Error-Correcting Codes for Cryptography

Jon-Lark Kim

The CODING-A Lab
Department of Mathematics
Sogang University, Seoul, Korea

http://maths.sogang.ac.kr/jlkim
Email: jlkim@sogang.ac.kr

PROOFS, Busan Korea
September 27, 2014
Outline

- Introduction to Coding Theory
- Complementary Information Set (CIS) Codes
- General Constructions Including SRG and DRT.
- Classification of CIS Codes of Lengths $\leq 12$
- Optimal CIS Codes of Lengths $\leq 130$
- Long CIS Codes
- Higher-Order CIS Codes
- Conclusion and Open Problems
Overview

Information Theory Basic

Coding Theory
  - Linear Algebra
  - Combinatorics
  - Algebra

Cryptography
  - Number Theory
  - Probability
  - Complexity
Father of Information Theory

Figure: Claude Shannon (1916-2001)

Shannon’s two foundational papers from Bell System Technical Journal:
“A Mathematical Theory of Communication” on Information Theory (1948)
“Communication Theory of Secrecy Systems” on Cryptography (1949)
What is a code?

- Let $A$ be a finite alphabet. Usually $A = \mathbb{Z}_2, \mathbb{Z}_p$ (in general $\mathbb{F}_q, \mathbb{Z}_m$, chain rings, Galois rings, or Frobenius rings).

- $A^n := \{(x_1, \cdots, x_n) | x_i \in A\}$.

- An (error-correcting) code $C$ over $A$ is a subset of $A^n$ (with at least two elements).

- Elements of $C$ are called codewords.

- A code over $\mathbb{Z}_2$ is called a binary code.

- The weight of $x = (x_1, \cdots, x_n)$ is the number of nonzero coordinates, denoted by $wt(x)$. For example, $wt(0, 1, 2, 1, 0) = 3$.

- The Hamming distance $d(x, y)$ between $x, y \in A^n$ is $wt(x - y)$. For example, if $x = (1, 0, 0, 1, 0)$ and $y = (0, 0, 1, 0, 0)$, then their Hamming distance is 3.
Linear codes: most useful codes

- A linear code $C$ of length $n$ and dimension $k$ over $\mathbb{Z}_p$ is a $k$-dimensional subspace of $\mathbb{Z}_p^n$.
- We denote $C$ by an $[n, k]$ linear code over $\mathbb{Z}_p$.
- The minimum distance (weight) $d$ of a linear code $C$ is the minimum of $\text{wt}(x)$, $x \neq 0 \in C$.
- We denote it by an $[n, k, d]$ code. Given $n$ and $k$, $d$ can be at most $n - k + 1$ (Singleton’ bound).
- A set of $k$ columns of an $[n, k, d]$ code is called an information set if it is linearly independent.
How many errors can correct?

**Theorem**

Any \([n, k, d]\) linear code can correct up to \(t = \left\lfloor \frac{d-1}{2} \right\rfloor\) errors (by the nearest neighbor decoding).
Let $C$ be a linear $[n, k, d]$ code over finite field $GF(q)$ of length $n$, dimension $k$ and minimum distance $d$.

The Euclidean inner product of $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $GF(q)^n$ is $x \cdot y = \sum_{i=1}^{n} x_i y_i$.

The dual of $C$, denoted by $C^\perp$ is the set of vectors orthogonal to every codeword of $C$ under the Euclidean inner product.

If $C = C^\perp$, $C$ is called self-dual (sd), and if $C \subset C^\perp$, self-orthogonal.
• The weight enumerator of $C$ is the polynomial $W_C(X, Y) = \sum_{i=0}^{n} A_i X^{n-i} Y^i$, where $A_i$ is the number of codewords of weight $i$.

• A code $C$ is called formally self-dual (f.s.d.) if $W_{C^\perp}(x, y) = W_C(x, y)$.

• Of course any self-dual code is an f.s.d. code but an f.s.d. code is not necessarily self-dual.

• A code $C$ is divisible by $\delta$ provided all codewords have weights divisible by an integer $\delta$, called a divisor of $C$. 

