	5
7	$201 \oplus_3 022 = 220$	<u>21</u>220	7
8	$201 \oplus_3 002 = 200$	<u>22</u>200	6

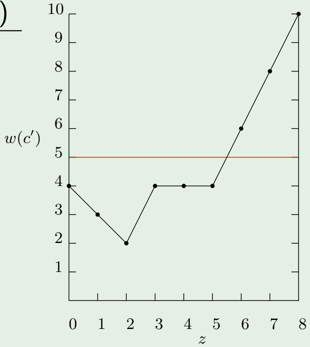


Encoding (Cont'd)

Example (3)

For $q = 3$, $k = 3$, consider the sequence 220; Graph does not go through β. Therefore we need to refine the algorithm.

z	$\mathbf{x} \oplus_q \mathbf{b}(z) = \mathbf{y}$	$\mathbf{c}' = [\mathbf{g}\|\mathbf{y}]$	$w(\mathbf{c}')$
0	$220 \oplus_3 000 = 220$	$\underline{00}220$	4
1	$220 \oplus_3 100 = 020$	$\underline{01}020$	3
2	$220 \oplus_3 110 = 000$	$\underline{02}000$	2
3	$220 \oplus_3 111 = 001$	$\underline{12}001$	4
4	$220 \oplus_3 211 = 101$	$\underline{11}101$	4
5	$220 \oplus_3 221 = 111$	$\underline{10}111$	4
6	$220 \oplus_3 222 = 112$	$\underline{20}112$	6
7	$220 \oplus_3 022 = 212$	$\underline{21}212$	8
8	$220 \oplus_3 002 = 222$	$\underline{22}222$	10

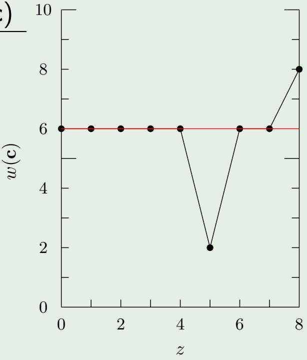


Encoding (Cont'd)

Example (Encoding)

For $q = 3$, $k = 3$, consider the sequence 212; Extra digit $u \in [0, q-1]$ and $u = \beta - w(\mathbf{c}')$.

z	$\mathbf{x} \oplus_q \mathbf{b}(z) = \mathbf{y}$	$\mathbf{c} = [u\|\mathbf{g}\|\mathbf{y}]$	$w(\mathbf{c})$
0	$212 \oplus_3 000 = 212$	100212	6
1	$212 \oplus_3 100 = 012$	201012	6
2	$212 \oplus_3 110 = 022$	002022	6
3	$212 \oplus_3 111 = 020$	112020	6
4	$212 \oplus_3 211 = 120$	111120	6
5	$212 \oplus_3 221 = 100$	010100	2
6	$212 \oplus_3 222 = 101$	220101	6
7	$212 \oplus_3 022 = 201$	021201	6
8	$212 \oplus_3 002 = 211$	022211	8



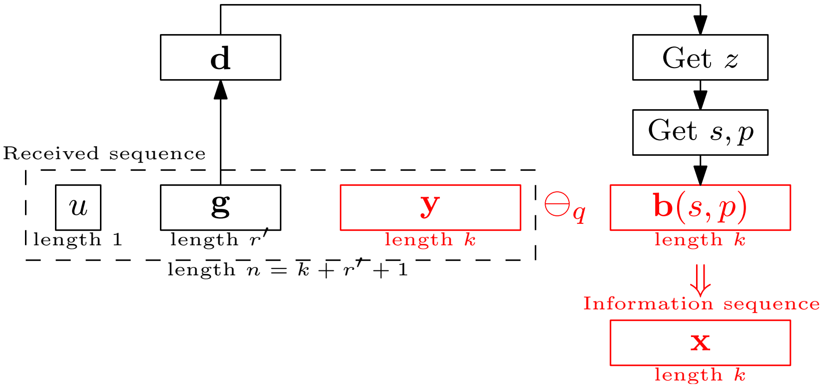
Encoding (Cont'd)

Encoding algorithm: Balance the sequence by finding the correct Gray code prefix:

1. Incrementing through z, determine the balancing sequence $\mathbf{b}(s, p)$ and add it to the information sequence $\mathbf{x}$ to obtain $\mathbf{y}$.
2. For each increment, convert z into base q over r' symbols and determine the corresponding Gray code sequence, $\mathbf{g}$.
3. Set $u = \beta - w(\mathbf{c}')$, provided that $u \in \{0, 1, \ldots, q-1\}$, otherwise set $u = 0$.
4. Continue incrementing z until the weight of u, $\mathbf{g}$ and $\mathbf{y}$ together is equal to β.

Decoding

The figure below shows the decoding process.



Decoding

Example (Decoding)

Consider the case $q = 3$, $n = 13$, where a sequence was encoded as 1012000122022, with a $(3, 3)$-Gray code.

