Asymptotically Optimal Covering Designs

Daniel M. Gordon* Greg Kuperberg† Oren Patashnik*

Joel H. Spencer‡

10 November 1995

Abstract

A (v, k, t) *covering design*, or *covering*, is a family of k-subsets, called *blocks*, chosen from a v-set, such that each t-subset is contained in at least one of the blocks. The number of blocks is the covering's *size*, and the minimum size of such a covering is denoted by $C(v, k, t)$. It is easy to see that a covering must contain at least $\binom{v}{t}/\binom{k}{t}$ blocks, and in 1985 Rödl [5] proved a long-standing conjecture of Erdős and Hanani [3] that for fixed k and t, coverings of size $\binom{v}{t}/\binom{k}{t}(1 + o(1))$ exist (as $v \to \infty$).

An earlier paper by the first three authors [4] gave new methods for constructing good coverings, and gave tables of upper bounds on $C(v, k, t)$ for small v, k, and t. The present paper shows that two of those constructions are asymptotically optimal: For fixed k and t, the size of the coverings constructed matches Rödl's bound. The paper also makes the $o(1)$ error bound explicit, and gives some evidence for a much stronger bound.

1   Introduction

Let the covering number $C(v, k, t)$ denote the smallest number of k-subsets of a v-set that cover all t-subsets. The best general lower bound on $C(v, k, t)$, due to Schönheim [7], comes from the following inequality:

---

*Center for Communications Research, 4320 Westerra Ct., San Diego, CA 92121

†Department of Mathematics, Yale University, New Haven, CT 06520

‡Courant Institute, NYU, New York, NY 10012

Theorem 1
$$C(v, k, t) \geq \left\lceil \frac{v}{k} C(v-1, k-1, t-1) \right\rceil.$$

Iterating this gives the Schönheim bound

$$C(v, k, t) \geq \left\lceil \frac{v}{k} \left\lceil \frac{v-1}{k-1} \cdots \left\lceil \frac{v-t+1}{k-t+1} \right\rceil \cdots \right\rceil \right\rceil.$$

The best general upper bound on $C(v, k, t)$ is due to Rödl [5]: The *density* of a covering is the average number of blocks containing a t-set. The minimum density is $C(v, k, t)\binom{k}{t}/\binom{v}{t}$, and is obviously at least 1. Rödl showed that for k and t fixed there exist coverings with density $1 + o(1)$ as v gets large.

This paper shows that two of our constructions [4] match the bound of Rödl's theorem. One of the constructions gives an easier proof of the theorem than Rödl's original proof [5]. The other construction provides a computationally efficient version of Rödl's theorem. In Section 2 we review the two constructions. In Section 3 we show that the first one, which uses a greedy algorithm, is asymptotically optimal. And in Section 4 we show that the second one, which constructs an induced covering from a finite-geometry covering, is also asymptotically optimal, and that it is computationally efficient as well.

Theorem 3 (in Section 3) is a special case of a main result of the fourth author [8]; Rödl and Thoma [6] gave another proof of that result. We present the proof here to keep the paper self-contained and to provide an explicit error bound for use in Section 4.

2 Covering Constructions

Here we summarize two methods for constructing asymptotically optimal coverings. Our previous paper [4] gives more details, as well as computational results for small v, k, and t.

2.1 Greedy Coverings

Algorithm 1 Random Greedy (v, k, t) Covering

1. Fix a random ordering of the k-sets of a v-set.

2. Choose the earliest k-set containing no already-covered t-set.

3. Repeat Step 2 until no k-set can be chosen.

4. Cover the remaining t-sets with one k-set each.

This greedy algorithm is a little different from our previous one. That algorithm uses one of four possible orderings in Step 1: lexicographic, colex, Gray code, or random. Also, it chooses in Step 2 the earliest k-set that contains the most still-uncovered t-sets; thus it continues with Steps 2 and 3 instead of cutting out to Step 4. That algorithm produces slightly better coverings in practice, but is harder to analyze than the algorithm here.

