Séminaire BOURBAKI juin 2015
67ème année, 2014-2015, n^{o} 1100

DESIGNS EXIST!
[after Peter Keevash]

by **Gil KALAI**

INTRODUCTION

A set S of q-subsets of an n-set X is a *design* with parameters (n, q, r, λ) if every r-subset of X belongs to exactly λ elements of S. (The elements of S are called *blocks* and designs are also referred to as *block designs*.) There are some necessary *divisibility conditions* for the existence of such a design, namely that

$$\binom{q-i}{r-i} \text{ divides } \lambda\binom{n-i}{r-i}, \quad 0 \le i \le r-1. \tag{1}$$

To see that the divisibility conditions are necessary, fix any i-subset I of X and consider the sets in S that contain I.

The following result was conjectured in the 19th century and was recently proved by Peter Keevash.

THEOREM 0.1 ([Kee14]). — *For fixed q, r, and λ, there exist $n_0(q, r, \lambda)$ such that if $n > n_0(q, r, \lambda)$ satisfies the divisibility conditions (1) then a design with parameters (n, q, r, λ) exists.*

In other words, for fixed q, r, and λ, the divisibility conditions are sufficient apart from a finite number of exceptional values of n.

A case of special interest is when $\lambda = 1$. A design of parameters $(n, q, r, 1)$ is called a *Steiner system* of parameters (n, q, r). The question if Steiner systems of given parameters exist goes back to works of Plücker, Kirkman, and Steiner. Until Keevash's result not a single Steiner system for $r > 5$ was known to exist.

The presentation of Keevash's work in this paper is based on Keevash's original paper [Kee14], his Jerusalem videotaped lectures [KeeV], and his subsequent paper [Kee15]. It is also based on lecture notes and personal explanations by Jeff Kahn.

1. REGULARITY, SYMMETRY AND RANDOMNESS

1.1. Between regularity and symmetry

An object is "regular" if it looks locally the same (for a certain notion of "locality,"), and it is "symmetric" if it admits a transitive group action (on its "local" pieces). The interplay between regularity and symmetry is of interest in several parts of mathematics. For example, a regular graph is a graph where every vertex is adjacent to the same number of neighbors, and every graph whose group of automorphisms act transitively on its vertices is regular. Regular graphs need not be symmetric (most of them have no non-trivial automorphisms), but there are still various connections between general regular graphs and symmetry. Another example: manifolds are regular objects and Lie groups are symmetric types of manifolds; also here there are rich connections between regularity and symmetry.

Designs are regular objects. (The local pieces can be regarded as the r-subsets of the ground set.) You can get them from groups acting transitively on r-sets.

PROPOSITION 1.1. — *Let Γ be a r-transitive permutation group. Then the orbit of a set of size q is a block design so that every set of size r belongs to the same number of blocks.*

However, it follows from the classification of finite simple groups that

THEOREM 1.2. — *Let Γ be a r-transitive permutation group, $r > 5$, then G is A_n or S_n.*

1.2. The probabilistic method and quasi-randomness

The proof of the existence of designs is probabilistic. In order to prove the existence of objects of some kind satisfying a property **P**, one proves that for a suitable probability distribution on all objects of this kind there is a positive probability for property **P** to hold. The probabilistic method is of central importance in combinatorics (and other areas) [AS00]. Keevash defines a complicated combinatorial process with random ingredients for building a design, and shows that with positive probability it leads to the desired construction.

Quasi-randomness refers to deterministic properties of mathematical structures which allow them to behave (for certain restricted purposes) "as if they were random." [(1)] Quasi random properties of primes are, of course, of much importance. In graph theory, a sequence of graphs G_n (where G_n has n vertices) is quasi-random if the number of induced 4-cycles C_4 is $\frac{1}{64}\binom{n}{4}(1 + o(1))$. This important notion was introduced and studied in important papers by Thomasson [Tho85] and by Chung, Graham, and Wilson [CGW89] (with extensions to hypergraphs by Chung and Graham [CG89]). A sequence

1. There are even cases that quasi-randomness of some kind can be attributed to arbitrary structures as is the case in Szemerédi regularity lemma which describes quasi-random structure on arbitrary graphs.

of subsets $A_n \subset \{1, 2, \ldots, n\}$ can be called quasi-random (for certain purposes), if the maximum Fourier coefficients of 1_{A_n} tends to zero with n.