- The first symbol 1 is dropped, then the Gray code sequence is extracted as 012, and decoded as 010.
- Thus $z = 3$, leading to $s = 0$ and $p = 3$, and $\mathbf{b}(0, 3) = 111000000$.
- Finally, the information sequence is recovered as $\mathbf{x} = \mathbf{y} \ominus_q \mathbf{b}(s, p) = 000122022 \ominus_3 111000000 = 222122022.$

Redundancy

The information sequence length k as a function of r for different constructions.

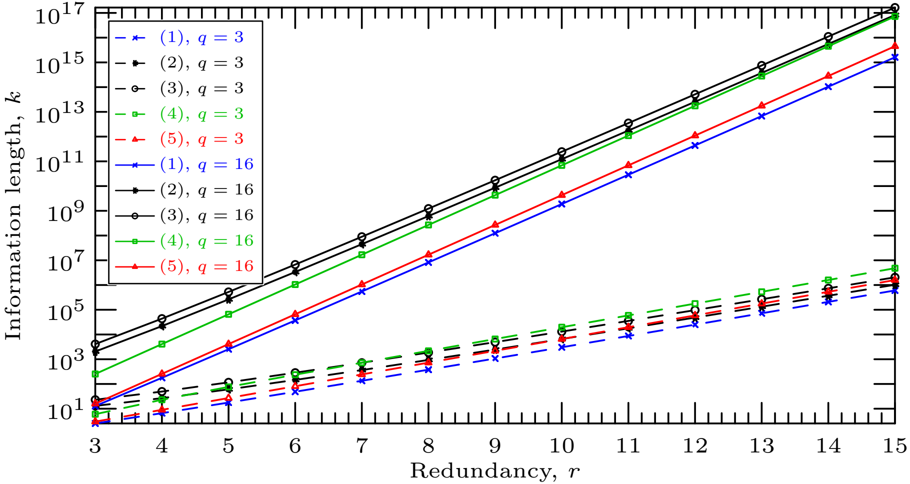
- (1) $k \leq q^{r-1}\sqrt{\frac{6}{\pi r(q^2-1)}}$ [9]
- (2) $k \leq \frac{q^r-1}{r-1}$ [10]
- (3) $k \leq 2\frac{q^r-1}{r-1} - r$ [10]
- (4) $k = q^{r-1} - r$ [11]
- (5) $k = q^{r-2}$ (This is our construction.)

[9] T. G. Swart and J. H. Weber, "Efficient balancing of q-ary sequences with parallel decoding," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009.

[10] R. M. Capocelli, L. Gargano and U. Vaccaro, "Efficient q-ary immutable codes," *Discrete Applied Mathematics*, 1991.

[11] L. G. Tallini and U. Vaccaro, "Efficient m-ary immutable codes," *Discrete Applied Mathematics*, 1999.

Redundancy


10^{17}
10^{15}
10^{13}
10^{11}
10^{9}
10^{7}
10^{5}
10^{3}
10^{1}
Information length, k
(1), $q = 3$
(2), $q = 3$
(3), $q = 3$
(4), $q = 3$
(5), $q = 3$
(1), $q = 16$
(2), $q = 16$
(3), $q = 16$
(4), $q = 16$
(5), $q = 16$
3 4 5 6 7 8 9 10 11 12 13 14 15
Redundancy, r

Complexity

- Previous schemes discussed in [12] and [13] require $\mathcal{O}(qk \log_q k)$ digit operations for the encoding and decoding processes.
- The scheme in [14] requires $\mathcal{O}(qk \log_q k)$ digit operations for the encoding and $\mathcal{O}(k)$ digit operations for the decoding.
- Our decoding process requires $\mathcal{O}(k + \log_q k)$ digit operations. Fast parallel decoding, after $\mathbf{b}(s, p)$ has been determined from the Gray code.
- Encoding takes longer than decoding. In the worst case where the kq-th balancing sequence and Gray code result in balancing, $\mathcal{O}(qk \log_q k)$ digit operations are needed.

[12] R. M. Capocelli, L. Gargano and U. Vaccaro, "Efficient q-ary immutable codes," *Discrete Applied Mathematics*, vol. 33, 1991.

[13] L. G. Tallini and U. Vaccaro, "Efficient m-ary immutable codes," *Discrete Applied Mathematics*, 1999.

[14] T. G. Swart and J. H. Weber, "Efficient balancing of q-ary sequences with parallel decoding," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009.

Conclusion

- A simple algorithm was presented to encode and decode balanced q-ary information sequences of length k, where $k = q^t$. This is based on a Gray code prefix that encodes the balancing index.
- Both the balancing and Gray code algorithms are efficient as only simple addition and subtraction operations are used, and no lookup tables are needed.
- The majority of the decoding algorithm can also be performed in parallel.
- As future work, this algorithm will be extended to q-ary sequences of length k, where $k \neq q^t$.
- An investigation into whether the extra symbol u can be eliminated for certain values of k and q has to be explored.

Thanks for your attention!

"We cannot solve our problems with the same thinking we used when we created them." Albert Einstein

QUESTIONS AND COMMENTS