On t-Covering Arrays

Sosina Martirosyan and Tran van Trung

SOSINA MARTIROSYAN
Institute for Experimental Mathematics, University of Duisburg-Essen
Ellernstrasse 29, 45326 Essen, Germany
sosina@exp-math.uni-essen.de

TRAN VAN TRUNG
Institute for Experimental Mathematics, University of Duisburg-Essen
Ellernstrasse 29, 45326 Essen, Germany
trung@exp-math.uni-essen.de

Abstract

This paper concerns construction methods for t-covering arrays. Firstly, a construction method using perfect hash families is discussed by combining with recursion techniques and error-correcting codes. In particular, by using algebraic-geometric codes for this method we obtain infinite families of t-covering arrays which are proved to be better than currently known probabilistic bounds for covering arrays. Secondly, inspired from a result of Roux and also from a recent result of Chateauneuf and Kreher for 3-covering arrays, we present several explicit constructions for t-covering arrays, which can be viewed as generalizations of their results for t-covering arrays.

Keywords t-covering arrays, orthogonal arrays, perfect hash families, algebraic-geometric codes.

1 Introduction

A t-covering array, denoted $\mathsf{CA}(N; t, k, v)$, is a $k \times N$-array with entries from a set of $v \geq 2$ symbols such that each $t \times N$-subarray contains each ordered t-tuple of symbols at least once as a column.

Let $\mathsf{CAN}(t, k, v)$ denote the minimum number N such that a $\mathsf{CA}(N; t, k, v)$ exists, i.e.,

$$\mathsf{CAN}(t, k, v) = \min\{N : \exists\ \mathsf{CA}(N; t, k, v)\}.$$

Then $\mathsf{CAN}(t, k, v)$ is called the *covering array number.*

Covering arrays can be viewed as a generalization of orthogonal arrays. In fact, if we require that each $t \times N$-subarray contains each ordered t-tuple of symbols in exactly λ times as a column, then we have an *t-orthogonal array*, denoted $\mathsf{OA}_\lambda(t,k,v)$. In this case we have $N = \lambda v^t$. Thus, an $\mathsf{OA}_\lambda(t,k,v)$ is a $\mathsf{CA}(\lambda v^t; t,k,v)$. In particular, if there is an $\mathsf{OA}_1(t,k,v)$, then $\mathsf{CAN}(t,k,v) = v^t$. For instance, an $\mathsf{OA}_1(t,t+1,v)$ exists for all t and v, see e.g. [12]; also, for any prime power q and any $t < q$, $\mathsf{OA}_1(t,q+1,q)$ exists [2]. Therefore, $\mathsf{CAN}(t,t+1,v) = v^t$ and $\mathsf{CAN}(t,q+1,q) = q^t$.

A main problem of covering arrays is to minimize N for given values t, k, v, or equivalently to maximize k for given values t, v, N. The case $t = 2$ has been studied by several authors, see for instance [8], [13], [14], [15], [21]. The case $t = 3$ can be found in [4], [5], [6], [16], [17]. Upper bounds on the number of columns N for t-covering arrays are given in [11]. However, very little are known for t-covering arrays with $t \geq 4$. This paper is concerned with t-covering arrays for an abitrary value t. Our interest in this paper is in constructing t-covering arrays using combinatorial techniques and in establishing bounds on the covering array number $\mathsf{CAN}(t,k,v)$. In particular, we present constructions of good classes of t-covering arrays using recursive methods and perfect hash families. We then show several explicit constructions of covering arrays for $t \geq 4$ from other covering arrays and thus obtain new bounds for t-covering arrays in the spirit of the results for 3-covering arrays of Roux [16], Chateauneuf and Kreher [6].

2 Preliminaries

The following basic facts on $\mathsf{CAN}(t,k,v)$ can be found in [6]. Let A be a $\mathsf{CA}(N;t,k,v)$ with entries from a set V.

Symbol-fusing. If a symbol x is replaced with any symbol in $V \setminus \{x\}$, wherever x occurs in the array A, then the resulting array is a $CA(N;t,k,v-1)$. Thus

$$\mathsf{CAN}(t,k,v-1) \leq \mathsf{CAN}(t,k,v).$$

Row-deleting. If any row of A is deleted, then the remaining rows form a $CA(N;t,k-1,v)$. Hence

$$\mathsf{CAN}(t,k-1,v) \leq \mathsf{CAN}(t,k,v).$$

Derived array. Note that if $x \in V$ appears M times in row i of A, then $M \geq v^{t-1}$. Removing all columns of A not having x on row i and then deleting row i form a $\mathsf{CA}(M;t-1,k-1,v)$. Therefore

$$\mathsf{CAN}(t,k,v) \geq v \cdot \mathsf{CAN}(t-1,k-1,v).$$

We prove a simple lemma which shows rough lower and upper bounds for $\mathsf{CAN}(t,k,v)$ for certain values of k.

Lemma 2.1 *For any $v \geq 2$, $t \geq 2$ we have*

$$v^t \leq \mathsf{CAN}(t,k,v) \leq 2^t \cdot v^t - 1,$$

where $k \leq 2^n$ and n is the smallest integer such that $v \leq 2^n$.

Proof. An obvious lower bound is

$$v^t \leq \mathsf{CAN}(t,k,v),$$

and this bound is reached if $v = q$ is a prime power and $k \leq q+1$ because an orthogonal array $\mathsf{OA}_1(t, q+1, q)$ exists [2]. If v is not a prime power, then $2^{n-1} < v < 2^n$ for a certain integer n. Now take $\mathsf{CA}(N; t, 2^n, 2^n) = \mathsf{OA}_1(t, 2^n, 2^n)$. Then $N = 2^{2nt}$. Using the symbol-fusing methode one gets a $\mathsf{CA}(N; t, 2^n, v)$. Since $N = 2^t \cdot 2^{(n-1)t} < 2^t \cdot v^t$, we have $N \leq 2^t v^t - 1$. ■

In [16], a Ph.D. dissertation, Roux shows the following theorem, (see also [17]).

Theorem 2.2 (Roux [16])

$$\mathsf{CAN}(3, 2k, 2) \leq \mathsf{CAN}(3, k, 2) + \mathsf{CAN}(2, k, 2).$$

Thus, Roux's theorem gives an upper bound for 3-covering array for $v = 2$.

