2012 International Conference on Innovation and Information Management (ICIIM 2012)
IPCSIT vol. 36 (2012) © (2012) IACSIT Press, Singapore

The New Result of Constant-Weight Code Research Based on Construction Algorithm

XuanDong Sun [1,2,a] and Xiaojing Wang [2,b +]

[1.] Chengdu Institute of Computer Application, Chengdu, China
[2.] Institute of computer, Guangdong University of Technology,Guangzhou, China)

Abstract. Group-divisible constant-weight code is the important research field of error-correct code. We specially put forward a calculation method based on constructional algorithm, which is quite different from the widely popularized theoretic deductive method at present. And we have also developed an experimental platform, which is based on this kind of algorithmic path. Our experimental platform has produced a huge number of very interesting data, most of which are very close to the best ones known by the world while part of which are even the same as the best ones. And we also produced a number of codes that no one has achieved at present, which advances the research on constant-weight codes that has been stagnant for many years.

Keywords: Error Correct Code; Group-Divisible Constant-Weight Code; Constructional Algorithm

1. Introduction

Error control coding, also known as error-correcting coding, is the most fundamental and most important means to improve the reliability of information transmission and storage, and is one of the core fundamental technologies in the field of computer and information security. Recently, the combination of error control coding with cryptography also plays a decisive role in information security[1]. In addition, the error control coding idea and technology also have extensive and important applications in many other areas of the IT technology. [2]

Group-divisible constant-weight codes are one of the most basic code types in error-correcting codes, and are of great significance in theory. It attracts much attention, and the world has made unremitting research on it over the past 50 years. However, compared with other code types, it still lacks effective quantitative theories and constructive methods currently, so up to now little progress has been made, hardly able to take one more step[3].

Different from the widespread theoretic deductive methods at present, based on years of research, we specially put forward a research method of the group-divisible constant-weight code based on calculation path, and develop a software experimental platform about the distance nature of block codes. Our experimental platform generates a large amount of varied and interesting data, which greatly expands the last scope of visual imagination, of which many computing results are close to or reach the currently known results in the world deduced from other theoretical approaches, and some are first discovered.

2. Preliminaries

At present, the studies that have much practical value in the world mainly focus on group-divisible constant-weight codes over GF(2), and this paper is also restricted to the discussion on codes over GF(2).

[+] Corresponding author
E-mail address: [a]sunxuandong@163.com, [b]wxjgxlwmd@sina.com

Therefore, in this paper, the range symbol of the code word in a block code is 0 or 1, but the symbol's calculation principles and the following definitions can be completely extended to codes over GF(q) (q is a prime).

Definition 1: Hamming Distance. For two code words of equal length whose element values are 0 or 1, the number of their corresponding elements which are different is known as the hamming distance between the two code words.

Definition 2: Minimum Hamming Distance. It means the minimum of the Hamming distance between all possible pairs of code words of equal length whose element values are 0 or 1.

Definition 3: Constant-weight Code. It is a code in block codes where all code words share the same weight, namely the number of 1 in code words is the same, recorded as (n,w). n represents the number of 1; n-w represents the number of 0; w is the weight of the code word.

We define that A(n,d，w) is the maximum number of code words whose minimum Hamming distances ≥d in block codes (n,w). At this time, the information bits of the block codes A(n,d，w) are $\log_2$ A(n,d，w), recorded as k.

Definition 4: Code Rate. In block codes (n,k), the code rate R=k/n.

Theorem 1: Channel Coding Theorem. Each channel has a defined channel capacity C. For any code rate R which is less than C, each channel has block codes and (n,k,m) convolutional codes with the rate of R and the length of n. If the maximum likelihood decoding is used, with the increase in the code length, its decoding error probability p can be arbitrarily small.

Theorem 2: For any (n,k) block code, in its code words, if we are to:

(1) detect e random errors, it requires that the code word's minimum Hamming distance d≥e+1;

(2) correct t random errors, it requires that the code word's minimum Hamming distance d≥2t+1;

(3) correct t random errors, and detect e random errors, it requires that the code word's minimum Hamming distance d≥e+t+1.

Therefore, when constructing a block code, it is hoped that as d is great enough, the code rate can be also as great as possible. The greater the d, the stronger the error-correcting function; the greater the code rate, the better the performance of block codes.