2.2 Induced Finite Geometry Coverings

The k-flats of an affine or projective geometry form a covering. For this paper, we restrict our attention to the hyperplanes of an affine geometry, which form an optimal covering:

Theorem 2 *For a prime power q and integer $t > 1$, the hyperplanes of the affine geometry $\mathrm{AG}(t, q)$ are a (q^t, q^{t-1}, t) covering of size*

$$C(q^t, q^{t-1}, t) = \frac{q^{t+1} - q}{q - 1}.$$

The density of such a covering is

$$\frac{q^{t+1} - q}{q - 1} \binom{q^{t-1}}{t} \bigg/ \binom{q^t}{t} = 1 + O(q^{-1}).$$

Algorithm 2 Induced (v, k, t) Covering

1. Choose a prime p with $p^t > v$, and an integer ℓ, as specified later.

2. Precompute (ℓ', k, t) coverings, for $\ell < \ell' < 9\ell$, using Algorithm 1.

3. Choose v points of the $\mathrm{AG}(t, p)$ at random.

4. For each hyperplane, find its intersection with the v points; let ℓ' be the size of the intersection.
   (a) If $\ell < \ell' < 9\ell$, add the blocks of the (ℓ', k, t) covering on those points to the (v, k, t) covering.
   (b) If $\ell' \leq \ell$ or $\ell' \geq 9\ell$, trivially add $\binom{\ell'}{t}$ blocks to the (v, k, t) covering.

3

The new blocks each have k elements, and together they cover all t-sets, so they form a (v, k, t) covering. The blocks of the affine covering and their intersection with the v-set may quickly be computed by solving linear equations over $GF(q)$.

This construction, too, differs slightly from our earlier version [4]. In that paper, we construct (ℓ', k, t) coverings for all $\ell' < v$ by whatever construction gives the best results, and then always use Step 4a. That results in better coverings in practice, but is harder to analyze.

3 Greedy Coverings and Rödl's Bound

The usual proofs of Rödl's theorem (Rödl [5] or Alon and Spencer [1]) seem nonconstructive; however, they are actually analyses of a covering algorithm, similar to the greedy algorithm with random ordering, that constructs a covering in two steps. First, it chooses a sequence of Rödl nibbles, each of which is a small, random collection of k-sets that do not contain any t-set contained in any previous Rödl nibble. Second, when there is no longer room for a nibble, it chooses a separate k-set for each remaining uncovered t-set.

The main difference between the k-sets chosen in the sequence of Rödl nibbles and those chosen by the greedy algorithm in Steps 2 and 3 is that two k-sets in the same nibble may intersect each other in a t-set. This difference seems small, hence it is natural to conjecture that the greedy algorithm, too, meets Rödl's bound. It does:

Theorem 3 *For fixed k and t, the greedy algorithm with random ordering produces a covering with expected density $1 + o(1)$ as $v \to \infty$.*

The proof of Theorem 3 will proceed in several steps, along the lines of Spencer [8].

3.1 The Continuous Model

Model the execution of the greedy algorithm as a Poisson process; that is, a given k-set is chosen between time τ and $\tau + \delta$ with probability asymptotic to $\delta / \binom{v-t}{k-t}$ as $\delta \to 0$, and the probabilities of any two k-sets being chosen in any two time intervals are independent. The process begins at time 0 and lasts forever. If a k-set chosen by the process at some time τ contains

4

any previously covered t-set, it *fails* at time τ, otherwise it *succeeds* and its t-sets are considered covered after time τ. The k-set thus fails at any time subsequent to τ it is chosen.