- Let $C$ have generator matrix

$$G = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0
\end{bmatrix}$$

- $C$ is the famous extended Hamming $[8, 4]$ code with minimum distance $d = 4$.
- $C$ is self-dual.
- Weight Distribution: $A_0 = 1$, $A_4 = 14$, $A_8 = 1$.
- Divisor $\delta = 4$. 
Why Self-dual codes?

- One of the most interesting classes of linear codes
- Connections with group theory, design theory, Euclidean lattices, modular forms, quantum codes
- Many optimal linear codes are often self-orthogonal/self-dual.
- They are also asymptotically good.
Why Self-dual codes?

- One of the most interesting classes of linear codes
- Connections with group theory, design theory, Euclidean lattices, modular forms, quantum codes
- Many optimal linear codes are often self-orthogonal/self-dual.
- They are also asymptotically good.

Question: Is there an interesting superclass of self-dual codes?
Complementary Information Set Codes

- A binary linear code of length $2n$ and dimension $n$ is called Complementary Information Set (CIS) with a partition $L, R$ if there is an information set $L$ whose complement $R$ is also an information set.


- We call the partition $[1..n], ..., [n + 1..2n]$ the systematic partition.
- Systematic self-dual codes are CIS with the systematic partition.
- It is also clear that the dual of a CIS code is CIS.
- Hence CIS codes are a natural generalization of self-dual codes.
Walsh Hadamard transform

- An vectorial Boolean function $F$ is any map from $\mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$.
- Its Walsh Hadamard transform of $F$ at $(a, b)$ is defined as
  \[ W_F(a, b) = \sum_{x \in \mathbb{F}_2^n} (-1)^{a \cdot x + b \cdot F(x)}, \]
  where $a \cdot x$ denotes the scalar product of vectors $a$ and $x$.
- If $f$ is a Boolean function with domain $\mathbb{F}_2^k$ and range $\mathbb{F}_2$, then the Fourier transform $\hat{f}$ of $f$ at $a$ is defined by
  \[ \hat{f}(a) = \sum_{x \in \mathbb{F}_2^k} f(x)(-1)^{a \cdot x} = \sum_{x \in \text{supp}(f)} (-1)^{a \cdot x}, \]
  where $\text{supp}(f)$ is the support of function $f$.
- We note that for $a \neq 0$,
  \[ W_{F_1}(a, b) = 0 \text{ if and only if } \hat{b} \cdot F_1(a) = 0. \]
Motivations

CIS codes have an application in cryptography, in the framework of counter-measures to side channel attacks on smartcards.
Motivations

• Assuming a systematic unrestricted code $C$ of length $2n$ of the form

$$C = \{(x, F(x)) | x \in \mathbb{F}_2^n\},$$

the vectorial Boolean function is constructed as the map $x \mapsto F(x)$.

• In that setting $C$ is CIS by definition iff $F$ is a bijection.

• When $C$ is a linear code, we can also consider a systematic generator matrix $(I, A)$ of the code, where $I$ is the identity matrix of order $n$ and $A$ is a square matrix of order $n$. Then $F(x) = xA$, and the CIS condition reduces to the fact that $A$ is nonsingular.
Motivations-continued

The physical implementation of cryptosystems on devices such as smart cards leaks information.
Motivations-continued

- This information can be used in differential power analysis or in other kinds of side channel attacks.
- These attacks can be disastrous if proper counter-measures are not included in the implementation.
- Until recently, it was believed that for increasing the resistance to attacks, new masks have to be added, thereby increasing the order of the countermeasure.
  

- Change the variable representation (say \( x \)) into randomized shares \( m_1, m_2, \ldots, m_{t+1} \) called masks such that \( x = m_1 + m_2 + \cdots + m_{t+1} \) where + is a group operation - in practice, the XOR.

- At the order \( t = 1 \), the masks are given by \((m_1, m_2) = (m_1, x + m_1)\). If both \( m_1 \) and \( x + m_1 \) are known, then \( x \) is obtained, hence not secure.
Motivations-continued

- It is shown that another option consists in encoding the some of masks, which is much less costly than adding fresh masks.
  

- For example, at the order $t = 1$, using a vectorial Boolean function $F$, we consider the ordered pair $(F(m_1), x + m_1)$.

- Notably, it is demonstrated that the same effect as adding several masks can be obtained by the encoding of one single mask.