Block codes without weight limits have many practical applications in error-correcting codes, and group-divisible constant-weight codes were generally regarded as a purely theoretical issue in the past. At present, it is believed that these two codes are both important in practical use, and are applied integratedly in fiber-optic CDMA (code-division mutiple-access) system, no response channel conflict protocol design, automatic answering error-correcting system, etc. [1]

3. Recent Advances in Group-Divisible Constant-Weight Codes

Binary group-divisible constant-weight codes (n,d,w) are a set of vectors with the length of n. In each vector, the number of 1 is w, and the number of 0 is n-w. The Hamming distances of any two different vectors are not less than d. Given the three parameters n,w,d, how shall we get the maximum number of vectors in binary group-divisible constant-weight codes (n,d,w)? Although this problem has been studied for nearly 40 years, it is still one of the fundamental problems and puzzles in coding theory.

The first A(n,d,w) upper bound table was compiled by Mac Williams and Sloane in 1977, where the table for n≤24 and d≤10 was improved in 1978. Honkala's Licentiate also published a table, which was improved to n≤27 and d≤12. So far, little progress has been made on the A(n,d,w) upper bound table. On the contrary, the A(n,d,w) lower bound table has been improved rapidly. In 1980, Graham and Sloane published a lower bound table for n≤25. In 1990, Brouwer, Shearer, Sloane and Smith published a lower bound table for n≤28 and d≤18, which has the best results up to now. [4]

In the A(n,d,w) lower bound table maintained by E. M. Rains and N. J. A. Sloane from AT&T Labs-Research, code words are constructed generally by means of mathematical derivation, so as to obtain a lower

bound when n,w and d are given. [5] This table has the best results of A(n,d,w) lower bounds in the world today (until 2010/12/15).

Under normal circumstances, the A(n,d,w) lower bounds when parameters are given can be obtained through constructing a set of code words. Erik Agrell （IEEE member）, Alexander Vardy,(Fellow, IEEE), Kenneth Zeger（Fellow IEEE）and others found that, in general, the simpler the constructing method, the more code words were constructed, and the simpler the decoding algorithm. They listed in detail the new progress in A(n,w,d) upper and lower bounds in recent years in [3], and improved the upper bound table a lot. They raised 14 known upper bounds, and obtained 7 exact upper bounds when n and w are given.

4. The Construction Algorithm Of Large Distance Group-Divisible Constant-Weight Codes

People have done a lot of work on the approach that uses theoretical derivation to construct large distance code words, but the progress is still very slow. Our algorithm adopts the method based on computational experiments, in accordance with the nature of the proposed method by Erik Agrell, [3]and the results obtained are very satisfactory. Especially in the case of large distance code words, the calculated results are exactly the same as the current best results in the world. In the remaining cases, they are also very close, and there are still many results that nobody has obtained before, which fill in gaps in the work results of international colleagues.

Considering that under the conditions of backward computing facilities we still obtain the best results abroad, it indicates that our method has great potential, and there is still a lot of work worth deeply carrying out.

4.1. Arrangement and counting of constant-weight codes

For code words with the length of n and the distance of w, they can share at most C_n^w possible code words, and we made a study on the arrangement of these possible code words. We found that, the search results of different code word sets depend heavily on the order of these code words, so that we cannot realize computer-aided code word structure research. Only one of the symmetrical arrangements has the best properties——search strategies on it produce no influence or little influence on the search results of code word sets, which is a basis for our algorithm to be effectively achieved. The algorithm of this symmetrical arrangement is as follows:

```
begin
   If w=1 then /* weight is 1,generate w vector*/
   Begin
(C1 C2 …Cn) =(0,0, …0)
（i） for i=1 to n do
       Ci =1 ；Output (C1 C2 …Cn)；Ci =0
     end for
    end if
   if n=w then
      (C1 C2 …Cw) =(1,1, …1)
Output (C1 C2 …Cn)
   end if
   GenerateVector(n-1,w-1) /* recursion call */
   C1 =0
(ii) ( C2 …Cw+1) =(1,1, …1)
(iii)(Cw+2 C w+3 …Cn) =(0,0, …0)
  GenerateVector(n-1,w) /* recursion call */
```

End

(i) requires n steps, and (ii) and (iii) need a total of n steps. The recursive program will be run a total of C_n^w times, so the whole process will be run nC_n^w times.