Recently, Chateauneuf and Kreher [6] generalized Roux's theorem for any $v \geq 2$.

Theorem 2.3 (Chateauneuf and Kreher [6])

$$\mathsf{CAN}(3, 2k, v) \leq \mathsf{CAN}(3, k, v) + (v-1) \cdot \mathsf{CAN}(2, k, v).$$

3 A recursive construction of covering arrays using perfect hash families

A *t-perfect hash family* $\mathcal{H}$, denoted $\mathsf{PHF}(N; k, q, t)$, is a family of N functions $h : A \longrightarrow B$, where $|A| = k \geq |B| = q$, such that for any subset $X \subseteq A$ with $|X| = t$, there is at least one function $h \in \mathcal{H}$ such that h is injective on X.

Thus, a $\mathsf{PHF}(N; k, q, t)$ can be described as an $k \times N$-array $\mathcal{H}$ with entries from a set of q symbols such that for any set of t rows there is at least one column having different entries in this set of rows.

There is a simple direct construction of perfect hash families from error-correcting codes. An (N, k, d, q) code is a subset $\mathsf{C} \subseteq \mathsf{Q}^N$ with $|\mathsf{C}| = k$, $|\mathsf{Q}| = q$ such that the Hamming distance between any two distinct vectors in C is at least d.

Theorem 3.1 [1] *Suppose there is an (N, k, d, q) code C. Then there is a $\mathsf{PHF}(N; k, q, t)$ provided*

$$N > (N-d)\binom{t}{2}.$$

We describe a relationship between covering arrays and perfect hash families. Let $\mathsf{A} = (a_{i,j})$ denote the $k \times N$-matrix of a $\mathsf{CA}(N; t, k, v)$. For any two columns j_1 and j_2 of A, define

$$I(j_1, j_2) = |\{i : \ a_{i,j_1} = a_{i,j_2}\}|,$$

and

$$I(\mathsf{A}) = \max\{I(j_1, j_2) : \ j_1 \neq j_2\}.$$

Theorem 3.2 *Suppose there exists a $\mathsf{CA}(N; t, k, v)$.*

(i) *Then there exists a $\mathsf{PHF}(N; k, v, t)$ provided $t \leq v$*

(ii) *If $k/I(\mathsf{A}) > \binom{t'}{2}$, then there is a $\mathsf{PHF}(k; N, v, t')$.*

Proof. Let A denote the $k \times N$-array presented the $\mathsf{CA}(N; t, k, v)$. (i) It is obvious that A is a $\mathsf{PHF}(N; k, v, t)$ if $t \leq v$.

(ii) Taking the columns of A as codewords, we have a $(k, N, k - I(\mathsf{A}), v)$ code. Then apply Theorem 3.1. ■

When A is an $\mathsf{OA}_1(r, N, v)$, it is easy to see that $I(\mathsf{A}) = r - 1$. Thus we have

Corollary 3.3 *Suppose there is an $\mathsf{OA}_1(r, N, v)$. Then there exists a $\mathsf{PHF}(N; v^r, v, t)$ if $N/(r-1) > \binom{t}{2}$.*

It is well-known that there is an $\mathsf{OA}_1(r, q, q)$ for any prime power q and any integer r such that $2 \leq r \leq q$. Applying Corollary 3.3 gives

Corollary 3.4 *For any prime power q and any integer r such that $2 \leq r \leq q$, there exists a $\mathsf{PHF}(q; q^r, q, t)$ if $q/(r-1) > \binom{t}{2}$.*

A construction of covering arrays using perfect hash families is as follows.

Theorem 3.5 *Suppose there exists a $\mathsf{PHF}(s; k, m, t)$ and a $\mathsf{CA}(N; t, m, v)$. Then there is a $\mathsf{CA}(sN; t, k, v)$.*

We now use Corollary 3.4 and Theorem 3.5 to construct an infinite class of t-covering arrays with good asymptotic behavior.

Theorem 3.6 *Suppose there exists a* $\mathsf{CA}(N_0; t, q^{s_0}, v)$*, where* q *is a prime power and* $q^{s_0} > t(t-1)/2$*. Then there exists a* $\mathsf{CA}(N_0 R_i; t, q^{s_i}, v)$ *for all* $i \geq 0$*, where* $R_0 = 1$*, and*

$$\begin{aligned} R_i &= q^{s_{i-1}} R_{i-1}, \\ s_i &= s_{i-1} \lceil \frac{q^{s_{i-1}}}{\binom{t}{2}} \rceil \end{aligned}$$

for all $i \geq 1$.

Proof. We proceed by induction on i. For $i = 0$, the assertion is correct. Now assume $i \geq 1$. We apply Corollary 3.4 with q replaced by $q^{s_{i-1}}$ and

$$r = \lceil \frac{q^{s_{i-1}}}{\binom{t}{2}} \rceil.$$

The conditions

$$q^{s_{i-1}}/(r-1) > \binom{t}{2}$$

and $r \geq 2$ are satisfied. Thus, there is a $\mathsf{PHF}(q^{s_{i-1}}; q^{s_i}, q^{s_{i-1}}, t)$.

By induction, there exists a $\mathsf{CA}(N_0 R_{i-1}; t, q^{s_{i-1}}, v)$. Now applying Theorem 3.5 yields a $\mathsf{CA}(N_0 R_i; t, q^{s_i}, v)$. The proof is complete. ∎

Let $N_i = N_0 R_i$ and $k_i = q^{s_i}$. Then, by a similar argumentation as shown in [23] pp.196-197 it can be proved that

$$N_i \leq \frac{N_0 t^{2i_0}}{s_0 \log q} (t^2)^{\log^*(k_i)} (\log k_i)$$

for all $i > i_0$.

For any given values of k_0, v and t we can always construct a $\mathsf{CA}(N_0; t, k_0, v)$ for some N_0. Therefore, we have the following theorem.

Theorem 3.7 *For any positive integers* v *and* t *there is an infinite family of covering array* $\mathsf{CA}(N; t, k, v)$ *such that* N *is* $O((t^2)^{\log^*(k)}(\log k)$.