The ordering determined by the first-choosings of the k-sets in this process corresponds to the random ordering of the k-sets in the greedy algorithm, and the k-sets that have succeeded at time infinity correspond to the k-sets chosen by the greedy algorithm just prior to Step 4. Thus to prove the theorem it suffices to show that, at time infinity of the Poisson process, a given t-set is covered with probability asymptotic to 1. (Since if the proportion of t-sets covered at that point of the greedy algorithm goes to 1 then so does the density of the eventual covering.) We actually find the limit of this probability as $v \to \infty$ for every fixed τ, and we show that this limit goes to 1 as $\tau \to \infty$.

Fix a time τ and a t-set T. Based on the Poisson process above, we either define the *dependence tree* of (τ, T) or else declare it to be *aborted*. The tree is rooted, and has t-vertices and k-vertices—(τ', T') with $|T'| = t$ and (τ', K') with $|K'| = k$. We begin at time τ with the tree consisting only of its t-vertex root (τ, T), and we examine k-sets chosen by the process, proceeding backwards in time from τ toward 0.

There are three cases for a k-set K^* chosen at some time τ^*: if K^* does not contain any T' already in the tree then do nothing; if it contains two or more such T' then declare the tree to be aborted; if (the important case) it contains precisely one such T' then add (τ^*, K^*) as a child of (τ', T') and for every t-set $T^* \subset K^*$ *except* T' add (τ^*, T^*) as a child of (τ^*, K^*). We will say that T has given birth to K^* at time τ^*, and K^* immediately gives birth to all the T' nodes.

The tree, if defined, is finite; a child of a t-vertex is a k-vertex and vice versa. We label each vertex as follows. A t-vertex is *covered* if at least one of its children is accepted, else it is *uncovered*; a k-vertex is *accepted* if none of its children is covered, else it is *rejected*. Thus a childless (leaf) t-vertex is uncovered, and a unique labeling is defined inductively from the leaves up.

Example. Take $t = 2$; $k = 3$; $v = 10^{10}$; $\tau = 4.3$; $T = \{1, 2\}$. Suppose $\{1, 2, 3\}$ is chosen at time 3.7 and $\{2, 3, 4\}$ at time 1.2 and these are the only relevant chosen sets. The dependence tree of $(4.3, \{1, 2\})$ is shown in Figure 1. Two of the leaves $(1.2, \{2, 4\})$ and $(1.2, \{3.4\})$ are uncovered, thus their parent $(1.2, \{2, 3, 4\})$ is accepted, so $(3.7, \{2, 3\})$ is covered and $(3.7, \{1, 2, 3\})$ is re-
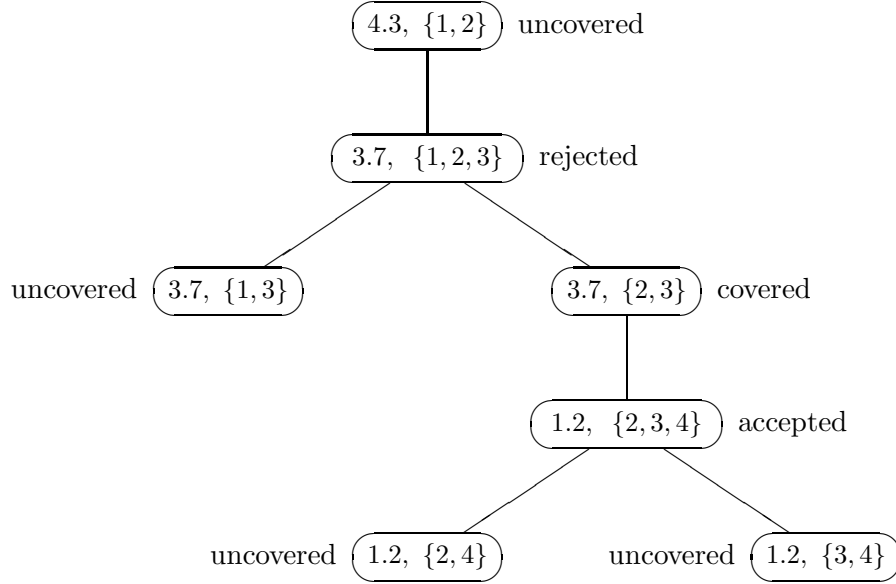
Figure 1: Example of a dependence tree.

jected and finally $(4.3, \{1, 2\})$ is uncovered. In the corresponding Poisson process, $\{2, 3, 4\}$ succeeds at time 1.2, thus $\{1, 2, 3\}$ fails at time 3.7, so no 3-set covering $\{1, 2\}$ is accepted by time 4.3. This example is consistent with the claim below.

