Partial Covering Arrays: Algorithms and Asymptotics

Kaushik Sarkar[1], Charles J. Colbourn[1], Annalisa De Bonis[2], and Ugo Vaccaro[2]

[1] CIDSE, Arizona State University, U.S.A.
[2] Dipartimento di Informatica, University of Salerno, Italy

Abstract. A covering array $\mathsf{CA}(N; t, k, v)$ is an $N \times k$ array with entries in $\{1, 2, \ldots, v\}$, for which *every* $N \times t$ subarray contains *each* t-tuple of $\{1, 2, \ldots, v\}^t$ among its rows. Covering arrays find application in interaction testing, including software and hardware testing, advanced materials development, and biological systems. A central question is to determine or bound $\mathsf{CAN}(t, k, v)$, the minimum number N of rows of a $\mathsf{CA}(N; t, k, v)$. The well known bound $\mathsf{CAN}(t, k, v) = O((t - 1)v^t \log k)$ is not too far from being asymptotically optimal. Sensible relaxations of the covering requirement arise when (1) the set $\{1, 2, \ldots, v\}^t$ need only be contained among the rows of *at least* $(1 - \epsilon)\binom{k}{t}$ of the $N \times t$ subarrays and (2) the rows of *every* $N \times t$ subarray need only contain a (large) *subset* of $\{1, 2, \ldots, v\}^t$. In this paper, using probabilistic methods, significant improvements on the covering array upper bound are established for both relaxations, and for the conjunction of the two. In each case, a randomized algorithm constructs such arrays in expected polynomial time.

1 Introduction

Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. Let $N, t, k,$ and v be integers such that $k \geq t \geq 2$ and $v \geq 2$. Let A be an $N \times k$ array where each entry is from the set $[v]$. For $I = \{j_1, \ldots, j_\rho\} \subseteq [k]$ where $j_1 < \ldots < j_\rho$, let A_I denote the $N \times \rho$ array in which $A_I(i, \ell) = A(i, j_\ell)$ for $1 \leq i \leq N$ and $1 \leq \ell \leq \rho$; A_I is the projection of A onto the columns in I.

A *covering array* $\mathsf{CA}(N; t, k, v)$ is an $N \times k$ array A with each entry from $[v]$ so that for each t-set of columns $C \in \binom{[k]}{t}$, each t-tuple $x \in [v]^t$ appears as a row in A_C. The smallest N for which a $\mathsf{CA}(N; t, k, v)$ exists is denoted by $\mathsf{CAN}(t, k, v)$.

Covering arrays find important application in software and hardware testing (see [22] and references therein). Applications of covering arrays also arise in experimental testing for advanced materials [4], inference of interactions that regulate gene expression [29], fault-tolerance of parallel architectures [15], synchronization of robot behavior [17], drug screening [30], and learning of boolean functions [11]. Covering arrays have been studied using different nomenclature, as qualitatively independent partitions [13], t-surjective arrays [5], and (k, t)-universal sets [19], among others. Covering arrays are closely related to hash families [10] and orthogonal arrays [8].

2 Background and Motivation

The exact or approximate determination of $\mathsf{CAN}(t, k, v)$ is central in applications of covering arrays, but remains an open problem. For fixed t and v, only when $t = v = 2$ is $\mathsf{CAN}(t, k, v)$ known precisely for infinitely many values of k. Kleitman and Spencer [21] and Katona [20] independently proved that the largest k for which a $\mathsf{CA}(N; 2, k, 2)$ exists satisfies $k = \binom{N-1}{\lceil N/2 \rceil}$. When $t = 2$, Gargano, Kőrner, and Vaccaro [13] establish that

$$\mathsf{CAN}(2, k, v) = \frac{v}{2} \log k (1 + \mathrm{o}(1)). \tag{1}$$

(We write log for logarithms base 2, and ln for natural logarithms.) Several researchers [2,5,14,16] establish a general asymptotic upper bound on $\mathsf{CAN}(t, k, v)$:

$$\mathsf{CAN}(t, k, v) \leq \frac{t-1}{\log \frac{v^t}{v^t - 1}} \log k (1 + \mathrm{o}(1)). \tag{2}$$

A slight improvement on (2) has recently been proved [12,28]. An (essentially) equivalent but more convenient form of (2) is:

$$\mathsf{CAN}(t, k, v) \leq (t-1) v^t \log k (1 + o(1)). \tag{3}$$

A lower bound on $\mathsf{CAN}(t, k, v)$ results from the inequality $\mathsf{CAN}(t, k, v) \geq v \cdot \mathsf{CAN}(t-1, k-1, v)$ obtained by derivation, together with (1), to establish that $\mathsf{CAN}(t, k, v) \geq v^{t-2} \cdot \mathsf{CAN}(2, k-t+2, v) = v^{t-2} \cdot \frac{v}{2} \log(k-t+2)(1+\mathrm{o}(1))$. When $\frac{t}{k} < 1$, we obtain:

$$\mathsf{CAN}(t, k, v) = \Omega(v^{t-1} \log k). \tag{4}$$

Because (4) ensures that the number of rows in covering arrays can be considerable, researchers have suggested the need for relaxations in which not all interactions must be covered [7,18,23,24] in order to reduce the number of rows. The practical relevance is that each row corresponds to a test to be performed, adding to the cost of testing.

