A construction of almost Steiner systems

Asaf Ferber* Rani Hod† Michael Krivelevich‡ Benny Sudakov§

September 25, 2013

Abstract

Let n, k, and t be integers satisfying $n > k > t \geq 2$. A Steiner system with parameters t, k, and n is a k-uniform hypergraph on n vertices in which every set of t distinct vertices is contained in exactly one edge. An outstanding problem in Design Theory is to determine whether a nontrivial Steiner system exists for $t \geq 6$.

In this note we prove that for every $k > t \geq 2$ and sufficiently large n, there exists an almost Steiner system with parameters t, k, and n; that is, there exists a k-uniform hypergraph on n vertices such that every set of t distinct vertices is covered by either one or two edges.

1 Introduction

Let n, k, t, and λ be positive integers satisfying $n > k > t \geq 2$. A t-(n, k, λ)-*design* is a k-uniform hypergraph $\mathcal{H} = (X, \mathcal{F})$ on n vertices with the following property: every t-set of vertices $A \subset X$ is contained in exactly λ edges $F \in \mathcal{F}$. The special case $\lambda = 1$ is known as a *Steiner system* with parameters t, k, and n, named after Jakob Steiner who pondered the existence of such systems in 1853. Steiner systems, t-designs[1] and other combinatorial designs turn out to be useful in a multitude of applications, e.g., in coding theory, storage systems design, and wireless communication. For a survey of the subject, the reader is referred to [2].

A counting argument shows that a Steiner triple system — that is, a 2-$(n, 3, 1)$-design — can only exist when $n \equiv 1$ or $n \equiv 3 \pmod 6$. For every such n, this is achieved via constructions based on symmetric idempotent quasigroups. Geometric constructions over finite fields give rise to some further infinite families of Steiner systems with $t = 2$ and $t = 3$. For instance, for a prime power q and an integer $m \geq 2$, affine geometries yield 2-$(q^m, q, 1)$-designs, projective geometries yield 2-$\big(q^m + \cdots + q^2 + q + 1, q + 1, 1\big)$-designs and spherical geometries yield 3-$(q^m + 1, q, 1)$-designs.

For $t = 4$ and $t = 5$, only finitely many nontrivial constructions of Steiner systems are known; for $t \geq 6$, no constructions are known at all.

Before stating our result, let us extend the definition of t-designs as follows. Let n, k, and t be positive integers satisfying $n > k > t \geq 2$ and let Λ be a set of positive integers. A t-(n, k, Λ)-design is a k-uniform hypergraph $\mathcal{H} = (X, \mathcal{F})$ on n vertices with the following property: for every t-set of vertices $A \subset X$, the number of edges $F \in \mathcal{F}$ that contain A belongs to Λ. Clearly, when $\Lambda = \{\lambda\}$ is a singleton, a t-$(n, k, \{\lambda\})$-design coincides with a t-(n, k, λ)-design as defined above.

Not able to construct Steiner systems for large t, Erdős and Hanani [3] aimed for large partial Steiner systems; that is, t-$(n, k, \{0, 1\})$-designs with as many edges as possible. Since a Steiner

*School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: ferberas@tau.ac.il.

†School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: rani.hod@cs.tau.ac.il. Research supported by an ERC advanced grant.

‡School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: krivelev@tau.ac.il. Research supported in part by USA-Israel BSF Grant 2010115 and by grant 912/12 from the Israel Science Foundation.

§Department of Mathematics, ETH, 8092 Zurich, Switzerland and Department of Mathematics, UCLA, Los Angeles, CA 90095. Email: bsudakov@math.ucla.edu. Research supported in part by NSF grant DMS-1101185 and by a USA-Israel BSF grant.