Theorem 3.5 becomes to be powerful when algebraic-geometric (AG) codes are used. The idea is to derive good classes of perfect hash families from AG codes by Theorem 3.1, and then apply Theorem 3.5. As a paradigmatic example we consider the class of linear AG codes defined on the Garcia-Stichtenoth (G-S) curves [9, 10]. The nth curve $\mathcal{X}_n$ over $\mathbb{F}_{q^2}$ in the sequence of Garcia-Stichtenoth curves is defined by the equations

$$x_i^q + x_i = \frac{x_{i-1}^q}{x_{i-1}^{q-1} + 1}, \quad i = 1, 2, \ldots, n.$$

The number of rational points of $\mathcal{X}_n$ is more than $q^n(q^2-q)$ and the genus g_n of $\mathcal{X}_n$ is less than q^{n+1}. The "one-point" AG codes constructed on the G-S curve is as follows: Let $\mathcal{P} = \{P_1, \ldots, P_N, P\}$ be $N+1$ distinct $\mathbb{F}_{q^2}$-rational points and let $L(mP)$ be the $\mathbb{F}_{q^2}$-vector space consisting of all functions defined on the curve such that the only pole of any $f \in L(mP)$ is P and the pole order is at most m. Define an evaluation map

$$\theta : L(mP) \longrightarrow \mathbb{F}_{q^2}^N$$

$$f \mapsto (f(P_1), \ldots, f(P_N)).$$

Then, the image $\mathsf{C} = Im\theta$ is referred to as a "one-point" AG code. Now, take

$$N = q^n(q^2-q),$$

$$2g_n - 2 < m < N.$$

Then C is a linear code with parameters $(N, q^{2\ell}, d, q^2)$, where $\ell = m - g_n + 1$ and $d \geq q^n(q^2-q) - m$. Thus, $q^{n+1} \leq \ell \leq q^{n+2} - 2q^{n+1} + 1$. We will write $\ell = \lceil uq^{n+1} \rceil$, where u is a real number satisfying $1 \leq u \leq q-2$. So, $d \geq q^n(q^2-q) - \lceil (u+1)q^{n+1} \rceil + 2$.

The parameters of C are then

$$(q^n(q^2-q), q^{2\lceil uq^{n+1} \rceil}, d, q^2)$$

Applying Theorem 3.1 to $\mathcal{C}$ we obtain the following result.

Theorem 3.8 *For every prime power q and any integer $n \geq 1$, there exists a* $\mathsf{PHF}(N; k, q^2, t)$, *where*

$N = q^{n+1}(q-1)$,
$k = q^{2\lceil uq^{n+1} \rceil}$,
u is a real number with $1 \leq u \leq q-2$, and
$t = \lceil \frac{1}{2}(1 + \sqrt{1 + \frac{8}{u+1}(q-1)}) \rceil$.

Now, combining Theorem 3.5 and Theorem 3.8 we can prove the following result.

Theorem 3.9 *For every given integers $t, v \geq 2$, and for any integer $n \geq 1$, there exists a covering array* $\mathsf{CA}(N; t, k, v)$, *where*

$N = N_0.(q-1)q^{n+1}$, *N_0 is a constant,*
$k = q^{2\lceil uq^{n+1} \rceil}$, *$q$ is a prime power such that* $q \geq \frac{t(t-1)(u+1)}{2} + 1$,
and u is a real number with $1 \leq u \leq q-2$.

Moreover, we have $N = O(\log k)$.

Proof. Let $t, v \geq 2$ be given integers. Let q be the smallest prime power such that $t = \lceil \frac{1}{2}(1 + \sqrt{1 + \frac{8}{u+1}(q-1)}) \rceil$, with $1 \leq u \leq q-2$, as shown in Theorem 3.8. A simple observation shows that we can always construct a $\mathsf{CA}(N_0; t, q^2, v)$ explicitly for a certain value N_0. Applying Theorem 3.5 and Theorem 3.8 yields the covering arrays with parameters as claimed. ■

It should be noticed that the first low-complexity algorithm for constructing "one-point" AG codes on G-S curves has a runtime upper-bounded by $(N \log_q N)^3$, where N is the length of the code and the complexity is measured in terms of multiplications and divisions over the finite field $\mathbb{F}_{q^2}$ [18]. The complexity of constructing t-covering arrays in Theorem 3.9 is, therefore, polynomial in N. The covering arrays in Theorem 3.7, however, can be viewed as an explicitly constructed family.

The following probabilistic upper bound for $\mathsf{CAN}(t, k, v)$ is due to Godbole *et al* [11].

Theorem 3.10 (Godbole, Skipper, Sunley [11])

$$\mathsf{CAN}(t, k, v) \leq \frac{(t-1)\log k}{\log\left(\frac{v^t}{v^t - 1}\right)}\{1 + o(1)\},$$

as $k \to \infty$.

It turns out that the explicit constructed covering arrays in Theorem 3.9 yield much better results compared to Godbole-Skipper-Sunley bound. To see it we consider e.g. the case with a square prime power $v = q^2$. For any given $t \geq 2$ and any prime power q satisfying the conditions of Theorem 3.9 choose a real number $1 \leq u \leq q-2$ such that $\frac{(q-1)}{(u+1)} = \binom{t}{2}$. By taking a $\mathsf{CA}(N_0; t, q^2, q^2)$ with $N_0 = q^{2t}$, Theorem 3.9 gives a $\mathsf{CA}(N; t, k, q^2)$ with $N = q^{2t}(q-1)q^{n+1}$ and $k = q^{2\lceil uq^{n+1} \rceil}$. Thus $N \approx \frac{q^{2t}(q-1)}{2u \ln q} \ln k$. For these t and k, the Godbole-Skipper-Sunley bound gives $\mathsf{CAN}(t, k, v) \leq \frac{(t-1)}{\ln \frac{q^{2t}}{q^{2t}-1}} \ln k\{1 + o(1)\}$. Let $\alpha = \frac{q^{2t}(q-1)}{2u \ln q}$ and $\beta = \frac{(t-1)}{\ln \frac{q^{2t}}{q^{2t}-1}}$. Then

$$\begin{aligned} \frac{\alpha}{\beta} &= \frac{(q-1)q^{2t} \ln \frac{q^{2t}}{q^{2t}-1}}{2u(t-1)\ln q} \\ &\approx \frac{(u+1)t}{4u \ln q} \\ &\approx \frac{t}{4 \ln q} \end{aligned}$$

by taking into account $q^{2t} \ln \frac{q^{2t}}{q^{2t}-1} \approx 1$. Thus $\frac{\alpha}{\beta} < 1$ for $q \geq e^{\frac{t}{4}}$. This shows that sizes of arrays from Theorem 3.9 with $v = q^2$ are better than Godbole-Skipper-Sunley bounds.