For example, an array *covers a t-set of columns* when it covers each of the v^t interactions on this t-set. Hartman and Raskin [18] consider arrays with a fixed number of rows that cover the *maximum* number of t-sets of columns. A similar question was also considered in [24]. In [23,24] a more refined measure of the (partial) coverage of an $N \times k$ array A is introduced. For a given $q \in [0, 1]$, let $\alpha(A, q)$ be the number of $N \times t$ submatrices of A with the property that at least qv^t elements of $[v]^t$ appear in their set of rows; the (q, t)-*completeness* of A is $\alpha(A, q)/\binom{k}{t}$. Then for practical purposes one wants "high" (q, t)-completeness with few rows.

In these works, no theoretical results on partial coverage appear to have been stated; earlier contributions focus on experimental investigations of heuristic construction methods. Our purpose is to initiate a mathematical investigation of arrays offering "partial" coverage. More precisely, we address:

- Can one obtain a significant improvement on the upper bound (3) if the set $[v]^t$ is only required to be contained among the rows of *at least* $(1 - \epsilon)\binom{k}{t}$ subarrays of A of dimension $N \times t$?
- Can one obtain a significant improvement if, among the rows of *every* $N \times t$ subarray of A, only a (large) *subset* of $[v]^t$ is required to be contained?
- Can one obtain a significant improvement if the set $[v]^t$ is only required to be contained among the rows of *at least* $(1 - \epsilon)\binom{k}{t}$ subarrays of A of dimension $N \times t$, **and** among the rows of each of the $\epsilon\binom{k}{t}$ subarrays that remain, a (large) *subset* of $[v]^t$ is required to be contained?

We answer these questions both theoretically and algorithmically in the following sections.

3   Partial Covering Arrays

When $1 \leq m \leq v^t$, a *partial m-covering array*, $\mathsf{PCA}(N; t, k, v, m)$, is an $N \times k$ array A with each entry from $[v]$ so that for each t-set of columns $C \in \binom{[k]}{t}$, at least m distinct tuples $x \in [v]^t$ appear as rows in A_C. Hence a covering array $\mathsf{CA}(N; t, k, v)$ is precisely a partial v^t-covering array $\mathsf{PCA}(N; t, k, v, v^t)$.

Theorem 1. *For integers t, k, v, and m where $k \geq t \geq 2$, $v \geq 2$ and $1 \leq m \leq v^t$ there exists a $\mathsf{PCA}(N; t, k, v, m)$ with*

$$N \leq \frac{\ln\left\{\binom{k}{t}\binom{v^t}{m-1}\right\}}{\ln\left(\frac{v^t}{m-1}\right)}. \tag{5}$$

.

Proof. Let $r = v^t - m + 1$, and A be a random $N \times k$ array where each entry is chosen independently from $[v]$ with uniform probability. For $C \in \binom{[k]}{t}$, let B_C denote the event that at least r tuples from $[v]^t$ are missing in A_C. The probability that a particular r-set of tuples from $[v]^t$ is missing in A_C is $\left(1 - \frac{r}{v^t}\right)^N$. Applying the union bound to all r-sets of tuples from $[v]^t$, we obtain $\Pr[B_C] \leq \binom{v^t}{r}\left(1 - \frac{r}{v^t}\right)^N$. By linearity of expectation, the expected number of t-sets C for which A_C misses at least r tuples from $[v]^t$ is at most $\binom{k}{t}\binom{v^t}{r}\left(1 - \frac{r}{v^t}\right)^N$. When A has at least $\frac{\ln\left\{\binom{k}{t}\binom{v^t}{m-1}\right\}}{\ln\left(\frac{v^t}{m-1}\right)}$ rows this expected number is less than 1. Therefore, an array A exists with the required number of rows such that for all $C \in \binom{[k]}{t}$, A_C misses at most $r - 1$ tuples from $[v]^t$, i.e. A_C covers at least m tuples from $[v]^t$. $\square$

Theorem 1 can be improved upon using the Lovász local lemma.