[1]That is, t-(n, k, λ)-designs for some parameters n,k, and λ.

system has exactly $\binom{n}{t}/\binom{k}{t}$ edges, they conjectured the existence of partial Steiner systems with $(1-o(1))\binom{n}{t}/\binom{k}{t}$ edges. This was first proved by Rödl [8] in 1985, with further refinements [4, 5, 6] of the $o(1)$ term, as stated in the following theorem:

Theorem 1 (Rödl)**.** *Let k and t be integers such that $k > t \geq 2$. Then there exists a partial Steiner system with parameters t, k, and n covering all but $o\left(n^t\right)$ of the t-sets.*

Theorem 1 can also be rephrased in terms of a covering rather than a packing; that is, it asserts the existence of a system with $(1+o(1))\binom{n}{t}/\binom{k}{t}$ edges such that every t-set is covered at least once (see, e.g., [1, page 56]). Nevertheless, some t-sets might be covered multiple times (perhaps even $\omega(1)$ times). It is therefore natural to ask for t-$(n,k,\{1,\ldots,r\})$-designs, where r is as small as possible. The main aim of this short note is to show how to extend Theorem 1 to cover all t-sets at least once but at most twice.

Theorem 2. *Let k and t be integers such that $k > t \geq 2$. Then, for sufficiently large n, there exists a t-$(n,k,\{1,2\})$-design.*

Our proof actually gives a stronger result: there exists a t-$(n,k,\{1,2\})$-design with $(1+o(1))\binom{n}{t}/\binom{k}{t}$ edges.

2 Preliminaries

In this section we present results needed for the proof of Theorem 2.

Given a t-$(n,k,\{0,1\})$-design $\mathcal{H} = (X,\mathcal{F})$, we define the *leave hypergraph* $(X,\mathcal{L}_{\mathcal{H}})$ to be the t-uniform hypergraph whose edges are the t-sets $A \subset X$ not covered by any edge $F \in \mathcal{F}$.

Following closely the proof of Theorem 1 appearing in [4], we recover an extended form of the theorem, which is a key ingredient in the proof of our main result.

Theorem 3. *Let k and t be integers such that $k > t \geq 2$. There exists a constant $\varepsilon = \varepsilon(k,t) > 0$ such that for sufficiently large n, there exists a partial Steiner system $\mathcal{H} = (X,\mathcal{F})$ with parameters t, k, and n satisfying the following property:*

(♣) *For every $0 \leq \ell < t$, every set $X' \subset X$ of size $|X'| = \ell$ is contained in $O\left(n^{t-\ell-\varepsilon}\right)$ edges of the leave hypergraph.*

We also make use of the following probabilistic tool.

Talagrand's inequality. In its general form, Talagrand's inequality is an isoperimetric-type inequality for product probability spaces. We use the following formulation from [7, pages 232–233], suitable for showing that a random variable in a product space is unlikely to overshoot its expectation under two conditions:

Theorem 4 (Talagrand)**.** *Let $Z \geq 0$ be a non-trivial random variable, which is determined by n independent trials $T_1,\ldots,T_n$. Let $c > 0$ and suppose that the following properties hold:*

i. (c-Lipschitz) changing the outcome of one trial can affect Z by at most c, and

ii. (Certifiable) for any s, if $Z \geq s$ then there is a set of at most s trials whose outcomes certify that $Z \geq s$.

Then $\Pr[Z > t] < 2\exp\left(-t/16c^2\right)$ *for any* $t \geq 2\mathbb{E}[Z] + 80c\sqrt{\mathbb{E}[Z]}$.

3 Proof of the main result

In this section we prove Theorem 2.

3.1 Outline

The construction is done in two phases:

I. Apply Theorem 3 to get a t-$(n, k, \{0, 1\})$-design $\mathcal{H} = (X, \mathcal{F})$ with property (♣) with respect to some $0 < \varepsilon < 1$.

II. Build another t-$(n, k, \{0, 1\})$-design $\mathcal{H}' = (X, \mathcal{F}')$ that covers the uncovered t-sets $\mathcal{L}_{\mathcal{H}}$.

Combining both designs, we get that every t-set is covered at least once but no more than twice; namely $(X, \mathcal{F} \cup \mathcal{F}')$ is a t-$(n, k, \{1, 2\})$-design, as required.