Quasi randomness is a central concept in modern combinatorics and two examples are its usage in Szemerédi's theorem and many of its extensions, and in the study of expanders and Ramanujan graphs. The first move made by Keevash is to vastly extend the situation in discussing very general decomposition of quasirandom hypergraphs.

2. KEEVASH'S RESULTS: EDGE-DECOMPOSITION OF QUASI-RANDOM HYPERGRAPHS AND THE NUMBER OF DESIGNS

Keevash vastly extended theorem 0.1 to describe sufficient conditions for general H-decompositions of quasirandom hypergraphs. We will describe in this section the general results closely following [Kee14, Section 1.1].

A hypergraph G consists of a vertex set $V(G)$ and an edge set $E(G)$, of subsets of $V(G)$. If every edge has size r we say that G is an r-uniform hypergraph. For $I \subset V(G)$, the *link* $G(I)$ of R is the $(r - |I|)$-uniform hypergraph

$$G(I) = \{S \subset V(G) \backslash I : I \cup S \in E(G)\}.$$

For an r-uniform hypergraph H, an H-decomposition of G is a partition of $E(G)$ into sub-hypergraphs isomorphic to H. Let K_r^q be the complete r-uniform hypergraph on q vertices, namely, an r-uniform hypergraph whose edges are all r-subsets of a set of size q. A Steiner system with parameters (n, q, r) is equivalent to a K_r^q-decomposition of K_r^n

The next definition generalizes the necessary divisibility conditions described above. Suppose G is an r-uniform hypergraph. We say that G is K_q^r-*divisible* if $\binom{q-i}{r-i}$ divides $|G(I)|$ for any i-set $I \subset V(G)$, for all $0 \leq i < r$.

We come now to a crucial notion of quasirandomness. Suppose G is an r-uniform hypergraph on n vertices. We say that G is (c, h)-*typical* if there is some $p > 0$ such that for any set A of $(r-1)$-subsets of $V(G)$ with $|A| \leq h$ we have

$$(1 - c)p^{|A|}n \leq |\cap_{S \in A} G(S)| \leq (1 + c)p^{|A|}n. \tag{2}$$

Keevash's main theorem (still in a somewhat simplified form) is

THEOREM 2.1. — *Let $1/n \ll c \ll d$, $r \leq q \ll h$. Suppose that G is a K_q^r divisible (c, h)-typical r-uniform hypergraph on n vertices with $|G| > dn^r$. Then G has a K_q^r-decomposition.*

Theorem 0.1 follows by applying Theorem 2.1 with $G = K_n^r$. Thus for fixed values of q and r and large values of n, the divisibility conditions are sufficient for the existence of Steiner systems. The next theorem, also from [Kee14], gives a good asymptotic estimate for the number of Steiner systems.

THEOREM 2.2. — *The number $S(n,q,r)$ of Steiner systems with parameters (n,q,r) (where n satisfies the divisibility conditions) satisfies*

$$\log S(n,q,r) = (1+o(1))\binom{q}{r}^{-1}\binom{n}{r}(q-r)\log n. \tag{3}$$

Theorem 2.2 also follows from a more general result for counting decomposition of (c,h)-typical hypergraphs.

For Steiner triple systems tighter asymptotic estimates were known [Wil74], and it is desirable to give asymptotic estimates for $S(n,q,r)$ itself. In [Kee15] Keevash sketches a proof of much tighter asymptotic estimates for $S(n,q,r)$. It is based on a finer application of his method for getting better lower bounds, and on upper bounds obtained by the entropy method following the work of Linial and Luria [LL15+].

3. SOME HISTORY, TWO LANDMARKS, AND A RELAXATION

We will give now a brief description of the history of designs based on [Wil03]. Our discussion is centered around (and hence biased toward) general existence theorems. It is perhaps right to start the history with Kirkman. The earliest general existence result is given in Kirkman's 1847 paper where he constructed a Steiner triple system (as called today) for every n which is 1 or 3 modulo 6. The prehistory is even earlier. Plücker encountered Steiner triple systems in 1830 while working on plane cubic curves. Woolhouse asked about the number of Steiner triple systems in the "Lady's and Gentleman's Diary" edited by him in 1844 (and again in 1846). Combinatorial designs are closely related to mathematical constructions that were studied since ancient times like Latin squares and Greco-Latin squares. Steiner, unaware of Kirkman's work, posed the question on the existence of Steiner triple systems in 1853 (leading to a solution by Reiss published in 1859).