Lemma 1. *(Lovász local lemma; symmetric case) (see [1]) Let $A_1, A_2, \ldots, A_n$ events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all other events A_j except for at most d, and that $\Pr[A_i] \leq p$ for all $1 \leq i \leq n$. If $ep(d + 1) \leq 1$, then $\Pr[\cap_{i=1}^n \bar{A}_i] > 0$.*

Lemma 1 provides an upper bound on the probability of a "bad" event in terms of the dependence structure among such bad events, so that there is a guaranteed outcome in which all "bad" events are avoided. This lemma is most useful when there is limited dependence among the "bad" events, as in the following:

Theorem 2. *For integers t, k, v and m where $v, t \geq 2$, $k \geq 2t$ and $1 \leq m \leq v^t$ there exists a $\mathsf{PCA}(N; t, k, v, m)$ with*

$$N \leq \frac{1 + \ln\left\{t\binom{k}{t-1}\binom{v^t}{m-1}\right\}}{\ln\left(\frac{v^t}{m-1}\right)}. \tag{6}$$

Proof. When $k \geq 2t$, each event B_C with $C \in \binom{[k]}{t}$ (that is, at least $v^t - m + 1$ tuples are missing in A_C) is independent of all but at most $\binom{t}{1}\binom{k-1}{t-1} < t\binom{k}{t-1}$ events in $\{B_{C'} : C' \in \binom{[k]}{t} \setminus \{C\}\}$. Applying Lemma 1, $\Pr[\wedge_{C \in \binom{[k]}{t}} \overline{B_C}] > 0$ when

$$\mathrm{e}\binom{v^t}{r}\left(1 - \frac{r}{v^t}\right)^N t\binom{k}{t-1} \leq 1. \tag{7}$$

Solve (7) to obtain the required upper bound on N. $\qquad\square$

When $m = v^t$, apply the Taylor series expansion to obtain $\ln\left(\frac{v^t}{m-1}\right) \geq \frac{1}{v^t}$, and thereby recover the upper bound (3). Theorem 2 implies:

Corollary 1. *Given $q \in [0, 1]$ and integers $2 \leq t \leq k$, $v \geq 2$, there exists an $N \times k$ array on $[v]$ with (q, t)-completeness equal to 1 (i.e., maximal), whose number N of rows satisfies*

$$N \leq \frac{1 + \ln\left\{t\binom{k}{t-1}\binom{v^t}{qv^t-1}\right\}}{\ln\left(\frac{v^t}{qv^t-1}\right)}.$$

Rewriting (6), setting $r = v^t - m + 1$, and using the Taylor series expansion of $\ln\left(1 - \frac{r}{v^t}\right)$, we get

$$N \leq \frac{1 + \ln\left\{t\binom{k}{t-1}\binom{v^t}{r}\right\}}{\ln\left(\frac{v^t}{v^t-r}\right)} \leq \frac{v^t(t-1)\ln k}{r}\left\{1 - \frac{\ln r}{\ln k} + o(1)\right\}. \tag{8}$$

Hence when $r = v(t - 1)$ (or equivalently, $m = v^t - v(t - 1) + 1$), there is a partial m-covering array with $\Theta(v^{t-1}\ln k)$ rows. This matches the lower bound (4) asymptotically for covering arrays by missing, in each t-set of columns, *no more than* $v(t - 1) - 1$ of the v^t possible rows.

The dependence of the bound (6) on the number of v-ary t-vectors that must appear in the t-tuples of columns is particularly of interest when test suites are run sequentially until a fault is revealed, as in [3]. Indeed the arguments here may have useful consequences for the rate of fault detection.

Algorithm 1: Moser-Tardos type algorithm for partial m-covering arrays.

Input: Integers N, t, k, v and m where $v, t \geq 2$, $k \geq 2t$ and $1 \leq m \leq v^t$

Output: A : a $\mathsf{PCA}(N; t, k, v, m)$

1 Let $N := \dfrac{1 + \ln\left\{ t\binom{k}{t-1}\binom{v^t}{m-1} \right\}}{\ln\left(\frac{v^t}{m-1}\right)}$;

2 Construct an $N \times k$ array A where each entry is chosen independently and uniformly at random from $[v]$;

3 repeat

4   Set *covered* := true;

5   **for** *each column t-set* $C \in \binom{[k]}{t}$ **do**

6   **if** A_C *does not cover at least m distinct t-tuples* $x \in [v]^t$ **then**

7   Set *covered* := false;

8   Set *missing-column-set* := C;

9   **break**;

10   **end**

11   **end**

12   **if** *covered* = *false* **then**

13   Choose all the entries in the t columns of *missing-column-set* independently and uniformly at random from $[v]$;

14   **end**

15 until *covered* = *true*;

16 Output A;

Lemma 1 and hence Theorem 2 have proofs that are non-constructive in nature. Nevertheless, Moser and Tardos [26] provide a randomized algorithm with the same guarantee. Patterned on their method, Algorithm 1 constructs a partial m-covering array with exactly the same number of rows as (6) in expected polynomial time. Indeed, for fixed t, the expected number of times the resampling step (line 13) is repeated is linear in k (see [26] for more details).