We now describe how to build $\mathcal{H}'$. For a set $A \subset X$, denote by $\mathcal{T}_A = \{C \subseteq X : |C| = k \text{ and } A \subseteq C\}$ the family of possible continuations of A to a subset of X of cardinality k. Note that $\mathcal{T}_A = \varnothing$ when $|A| > k$.

Consider the leave hypergraph $(X, \mathcal{L}_{\mathcal{H}})$. Our goal is to choose, for every uncovered t-set $A \in \mathcal{L}_{\mathcal{H}}$, a k-set $A' \in \mathcal{T}_A$ such that $|A' \cap B'| < t$ for every two distinct $A, B \in \mathcal{L}_{\mathcal{H}}$. This ensures that the obtained hypergraph $\mathcal{H}' = (X, \{A' : A \in \mathcal{L}_{\mathcal{H}}\})$ is indeed a t-$(n, k, \{0, 1\})$-design.

To this aim, for every $A \in \mathcal{L}_{\mathcal{H}}$ we introduce intermediate lists $\mathcal{R}_A \subseteq \mathcal{S}_A \subseteq \mathcal{T}_A$ that will help us control the cardinalities of pairwise intersections when choosing $A' \in \mathcal{R}_A$. First note that we surely cannot afford to consider continuations that fully contain some other $B \in \mathcal{L}_{\mathcal{H}}$, so we restrict ourselves to the list

$$\mathcal{S}_A = \mathcal{T}_A \setminus \bigcup \{\mathcal{T}_B : B \in \mathcal{L}_H \setminus \{A\}\} = \{C \in \mathcal{T}_A : B \nsubseteq C \text{ for all } B \in \mathcal{L}_H, B \neq A\}.$$

Note that, by definition, the lists $\mathcal{S}_A$ for different A are disjoint. Next, choose a much smaller sub-list $\mathcal{R}_A \subseteq \mathcal{S}_A$ by picking each $C \in \mathcal{S}_A$ to $\mathcal{R}_A$ independently at random with probability $p = n^{t-k+\varepsilon/2}$ (we can of course assume here and later that $\varepsilon < 1 \leq k - t$, and thus $0 < p < 1$). Finally, select $A' \in \mathcal{R}_A$ that has no intersection of size at least t with any $C \in \mathcal{R}_B$ for any other $B \in \mathcal{L}_{\mathcal{H}}$. If there is such a choice for every $A \in \mathcal{L}_{\mathcal{H}}$, we get $|A' \cap B'| < t$ for distinct $A, B \in \mathcal{L}_{\mathcal{H}}$, as requested.

3.2 Details

We start by showing that the lists $\mathcal{S}_A$ are large enough.

Claim 5. For every $A \in \mathcal{L}_{\mathcal{H}}$ we have $|\mathcal{S}_A| = \Theta\left(n^{k-t}\right)$.

Proof. Fix $A \in \mathcal{L}_{\mathcal{H}}$. Obviously $|\mathcal{T}_A| = \binom{n-t}{k-t} = \Theta\left(n^{k-t}\right)$. Since $\mathcal{S}_A \subseteq \mathcal{T}_A$, it suffices to show that $|\mathcal{T}_A \setminus \mathcal{S}_A| = o\left(n^{k-t}\right)$.

Writing $\mathcal{L}_H \setminus \{A\}$ as the disjoint union $\bigcup_{\ell=0}^{t-1} \mathcal{B}_\ell$, where $\mathcal{B}_\ell = \{B \in \mathcal{L}_H : |A \cap B| = \ell\}$, we have

$$\begin{aligned} \mathcal{T}_A \setminus \mathcal{S}_A &= \{C \in \mathcal{T}_A : \exists B \in \mathcal{L}_H \setminus \{A\} \text{ such that } C \in \mathcal{T}_B\} \\ &= \bigcup_{\ell=0}^{t-1} \{C \in \mathcal{T}_A : \exists B \in \mathcal{B}_\ell \text{ such that } C \in \mathcal{T}_B\} \\ &= \bigcup_{\ell=0}^{t-1} \bigcup_{B \in \mathcal{B}_\ell} \mathcal{T}_{A \cup B}. \end{aligned}$$

