Complementary information set codes from SRG, DRT, and matroids

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

Mathematics of Communications:
Sequences, Codes and Designs
BIRS, Canada
Jan. 25-30 2015

Outline

- Complementary Information Set (CIS) Codes
- General Constructions Including SRG and DRT.
- Classification of CIS Codes of Lengths ≤ 12
- Optimal CIS Codes of Lengths ≤ 130
- Long CIS Codes
- Higher-Order CIS Codes
- Conclusion and Open Problems

Linear codes: most useful codes

- A linear $[n, k]$ code C of length n and dimension k over GF(p) is called self-dual if $\mathcal{C} = \mathcal{C}^{\perp}$.
- A set of k columns of an $[n, k, d]$ code is called an information set if it is linearly independent.
- Observation: If C is a self-dual $[2n, n]$ code, then there are two disjoint information sets whose union is the set of n coordinates.

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 partitition L, R if there is an information set L whose complement R is also an information set.
  [Ref: Claude Carlet, Philippe Gaborit, Jon-Lark Kim, and Patrick Sole, "A new class of codes for Boolean masking of cryptographic computations", *IEEE Trans. Inform. Theory,* VOL. 58, NO. 9, Sep. 2012, pp. 6000-6011.]
- We call the partition $[1..n], \ldots, [n+1..2n]$ the **systematic partition.**
- Systematic self-dual codes are CIS with the systematic partition.
- Hence CIS codes are a natural generalization of self-dual codes.

Walsh Hadamard transform

- An vectorial Boolean function F is any map from $\mathbb{F}_2^n \to \mathbb{F}_2^n$.
- 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

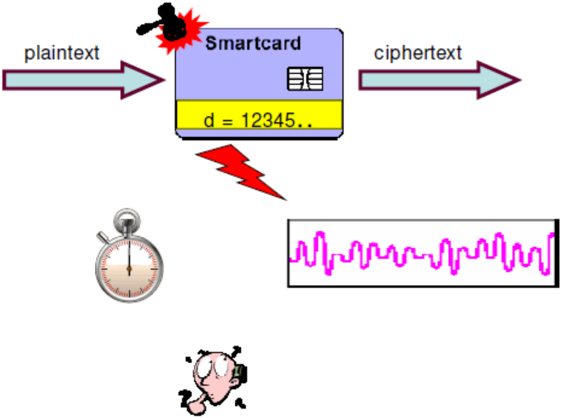
$$\hat{f}(a) = \sum_{x \in \mathbb{F}_2^k} f(x)(-1)^{a \cdot x} = \sum_{x \in \mathrm{supp}(f)} (-1)^{a \cdot x},$$

  where $\mathrm{supp}(f)$ is the support of function f.
- We note that for $a \neq 0$,

$$W_{F_1}(a,b) = 0 \text{ if and only if } \widehat{b \cdot F_1}(a) = 0. \qquad (1)$$

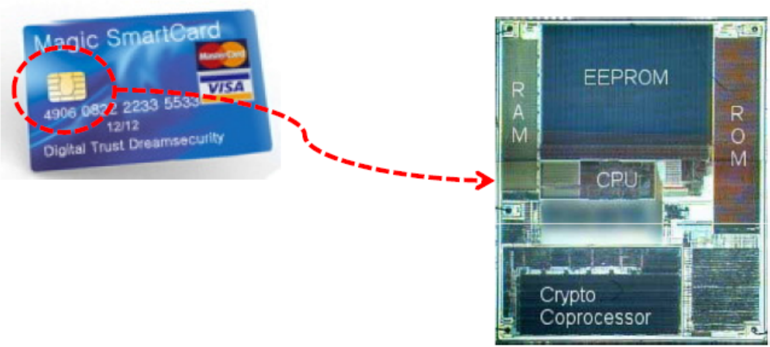
Motivations

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



Motivations-continued

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



Graph Correlation Immune Functions

- The method, called leakage squeezing, uses vectorial Boolean functions - more precisely, permutations $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$, 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, S. Guilley, C. Carlet and J.-L. Danger. 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.