4   Almost Partial Covering Arrays

For $0 < \epsilon < 1$, an ϵ-*almost partial m-covering array*, $\mathsf{APCA}(N; t, k, v, m, \epsilon)$, is an $N \times k$ array A with each entry from $[v]$ so that for at least $(1 - \epsilon)\binom{k}{t}$ column t-sets $C \in \binom{[k]}{t}$, A_C covers at least m distinct tuples $x \in [v]^t$. Again, a covering array $\mathsf{CA}(N; t, k, v)$ is precisely an $\mathsf{APCA}(N; t, k, v, v^t, \epsilon)$ when $\epsilon < 1/\binom{k}{t}$. Our first result on ϵ-*almost* partial m-covering arrays is the following.

Theorem 3. *For integers t, k, v, m and real ϵ where $k \geq t \geq 2$, $v \geq 2$, $1 \leq m \leq v^t$ and $0 \leq \epsilon \leq 1$, there exists an $\mathsf{APCA}(N; t, k, v, m, \epsilon)$ with*

$$N \leq \frac{\ln\left\{ \binom{v^t}{m-1} / \epsilon \right\}}{\ln\left(\frac{v^t}{m-1}\right)}. \tag{9}$$

Proof. Parallelling the proof of Theorem 1 we compute an upper bound on the expected number of t-sets $C \in \binom{[k]}{t}$ for which A_C misses at least r tuples $x \in [v]^t$. When this expected number is at most $\epsilon\binom{k}{t}$, an array A is guaranteed to exist with at least $(1-\epsilon)\binom{k}{t}$ t-sets of columns $C \in \binom{[k]}{t}$ such that A_C misses at most $r-1$ distinct tuples $x \in [v]^t$. Thus A is an $\mathsf{APCA}(N; t, k, v, m, \epsilon)$. To establish the theorem, solve the following for N:

$$\binom{k}{t}\binom{v^t}{r}\left(1 - \frac{r}{v^t}\right)^N \leq \epsilon\binom{k}{t}.$$

$\square$

When $\epsilon < 1/\binom{k}{t}$ we recover the bound from Theorem 1 for partial m-covering arrays. In terms of (q, t)-completeness, Theorem 3 yields the following.

Corollary 2. *For $q \in [0, 1]$ and integers $2 \leq t \leq k$, $v \geq 2$, there exists an $N \times k$ array on $[v]$ with (q, t)-completeness equal to $1 - \epsilon$, with*

$$N \leq \frac{\ln\left\{\binom{v^t}{m-1}/\epsilon\right\}}{\ln\left(\frac{v^t}{m-1}\right)}.$$

When $m = v^t$, an ϵ-almost covering array exists with $N \leq v^t \ln\left(\frac{v^t}{\epsilon}\right)$ rows. Improvements result by focussing on covering arrays in which the symbols are acted on by a finite group. In this setting, one chooses orbit representatives of rows that collectively cover orbit representatives of t-way interactions under the group action; see [9], for example. Such group actions have been used in direct and computational methods for covering arrays [6,25], and in randomized and derandomized methods [9,27,28].

We employ the sharply transitive action of the cyclic group of order v, adapting the earlier arguments using methods from [28]:

Theorem 4. *For integers t, k, v and real ϵ where $k \geq t \geq 2$, $v \geq 2$ and $0 \leq \epsilon \leq 1$ there exists an $\mathsf{APCA}(N; t, k, v, v^t, \epsilon)$ with*

$$N \leq v^t \ln\left(\frac{v^{t-1}}{\epsilon}\right). \tag{10}$$

Proof. The action of the cyclic group of order v partitions $[v]^t$ into v^{t-1} orbits, each of length v. Let $n = \lfloor \frac{N}{v} \rfloor$ and let A be an $n \times k$ random array where each entry is chosen independently from the set $[v]$ with uniform probability. For $C \in \binom{[k]}{t}$, A_C *covers the orbit* X if at least one tuple $x \in X$ is present in A_C. The probability that the orbit X is not covered in A is $\left(1 - \frac{v}{v^t}\right)^n = \left(1 - \frac{1}{v^{t-1}}\right)^n$. Let D_C denote the event that A_C does not cover at least one orbit. Applying the union bound, $\Pr[D_C] \leq v^{t-1}\left(1 - \frac{1}{v^{t-1}}\right)^n$. By linearity of expectation, the expected number of column t-sets C for which D_C occurs is at most $\binom{k}{t}v^{t-1}\left(1 - \frac{1}{v^{t-1}}\right)^n$. As earlier, set this expected value to be