As examples we consider several values of v.

For $v = 3^2$, $t = 2$ $u = 1$ we have $\alpha = 73.729$ and $\beta = 80.498$

For $v = 7^2$, $t = 3$ and $u = 1$ we have $\alpha = 181378.878$ and $\beta = 235296.999$.

For $v = 13^2$ and $t = 4$ and $u = 1$ we have $\alpha = 1908179711.915$ and $\beta = 2447192161.523$.

Since $\frac{\alpha}{\beta} \to 0$ as $q \to \infty$, the Godbole-Skipper-Sunley bound becomes weak. For instance, if $v = 2^{32}$, $t = 2$, $u = 2^{16} - 2$ we have $\alpha \approx 8,3 * 10^{17}$ whereas $\beta = 25 * 10^{18}$. Thus, α is about 30 times smaller than β.

4 Constructions of Roux's type for t-covering arrays

The constructions in the previous section provide classes of covering arrays with good asymptotic behavior, when $k \to \infty$ and v, t are fixed. In this section we focus on construction techniques that can be used to improve the results for small values of k.

With Theorem 2.2 Roux shows an interesting bound for binary 3-covering array, i.e. $v = 2$. This bound is recently generalized by Chateauneuf and Kreher to any $v \geq 2$, as presented in Theorem 2.3. The idea is to construct a $\mathsf{CA}(3, 2k, v)$ using a $\mathsf{CA}(3, k, v)$ and a $\mathsf{CA}(2, k, v)$.

Remark 4.1 We want to make a remark that Theorem 4.7. of Chateauneuf and Kreher [6] p.231 is incorrect. Theorem 4.7. [6] states that one obtains

$$\lim_{k\to\infty} \frac{\mathsf{CAN}(3, k, v)}{\log k} = \binom{v}{2}$$

from

$$\mathsf{CAN}(3, 2k, v) \leq \mathsf{CAN}(3, k, v) + (v - 1)\mathsf{CAN}(2, k, v), \qquad (*)$$

and

$$\lim_{k\to\infty} \frac{\mathsf{CAN}(2, k, v)}{\log_2 k} = \frac{v}{2} \qquad (**)$$

In fact, it can be shown from $(*)$ and $(**)$ that

$$\lim_{k\to\infty} \frac{\mathsf{CAN}(3, k, v)}{\log k} = \infty.$$

In this section we discuss several constructions of $\mathsf{CA}(t, 2k, v)$ using $\mathsf{CA}(s, k, v)$ for $s \leq t$ in the spirit of Roux, Chateauneuf and Kreher.

4.1 4-Covering arrays

Let D be a $\mathsf{CA}(N_1; 2, v, v)$ with entries $d_{j,i} \in V = \{1, \ldots, v\}$. Let $\mathcal{F}_{\mathsf{D}} = \{f_1, \ldots, f_{N_1}\}$ be a set of mappings derived from D as follows. For each $i = 1, \ldots, N_1$ define

$$f_i : V \longrightarrow V$$

by

$$f_i(j) = d_{j,i}.$$

Thus f_i maps the vector $(1, \dots, v)^T$ to the i−th column of D, i.e., $f_i(j) = d_{j,i}$.

Remark 4.2 The family $\mathcal{F}_{\mathsf{D}}$ has the following property. For any given two pairs (x, y) and (z, w) with $x, y, z, w \in V$ and $x \neq y$, there is at least an $f_i \in \mathcal{F}_{\mathsf{D}}$ such that $f_i(x) = z$ and $f_i(y) = w$. This is because D is a $\mathsf{CA}(N_0; 2, v, v)$.

In the following theorem we give a bound for 4-covering arrays by means of a direct construction.

Theorem 4.3 *For any $v \geq 2$ we have*

$$\mathsf{CAN}(4, 2k, v) \leq \mathsf{CAN}(4, k, v) + (v-1)\mathsf{CAN}(3, k, v) + 2\mathsf{CAN}(2, v, v)\mathsf{CAN}(2, k, v).$$

Proof. Let A be a $\mathsf{CA}(N_4; 4, k, v)$, B be a $\mathsf{CA}(N_3; 3, k, v)$, C be a $\mathsf{CA}(N_2; 2, k, v)$, and D be a $\mathsf{CA}(N_1; 2, v, v)$, all on the symbol set $V = \{1, 2, \dots, v\}$. Let $\mathcal{F}_{\mathsf{D}} = \{f_1, f_2, \dots, f_{N_1}\}$ be the set of mappings derived from D as defined above. Finally, let $\pi = (1, 2, \dots, v)$ be a cyclic permutation on the symbol set V. Define

$$\mathsf{E}_1 = \begin{array}{||c||} \hline \mathsf{A} \\ \hline \mathsf{A} \\ \hline \end{array}$$

$$\mathsf{E}_2 = \begin{array}{||c|c|c|c||} \hline \mathsf{B} & \mathsf{B} & \dots & \mathsf{B} \\ \hline \mathsf{B}^{\pi^1} & \mathsf{B}^{\pi^2} & \dots & \mathsf{B}^{\pi^{v-1}} \\ \hline \end{array}$$

$$\mathsf{E}_3 = \begin{array}{||c|c|c|c||} \hline \mathsf{C} & \mathsf{C} & \dots & \mathsf{C} \\ \hline \mathsf{C}^{f_1} & \mathsf{C}^{f_2} & \dots & \mathsf{C}^{f_{N_1}} \\ \hline \end{array}$$

$$\mathsf{E}_4 = \begin{array}{||c|c|c|c||} \hline \mathsf{C}^{f_1} & \mathsf{C}^{f_2} & \dots & \mathsf{C}^{f_{N_1}} \\ \hline \mathsf{C} & \mathsf{C} & \dots & \mathsf{C} \\ \hline \end{array}$$

where B^{π^i} and C^{f_j} are the arrays obtained by applying π^i and f_j to the symbols of B and C, respectively.

Construct an array E as follows:

$\mathsf{E} =$	E_1	E_2	E_3	E_4

E is therefore an $2k \times N$-array, where $N = N_4 + (v-1)N_3 + 2N_2N_1$.