It is common to start the story of designs with Kirkman's schoolgirl problem. Kirkman proposed in 1850, again in the "Lady's and Gentleman's Diary," his famous problem on "fifteen young ladies," with solutions by himself, Cayley, Anstice, Pierce, and others. There are fifteen schoolgirls who take their daily walks in rows of threes. It is required to arrange them daily for a week so that no two schoolgirls will walk in the same row more than once.

Kirkman's question can be asked in greater generality for every $n = 3 (\mathrm{mod}\, 6)$ and a partial solution was offered in 1852 by Spottiswoode. The general question was settled independently by Lu Jiaxi (a schoolteacher from China) in the mid '60s and by Ray-Chaudhuri and Wilson in 1972. Sylvester asked (as reported by a 1850 paper by Cayley) if we can divide all $\binom{15}{3}$ triples into 13 different solutions of Kirkman's problem and this was settled by Denniston in 1974. Sylvester's question for general n is still open.

FIGURE 1. Kirkman's problem.

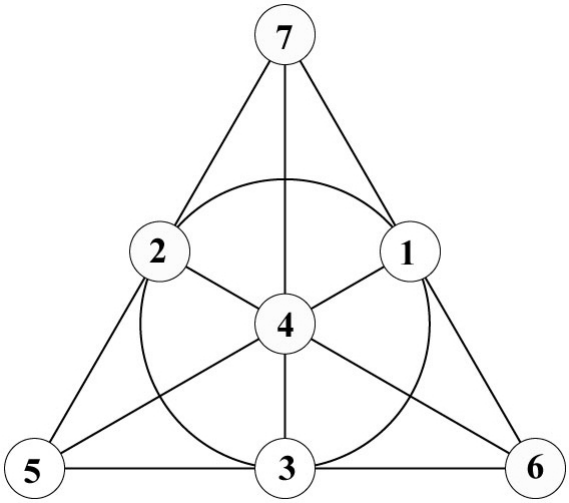

FIGURE 2. Fano plane.

3.1. Designs, finite geometries, statistics, groups, and codes

Finite projective planes. — A Steiner system of type $(q^2+q+1, q+1, 2)$ is equivalent to a finite projective plane of order q. The ground set in the set of points and the blocks correspond to lines. Being a design means that every two points belong to a single line. Finite projective planes (and spaces of higher dimensions) of prime orders were constructed by Fano in 1892. A Fano plane is a finite projective plane over a field with two elements.

FIGURE 3. Statistical experiment based on a Latin square (constructed by R. Baily, picture by S. Welham).

Designs and statistics. — In the first half of the 20th century, combinatorial designs played an important role in experimental designs in statistics. To demonstrate the connection consider the following question taken from [BC09]:

"Seven different makes of fertilizer are to be tested in an experiment. Twenty-one plots of land are available for the experiment, three plots on each of seven farms in different parts of the country. We can apply one fertilizer to each plot, then grow a crop on all the plots and measure the yield. How should we allocate fertilizers to the plots?"

If we put fertilizer 1 on the three plots on Farm A, fertilizer 2 on the plots on Farm B, etc., then any difference in yield between these plots could be the result of the fertilizer, but could equally be the result of differences in fertility, soil structure or climate between the farms. It turns out that an optimal design would arise from a Steiner triple system of seven triples which corresponds to the Fano plane: Regard each fertilizer as a point in the Fano plane, and each farm as a line, and assign to the three plots on each farm the three fertilizers on the corresponding line.

Combinatorial designs play an important role in designing experiments in Statistics. Figure 3 shows an experiment at Rothamsted Experimental Station, the place where the famous statistician R. A. Fisher did his best work.

Fisher's inequality. — There is an important result by Fisher later rediscovered and generalized in different forms by Hanani, Erdős, and deBruijn which asserts that the number b of blocks in a design is at least n. A beautiful argument whose early roots go

to Bose is to use linear algebra: Consider the $n \times b$ incidence matrix of a design. We claim that the columns $v_1, v_2, \ldots, v_n$ are always linearly independent. Indeed for some $x > y$,

$$\Big\langle \sum \alpha_i v_i, \sum \alpha_i v_i \Big\rangle = x \sum_i \alpha_i^2 + y \sum_{i \neq j} \alpha_i \alpha_j$$
$$= (x-y) \sum_i \alpha_i^2 + y \Big(\sum_i \alpha_i \Big)^2,$$

which can vanish only when all α_is vanish. For Steiner systems of parameters $(n, q+1, 2)$ the number of blocks is $\binom{n}{2} / \binom{q+1}{2}$. Fisher inequality is equivalent, in this case, to $n \geq q^2 + q + 1$, and finite projective planes are the extremal cases.