at most $\epsilon\binom{k}{t}$ and solve for n. An array exists that covers all orbits in at least $(1-\epsilon)\binom{k}{t}$ column t-sets. Develop this array over the cyclic group to obtain the desired array. $\qquad\square$

As in [28], further improvements result by considering a group, like the Frobenius group, that acts sharply 2-transitively on $[v]$. When v is a prime power, the *Frobenius group* is the group of permutations of $\mathbb{F}_v$ of the form $\{x \mapsto ax + b : a, b \in \mathbb{F}_v, a \neq 0\}$.

Theorem 5. *For integers t, k, v and real ϵ where $k \geq t \geq 2$, $v \geq 2$, v is a prime power and $0 \leq \epsilon \leq 1$ there exists an $\mathsf{APCA}(N; t, k, v, v^t, \epsilon)$ with*

$$N \leq v^t \ln\left(\frac{2v^{t-2}}{\epsilon}\right) + v. \tag{11}$$

Proof. The action of the Frobenius group partitions $[v]^t$ into $\frac{v^{t-1}-1}{v-1}$ orbits of length $v(v-1)$ (full orbits) each and 1 orbit of length v (a short orbit). The short orbit consists of tuples of the form $(x_1, \ldots, x_t) \in [v]^t$ where $x_1 = \ldots = x_t$. Let $n = \lfloor \frac{N-v}{v(v-1)} \rfloor$ and let A be an $n \times k$ random array where each entry is chosen independently from the set $[v]$ with uniform probability. Our strategy is to construct A so that it covers all full orbits for the required number of arrays $\{A_C : C \in \binom{[k]}{t}\}$. Develop A over the Frobenius group and add v rows of the form $(x_1, \ldots, x_k) \in [v]^t$ with $x_1 = \ldots = x_k$ to obtain an $\mathsf{APCA}(N; t, k, v, v^t, \epsilon)$ with the desired value of N. Following the lines of the proof of Theorem 4, A covers all full orbits in at least $(1-\epsilon)\binom{k}{t}$ column t-sets C when

$$\binom{k}{t}\frac{v^{t-1}-1}{v-1}\left(1 - \frac{v-1}{v^{t-1}}\right)^n \leq \epsilon\binom{k}{t}.$$

Because $\frac{v^{t-1}-1}{v-1} \leq 2v^{t-2}$ for $v \geq 2$, we obtain the desired bound. $\qquad\square$

Using group action when $m = v^t$ affords useful improvements. Does this improvement extend to cases when $m < v^t$? Unfortunately, the answer appears to be no. Consider the case for $\mathsf{PCA}(N; t, k, v, m)$ when $m \leq v^t$ using the action of the cyclic group of order v on $[v]^t$. Let A be a random $n \times k$ array over $[v]$. When $v^t - vs + 1 \leq m \leq v^t - v(s-1)$ for $1 \leq s \leq v^{t-1}$, this implies that for all $C \in \binom{[k]}{t}$, A_C misses at most $s-1$ orbits of $[v]^t$. Then we obtain that $n \leq \left(1 + \ln\left(t\binom{k}{t-1}\binom{v^{t-1}}{s}\right)\right) / \ln\left(\frac{v^{t-1}}{v^{t-1}-s}\right)$. Developing A over the cyclic group we obtain a $\mathsf{PCA}(N; t, k, v, m)$ with

$$N \leq v\frac{1 + \ln\left\{\binom{k}{t-1}\binom{v^{t-1}}{s}\right\}}{\ln\left(\frac{v^{t-1}}{v^{t-1}-s}\right)} \tag{12}$$