Consider 4 rows $r_1,\ r_2,\ r_3,\ r_4$ of E.

1. If $r_1,\ r_2,\ r_3,\ r_4$ include 4 distinct rows of A, then all quadruples occur on these rows among the columns of E_1.
2. If $r_1 < r_2 < r_3 \leq k < r_4 = r_1 + k$ or $r_1 \leq k < r_2 = r_1 + k < r_3 < r_4$, then all quadruples of the form $(x, y, w, x)^T$ for any x, y, w occur on these rows among the columns of E_1 and quadruples $(x, y, w, z)^T$ with $x \neq z$ occur in E_2.
3. If $r_1 < r_2 \leq k < r_3 = r_1 + k < r_4$, then we have two subcases.
   - 3.1. $r_4 \neq r_2 + k$. Quadruples of the form $(x, y, x, z)^T$ for any x, y, z occur among the columns of E_1. Let $r_4' = r_4 - k$. Then $r_1, r_2, r_4' \leq k < r_3 = r_1 + k$. For any quadruple of the form $(x, y, x', z)^T$ with $x' \neq x$, we have $x' = x^{\pi^i}$ for some i. Hence there is a column in E_2 containing x in row r_1, y in row r_2, $z^{(\pi^i)^{-1}}$ in row r_4', and $x' = x^{\pi^i}$ in row r_3. Therefore, $(x, y, x', z)^T$ appears in that column on the rows r_1, r_2, r_3, r_4.
   - 3.2. $r_4 = r_2 + k$. Quadruples of the form $(x, y, w, z)^T$ with $x \neq y$ for any w, z occur on the rows r_1, r_2, r_3, r_4 among the columns of E_3, because there exists an f_i such that $x^{f_i} = w$ and $y^{f_i} = z$; similarly quadruples $(x, y, w, z)^T$ with $w \neq z$ is covered by E_4; quadruples of the form $(x, x, y, y)^T$ for every x and y occur among the columns of E_3 and E_4.

Therefore, E is a coverring array $\mathsf{CA}(N; 4, 2k, v)$ with $N = N_4 + (v-1)N_3 + 2N_2N_1$, as required.

■

If $v = q$ is a prime power, then a $\mathsf{CA}(q^2; 2, q, q)$ exists. Hence, the bound in Theorem 4.3 can be strenghened and we obtain:

Corollary 4.4 *For any prime power $q \geq 2$ we have*

$$\mathsf{CAN}(4, 2k, q) \leq \mathsf{CAN}(4, k, q) + (q-1) \cdot \mathsf{CAN}(3, k, q) + 2q^2 \cdot \mathsf{CAN}(2, k, q).$$

It can be observed from the proof of Theorem 4.3 that we can even construct better covering arrays in several cases by chosing the arrays A, C, D more carefully. These cases are listed in the following proposition.

Proposition 4.5 *The construction in Theorem 4.3 still works if any of arrays* A, C *and* D *is chosen as follows:*

1. C *is a* $k \times N_2$*-array with entries from a set of* v *symbols such that each* $2 \times N$*-subarray contains each ordered* 2*-tuple of not equal symbols at least once as a column.*
2. *In the binary alphabet case,* D *is a* 2×2 *array whose rows are both equal to* $\{0, 1\}$.
3. *In the case* $k < 4$, A *is the same as the array* B.

From Proposition 4.5 (2) and Theorem 4.3 we obtain the following corollary.

Corollary 4.6

$$\mathsf{CAN}(4, 2k, 2) \leq \mathsf{CAN}(4, k, 2) + \mathsf{CAN}(3, k, 2) + 4\mathsf{CAN}(2, k, 2).$$

Theorem 4.3 together with Proposition 4.6 gives the following example.

Example 4.7 $\mathsf{CA}(28; 4, 6, 2)$

0001110100011101000100111100
0010101100101011000010111010
0100011101000111000001111001
0001110111100010100000100111
0010101111010100010000010111
0100011110111000001000001111

The next example is a $\mathsf{CA}(25; 4, 6, 2)$ which is obtained from this $\mathsf{CA}(28; 4, 6, 2)$ by removing the 1st, 9th and 16th columns.

Example 4.8 $\mathsf{CA}(25; 4, 6, 2)$

0011101001110000100111100
0101011010101000010111010
1000111100011000001111001
0011101110001100000100111
0101011101010010000010111
1000111011100001000001111

Therefore $\mathsf{CAN}(4,6,2) \leq 25$.

Example 4.9 $\mathsf{CA}(40;4,8,2)$

We take the arrays A, B, C *and* D *as follows:*

$$\mathsf{A} = \begin{array}{|c|}\hline 0001110100011101 \\ 0010101100101011 \\ 0100011101000111 \\ 0001110111100010 \\ \hline\end{array} \qquad \mathsf{B} = \begin{array}{|c|}\hline 00011101 \\ 00101011 \\ 01000111 \\ 01110001 \\ \hline\end{array}$$

A *is a 4-covering array,* B *is a 3-covering array.*

$$\mathsf{C} = \begin{array}{|c|}\hline 1000 \\ 0100 \\ 0010 \\ 0001 \\ \hline\end{array} \qquad \mathsf{D} = \begin{array}{|c|}\hline 01 \\ 01 \\ \hline\end{array}$$

C *is an array defined from Proposition 4.5 (case 1) and* D *is the array from Proposition 4.5 (case 2).*

Applying Theorem 4.3 to A, B, C *and* D *yields a* $\mathsf{CA}(40;4,8,2)$ E *as follows:*

0001110100011101	00011101	1000	1000	0000	1111
0010101100101011	00101011	0100	0100	0000	1111
0100011101000111	01000111	0010	0010	0000	1111
0001110111100010	01110001	0001	0001	0000	1111
0001110100011101	11100010	0000	1111	1000	1000
0010101100101011	11010100	0000	1111	0100	0100
0100011101000111	10111000	0000	1111	0010	0010
0001110111100010	10001110	0000	1111	0001	0001

Example 4.10 $\mathsf{CA}(37;4,8,2)$

It can be shown that by deleting the 1st, 17th and 24th columns from the $\mathsf{CA}(40;4,8,2)$ in Example 4.9 we obtain a $\mathsf{CA}(37;4,8,2)$.