4.2. The generation of large distance code word subsets that meet the conditions

Based on the algorithm in 3.1 and combining the properties of code word sets, we can get an algorithm which may generate large distance code word subsets that meet the conditions:

```
define set S //S save generate subset of vector
empty(s);
for  i=1 to Cnw do
  Ci=generateVector(n,w,i);//Ci is vector generated by Algorithm 3.1
  for j=1 to length(S) do
     if disance(Ci,Cj)< d then //Ci can not push into S
        break;
  end for
  if j=length(S)+1 then
     push (S,Ci)
  end if
end for
```

This algorithm is run a total of n | S | C_n^w steps, generally | S | >n, which thus corresponds to $n^2C_n^w$ steps. This is the minimum computational complexity which can only be achieved by exhaustion algorithmic path by far, while other methods all fail to complete the exact search.

5. The Latest Experimental Data and Its Properties

A large amount of data is produced from the experimental platform of large distance code words written based on the algorithm of 3.2. We made a detailed analysis on the data, and compared it with the current best results in the world. At last, we obtained many meaningful properties of group-divisible constant-weight codes.

In most cases, our operational results are very close to the current best results in the world. In the cases when the difference of w and n is very small, or n, w and d are 2^i, our results are equal to the current international best results. Especially when d is great, we can get a lot of results that no one has obtained at present. We use A(n,w,d) to indicate the current international best results, and A*(n,w,d) is used to indicate our results. A^(n,w,d) is used to indicate our results that have not been obtained by others.(until now, the current international best results are shown in [5]) The following are some comparisons:

A(6,3,4)=4 A*(6,3,4)=4

A(7,3,4)=7 A*(7,3,4)=7

A(16,8,4)=1170 A*(16,8,4)=918

A(20,10,4)=13452 A*(12,6,4)=6140

A(16,8,8)=32 A*(16,8,8)=32

A(32,16,16)=64 A*(32,16,16)=64

……

A(38,8,8)=2997 A*(38,8,8)=2439

A^(32,7,10)=46

Based on the research on data that have been obtained, we construct a number of A(n,w,d) values that have not been obtaned in the world today especially when d is relatively large. The larger the code distance, the greater its role in information security, but the more difficult is the code constructing. For example, A(64,32,32)A(128,64,64),A(256,128,128), etc.

Due to space limitations, many new results have not been published in this article. Please contact with the author by email.

6. Summary

Error-correcting coding is the core technology in the field of computer and information security, and group-divisible constant-weight codes are the most important and the most fundamental in error-correcting codes. We put forward a calculation-based group-divisible constant-weight code construction method which is different from the current widely used method of theoretical derivation and mathematical calculation, and develop an experimental platform based on this algorithm. The experimental platform produces a lot of meaningful experimental data, which approaches and reaches the current international best results. We actually have also produced a number of codes that no one has achieved at present, which advances the research on constant-weight codes that has been stagnant for many years, and fills in gaps in this area in the world. It suggests that we can find more and better constant-weight codes based on this kind of algorithmic path, and this research path has much practical value and great potential for further development. Based on the test platform of the PC version that has been completed, we are trying for the experimental platform based on parallel algorithm and grid computing, hoping to obtain more new results more quickly.

7. Acknowledgment

This work is supported by the Industry-Education-Research Cooperation Project of Guangdong Province and the Ministry of Education (No.2011A090200068) .

8. References

[1] X. M. Wang and G. Z. Xiao. Error-correcting Codes—principles and methods[M]. Xi'an: Xidian University Press,2001：120－125.

[2] X. M. Wang, W. P. Ma and C. K. Wu. Theory of error-correcting codes[M]. Beijing: People's Posts and Telecom Press, 2001：13-15.

[3] E. Agrell, A. Vardy and K. Zeger, Bounds for Constant-Weight Codes[J]. IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 46, NO. 7. NOVEMBER 2000:2373-2381

[4] S. Verdii, Fellow, IEEE, and Victor K. Wei, Member IEEE. [J] Explicit Construction of Optimal Constant-Weight Codes for Identification via Channels， IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 39, NO. 1, JANUARY 1993: 30-36.

http://www.research.att.com/~njas/codes/Andw/

[5] I. Gashkov . Optimal Constant Weight Codes.[M] Springer Berlin / Heidelberg:Lecture Notes in Computer Science, 2006: 912-915.

[6] D. H. Wu and P. Z. Fan .Constructions of optimal quaternary constant weight codes via group divisible designs[J].Discrete Mathematics, volume.309, 2009, Pages 6009-6013.