Figure 1 compares (12) and (6). In Figure 1a we plot the size of the partial m-covering array as obtained by (12) and (6) for $v^t - 6v + 1 \leq m \leq v^t$ and
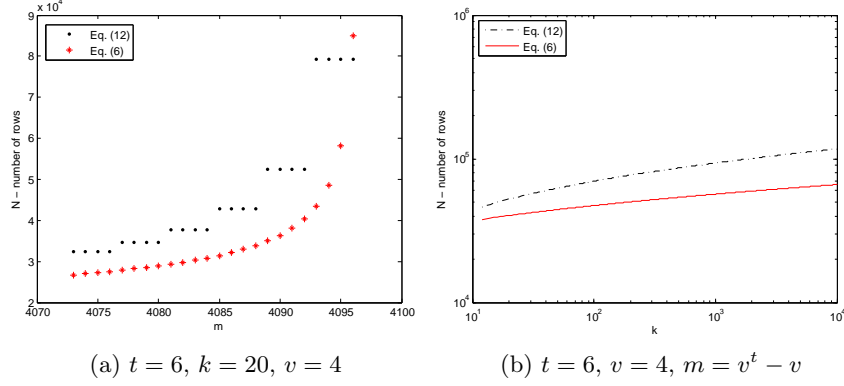
(a) $t = 6$, $k = 20$, $v = 4$ (b) $t = 6$, $v = 4$, $m = v^t - v$

Fig. 1: Comparison of (12) and (6). Figure (a) compares the sizes of the partial m-covering arrays when $v^t - 6v + 1 \le m \le v^t$. Except for $m = v^t = 4096$ the bound from (6) outperforms the bound obtained by assuming group action. Figure (b) shows that for $m = v^t - v = 4092$, (6) outperforms (12) for all values of k.

$t = 6$, $k = 20$, $v = 4$. Except when $m = v^t = 4096$, the covering array case, (6) outperforms (12). Similarly, Figure 1b shows that for $m = v^t - v = 4092$, (6) consistently outperforms (12) for all values of k when $t = 6$, $v = 4$. We observe similar behavior for different values of t and v.

Next we consider even stricter coverage restrictions, combining Theorems 2 and 4.

Theorem 6. *For integers t, k, v, m and real ϵ where $k \ge t \ge 2$, $v \ge 2$, $0 \le \epsilon \le 1$ and $m \le v^t + 1 - \frac{\ln k}{\ln(v/\epsilon^{1/(t-1)})}$ there exists an $N \times k$ array A with entries from $[v]$ such that*

1. *for each $C \in \binom{[k]}{t}$, A_C covers at least m tuples $x \in [v]^t$,*
2. *for at least $(1 - \epsilon)\binom{k}{t}$ column t-sets C, A_C covers all tuples $x \in [v]^t$,*
3. *$N = O(v^t \ln\left(\frac{v^{t-1}}{\epsilon}\right))$.*

Proof. We vertically juxtapose a partial m-covering array and an ϵ-almost v^t-covering array. For $r = \frac{\ln k}{\ln(v/\epsilon^{1/(t-1)})}$ and $m = v^t - r + 1$, (8) guarantees the existence of a partial m-covering array with $v^t \ln\left(\frac{v^{t-1}}{\epsilon}\right)\{1 + o(1)\}$ rows. Theorem 4 guarantees the existence of an ϵ-almost v^t-covering array with at most $v^t \ln\left(\frac{v^{t-1}}{\epsilon}\right)$ rows. $\qquad\square$

Corollary 3. *There exists an $N \times k$ array A such that:*

1. *for any t-set of columns $C \in \binom{[k]}{t}$, A_C covers at least $m \le v^t + 1 - v(t - 1)$ distinct t-tuples $x \in [v]^t$,*

2. *for at least* $\left(1 - \frac{v^{t-1}}{k^{1/v}}\right)\binom{k}{t}$ *column t-sets C, A_C covers all the distinct t-tuples* $x \in [v]^t$.

3. $N = O(v^{t-1} \ln k)$.

Proof. Apply Theorem 6 with $m = v^t + 1 - \frac{\ln k}{\ln(v/\epsilon^{1/(t-1)})}$. There are at most $\frac{\ln k}{\ln(v/\epsilon^{1/(t-1)})} - 1$ missing t-tuples $x \in [v]^t$ in the A_C for each of the at most $\epsilon\binom{k}{t}$ column t-sets C that do not satisfy the second condition of Theorem 6. To bound from above the number of missing tuples to a certain small function $f(t)$ of t, it is sufficient that $\epsilon \le v^{t-1}\left(\frac{1}{k}\right)^{\frac{t-1}{f(t)+1}}$. Then the number of missing t-tuples $x \in [v]^t$ in A_C is bounded from above by $f(t)$ whenever ϵ is not larger than

$$v^{t-1}\left(\frac{1}{k}\right)^{\frac{t-1}{f(t)+1}} \tag{13}$$

On the other hand, in order for the number $N = O\left(v^{t-1}\ln\left(\frac{v^{t-1}}{\epsilon}\right)\right)$ of rows of A to be asymptotically equal to the lower bound (4), it suffices that ϵ is not smaller than

$$\frac{v^{t-1}}{k^{\frac{1}{v}}}. \tag{14}$$

When $f(t) = v(t-1) - 1$, (13) and (14) agree asymptotically, completing the proof. $\square$

Once again we obtain a size that is $O(v^{t-1}\log k)$, a goal that has not been reached for covering arrays. This is evidence that even a small relaxation of covering arrays provides arrays of the best sizes one can hope for.