Nonexistence of CIS codes

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

Proof

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 κ, λ both even and μ odd and if $M = A + I$
- DRT of odd order with κ, μ odd and λ even and if $M = A$
- SRG of odd order with κ even and λ, μ both odd and if $M = A + J$
- DRT of odd order with κ even and λ, μ 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 a 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 ≤ 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 $g_n/n!$.

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 $e_2 = 2 < g_2/2! = 3$.

Counting formula similar to mass formula

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

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

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

Summary: Classification of all CIS codes of lengths up to 12

Here (a, b, c) denotes the number a, b, c of sd, non-sd fsd, and none of them, respectively.

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

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 ≤ 130 with record distances

Theorem

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

$2n$	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
d	2*	2*	3*	4*	4*	4*	4*	5*	6*	6*	7*	8*	7*	8*	8*
code	dc	dc	∼dc	sd	dc	sd	sd	∼dc	∼dc	nfsd	id	sd	fsd	dc	dc

$2n$	102	104	106	108	110	112	114	116	118	120	122	124	126	128
d	19	20	19	20	18	19	20	20	20	20	20	21	21	22
code	bk	bk	qdc	bk	bk	bk	bk	sc	sc	sd	qdc	xqdc	bk	bk

CIS codes are asymptotically good

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

Proposition

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

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.
- Reference: "Higher-order CIS codes" by Claude Carlet, Finley Freibert, Sylvain Guilley, Michael Kiermaier, Jon-Lark Kim, Patrick Solé, IEEE Trans. Information Theory, VOL. 60, No. 9, Sep. 2014, pp. 5283 - 5295.

3-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 $\widehat{b \cdot F_1}(a) = 0$ or $\widehat{c \cdot F_2}(a) = 0$, 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.

Characterization of CIP

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}

$$C(F_1, F_2) = \{(x + y, F_1(x), F_2(y)) |\, x,\ y \in \mathbb{F}_2^k\} \tag{3}$$

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

$$C(F_1, F_2)^\perp = \{(u, G_1(u), G_2(u)) | \, u \in \mathbb{F}_2^k\}$$

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 $C(F_1, F_2)$ being the set of words $(x + y, F_1(x), F_2(y))$, with $x, y \in \mathbb{F}_2^k$, its dual $C^\perp$ is the set of words (u, v, w) such that

$$\begin{aligned}&(x + y) \cdot u + F_1(x) \cdot v + F_2(y) \cdot w\\ &= x \cdot (u + F_1^*(v)) + y \cdot (u + F_2^*(w))\\ &= 0 \text{ for every} x, y \in \mathbb{F}_2^k.\end{aligned}$$

Hence $C^\perp$ is the set of words (u, v, w) such that $u = F_1^*(v)$, $u = F_2^*(w)$ so that $v = (F_1^*)^{-1}(u) = G_1(u)$, $w = (F_2^*)^{-1}(v) = G_2(u)$. The result follows.

Building up construction

Proposition

Suppose that C is a t-CIS $[tk, k]$ code C with generator matrix $G = (A_1\ A_2\ \cdots A_t)$, where each A_j $(1 \le j \le t)$ is an invertible $k \times k$ matrix. Let $A_j(\mathbf{r_i})$ $(1 \le i \le k)$ denote the ith row of the matrix A_j. Then for any vectors $\mathbf{x}_j \in \mathbb{F}_2^n$ $(1 \le j \le t)$ and $y_{ij} \in \mathbb{F}_2$ $(1 \le i \le k, 1 \le j \le t)$, the following matrix G_1 generates a t-CIS $[t(k+1), k+1]$ code C_1.

$$G_1 = \left(\begin{array}{c|c|c|c|c|c|c}
z_1 & \mathbf{x}_1 & z_2 & \mathbf{x}_2 & \cdots & z_t & \mathbf{x}_t \\
\hline
y_{11} & A_1(\mathbf{r_1}) & y_{12} & A_2(\mathbf{r_1}) & \cdots & y_{1t} & A_t(\mathbf{r_1}) \\
y_{21} & A_1(\mathbf{r_2}) & y_{22} & A_2(\mathbf{r_2}) & \cdots & y_{2t} & A_t(\mathbf{r_2}) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
y_{k1} & A_1(\mathbf{r_k}) & y_{k2} & A_2(\mathbf{r_k}) & \cdots & y_{kt} & A_t(\mathbf{r_k})
\end{array}\right) \tag{4}$$

where for each j $(1 \le j \le t)$, $\mathbf{x}_j$ satisfies $\mathbf{x}_j = \sum_{i=1}^{k} c_{ij} A_j(\mathbf{r_i})$ for uniquely determined c_{ij}'s $(c_{ij} = 0, 1)$ and z_j satisfies $z_j = 1 + \sum_{i=1}^{k} c_{ij} y_{ij}$.

t-CIS Partition Algorithm

t-CIS Partition Algorithm based on Edmonds' matroid base packing theorem.