Designs, groups, and codes. — We already mentioned that there are no t-transitive groups for $t > 5$ other than S_n and A_n. The only 4-transitive groups other than A_n and S_n are the Mathieu groups, M_{24} (5-transitive), M_{23}, M_{12} (5-transitive), and M_{11}. The Mathieu groups were introduced in 1861 and 1873, and they are closely related to designs. Indeed M_{24} and M_{12} can be described as the automorphism groups of Steiner systems. In 1938 Witt described the group M_{24} as the automorphism group of the *Witt design*, which is a Steiner system of parameters $(24, 8, 5)$, thus giving a definite existence proof. Mathieu groups and Witt designs are closely related to the Golay error-correcting codes, discovered in 1949, which have much practical use.

3.2. Wilson's and Teirlinck's constructions

In the early 1960s Hanani solved the question on the existence of designs for $(q, r) = (4, 2), (4, 3)$ and $(5, 3)$. The existence conjecture for the case $r = 2$, namely when every pair of elements belong to λ blocks, was solved by Wilson in 1972. Such designs are called *pairwise balanced designs.*

THEOREM 3.1 (Wilson [Wil72a, Wil72b, Wil75]). — *If n satisfies the divisibility conditions and is large enough then designs of parameters $(n, k, 2, \lambda)$ exist.*

A review of the crucial methods needed for building pairwise balanced designs is given in [Wil74b]. Bose's "method of differences" from 1939 is of great importance. Given an abelian group G, a *simple difference family* is a family of q-subsets such that every non-zero element of G occurs exactly once as the difference of two elements in one of these sets. (When the family consists of a single set it is called a *difference set.*) The family of all cosets of the sets in a simple difference family is a pairwise balanced design. For example, $\{\{1, 3, 9\}, \{2, 6, 5\}\}$ is a difference set in Z_{13}. Wilson used a probabilistic argument [Wil72b, Wil74b] to demonstrate that given q, simple difference families of q-sets always exist for certain sufficiently large elementary abelian groups. In addition to this construction, Wilson introduced a variety of ingenious inductive constructions of designs with additional (geometric-like) structure, rich enough to enable Theorem 3.1.

In 1987 Teirlinck showed that for a large λ depending on q and r (but not on n) designs exist!

THEOREM 3.2 (Teirlinck [Tei87]). — *For every q and r there is $\lambda = \lambda(q, r)$ such that designs of parameters (n, q, r, λ) exist.*

In his short and rich paper Teirlinck introduced more general combinatorial structures and stronger regularity conditions that allow an inductive argument to go through. While λ is large, Teirlinck's example has the additional property of not having repeated blocks. The proof is tour de force and for me it still remains quite mysterious.

3.3. The necessary conditions for designs are sufficient for something

We can regard the question of finding a design as an integer programming question. We need to find 0-1 solutions to a system of equations: The variables α_S are associated to q-subsets S of $[n]$ and for every r-set we have an equation $\sum_{R \subset S} \alpha_S = 1$. It is a common practice to look at a generalized notion of solution and Wilson [Wil74] and Graver and Jurkat [GJ73] asked for integral solutions. [(2)] They proved:

THEOREM 3.3 (Wilson [Wil73]; Graver and Jurkat [GJ73])
For every n, q, r, λ, if n satisfies the divisibility conditions then integral designs exist.

The proof of this result is not difficult and the existence of integral designs plays an important role in Keevash's work.

4. THE GREEDY RANDOM METHOD

4.1. A probabilistic heuristic and a simple application of the probabilistic method

Consider a q-hypergraph with n vertices and $b = \binom{n}{r}/\binom{q}{r}$ edges. Write $a = \binom{n}{q}$. There are altogether $\binom{a}{b}$ such hypergraphs. Given a set R of r vertices what is the distribution of the number of edges containing R? This is a Poisson distribution of parameter 1. The probability that R is contained in a unique edge is $1/e$. If these probabilities were statistically independent we could conclude that Steiner triple systems of parameters n, q, r exist and that their number is $\binom{a}{b}e^{-b}$. We will refer to this argument and estimates it gives as the *probabilistic heuristic*.

2. Another generalized notion of solution, the linear programming relaxation, replaces the 0-1 variables by real numbers in the interval [0,1]. This is useful to several related combinatorial packing and covering problems. I am not aware of it being used for our problem.