It should be noted that by a computer search [19, 20] Sherwood has constructed a $\mathsf{CA}(28;4,6,2)$ and a $\mathsf{CA}(40;4,8,2)$. Also, a $\mathsf{CA}(31;4,8,2)$ has been found, as reported in [20].

4.2 5-Covering arrays

We prove the following theorem.

Theorem 4.11 *For any $v \geq 3$ we have*

$$\mathsf{CAN}(5, 2k, v) \leq \mathsf{CAN}(5, k, v) + (v-1)\mathsf{CAN}(4, k, v) + [6v(v-1) + 2\mathsf{CAN}(2, v, v)]\mathsf{CAN}(3, k, v).$$

Proof. Let A be a $\mathsf{CA}(N_5; 5, k, v)$, B be a $\mathsf{CA}(N_4; 4, k, v)$, C be a $\mathsf{CA}(N_3; 3, k, v)$, and D be a $\mathsf{CA}(N_1; 2, v, v)$, all on the symbol set $V = \{1, 2, \ldots, v\}$. Again let $\mathcal{F}_{\mathsf{D}} = \{f_1, f_2, \ldots, f_{N_1}\}$ be the set of mappings defined from a $\mathsf{CA}(N_1; 2, v, v)$ as in Section 4. Also, let $\pi = (1, 2, \ldots, v)$ be a cyclic permutation on the symbol set V.

We define three families of mappings from V into V as follows:

(i). Let $\mathcal{G} = \{g_{a,b} : \; V \longrightarrow V : \; a, b \in V, \; a \neq b\}$, where

$$g_{a,b}(x) = \begin{cases} a & \text{if } x = a \\ b & \text{if } x \neq a \end{cases}$$

(ii). Let $\bar{\mathcal{G}} = \{\bar{g}_{a,b} : \; V \longrightarrow V : \; a, b \in V, \; a \neq b\}$, where

$$\bar{g}_{a,b}(x) = \begin{cases} a & \text{if } x = b \\ b & \text{if } x \neq b \end{cases}$$

(iii). Let $\mathcal{H} = \{h_{a,b} : \; V \longrightarrow V : \; a, b \in V, \; a \neq b\}$, where

$$h_{a,b}(x) = \begin{cases} a & \text{if } x \neq a \text{ or } x \neq b \\ b & \text{if } x = a \text{ or } x = b \end{cases}$$

Define

$$\mathsf{E}_1 = \begin{array}{||c||} \hline \mathsf{A} \\ \hline \mathsf{A} \\ \hline \end{array}$$

$$\mathsf{E}_2 = \begin{array}{||c|c|c|c||} \hline \mathsf{B} & \mathsf{B} & \ldots & \mathsf{B} \\ \hline \mathsf{B}^{\pi^1} & \mathsf{B}^{\pi^2} & \ldots & \mathsf{B}^{\pi^{v-1}} \\ \hline \end{array}$$

$$\mathsf{E}_3 = \begin{array}{||c|c|c|c||c|c|c|c||} \hline \mathsf{C} & \mathsf{C} & \ldots & \mathsf{C} & \mathsf{C}^{f_1} & \mathsf{C}^{f_2} & \ldots & \mathsf{C}^{f_{N_1}} \\ \hline \mathsf{C}^{f_1} & \mathsf{C}^{f_2} & \ldots & \mathsf{C}^{f_{N_1}} & \mathsf{C} & \mathsf{C} & \ldots & \mathsf{C} \\ \hline \end{array}$$

$\mathsf{E}_4 =$

C	C	$\dots$	C	$\mathsf{C}^{g_{1,2}}$	$\mathsf{C}^{g_{1,3}}$	$\dots$	$\mathsf{C}^{g_{v,v-1}}$
$\mathsf{C}^{g_{1,2}}$	$\mathsf{C}^{g_{1,3}}$	$\dots$	$\mathsf{C}^{g_{v,v-1}}$	C	C	$\dots$	C

$\mathsf{E}_5 =$

C	C	$\dots$	C	$\mathsf{C}^{\bar{g}_{1,2}}$	$\mathsf{C}^{\bar{g}_{1,3}}$	$\dots$	$\mathsf{C}^{\bar{g}_{v,v-1}}$
$\mathsf{C}^{\bar{g}_{1,2}}$	$\mathsf{C}^{\bar{g}_{1,3}}$	$\dots$	$\mathsf{C}^{\bar{g}_{v,v-1}}$	C	C	$\dots$	C

$\mathsf{E}_6 =$

C	C	$\dots$	C	$\mathsf{C}^{h_{1,2}}$	$\mathsf{C}^{h_{1,3}}$	$\dots$	$\mathsf{C}^{h_{v,v-1}}$
$\mathsf{C}^{h_{1,2}}$	$\mathsf{C}^{h_{1,3}}$	$\dots$	$\mathsf{C}^{h_{v,v-1}}$	C	C	$\dots$	C

Construct an array E as follows.

$\mathsf{E} =$

E_1	E_2	E_3	E_4	E_5	E_6

Let r_1, r_2, r_3, r_4, r_5 be 5 rows of E. Because of the symmetry of E we need to consider the following cases.

1. If $r_1,\ r_2,\ r_3,\ r_4,\ r_5$ satisfy $r_i \neq r_j + k,\ i \neq j$ and $i, j = 1, 2, 3, 4, 5$, then all 5-tuples occur on these rows among the columns of E_1.

2. If $r_1 < r_2 < r_3 < r_4 \leq k < r_5 = r_1 + k$, then 5-tuples of the form $(a, b, c, d, a)^T$ occur on these rows among the columns of E_1, and all 5-tuples $(a, b, c, d, a')^T$ with $a' \neq a$ appear in the columns of E_2.

3. Assume $r_1 < r_2 < r_3 \leq k < r_4 < r_5$, $r_4 = r_1 + k$ and $r_5 \neq r_i + k$ for all $i = 1, 2, 3$. Consider a 5-tuple $X = (a, b, c, a', e)^T$. If $a = a'$, then X is covered by E_1. Now assume $a \neq a'$. As B is a 4-covering array, all $(v-1)$ quadruples $(a, b, c, e_1)^T, \dots, (a, b, c, e_{v-1})^T$ with $e \neq e_i$, appear on the rows $r_1,\ r_2,\ r_3,\ r_5 - k$ among the columns. Thus, for each π^i, there is a e_j such that $\pi^i(e_j) = e$. Further, $\pi^i(a) = a_i$ with $a \neq a_i$. It follows that all 5-tuples $(a, b, c, a_1, c), (a, b, c, a_2, c), \dots, (a, b, c, a_{v-1}, c)$, where and $a_i \neq a_j$ for $i \neq j$, appear in the columns corresponding to rows $r_1,\ r_2,\ r_3,\ r_4,\ r_5$ in E_2.