Next we consider the efficient construction of the arrays whose existence is ensured by Theorem 6. Algorithm 2 is a randomized method to construct an $\mathsf{APCA}(N; t, k, v, m, \epsilon)$ of a size N that is very close to the bound of Theorem 3. By Markov's inequality the condition in line 9 of Algorithm 2 is met with probability at most $1/2$. Therefore, the expected number of times the loop in line 2 repeats is at most 2.

To prove Theorem 3, t-wise independence among the variables is sufficient. Hence, Algorithm 2 can be derandomized using t-wise independent random variables. We can also derandomize the algorithm using the method of conditional expectation. In this method we construct A by considering the k columns one by one and fixing all N entries of a column. Given a set of already fixed columns, to fix the entries of the next column we consider all possible v^N choices, and choose one that provides the maximum conditional expectation of the number of column t-sets $C \in \binom{[k]}{t}$ such that A_C covers at least m tuples $x \in [v]^t$. Because $v^N = O(\mathsf{poly}(1/\epsilon))$, this derandomized algorithm constructs the desired array in polynomial time. Similar randomized and derandomized strategies can be applied to construct the array guaranteed by Theorem 4. Together with Algorithm 1 this implies that the array in Theorem 6 is also efficiently constructible.

Algorithm 2: Randomized algorithm for ϵ-almost partial m-covering arrays.

Input: Integers N, t, k, v and m where $v, t \geq 2$, $k \geq 2t$ and $1 \leq m \leq v^t$, and real $0 < \epsilon < 1$

Output: A : an $\mathsf{APCA}(N; t, k, v, m, \epsilon)$

1 Let $N := \dfrac{\ln\left\{2\binom{v^t}{m-1}/\epsilon\right\}}{\ln\left(\frac{v^t}{m-1}\right)}$;

2 repeat

3   Construct an $N \times k$ array A where each entry is chosen independently and uniformly at random from $[v]$;

4   Set *isAPCA*:= true;

5   Set *defectiveCount*:= 0;

6   **for** *each column t-set $C \in \binom{[k]}{t}$* **do**

7   **if** *A_C does not cover at least m distinct t-tuples $x \in [v]^t$* **then**

8   Set *defectiveCount*:= *defectiveCount* $+ 1$;

9   **if** defectiveCount $> \lfloor\epsilon\binom{k}{t}\rfloor$ **then**

10   Set *isAPCA*:= false;

11   **break**;

12   **end**

13   **end**

14   **end**

15 until *isAPCA* $= true$;

16 Output A;

---

5   Final Remarks

We have shown that by relaxing the coverage requirement of a covering array somewhat, powerful upper bounds on the sizes of the arrays can be established. Indeed the upper bounds are substantially smaller than the best known bounds for a covering array; they are of the same order as the *lower* bound for $\mathsf{CAN}(t, k, v)$. As importantly, the techniques not only provide asymptotic bounds but also randomized polynomial time construction algorithms for such arrays.

Our approach seems flexible enough to handle variations of these problems. For instance, some applications require arrays that satisfy, for different subsets of columns, different coverage or separation requirements [8]. In [16] several interesting examples of combinatorial problems are presented that can be unified and expressed in the framework of S-constrained matrices. Given a set of vectors S each of length t, an $N \times k$ matrix M is *S-constrained* if for every t-set $C \in \binom{[k]}{t}$, M_C contains as a row each of the vectors in S. The parameter to optimize is, as usual, the number of rows of M. One potential direction is to ask for arrays that, in every t-tuple of columns, cover at least m of the vectors in S, or that all vectors in S are covered by all but a small number of t-tuples of columns. Exploiting the structure of the members of S appears to require an extension of the results developed here.