Of course, there is neither statistical independence nor good reasons to think that lack of independence is not devastating. [3] Indeed the probabilistic heuristic is completely blind to the divisibility conditions.

What we can do is to choose c edges at random for some $c < b$ so that every r-set is included in *at most* one of them. Doing so shows that we can find $b/(1 + \log \binom{k}{r})$ edges so that every set of size r is included in at most one edge. Here we do not need randomness and we can just add edges in a greedy way. Erdős and Hanani conjectured in 1963 that there are $b(1 - o(1))$ edges so that every r-set is covered at most once, or, equivalently, there are $b(1 + o(1))$ edges so that every r-set is covered at least once.

4.2. Rödl nibble and approximate designs

The greedy-random method (it is also called in the literature "incremental-random method" and "semi-random method") is based on the idea of adding to our desired objects elements in small chunks (or even one at a time). The general idea can be traced to works of Ajtai-Komlos-Szemerédi on Ramsey numbers. In [Rod85] Rödl used a certain greedy-random process known as the Rödl nibble to prove the Erdős -Hanani conjecture.

THEOREM 4.1 (Rödl [Rod85]). — *For every fixed q and r there exists a nearly Steiner system of parameters (n, q, r), namely a system of $(1 + o(1))\binom{n}{r}\binom{q}{r}^{-1}$ q-subsets of $[n]$ such that every r set is included in at least one block in the system.*

The idea is this. You choose at random ϵb blocks and show that (with high probability) they form a very efficient covering of the r-sets they cover. Then you show that (again, with high probability) both the hypergraph of unused q-blocks and the the hypergraph of uncovered r-sets are quasirandom. This allows you to proceed until reaching $(1 - o(1))\binom{n}{r}\binom{q}{r}^{-1}$ q-subsets of $[n]$ such that every r set is included in at most one block in the system. (At this point you can add arbitrary blocks to cover the remaining r-sets.) It was later discovered that a variant of this process where you add one block at a time also works.

4.3. Pippinger-Spencer and beyond

The greedy random method and, in particular, variants of the Rödl nibble had important applications in combinatorics over the years. A general framework for the Rödl nibble was laid by Frankl and Rödl [FR85] with the definite result given by Pippinger and Spencer [PS89].

We consider an auxiliary hypergraph: the vertices correspond to r-sets and the edges correspond to q-sets. For this hypergraph the task is to find a large matching —

3. We note that an important theme in the "probabilistic method" is to prove (when statistical independence does not apply) that certain rare events still have positive probability. There are various methods that were developed for this purpose [AS00].

a collection of pairwise disjoint edges —, or a small covering — a collection of edges covering all vertices.

The result of Pippinger and Spencer asserts that it is enough that

(i) all vertices have the same degree d (or roughly the same degree) and
(ii) every pair of vertices are included in at most $o(d)$ edges.

This level of abstraction is crucial for some further important applications of Rödl's method [Kah96, Spe95, KR97, Kah00, Vu00] and forms a crucial ingredient of Keevash's proof.

4.4. The precise limits of the nibble

The greedy random method has become a standard part of the toolkit in probabilistic and extremal combinatorics. We can ask first, what is the limit of the method and second, can we use it to get perfect constructions which are well above the limit. As for the first question we can start (and limit ourselves) with the simple example of Steiner triple systems. (Much less is known for the general problem.) Suppose that we add at random edge disjoint triangles one after the other. Of course when the remaining edges form a triangle-free graph the process will end. But how many edges can we expect at this stage? (Or in other words, how many edges are left uncovered by triangles.) Bohman, Frieze and Lubetzky [BFL15] proved that the answer is $\theta(n^{2/3})$, confirming a conjecture by Bollobás and Erdős from 1990. The paper contains also an interesting discussion of the history of the problem. [4]

4.5. Using the nibble and similar probabilistic methods for perfect constructions

The next issue is if such probabilistic constructions could be the basis for perfect constructions. Keevash's theorem gives a spectacular and surprising yes answer for designs. But there were interesting earlier examples. Some are described in Endré Szémeredi's paper [Sze12] "Is laziness paying off? ("Absorbing" method)." The abstract of the paper reads "We are going to mention some embedding problems. The main ideas of most of the solution of these problems are that do not work too much. Finish the job almost, then use some kind of absorbing configuration (that is what we call laziness)." An earlier example for this approach is [RRS06].