Note that for all $0 \leq \ell < t$ and for all $B \in \mathcal{B}_\ell$, $|A \cup B| = 2t - \ell$ and thus $|\mathcal{T}_{A \cup B}| = \binom{n-2t+\ell}{k-2t+\ell} \leq n^{k-2t+\ell}$. Moreover, $|\mathcal{B}_\ell| = \binom{t}{\ell} \cdot O\left(n^{t-\ell-\varepsilon}\right) = O\left(n^{t-\ell-\varepsilon}\right)$ by Property (♣). Thus,

$$|\mathcal{T}_A \setminus \mathcal{S}_A| \leq \sum_{\ell=0}^{t-1} |\mathcal{B}_\ell| \, n^{k-2t+\ell} = \ell \cdot O\left(n^{k-t-\varepsilon}\right) = o\left(n^{k-t}\right),$$

establishing the claim. □

Recall that the sub-list $\mathcal{R}_A \subseteq \mathcal{S}_A$ was obtained by picking each $C \in \mathcal{S}_A$ to $\mathcal{R}_A$ independently at random with probability $p = n^{t-k+\varepsilon/2}$. The next claim shows that $\mathcal{R}_A$ typically contains many k-sets whose pairwise intersections are exactly A. This will be used in the proof of Claim 7.

Claim 6. Almost surely (i.e., with probability tending to 1 as n tends to infinity), for every $A \in \mathcal{L}_H$, the family $\mathcal{R}_A$ contains a subset $\mathcal{Q}_A \subseteq \mathcal{R}_A$ of size $\Theta\left(n^{\varepsilon/3}\right)$ such that $C_1 \cap C_2 = A$ for every two distinct $C_1, C_2 \in \mathcal{Q}_A$.

Proof. Fix $A \in \mathcal{L}_H$. Construct $\mathcal{Q}_A$ greedily as follows: start with $\mathcal{Q}_A = \varnothing$; as long as $|\mathcal{Q}_A| < n^{\varepsilon/3}$ and there exists $C \in \mathcal{R}_A \setminus \mathcal{Q}_A$ such that $C \cap C' = A$ for all $C' \in \mathcal{Q}_A$, add C to $\mathcal{Q}_A$. It suffices to show that this process continues $\lfloor n^{\varepsilon/3} \rfloor$ steps.

If the process halts after $s < \lfloor n^{\varepsilon/3} \rfloor$ steps, then every k-tuple of $\mathcal{R}_A$ intersects one of the s previously chosen sets in some vertex outside A. This means that there exists a subset $X_A \subset X$ of cardinality $|X_A| = kn^{\varepsilon/3}$ (X_A contains the union of these s previously picked sets) such that none of the edges C of $\mathcal{S}_A$ satisfying $C \cap X_A = A$ is chosen into $\mathcal{R}_A$. For bounding the number of such edges in $\mathcal{S}_A$ from bellow, we need to subtract from $|\mathcal{S}_A|$ the number of edges $C \in \mathcal{T}_A$ with $C \cap X_A \neq A$. The latter can be bounded (from above) by $\sum_{i=1}^{k-t} \binom{|X_A|}{i}\binom{n-|X_A|}{k-t-i}$ (choose $1 \leq i \leq k-t$ vertices from X_A, other than A, to be in C, and then choose the remaining $k-t-i$ vertices from $X \setminus X_A$). Since $|\mathcal{S}_A| = \Theta\left(n^{k-t}\right)$, and since

$$\sum_{i=1}^{k-t} \binom{|X_A|}{i}\binom{n-|X_A|}{k-t-i} \leq \sum_{i=1}^{k-t} \Theta(n^{i\varepsilon/3}) \cdot n^{k-t-i} = o(n^{k-t}),$$