Graph Correlation Immune Functions

- This method, called leakage squeezing, uses vectorial Boolean functions - more precisely, permutations $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$, such that, given some integer $d$ as large as possible, for every pair of vectors $a, b \in \mathbb{F}_2^n$ such that $(a, b)$ is nonzero and has Hamming weight $< d$, the value of the Walsh Hadamard transform of $F$ at $(a, b)$, is null.

- We call such functions $d$-GCI, for Graph Correlation Immune.

- Thus a $d$-GCI function is a protection against an attack of order $d$.

Proposition (Maghebi, et. al, 2011)
The existence of a linear $d$-GCI function of $n$ variables is equivalent to the existence of a CIS code of parameters $[2n, n, \geq d]$ with the systematic partition.
General construction

Lemma
If a \([2n, n]\) code \(C\) has generator matrix \((I, A)\) with \(A\) invertible then \(C\) is CIS with the systematic partition. Conversely, every CIS code is equivalent to a code with a generator matrix in that form.

In particular this lemma applies to systematic self dual codes whose generator matrix \((I, A)\) satisfies \(AA^T = I\).

Lemma
Let \(f(x)\) be a polynomial over \(\mathbb{F}_2\) of degree less than \(n\). Then, \(\gcd(f(x), x^n - 1) = 1\) if and only if the circulant matrix generated by \(f(x)\) has \(\mathbb{F}_2\)-rank \(n\).
Proposition

The double circulant code whose generator matrix is represented by \((1, f(x))\) satisfying Lemma is a CIS code.

Proposition

If a \([2n, n]\) code \(C\) has generator matrix \((I, A)\) with \(rk(A) < n/2\) then \(C\) is not CIS.
Rank criterion for linear codes

Theorem
Let $\Sigma$ denote the set of columns of the generator matrix of a $[2n, n]$ linear code $C$.
$C$ is CIS iff $\forall B \subseteq \Sigma$, $rk(B) \geq |B|/2$.

The proof uses matroid theory and Edmonds’ matroid base packing theorem: A matroid on a set $S$ contain $k$ disjoint bases iff

$$\forall U \subseteq S, k(rk(S) - rk(U)) \leq |S \setminus U|.$$ 

Apply to the matroid of the columns of the generator matrix under linear dependence, with

$$S = \Sigma, k = 2, rk(\Sigma) = n, |\Sigma| = 2n.$$
SRG and DRT

- Let $A$ be an integral matrix with $0, 1$ valued entries.
- We say that $A$ is the adjacency matrix of a strongly regular graph (SRG) of parameters $(n, \kappa, \lambda, \mu)$ if $A$ is symmetric, of order $n$, verifies $AJ = JA = \kappa J$ and satisfies
  \[ A^2 = \kappa I + \lambda A + \mu(J - I - A) \]
- We say that $A$ is the adjacency matrix of a doubly regular tournament (DRT) of parameters $(n, \kappa, \lambda, \mu)$ if $A$ is skew-symmetric, of order $n$, verifies $AJ = JA = \kappa J$ and satisfies
  \[ A^2 = \lambda A + \mu(J - I - A) \]
where $I, J$ are the identity and all-one matrices of order $n$. 
**CIS codes from SRG and DRT**

In the next result we identify $A$ with its reduction mod 2.

**Proposition**
Let $C$ be the linear binary code of length $2n$ spanned by the rows of $(I, M)$. With the above notation, $C$ is CIS if $A$ is the adjacency matrix of a

- SRG of odd order with $\kappa, \lambda$ both even and $\mu$ odd and if $M = A + I$
- DRT of odd order with $\kappa, \mu$ odd and $\lambda$ even and if $M = A$
- SRG of odd order with $\kappa$ even and $\lambda, \mu$ both odd and if $M = A + J$
- DRT of odd order with $\kappa$ even and $\lambda, \mu$ both odd and if $M = A + J$
Quadratic Double Circulant Codes

Let $q$ be an odd prime power. Let $Q$ be the $q$ by $q$ matrix with zero diagonal and $q_{ij} = 1$ if $j - i$ is a square in $GF(q)$ and zero otherwise. (This $Q$ is a modified Jacobsthal matrix.)

**Corollary**

If $q = 8j + 5$ then the span of $(I, Q + I)$ is CIS. If $q = 8j + 3$ then the span of $(I, Q)$ is CIS.