Claim *Suppose the dependence tree of (τ, T) for some τ and T is defined. Then (τ, T) is covered if and only if T is covered by the Poisson process.*

Proof of claim. If T is covered in the Poisson process by a k-set K, then K succeeded at some time τ^*. Thus no k-set containing any of the t-sets covered by K was chosen before τ^*, and (τ^*, K) is accepted, hence (τ, T) is covered. Conversely, suppose that (τ, T) is covered in its dependence tree. Then it has an accepted child. It might have several accepted children, but since the tree is defined, the k-sets of these children can intersect only in T. The earliest such k-set succeeded, so it covers T. That establishes the claim.

6

3.2 The Idealized Tree

The process above is still difficult to analyze directly, so we will define for a fixed τ an idealized process and an idealized tree, analogous to the Poisson process and dependence tree. We will show that the idealized trees behave like the dependence trees, and then find the probability that the root of an idealized tree is covered.

The idealized tree has t-vertices and k-vertices, and consists at time τ just of a t-vertex root. Again, time goes backwards, from τ to 0. In the interval from τ_1 to $\tau_1 - \delta$ each t-vertex has probability asymptotic to δ of giving birth to a k-vertex, which then instantly gives birth to $D = \binom{k}{t} - 1$ new t-vertices. In a length δ interval each t-vertex has on average δD grandchildren (also t-vertices), so the expected number of t-vertices goes up by a factor of $1 + \delta D$. The expected number of t-vertices at time 0 is thus $(1 + \delta D)^{\tau/\delta} = e^{\tau D}(1 + O(\delta))$ as $\delta \to 0$, hence with probability 1 the idealized tree is finite. The notions of covered, uncovered, accepted, and rejected are defined on it as before.

We claim that the limit distribution of the dependence tree of (τ, T) as $v \to \infty$ is the distribution for the idealized tree. Consider a fixed idealized tree at time τ, and look at the dependence tree of (τ, T) from time τ_1 to $\tau_1 - \delta$ given that at τ_1 it matches the idealized tree. The number of t-sets in the tree is $O(e^{\tau D})$, with probability asymptotic to 1, so the number of k-sets that contain more than one t-set already in the tree is $O(e^{2\tau D}v^{k-t-1})$, and thus the probability of aborting (i.e., that some such k-set is chosen) is $O(\delta e^{2\tau D}v^{-1})$. Therefore the total chance of aborting throughout the length τ interval is $O(\tau e^{2\tau D}v^{-1}) = o(1)$ for $\tau \leq (\ln v)/(2 + \epsilon)D$, for any fixed $\epsilon > 0$.

For each T' in the tree, the number of k-sets that contain T' and no other t-set in the tree is asymptotically $\binom{v-t}{k-t}$, so T' has a (k-vertex) child with probability asymptotic to δ, as in the idealized version. Hence the two distributions are the same, as claimed.