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.

t-CIS Partition Algorithm: Continued

1. Let $\{I_1, \ldots, I_t\}$ be a set of labeled disjoint independent subsets of M. (Note that each I_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 $x \in M \setminus \bigcup_{1 \leq i \leq t} I_i$.
3. While $\bigcup_{1 \leq i \leq t} I_i \subsetneq M$ do:
   - **3.1** Initialize $S_0 := M$. For $j > 0$, recursively define $S_j := \text{span}(I_{j'} \cap S_{j-1})$, where $j' = ((j-1) \bmod t) + 1$. 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 $I_{j'} \cup \{x\}$ is independent. If so then replace $I_{j'}$ with the larger independent set and repeat the while loop with a new $x \in M \setminus \bigcup_{1 \leq i \leq t} I_i$.
   - **3.5** If $I_{j'} \cup \{x\}$ is dependent, then find the unique minimal dependent set $C \subset I_{j'} \cup \{x\}$ (accomplished by solving the matrix equation associated with finding the linear combination of columns in $I_{j'}$ that sum to x).
   - **3.6** Select any $x' \in C \setminus S_{j-1}$ and replace $I_{j'}$ with $I_{j'} \cup \{x\} \setminus \{x'\}$, 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 $\{I_1, \ldots, I_t\}$ 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 $n \leq 36$.

n	6	9	12	15	18	21	24	27	30	33	36	39
k	2	3	4	5	6	7	8	9	10	11	12	13
d	4	4	6	7	8	8	8	10	11	12	12	12
code	qc	qc	bk	bk	bk	bk*	bk*	bk	bk	bk	bk*	bk*

n	123	126	129	132	135	138	141	144	147	150	153	156	159	162
k	41	42	43	44	45	46	47	48	49	50	51	52	53	54
d	29	31	32	?	32	32	32	32	34	34	33	34	34	35
code	bk*	bk*	bk*	?	bk*	bk*	bk*	bk*	bk	bk*	bk	bk*	bk*	bk

n	165	168	171	174	177	180	183	186	189	192	195	198	201	204
k	55	56	57	58	59	60	61	62	63	64	65	66	67	68
d	36	36	36	36	36	38	38	38	40	41	42	42	42	41
code	bk*	bk*	bk*	bk*	bk*	bk	bk*	bk*	bk	bk	bk	bk*	bk*	bk

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.
- Open question: Does there exist a 4-CIS $[148, 37, 41]$ code?
- 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 ≤ 12 and give optimal or best known CIS codes of lengths ≤ 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.

Future Work

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.

References

[1] Claude Carlet, Philippe Gaborit, Jon-Lark Kim, and Patrick Sole, "A new class of codes for Boolean masking of cryptographic computations", *IEEE Trans. Inform. Theory,* VOL. 58, NO. 9, Sep. 2012, pp. 6000-6011.

[2] Claude Carlet, Finley Freibert, Sylvain Guilley, Michael Kiermaier, Jon-Lark Kim, Patrick Solé, "Higher-order CIS codes", *IEEE Trans. Inform. Theory,* VOL. 60, No. 9, Sep. 2014, pp. 5283 - 5295.

Advertisement

- SIAM Conference on Applied Algebraic Geometry
- When? 08/03/2015 - 08/07/2015
- Where? NIMS and KAIST, Daejon Korea
- Invited Speakers include M. Sudan and J. Walker etc.
  `http://camp.nims.re.kr`
- Contributed Talks Submission Dates: Dec/11/2014 - Feb/28/2015
- Email me if you want to speak at Minisymposium on Coding Theory and Cryptography