**Proof**

It is well-known that if $q = 4k + 1$ then $Q$ is the adjacency matrix of a SRG with parameters $(q, \frac{q-1}{2}, \frac{q-5}{4}, \frac{q-1}{4})$. If $q = 4k + 3$ then $Q$ is the adjacency matrix of a DRT with parameters $(q, \frac{q-1}{2}, \frac{q-3}{4}, \frac{q+1}{4})$. The result follows by the previous proposition.

The codes obtained in that way are Quadratic Double Circulant codes (Gaborit, 2002).
Existence of an optimal code that is not CIS

**Proposition**
If $C$ is a $[2n, n]$ code whose dual has minimum weight 1 then $C$ is not CIS.

**Proposition**
There exists at least one optimal binary code that is not CIS.

**Proof:**
The $[34, 17, 8]$ code described in the Magma package $BKLC(GF(2), 34, 17))$ (best known linear code of length 34 and dimension 17) is an optimal code (minimum weight 8 is the best possible minimum distance for such a code) which dual has minimum distance 1, and therefore is not CIS.
Classification of CIS codes of lengths \( \leq 12 \)

- Let \( n \geq 2 \) be an integer and \( g_n \) denote the cardinal of \( GL(n,2) \) the general linear group of dimension \( n \) over \( GF(2) \).
- It is well-known (see MacWilliams-Sloane’s book), that
  \[
  g_n = \prod_{j=0}^{n-1} (2^n - 2^j) .
  \]

**Proposition**

The number \( e_n \) of equivalence classes of CIS codes of dimension \( n \geq 2 \) is at most \( g_n/n! \).

**Proof:**

Every CIS code of dimension \( n \) is equivalent to the linear span of \((I, A)\) for some \( A \in GL(n,2) \). But the columns of such an \( A \) are pairwise linearly independent, hence pairwise distinct. Permuting the columns of \( A \) lead to equivalent codes.
Examples

• There is a unique CIS code in length 2 namely $R_2$ the repetition code of length 2.

• For $n = 2$, the $g_2 = 6$ invertible matrices reduce to three under column permutation: the identity matrix $I$ and the two triangular matrices $T_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, and $T_2 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$.

• The generator matrix $(I, I)$ spans the direct sum $R_2 \oplus R_2$, while the two codes spanned by $(I, T_1)$ and $(I, T_2)$ are equivalent to a code $C_3$, an isodual code which is not self dual. Thus $e_2 = 2 < \frac{g_2}{2!} = 3$. 
Shortening

The building up construction is known for binary self-dual codes. In this section, we extend it to CIS codes. We show that every CIS code can be constructed in this way.

**Lemma**
Given a $[2n, n]$ CIS code $C$ with generator matrix $(I_n|A)$ where $A$ is an invertible square matrix of order $n$, we can obtain a $[2(n - 1), n - 1]$ CIS code $C'$ with generator matrix $(I_{n-1}|A')$, where $A'$ is an invertible square matrix of order $n - 1$. 
Building up construction

Suppose that $C$ is a $[2n, n]$ CIS code $C$ with generator matrix $(I_n|A)$, where $A$ is an invertible matrix with $n$ rows $r_1, \ldots, r_n$. Then for any two vectors $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)^T$ the following matrix $G_1$ generates a $[2(n + 1), n + 1]$ CIS code $C_1$ with the systematic partition:

$$ G_1 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & z_1 & x \\ 0 & 1 & 0 & \cdots & 0 & y_1 & r_1 \\ 0 & 0 & 1 & \cdots & 0 & y_2 & r_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & y_n & r_n \end{pmatrix}, \quad (2) $$

where $c_i$’s satisfy $x = \sum_{i=1}^{n} c_i r_i$ and $z_1 = 1 + \sum_{i=1}^{n} c_i y_i$. 

Building up construction
Example

- Let us consider a [6, 3, 3] CIS code $C$ whose generator matrix is given below.

$$G = \begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1
\end{pmatrix}.$$  

- In order to apply the building-up construction, we take for example $x = (1, 1, 0)$ and $y = (1, 1, 0)^T$. Then $c_1 = c_2 = 1$, $c_3 = 0$. Hence $z = 1$.