Now we compute the probability $P(\tau)$ that the root of an idealized tree at time τ is uncovered. In the interval from τ to $\tau - \delta$ of an idealized process, a t-vertex either does or does not give birth, with probabilities asymptotic to δ and $1 - \delta$ as $\delta \to 0$. In the former case, a k-vertex child is accepted with probability $P(\tau - \delta)^D$, because each t-vertex grandchild has independent probability $P(\tau - \delta)$ of being uncovered at time $\tau - \delta$, and thus is rejected

7

with probability $1 - P(\tau - \delta)^D$. Hence

$$P(\tau) \sim \delta(1 - P(\tau - \delta)^D)P(\tau - \delta) + (1 - \delta)P(\tau - \delta).$$

So $P(\tau - \delta) - P(\tau) \sim \delta P(\tau - \delta)^{D+1}$, which leads to the differential equation $P(\tau)' = -P(\tau)^{D+1}$ with the initial condition $P(0) = 1$. The solution is

$$P(\tau) = (\tau D + 1)^{-1/D}.$$

In particular $\lim_{\tau \to \infty} P(\tau) = 0$, so the root of an idealized tree at time infinity is covered with probability asymptotic to 1. Therefore, at time infinity of the Poisson process, a given t-set is covered with probability asymptotic to 1, and Theorem 3 is established.

3.3 Estimating The Error Term

The proof above shows that the greedy covering is optimal, but we have not estimated the error term. We conclude this section by giving a weak estimate, along with some evidence for a stronger conjecture.

Consider the state of the algorithm at time $\tau = O(\log v)$. First, notice that at this time of the Poisson process, the expected number of k-sets chosen is $O(v^t \log v)$. Thus in the greedy algorithm it suffices to examine just $O(v^t \log v)$ random k-sets before cutting out to Step 4. It takes only $O(v^t \log v)$ expected time and $O(v^t)$ space to generate those k-sets (Brassard and Kannan [2]), so this early abort strategy dramatically speeds up the algorithm, at negligible cost to the density of the covering:

Corollary 1 *The early-abort greedy algorithm produces a covering with expected density $1 + o(1)$ in time $O(v^t \log v)$.*

Second, at time $\tau = (\ln v)/(2 + \epsilon)D$ for any fixed $\epsilon > 0$, the probability of a t-set being uncovered is $P(\tau) = O((\log v)^{-1/D})$. Thus:

Corollary 2 *The expected density of a covering produced by the random greedy algorithm is $1 + O((\log v)^{-1/D})$, where $D = \binom{k}{t} - 1$.*

This bound is pessimistic. Figure 2 gives log-log plots for several (k, t) pairs, based on 1000 random greedy coverings per (v, k, t) triple for $v \leq 50$, and 10^{6-k} such coverings for $v > 50$. The apparent asymptotic linearity of
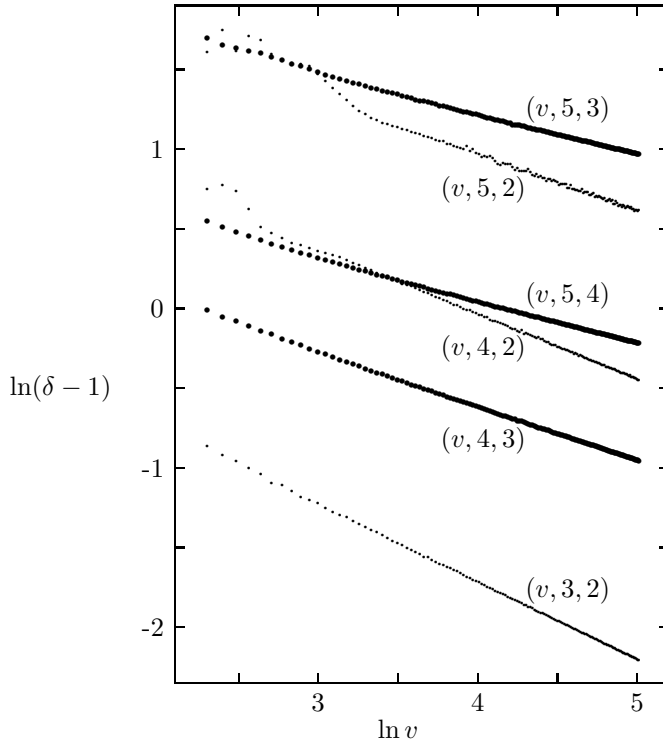
8

Figure 2: Average density δ of random greedy coverings.

the plots suggests that the expected density of a random greedy covering for k and t fixed is $1 + \Theta(v^{-\alpha})$, for some positive $\alpha = \alpha(k, t)$ as $v \to \infty$.