we obtain that the number of such edges in $\mathcal{S}_A$ is at least $|\mathcal{S}_A| - \sum_{i=1}^{k-t} \binom{|X_A|}{i}\binom{n-|X_A|}{k-t-i} = \Theta\left(n^{k-t}\right)$. It thus follows that the probability of the latter event to happen for a given A is at most

$$\binom{n}{kn^{\varepsilon/3}} (1-p)^{\Theta\left(n^{k-t}\right)}$$

(choose X_A first, and then require all edges of $\mathcal{S}_A$ intersecting X_A only at A to be absent from $\mathcal{R}_A$). The above estimate is clearly at most

$$n^{n^{\varepsilon/3}} \cdot e^{-\Theta\left(pn^{k-t}\right)} = \exp\left\{n^{\varepsilon/3} \ln n - \Theta\left(n^{\varepsilon/2}\right)\right\} < \exp\left\{-n^{\varepsilon/3}\right\}.$$

Taking the union bound over all ($\leq \binom{n}{t}$) choices of A establishes the claim. □

The last step is to select a well-behaved set $A' \in \mathcal{R}_A$. The next claim shows this is indeed possible.

Claim 7. Almost surely for every $A \in \mathcal{L}_H$ we can select $A' \in \mathcal{R}_A$ such that $|A' \cap C| < t$ for all $C \in \bigcup\{\mathcal{R}_B : B \in \mathcal{L}_H \setminus \{A\}\}$.

Proof. Fix $A \in \mathcal{L}_H$ and fix all the random choices which determine the list $\mathcal{R}_A$ such that it satisfies Claim 6. Let $\mathcal{Q}_A \subseteq \mathcal{R}_A$ be as provided by Claim 6 and let $\mathcal{R} = \bigcup\{\mathcal{R}_B : B \in \mathcal{L}_H \setminus \{A\}\}$ be the random family of all obstacle sets. Define the random variable Z to be the number of sets $A' \in \mathcal{Q}_A$ for which $|A' \cap C| \geq t$ for some $C \in \mathcal{R}$. Since $\mathcal{S}_A$ is disjoint from $\mathcal{S} = \bigcup\{\mathcal{S}_B : B \in \mathcal{L}_H \setminus \{A\}\}$, we can view $\mathcal{R}$ as a random subset of $\mathcal{S}$, with each element selected to $\mathcal{R}$ independently with probability $p = n^{t-k+\varepsilon/2}$. Thus Z is determined by $|\mathcal{S}|$ independent trials. We wish to show that Z is not too large via Theorem 4; for this, Z has to satisfy the two conditions therein.

1. If $C \in \mathcal{S}$ satisfies $|A' \cap C| \geq t$ for some $A' \in \mathcal{Q}_A$ then $C \setminus A$ must intersect $A' \setminus A$ (since $A \not\subseteq C$). However, the $(k-t)$-sets $\{A' \setminus A : A' \in \mathcal{Q}_A\}$ are pairwise disjoint (by the definition of $\mathcal{Q}_A$) so each C cannot rule out more than k different sets $A' \in \mathcal{Q}_A$. Thus Z is k-Lipschitz.

2. Assume that $Z \geq s$. Then, by definition, there exist distinct sets $A'_1, \ldots, A'_s \in \mathcal{Q}_A$ and (not necessarily distinct) sets $C_1, \ldots, C_s \in \mathcal{R}$ such that $|A'_i \cap C_i| \geq t$ for $i = 1, \ldots, s$. These are at most s trials whose outcomes ensure that $Z \geq s$; i.e., Z is certifiable.