- In fact, we get the extended Hamming [8, 4, 4] code whose generator matrix is given below.

$$G_1 = \begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 1
\end{pmatrix}.$$
Converse of the building up construction

Proposition
Any $[2n, n]$ CIS code $C$ is equivalent to a $[2n, n]$ CIS code with the systematic partition which is constructed from a $[2(n - 1), n - 1]$ CIS code by using the building up construction.
Proposition

Let $n \geq 2$. Let $\mathbf{C}$ be the set of all $[2n, n]$ CIS codes and let $S_{2n}$ act on $\mathbf{C}$ as column permutations of the codes in $\mathbf{C}$. Let $C_1, \ldots, C_s$ be representatives from every equivalence class of $\mathbf{C}$ under the action of $S_{2n}$. Let $\mathbf{C}_{sys}$ be the set of all $[2n, n]$ CIS codes with generator matrix $(I_n|A)$ with $A$ invertible. Suppose that each $C_i \in \mathbf{C}_{sys}$ ($1 \leq i \leq s$). Then we have

$$g_n = \sum_{j=1}^{s} |\text{Orb}_{S_{2n}}(C_j) \cap \mathbf{C}_{sys}|,$$

where $\text{Orb}_{S_{2n}}(C_j)$ denotes the orbit of $C_j$ under $S_{2n}$. 

Counting formula similar to mass formula
Classification of CIS codes of lengths 2, 4

We classify all CIS codes of lengths up to 12 up to equivalence using the building up method. It is easy to see that any CIS code has minimum distance $\geq 2$.

- $2n = 2$. It is clear that there is a unique CIS code of length 2, whose generator matrix is $[11]$.

- $2n = 4$. Applying Proposition (building-up) to the repetition code of generator matrix $[11]$, we show that there are exactly two CIS codes of length 4. Their generator matrices are ($I|A_{2,1}$) and ($I|A_{2,2}$), where

\[
A_{2,1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_{2,2} = T_2 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}
\]
Proposition

There are exactly six CIS codes of length 6. Only one code has $d = 3$ and the rest have $d = 2$.

$(l|A)$, where $A$ is one of the following.

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{pmatrix},
\begin{pmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 0 & 1 \\
\end{pmatrix},
\begin{pmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 1 \\
\end{pmatrix},
\begin{pmatrix}
1 & 0 & 0 \\
1 & 1 & 1 \\
1 & 1 & 0 \\
\end{pmatrix},
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 1 \\
\end{pmatrix},
\begin{pmatrix}
1 & 0 & 1 \\
0 & 0 & 1 \\
1 & 0 & 1 \\
\end{pmatrix}
\]
Summary: Classification of all CIS codes of lengths up to 12 in the order of sd, non-sd fsd, and none of them

<table>
<thead>
<tr>
<th>2n</th>
<th>$d = 2$</th>
<th>$d = 3$</th>
<th>$d = 4$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1 (1+0+0)</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2 (1+1+0)</td>
<td>1 (0+1+0)</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>5 (1+2+2)</td>
<td>4 (0+2+2)</td>
<td>1 (1+0+0)</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>22 (1+9+12)</td>
<td>35 (0+9+26)</td>
<td>4 (0+2+2)</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>156 (2+40+114)</td>
<td>565 (0+87+478)</td>
<td>41 (1+7+33)</td>
<td>195</td>
</tr>
<tr>
<td>12</td>
<td>2099 (2+318+1779)</td>
<td></td>
<td></td>
<td>2705</td>
</tr>
</tbody>
</table>

Recently, Finley Freibert (Ohio Dominican University) in his thesis has classified all CIS codes of length 14 and all CIS codes of length 16 and $d = 4$. 
CIS codes of lengths $\leq 130$ with record distances