To estimate α for each of the curves in Figure 2, we used the tails of the curves ($100 \le v \le 150$) for a least-squares fit to a straight line. That gave us rough estimates for the slopes $-\alpha(k, t)$, as indicated in Table 1. Those values suggest:

Conjecture *The expected density of a covering produced by the random greedy algorithm is $1 + \Theta(v^{-(k-t)/D})$, where $D = \binom{k}{t} - 1$.*

The following argument, though far from a proof, supports the conjecture.

Heuristic argument. Let $\alpha = (k - t)/D$. The conjecture is equivalent to the statement that there are $\Theta(v^{t-\alpha})$ expected t-sets not covered by a random greedy packing. (The first three steps of Algorithm 1 constitute the random greedy packing algorithm.) So consider the t-uniform hypergraph whose

9

k	t	α	$(k-t)/D$
3	2	0.484	1/2
4	2	0.407	2/5
4	3	0.332	1/3
5	2	0.344	1/3
5	3	0.241	2/9
5	4	0.256	1/4

Table 1: Estimates for $\alpha(k, t)$

edges are the t-sets still uncovered during the packing algorithm. Assume that this hypergraph looks like a random hypergraph with the same number of edges, and assume that the packing algorithm has managed to leave just $c_1 v^{t-\alpha}(1 + o(1))$ edges in the hypergraph, for some positive constant c_1. We show that a positive fraction of these edges—that is, $\Theta(v^{t-\alpha})$ in all—are contained in no k-clique, and hence can never be covered by the packing; this provides the $\Omega(v^{-\alpha})$ lower bound of the conjecture's error term.

Under the stated assumptions, the probability p that a given edge exists in the hypergraph is asymptotic to $c_1 t! \, v^{-\alpha} = c_2 v^{-\alpha}$, and the probability, for a given edge in the hypergraph and a given k-set containing that edge, that the other $\binom{k}{t} - 1 = D$ edges on those k vertices also exist is p^D. Therefore the expected number of k-cliques that contain the given edge is asymptotic to $p^D v^{k-t}/(k-t)! = c_3 v^{-\alpha D} v^{k-t} = c_3$, a positive constant. But this number of k-cliques is Poisson distributed, so is zero with probability asymptotic to e^{-c_3}, also a positive constant, thus a positive fraction of the edges are contained in no k-clique, as claimed. The matching $O(v^{-\alpha})$ upper bound follows from similar reasoning, and that completes the argument. It, together with our empirical data, makes the conjecture quite compelling.

4   Induced Coverings and Rödl's Bound

While the greedy algorithm produces good coverings, it works in time and space $\Theta(v^k)$. These can be reduced to time $O(v^t \log v)$ and space $O(v^t)$ using the early abort strategy of Corollary 1, but for larger values of v, k, and t, the induced covering algorithm is more practical, because it is faster.

Theorem 4 *For fixed k and t the expected density of an induced covering is $1 + o(1)$.*

Proof. For Step 1 of Algorithm 2 choose $\ell = \frac{1}{9}v^{1-1/t}$, and choose the prime p such that
$$4\ell \ \leq \ \frac{v-t}{p} \ \leq \ 8\ell\,.$$
Such a prime exists by Bertrand's Postulate, which states that there is always a prime between n and $2n$. These choices ensure that $p^t > v$, and that the affine (p^t, p^{t-1}, t) covering by hyperplanes has density $1 + O(v^{-1/t})$.

By Corollary 2 the precomputed (ℓ', k, t) greedy coverings of Step 2 have expected density $1 + O((\log v)^{-1/D})$. So by running $O(\log \ell)$ trials per precomputed covering, we can ensure, with probability greater than, for example, $1 - 1/\ell$, that all precomputed coverings have density $1 + O((\log v)^{-1/D})$.