Acknowledgements

References

1. Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2008.
2. B. Becker and H.-U. Simon. How robust is the n-cube? *Inform. and Comput.*, 77:162–178, 1988.
3. Renée C. Bryce, Yinong Chen, and Charles J. Colbourn. Biased covering arrays for progressive ranking and composition of web services. *Int. J. Simulation Process Modelling*, 3(1/2):80–87, 2007.
4. J. N. Cawse. Experimental design for combinatorial and high throughput materials development. *GE Global Research Technical Report*, 29:769–781, 2002.
5. Ashok K. Chandra, Lawrence T. Kou, George Markowsky, and Shmuel Zaks. On sets of boolean n-vectors with all k-projections surjective. *Acta Informatica*, 20(1):103–111, 1983.
6. M. A. Chateauneuf, C. J. Colbourn, and D. L. Kreher. Covering arrays of strength 3. *Des. Codes Crypt.*, 16:235–242, 1999.
7. Baiqiang Chen and Jian Zhang. Tuple density: a new metric for combinatorial test suites. In *Proceedings of the 33rd International Conference on Software Engineering, ICSE 2011, Waikiki, Honolulu , HI, USA, May 21-28, 2011*, pages 876–879, 2011.
8. C. J. Colbourn. Combinatorial aspects of covering arrays. *Le Matematiche (Catania)*, 58:121–167, 2004.
9. C. J. Colbourn. Conditional expectation algorithms for covering arrays. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 90:97–115, 2014.
10. Charles J. Colbourn. Covering arrays and hash families. In D. Crnkovič and V. Tonchev, editors, *Information Security, Coding Theory, and Related Combinatorics*, NATO Science for Peace and Security Series, pages 99–135. IOS Press, 2011.
11. Peter Damaschke. Adaptive versus nonadaptive attribute-efficient learning. *Machine Learning*, 41(2):197–215, 2000.
12. N. Francetić and B. Stevens. Asymptotic size of covering arrays: an application of entropy compression. *ArXiv e-prints*, March 2015.
13. L. Gargano, J. Körner, and U. Vaccaro. Sperner capacities. *Graphs and Combinatorics*, 9:31–46, 1993.
14. A. P. Godbole, D. E. Skipper, and R. A. Sunley. t-covering arrays: upper bounds and Poisson approximations. *Combinatorics, Probability and Computing*, 5:105–118, 1996.
15. N. Graham, F. Harary, M. Livingston, and Q.F. Stout. Subcube fault-tolerance in hypercubes. *Information and Computation*, 102(2):280 – 314, 1993.
16. Sylvain Gravier and Bernard Ycart. S-constrained random matrices. *DMTCS Proceedings*, 0(1), 2006.

17. A. Hartman. Software and hardware testing using combinatorial covering suites. In M. C. Golumbic and I. B.-A. Hartman, editors, *Interdisciplinary Applications of Graph Theory, Combinatorics, and Algorithms*, pages 237–266. Springer, Norwell, MA, 2005.

18. Alan Hartman and Leonid Raskin. Problems and algorithms for covering arrays. *Discrete Mathematics*, 284(13):149 – 156, 2004.

19. Stasys Jukna. *Extremal Combinatorics: With Applications in Computer Science*. Springer Publishing Company, Incorporated, 1st edition, 2010.

20. G. O. H. Katona. Two applications (for search theory and truth functions) of Sperner type theorems. *Periodica Math.*, 3:19–26, 1973.

21. D. Kleitman and J. Spencer. Families of k-independent sets. *Discrete Math.*, 6:255–262, 1973.

22. D. R. Kuhn, R. Kacker, and Y. Lei. *Introduction to Combinatorial Testing*. CRC Press, 2013.

23. D. R. Kuhn, I. D. Mendoza, R. N. Kacker, and Y. Lei. Combinatorial coverage measurement concepts and applications. In *Software Testing, Verification and Validation Workshops (ICSTW), 2013 IEEE Sixth International Conference on*, pages 352–361, March 2013.

24. J. R. Maximoff, M. D. Trela, D. R. Kuhn, and R. Kacker. A method for analyzing system state-space coverage within a t-wise testing framework. In *4th Annual IEEE Systems Conference*, pages 598–603, 2010.

25. K. Meagher and B. Stevens. Group construction of covering arrays. *J. Combin. Des.*, 13:70–77, 2005.

26. Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. *J. ACM*, 57(2):Art. 11, 15, 2010.

27. K. Sarkar and C. J. Colbourn. Two-stage algorithms for covering array construction. *submitted for publication*.

28. K. Sarkar and C. J. Colbourn. Upper bounds on the size of covering arrays. *ArXiv e-prints*, March 2016.

29. D. E. Shasha, A. Y. Kouranov, L. V. Lejay, M. F. Chou, and G. M. Coruzzi. Using combinatorial design to study regulation by multiple input signals: A tool for parsimony in the post-genomics era. *Plant Physiol.*, 127:1590–2594, 2001.

30. A. J. Tong, Y. G. Wu, and L. D. Li. Room-temperature phosphorimetry studies of some addictive drugs following dansyl chloride labelling. *Talanta*, 43(9):14291436, September 1996.