**Theorem**
There exist optimal or best-known CIS codes of lengths $2n \leq 130$.

<table>
<thead>
<tr>
<th>$2n$</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>2*</td>
<td>2*</td>
<td>3*</td>
<td>4*</td>
<td>4*</td>
<td>4*</td>
<td>4*</td>
<td>5*</td>
<td>6*</td>
<td>6*</td>
<td>7*</td>
</tr>
<tr>
<td>code</td>
<td>dc</td>
<td>dc</td>
<td>~dc</td>
<td>sd</td>
<td>dc</td>
<td>sd</td>
<td>sd</td>
<td>~dc</td>
<td>~dc</td>
<td>nfsd</td>
<td>id</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$2n$</th>
<th>102</th>
<th>104</th>
<th>106</th>
<th>108</th>
<th>110</th>
<th>112</th>
<th>114</th>
<th>116</th>
<th>118</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>19</td>
<td>20</td>
<td>19</td>
<td>20</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>code</td>
<td>bk</td>
<td>bk</td>
<td>qdc</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
<td>sc</td>
<td>sc</td>
<td>sd</td>
</tr>
</tbody>
</table>
Long CIS codes

We begin by a well-known fact from MacWilliams-Sloane.

**Lemma**
The number of invertible $n$ by $n$ matrices is $\sim c2^{n^2}$, with $c \approx 0.29$.

Denote by $B(n, d)$ the number of matrices $A$ such that $d$ columns or less of $(I, A)$ are linearly dependent. A crude upper bound on this function can be derived as follows.

**Lemma**
The quantity $B(n, d)$ is $\leq M(n, d)$ where

$$M(n, d) = \sum_{j=2}^{d} \sum_{t=1}^{j-1} \binom{n}{j-t} \binom{n}{t} t2^{n(n-1)}.$$
CIS codes are asymptotically good

Denote by $H(x) = -x \log_2 x - (1 - x) \log_2 (1 - x)$ the binary entropy function.

**Lemma**
The quantity $M(n, d)$ is dominated by $2^{n^2 - n} 2^{2nH(\delta)}$ when $d \sim 2\delta n$ with $0 < \delta < 1$.

**Proposition**
For each $\delta$ such that $H(\delta) < 0.5$ there are long CIS codes of relative distance $\delta$.

**Proof:**
Consider $(I, A)$ as the parity check matrix of the CIS code and combine the above lemmas to ensure that, asymptotically, $|GL(n, 2)| >> B(n, d)$ showing the existence of a CIS code of distance $> d$, for $n$ large enough.
Higher-order CIS codes

- The generator matrix of a $[tk, k]$ code is said to be in **systematic form** if these columns are at the first $k$ positions, that is, if it is blocked as $(I_k|A)$ with $I_k$ the identity matrix of order $k$.

- We call a systematic code of length $tk$ which admits $t$ pairwise disjoint information sets a **$t$-CIS (unrestricted) code**.

- Therefore, 2-CIS codes mean the above CIS codes.

A pair \((F_1, F_2)\) of permutations of \(\mathbb{F}_2^k\) forms a Correlation Immune Pair (CIP) of strength \(d\) if and only if for every \((a, b, c)\) such that \(a, b, c \in \mathbb{F}_2^k, a \neq 0\), and 
\[ w_H(a) + w_H(b) + w_H(c) \leq d, \]
we have \(b \cdot F_1(a) = 0\) or \(c \cdot F_2(a) = 0\), equivalently \(W_{F_1}(a, b) = 0\) or \(W_{F_2}(a, c) = 0\).

It expresses the fact that the leakage squeezing with two masks (i.e., \(t = 3\) shares) and two permutations \(F_1\) and \(F_2\) allows to resist high-order attacks of order \(d\).

We here give it the name of CIP of strength \(d\).
Equivalent form of CIP

The definition of a CIP of strength $d$ is equivalent to Condition (8) in the below reference:

$$\forall a \in \mathbb{F}_2^k, a \neq 0, \exists q, r \text{ such that}$$

$w_H(a) + q + r = d - 1,$

$$\forall b \in \mathbb{F}_2^k, w_H(b) \leq q \implies \hat{b} \cdot F_1(a) = 0,$$

$$\forall c \in \mathbb{F}_2^k, w_H(c) \leq r \implies c \cdot F_2(a) = 0.$$

We are now ready for the coding theoretic characterization of CIP.