Now select the v-set V as a random subset of the points in the affine covering, and consider a fixed t-set T of V. There are, on average, $1 + O(v^{-1/t})$ hyperplanes containing T; let P be one of them. The size of the intersection of V and $P - T$ has a hypergeometric distribution from 0 to $v - t$ with mean $M = (v-t)(p^{t-1}-t)/(p^t-t)$. For $p \geq 5$ we have $(v-t)/2p < M < (v-t)/p$, thus $2\ell < M < 8\ell$ by our choice of p. So the probability that the size of the intersection is at most ℓ or at least 9ℓ is $O(e^{-c\ell})$ for some $c > 0$.

This intersection, together with T itself, is replaced in the induced covering by an (ℓ', k, t) covering. If $\ell < \ell' < 9\ell$, then this covering has density $1 + O((\log v)^{-1/D})$. If ℓ' is outside this range, the covering has density $\binom{k}{t}$, but the probability of this event is $O(e^{-c\ell})$, so the total expected number of k-sets containing T coming from a given hyperplane containing T is $1 + O((\log v)^{-1/D})$, and the total expected number coming from all such hyperplanes is $(1 + O((\log v)^{-1/D}))(1 + O(v^{-1/t})) = 1 + O((\log v)^{-1/D})$. Thus the expected density of the induced covering is $1 + O((\log v)^{-1/D})$. ∎

Corollary 3 *The induced covering algorithm runs in time and space $O(v^t)$.*

Proof. By Corollary 1, precomputation takes time $O(\ell^{t+1} \log^2 \ell)$, which is $O(v^t)$ by our choice of ℓ. The number of hyperplanes is $O(p^t) = O(v)$ by our choice of p, so the time to compute the affine geometry is $O(v^2) = O(v^t)$. For each hyperplane, the work to find the intersection and convert it into an (ℓ', k, t) covering will vary, but the time per block is constant. Hence the

11

total time and space of the algorithm is dominated by the size of the (v, k, t) covering, which is also $O(v^t)$. ∎

Corollary 4 *The induced covering has expected density* $1 + O((\log v)^{-1/D})$.

Furthermore, if, as we conjecture, the greedy covering has expected density $1 + O(v^{-(k-t)/D})$, then the expected density of the induced covering improves to $1 + O(v^{-(k-t)/D}) + O(v^{-1/t}) = 1 + O(v^{-(k-t)/D})$.

The best way to use the induced covering algorithm in practice is to first find or make a large table of good coverings with small parameters using many different methods, and then use these for the (ℓ', k, t) coverings. We used that strategy to produce the induced coverings of our earlier paper [4].

References

[1] Noga Alon and Joel H. Spencer. *The Probabilistic Method*, section 9.4. Wiley, 1992.

[2] Gilles Brassard and Sampath Kannan. The generation of random permutations on the fly. *Information Processing Letters*, 28:207–212, 1988.

[3] P. Erdős and H. Hanani. On a limit theorem in combinatorical analysis. *Publicationes Mathematicae Debrecen*, 10:10–13, 1963.

[4] Daniel M. Gordon, Greg Kuperberg, and Oren Patashnik. New constructions for covering designs. *Journal of Combinatorial Designs*, 3:269–284, 1995.

[5] Vojtěch Rödl. On a packing and covering problem. *European Journal of Combinatorics*, 5:69–78, 1985.

[6] Vojtěch Rödl and Luboš Thoma. Asymptotic packing and the random greedy algorithm. *Random Structures and Algorithms*, to appear.

[7] J. Schönheim. On coverings. *Pacific Journal of Mathematics*, 14:1405–1411, 1964.

[8] Joel Spencer. Asymptotic packing via a branching process. *Random Structures and Algorithms*, 7:167–172, 1995.