Let us now calculate $\mathbb{E}[Z]$. Fix $A' \in \mathcal{Q}_A$ and let $Z_{A'}$ be the indicator random variable of the event $E_{A'} = \{\exists C \in \mathcal{R} : |A' \cap C| \geq t\}$. The only set in $\mathcal{L}_H$ fully contained in A' is A, so we can write $\mathcal{L}_H \setminus \{A\}$ as the disjoint union $\bigcup_{\ell=0}^{t-1} \mathcal{B}'_\ell$, where $\mathcal{B}'_\ell = \{B \in \mathcal{L}_H : |A' \cap B| = \ell\}$. For any $0 \leq \ell < t$ and $B \in \mathcal{B}'_\ell$, the number of bad sets (i.e., sets that will trigger $E_{A'}$) in $\mathcal{S}_B$ is

$$\begin{aligned}\left|\left\{C \in \mathcal{S}_B : \left|A' \cap C\right| \geq t\right\}\right| &\leq \left|\left\{C \in \mathcal{T}_B : \left|A' \cap C\right| \geq t\right\}\right| \\ &= \left|\left\{C \in \mathcal{T}_B : \left|\left(A' \cap C\right) \setminus B\right| \geq t - \ell\right\}\right| \\ &= \sum_{i=t-\ell}^{k-t} \binom{k-\ell}{i} \binom{n-k-t+\ell}{k-t-i} = O\left(n^{k-2t+\ell}\right),\end{aligned}$$

since C contains B, $|B| = t$, together with $i \geq t - \ell$ elements from $A' \setminus B$ and the rest from $X \setminus (A' \cup B)$. Each such bad set ends up in $\mathcal{R}_B$ with probability $p = n^{\varepsilon/2+t-k}$, so the expected number of bad sets in $\mathcal{R}_B$ is $O\left(n^{\varepsilon/2-t+\ell}\right)$. By Property (♣) we have $|\mathcal{B}'_\ell| = O\left(n^{t-\ell-\varepsilon}\right)$ and thus the total expected number of bad sets in $\mathcal{R}$ is $\ell \cdot O\left(n^{\varepsilon/2-\varepsilon}\right) = O\left(n^{-\varepsilon/2}\right)$. By Markov's inequality we have

$$\Pr\left[E_{A'}\right] = O\left(n^{-\varepsilon/2}\right).$$

Now, Z is a sum of $|\mathcal{Q}_A| = \Theta\left(n^{\varepsilon/3}\right)$ random variables $Z_{A'}$ and thus

$$\mathbb{E}[Z] = |\mathcal{Q}_A| \cdot \mathbb{E}\left[Z_{A'}\right] = |\mathcal{Q}_A| \cdot \Pr\left[E_{A'}\right] = O\left(n^{-\varepsilon/6}\right) = o(1).$$

Applying Theorem 4 with $t = |\mathcal{Q}_A| = \Theta\left(n^{\varepsilon/3}\right)$ and $c = k = O(1)$, we get that

$$\Pr\left[\text{all } A' \in \mathcal{R}_A \text{ are ruled out}\right] \leq \Pr\left[Z \geq |\mathcal{Q}_A|\right] < 2\exp\left(-\Omega\left(n^{\varepsilon/3}\right)\right).$$

Taking the union bound over all $|\mathcal{L}_H| \leq \binom{n}{t}$ choices of $A \in \mathcal{L}_H$ establishes the claim. □

References

[1] N. Alon and J. H. Spencer, **The Probabilistic Method**, 3rd edition, Wiley, New York, 2008.

[2] C. J. Colbourn and J. H. Dinitz, **Handbook of Combinatorial Designs**, 2nd edition, Chapman & Hall/ CRC, Boca Raton, 2006.

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

[4] D. A. Grable, More-than-nearly perfect packings and partial designs. *Combinatorica* **19**(2):221–239 (1999).

[5] J.-H. Kim, Nearly optimal partial Steiner systems. *Electronic Notes in Discrete Mathematics* **7**:74–77 (2001).

[6] A. V. Kostochka and V. Rödl, Partial Steiner systems and matchings in hypergraphs. *Random Structures & Algorithms* **13**(3–4):335–347 (1998).

[7] M. Molloy and B. Reed, **Graph Colouring and the Probabilistic Method**, Springer, Berlin, 2002.

[8] V. Rödl, On a packing and covering problem. *European Journal of Combinatorics* **6**:69–78 (1985).