**Theorem**
If $F_1$, $F_2$ are permutations of $\mathbb{F}_2^k$ then they form a CIP of strength $d$ if and only if the systematic code of length $3k$ and size $2^{2k}$

$$C(F_1, F_2) = \{(x + y, F_1(x), F_2(y)) | x, y \in \mathbb{F}_2^k\}$$

(4)

has dual distance at least $d + 1$. 
Theorem (Carlet, Danger, Guilley, Maghrebi)
If \( F_1, F_2 \) are linear permutations of \( \mathbb{F}_2^k \), then they form a CIP of strength \( d \) if and only if the \([3k, k]\) linear code
\[
\mathcal{C}(F_1, F_2) = \{(u, G_1(u), G_2(u)) | u \in \mathbb{F}_2^k\}
\]
is 3-CIS and has minimum distance at least \( d + 1 \).
Here \( G_1 = (F_1^*)^{-1}, G_2 = (F_2^*)^{-1} \) where \( F^* \) denotes the adjoint operator of \( F \), that is, the operator whose matrix is the transpose of that of \( F \).

Proof
The code \( \mathcal{C}(F_1, F_2) \) being the set of words \((x + y, F_1(x), F_2(y))\), with \( x, y \in \mathbb{F}_2^k \), its dual \( \mathcal{C}^\perp \) is the set of words \((u, v, w)\) such that
\[
(x + y) \cdot u + F_1(x) \cdot v + F_2(y) \cdot w
= x \cdot (u + F_1^*(v)) + y \cdot (u + F_2^*(w))
= 0 \text{ for every } x, y \in \mathbb{F}_2^k.
\]
Hence \( \mathcal{C}^\perp \) is the set of words \((u, v, w)\) such that \( u = F_1^*(v), u = F_2^*(w) \) so that
\[
v = (F_1^*)^{-1}(u) = G_1(u), w = (F_2^*)^{-1}(v) = G_2(u). \text{ The result follows.}
\]
Correlation Immune $t$-uple ($t$-CI) of strength $d$

More generally we make the following definition for $t \geq 2$. The $t$-uple $F_1, \cdots, F_t$ of permutations of $\mathbb{F}_2^k$ form a Correlation Immune $t$-uple ($t$-CI) of strength $d$ if and only if for every $(a_0, \cdots, a_t)$ such that $a_0 \neq 0$ and $w_H(a_0) + \cdots + w_H(a_t) \leq d$, we have that

$$\prod_{i=1}^{t} \widehat{a_i \cdot F_i(a_0)} = 0.$$
**t-CIS Partition Algorithm:**

An algorithm to determine if a given linear code is t-CIS.

**Input:** Begin with a binary \([tk, k]\) code \(C\).

**Output:** An answer of “Yes” if \(C\) is t-CIS (along with a column partition) and an answer of “No” if not.

1. Let \(\{l_1, \ldots, l_t\}\) be a set of labeled disjoint independent subsets of \(M\). (Note that each \(l_i\) \((1 \leq i \leq t)\) can be randomly assigned to each have order 1, or one may be given the first \(k\) indices of a standard form matrix \(G\).)

2. Select \(x \in M \setminus \bigcup_{1 \leq i \leq t} l_i\).

3. While \(\bigcup_{1 \leq i \leq t} l_i \subsetneq M\) do:
   3.1  Initialize \(S_0 := M\). For \(j > 0\), recursively define \(S_j := \text{span}(l_{j'} \cap S_{j-1})\), where \(j' = ((j - 1) \mod t) + 1\). Initialize \(j := 0\).
   3.2  For the current value of \(j\) check that \(|S_j| \leq t \cdot \text{rank}(S_j)\). If the inequality is false (it is immediately clear that Edmonds’ Theorem is violated), then exit the while loop and output the set \(S_j\) with an answer of “No.”
   3.3  If \(x \in S_j\), then set \(j := j + 1\) and go back to b).
   3.4  If \(x \notin S_j\), then check if \(l_{j'} \cup \{x\}\) is independent. If so then replace \(l_{j'}\) with the larger independent set and repeat the while loop with a new \(x \in M \setminus \bigcup_{1 \leq i \leq t} l_i\).
   3.5  If \(l_{j'} \cup \{x\}\) is dependent, then find the unique minimal dependent set \(C \subset l_{j'} \cup \{x\}\) (accomplished by solving the matrix equation associated with finding the linear combination of columns in \(l_{j'}\) that sum to \(x\)).
   3.6  Select any \(x' \in C \setminus S_{j-1}\) and replace \(l_{j'}\) with \(l_{j'} \cup \{x\} \setminus \{x'\}\), then set \(x := x'\) and repeat the while loop.