4. Assume $r_1 < r_2 < r_3 \leq k < r_4 < r_5$, $r_4 = r_1 + k$ and $r_5 = r_2 + k$. We need to consider different types of 5-tuples.

   (i) A 5-tuple of the form $(a, b, x, a, b)^T$ for any a, b, x is covered by E_1.

(ii) A 5-tuple of the form $(a, a, x, b, b)^T$ for any a, b, x is covered by E_2.

(iii) A 5-tuple of the form $(a, b, x, c, d)^T$ for any a, b, x, c, d and $a \neq b$ is covered by E_3. This is because C is a 3-covering array, there is at least one column of C containing the triple $(a, b, x)^T$ in the rows $r_1,\ r_2,\ r_3$ and there is an f_i such that $f_i(a) = c$ and $f_i(b) = d$. From now on we assume $c \neq d$.

(iv) Consider a 5-tuple of the form $(a, a, x, c, d)^T$ for any a, x, c, d and $c \neq d$. We have the following subcases.

(α) $x \neq a, c, d$. There is a column j of C containing the triple $(x, c, d)^T$ in the rows $r_3 + k,\ r_1 + k,\ r_2 + k$ of E_4. The column j of the block

$$\begin{array}{|c|}\hline \mathsf{C}^{g_{x,a}} \\ \hline \mathsf{C} \\ \hline \end{array}$$

contains the 5-tuple $(a, a, x, c, d)^T$ with $x \neq a, c, d$ in the rows $r_1,\ r_2,\ r_3,\ r_1 + k,\ r_2 + k$, because $g_{x,a}(x) = x$, $g_{x,a}(c) = a$, and $g_{x,a}(d) = a$.

(β) $x = a$. As C is a 3-covering array, there is a column j containing the triple $(c, d, c)^T$ in the rows $r_1{+}k,\ r_2{+}k,\ r_3{+}k$. Also there is a mapping f_i, $1 \leq i \leq N_1$, such that $f_i(c) = a$ and $f_i(d) = a$, by Remark 4.2. Therefore the column j of the block

$$\begin{array}{|c|}\hline \mathsf{C}^{f_i} \\ \hline \mathsf{C} \\ \hline \end{array}$$

in E_3 contains the 5-tuple $(a, a, a, c, d)^T$ in the rows $r_1,\ r_2,\ r_3,\ r_1{+} k,\ r_2 + k$.

(γ) $x \neq a$ and $x = c$. Again there is a column j of C containing the triple $(c, d, a)^T$ in the row $r_1 + k,\ r_2 + k,\ r_3 + k$. Hence the column j of the block

$$\begin{array}{|c|}\hline \mathsf{C}^{\bar{g}_{c,a}} \\ \hline \mathsf{C} \\ \hline \end{array}$$

in E_5 contains the 5-tuple $(a, a, c, c, d)^T$ in the rows $r_1,\ r_2,\ r_3,\ r_1{+} k,\ r_2 + k$.

(δ) $x = a = c$ (i.e. $a \neq d$). The 5-tuple $(a, a, a, a, d)^T$ is covered by a column of the block

$$\begin{array}{|c|}\hline \mathsf{C}^{f_i} \\ \hline \mathsf{C} \\ \hline \end{array}$$

with $f_i(a) = a$ and $f_i(d) = a$ in part E_3.

(θ) $x = c$ and $a = d$. Consider a column j of C containing the triple $(c,a,b)^T$ with $b \neq c, a$ in the rows r_1+k, r_2+k, r_3+k. The 5-tuple $(a,a,c,c,a)^T$ is contained in a column j corresponding to the rows r_1, r_2, r_3, r_1+k, r_2+k of the block

$\mathsf{C}^{h_{c,a}}$
C

of E_6. This is because $h_{c,a}(b) = c$, $h_{c,a}(c) = a$ and $h_{c,a}(a) = a$.

Hence E is a 5-covering array. The proof is complete by using $|\mathcal{G}| = |\bar{\mathcal{G}}| = |\mathcal{H}| = v(v-1)$. ∎

If $v = q$ is a prime power, then $N_1 = v^2$ by Lemma 2.1. Therefore we have

Corollary 4.12 *For any prime power $q \geq 3$ we have*

$$\mathsf{CAN}(5,2k,q) \leq \mathsf{CAN}(5,k,q) + (q-1)\mathsf{CAN}(4,k,q) + (8q^2-6q)\mathsf{CAN}(3,k,q).$$

4.3 t-Covering arrays for $t \geq 4$

Theorem 4.13 *For any integers $t \geq 4$ and $v \geq 2$ we have*

$$\mathsf{CAN}(t,2k,v) \leq \mathsf{CAN}(t,k,v) + (v-1)\mathsf{CAN}(t-1,k,v) + \sum_{i=2}^{t-2} \mathsf{CAN}(i,k,v)\mathsf{CAN}(t-i,k,v).$$

Proof. Let $\mathsf{A}_t, \mathsf{A}_{t-1}, \ldots, \mathsf{A}_2$ be $\mathsf{CA}(N_t;t,k,v), \mathsf{CA}(N_{t-1};t-1,k,v), \ldots, \mathsf{CA}(N_2;2,k,v)$, respectively.

Let $\mathsf{B}_i^{N_j}$ be the $k \times N_i.N_j$ array obtained from A_i by repeating each column N_j times, where $i,j = t-2,\ldots,2$ and $i+j=t$.

Let $\mathsf{C}_j^{N_i}$ be the $k \times N_i.N_j$ array obtained by concatenating N_i copies of A_j, where $i,j = t-2,\ldots,2$ and $i+j=t$.

Define

$\mathsf{E}_t =$

A_t
A_t

$\mathsf{E}_{t-1} =$

A_{t-1}	A_{t-1}	...	A_{t-1}
A_{t-1}^{π}	$\mathsf{A}_{t-1}^{\pi^2}$	...	$\mathsf{A}_{t-1}^{\pi^{v-1}}$

For $i = t-2,\ldots,2$, define

$$\mathsf{E}_i = \begin{array}{||c||} \hline \mathsf{B}_i^{N_{t-i}} \\ \hline \mathsf{C}_{t-i}^{N_i} \\ \hline \end{array}$$

Construct an array E as follows.

$$\mathsf{E} = \begin{array}{|c|c|c|c|c|} \hline \mathsf{E}_t & \mathsf{E}_{t-1} & \mathsf{E}_{t-2} & \ldots & \mathsf{E}_2 \\ \hline \end{array}$$

Let $r_1, \ldots, r_t$ be t rows of E. Without loss of generality it is enough to consider the following cases.

1. If $r_1, \ldots, r_t$ include t distinct rows of A_t, then all t-tuples occur on these rows among the columns of E_t.

2. If $r_1 < \ldots < r_{t-1} \leq k < r_t = r_1 + k$, then t-tuples of the form $(a_1, \ldots, a_{t-1}, a_1)^T$ is covered by E_t, and all t-tuples $(a_1, \ldots, a_{t-1}, a')^T$ with $a' \neq a_1$ appear in the columns of E_{t-1}.

3. For the remaining cases we can assume $r_1 < \ldots < r_i \leq k$ and $k < r_{i+1} < \ldots < r_t$, where $i = t-2, t-3, \ldots, 2$. Then for each $i = t-2, t-3, \ldots, 2$ and for any t-tuple $(a_1, a_2, \ldots, a_t)^T$ of symbols there is a column in E_i containing this t-tuple in the rows $r_1 < \ldots < r_i \leq k < r_{i+1} < \ldots < r_t$.

The proof is complete. ■

For $t = 4, 5$ and large values of k the construction in Theorem 4.13 yields a weaker upper bound on covering array number than the constructions for 4-, 5-covering arrays in Theorem 4.3 and Theorem 4.11. However, for a not very large k the construction in Theorem 4.13 provides better results, as shown in the following example.

Example 4.14

1. $\mathsf{CAN}(4, 8, 3) \leq 216$,

2. $\mathsf{CAN}(5, 10, 4) \leq 3840$,

3. $\mathsf{CAN}(5, 12, 5) \leq 11875$.

Note that from Corollary 4.4 and Corollary 4.12 we obtain a $\mathsf{CA}(297; 4, 8, 3)$, a $\mathsf{CA}(8448; 5, 10, 4)$ and a $\mathsf{CA}(26875; 5, 12, 5)$. Moreover, the probabilistic bound in [11] only shows the existence of a $\mathsf{CA}(904; 4, 8, 3)$, a $\mathsf{CA}(15940; 5, 10, 4)$ and a $\mathsf{CA}(54424; 5, 12, 5)$.

References

[1] N. Alon, Explicit construction of exponential sized families of k-independent sets, *Discrete Math.* **58** (1986), 191–193.

[2] K. A. Bush, Orthogonal arrays of index unity, *Ann. Math. Stat.* **23** (1952), 426–434.

[3] C. J. Colbourn and J. H. Dinitz, *CRC Handbook of Combinatorial Designs*, CRC Press, Inc., 1996.

[4] M. Chateauneuf, *Covering arrays*, PhD thesis, Michigan Technological University, 2000.

[5] M. Chateauneuf, C. J. Colbourn, and D. L. Kreher, Covering arrays of strength 3, *Designs, Codes and Cryptography* **16** (1999) 235–242.

[6] M. Chateauneuf and D. L. Kreher, On the State of Strength-Three Covering Arrays, *J. Combin. Designs***10** (2002), 217–238.

[7] P. Erdös, P. Frankl and Z. Füredi, Families of finite sets in which no set is covered by the union of r other, *Israel Journal of Mathematics* **51** (1985), 75–89.

[8] L. Gargano, J. Körner, and U. Vaccaro, Sperner capacities, *Graphs and Combinatorics* **9** (1993), 31–46.

[9] A. Garcia, H. Stichtenoth, A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vlădut bound, *Invent. Math.* **121** (1995), 211–222.

[10] A. Garcia, H. Stichtenoth, On asymptotic behavior of some towers of function fields over finite fields, *J. Number Theory* **61** (1996), 248–273.

[11] A. P. Godbole, D. E. Skipper, and R. A. Sunley, t-covering arrays: upper bounds and Poisson approximations, *Combinatorics, Probab.Comput.* **5** (1966), 105–117.

[12] A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal Arrays, Theory and applications*, Springer, 1999.

[13] G. O. H. Katona, Two applications (for search theory and truth functions) of Spencer type theorems, *Periodica Math. Hung.* **3** (1973), 19–26.

[14] D. J. Kleitman, J. Spencer, Families of k-independent sets, *Discr. Math.* **6** (1973), 255–262.

[15] A. Rényi, *Foundation of probability*, Wiley, New York, 1971.

[16] G. Roux, k-propriétés dans des tableaux de n colonnes; cas particulier de la k-surjectivité et de la k-permutivité, Ph.D. Dissertation, University of Paris 6, March 1987.

[17] N. J. A. Sloane, Covering arrays and intersection codes, *J. Combin Designs* **1** (1993), 51–63.

[18] K. W. Shum, I. Aleshnikov, P. V. Kumar, H. Stichtenoth, and V. Deolalikar, A Low-Complexity Algorithm for the Construction of Algebraic-Geometric Codes Better Than the Gilbert-Varshamov Bound, *IEEE Trans. Inform Theory* **47** (2001), 2225–2241.

[19] G. Sherwood, Constrained Arrays Test System (CATS), *AT&T Bell Laboratories, Murray Hill, NJ, User's Manual* (1992)

[20] G. Sherwood, Construction of orthogonal arrays and covering arrays using permutation groups,
`http://home.att.net/gsherwood/cover.htm`

[21] B. Stevens, L. Moura, and E. Mendelsohn, Lower bounds for transversal covers, *Designs, Codes and Cryptography* **15** (1998) 279–299.

[22] J. N. Staddon, D. R. Stinson and R. Wei, Combinatorial properties of frameproof and traceability codes, *IEEE Trans. Inform. Theory* **47** (2001), 1042–1049

[23] D. R. Stinson and R. Wei and L. Zhu, New constructions for perfect hash families and related structures using combinatorial designs and codes, *J. Combin. Designs* **8** (2000), 189–200.