4. End while loop. If the while loop was not exited early, then output the partition \(\{l_1, \ldots, l_t\}\) of \(M\) and answer “Yes.”
The table captions are as follows.

- $bk =$ obtained by the command $BKLC(GF(2), n, k)$ from Magma.
- $bk^* =$ same as $bk$ with successive zero columns of the generator matrix replaced in order by successive columns of the identity matrix of order $k$. Trivially the generator matrix of $bk$ has $< k$ zero columns.
- $qc =$ quasi-cyclic.

The following tables show that all 3-CIS codes of dimension 3 to 85 have the best known minimum distance among all linear $[n, k]$ codes, and in fact the best possible minimum distance for $n \leq 36$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>6</th>
<th>9</th>
<th>12</th>
<th>15</th>
<th>18</th>
<th>21</th>
<th>24</th>
<th>27</th>
<th>30</th>
<th>33</th>
<th>36</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>$d$</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>code</td>
<td>qc</td>
<td>qc</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
<td>bk*</td>
<td>bk*</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
<td>bk*</td>
<td>bk*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>123</th>
<th>126</th>
<th>129</th>
<th>132</th>
<th>135</th>
<th>138</th>
<th>141</th>
<th>144</th>
<th>147</th>
<th>150</th>
<th>153</th>
<th>156</th>
<th>159</th>
<th>162</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
</tr>
<tr>
<td>$d$</td>
<td>29</td>
<td>31</td>
<td>32</td>
<td>?</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>code</td>
<td>bk*</td>
<td>bk*</td>
<td>bk*</td>
<td>?</td>
<td>bk*</td>
<td>bk*</td>
<td>bk*</td>
<td>bk</td>
<td>bk</td>
<td>bk*</td>
<td>bk*</td>
<td>bk*</td>
<td>bk*</td>
<td>bk</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>165</th>
<th>168</th>
<th>171</th>
<th>174</th>
<th>177</th>
<th>180</th>
<th>183</th>
<th>186</th>
<th>189</th>
<th>192</th>
<th>195</th>
<th>198</th>
<th>201</th>
<th>204</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
<td>66</td>
<td>67</td>
<td>68</td>
</tr>
<tr>
<td>$d$</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>40</td>
<td>41</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>41</td>
</tr>
<tr>
<td>code</td>
<td>bk*</td>
<td>bk*</td>
<td>bk*</td>
<td>bk*</td>
<td>bk*</td>
<td>bk</td>
<td>bk*</td>
<td>bk*</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
<td>bk</td>
</tr>
</tbody>
</table>

We have checked that the best known linear $[132, 44, 32]$ code in the Magma database is not 3-CIS.
Optimal $t$-CIS codes with $5 \leq t \leq 256$

- For $1 \leq k \leq \lfloor 256/t \rfloor$ except for $k = 37$, we have checked that there are 4-CIS $[tk, k]$ codes that are either $bk$ or $bk^*$. We have checked that the best known linear $[148, 37, 41]$ code in the Magma database is not 4-CIS.
- For $5 \leq t \leq 256$ and $1 \leq k \leq \lfloor 256/t \rfloor$, all the best known codes in the Magma database have been checked. We conclude that there are $t$-CIS $[tk, k]$ codes that are either $bk$ or $bk^*$. 
Conclusion

We show the following.

• Introduce a new class of CIS codes.
• In length $2n$ these codes are, when in systematic form, in one to one correspondence with linear bijective vectorial Boolean functions in $n$ variables.
• Classify CIS codes of lengths $\leq 12$ and give optimal or best known CIS codes of lengths $\leq 130$ and discuss an asymptotic bound.
• Introduce $t$-CIS codes of rate $1/t$ with $t$ pairwise disjoint information sets and find optimal $t$-CIS codes.
For the future work,

- More generally, does the CIS property involves an upper bound on the minimum distance?
- Finally, it is worth studying CIS codes over other fields than $\mathbb{F}_2$, and also over $\mathbb{Z}_4$.
- More constructions and classifications of $t$-CIS codes are desired.
- For a connection of multiply constant-weight codes with PUFs, see ref [3].
References