4. There is, of course, a different important process to reach a triangle-free graph: add edges at random until any new edge will create a triangle. Analyzing such processes is a different story [Boh09, BK10]. Bohman proved that this process terminates, w.h.p. with a graph of $n^{3/2}\sqrt{\log n}$ edges.

5. KEEVASH'S PROOF: THE TEMPLATE, THE NIBBLE, THE OCTAHEDRON, AND THE SHUFFLE

In this section we will present a very rough outline of Keevash proof for a very special (but important) case. We will deal only with decompositions of graphs into edge-disjoint triangles. Of course, decomposing K_n into edge-disjoint triangles is precisely the problem of finding a Steiner triple system on $\{1, 2, \ldots, n\}$. Part of the proof consists of repeated reference (with complicated and subtle details) to the greedy-random method. These parts are quite difficult and long, however we will largely take them for granted.

The difficulty in applications of the greedy-random methods (and other probabilistic arguments) for packing problems is in the last stages. Once we packed a large number of objects the probabilistic arguments do not apply and some backtracking is needed. In some cases, a careful preprocessing of our combinatorial object can assist the required backtracking. In our case, we need an auxiliary collection of triangles called the *template* defined via a combination of algebra and probability. Both for the applications of the greedy random method and for the appellate and algebraic parts, the general case is more difficult than the special case of triangles.

5.1. Edge-decomposition into triangles

When is it possible to decompose the edges of a graph G into edge-disjoint triangles? We say that G is *trivisible* if (i) the number of edges in G is divisible by 3, and (ii) every vertex has an even degree. Next we define the *density* $d(G)$ of a graph as the number of edges divided by $\binom{n}{2}$ where n is the number of vertices. G is (c, t)-*typical* if for every set X of at most k vertices

$$(1 - c)d(G)^{|X|} \leq |\cap_{x \in X} N_x| \leq (1 + c)d(G)^{|X|}.$$

Here, N_x is the set of neighbors of a vertex v. We denote by V the set of vertices of G.

THEOREM 5.1 (Keevash). — *For every $d > 0$ there is $c > 0$ such that if G is trivisible, $(c, 16)$-typical and $d(G) > d$, then G admits a triangle-decomposition.*

Remark 5.2. — Bootstrapping this result Keevash proves that $(c, 2)$-typicality suffices. (Our level of description is not detailed enough to show how typicality is actually used.)

5.2. The template

We choose a so that $2^{a-2} < n \leq 2^{a-1}$. We consider a random map from V into $F_{2^a}^*$ (the non-zero elements of a field with 2^a element). We define the *template* T to be a collection of triangles — those triangles $\{x, y, z\}$ in G such that $x + y + z = 0$. We let $G^* = \cup T$, namely the union of all edges in triangles in T.

Note that the number of triangles in the template is roughly $K^3 d(G)^3 n^{-1} \binom{n}{3}$ where K is some constant between $1/2$ and $1/4$.

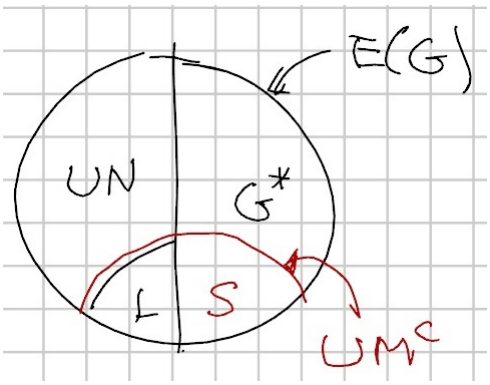

FIGURE 4. Step 2 (picture out of scale; the size of the right side is a small fraction of the whole).

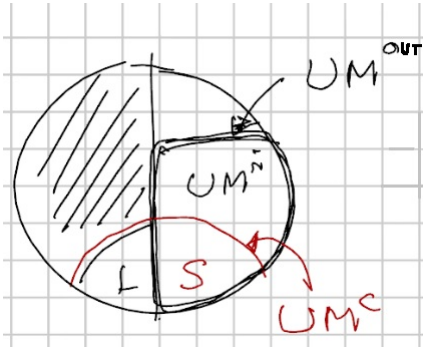

FIGURE 5. Step 3 (picture out of scale).

5.3. The plan

Plan: Start with an approximate triangle decomposition and apply a sequence of repairs.

5.4. Steps 1: nibble

We want to use the nibble (greedy random) method to find a collection N of edge-disjoint triangles whose union is most of $G\backslash G^*$. We will not modify N any further. In order for the method to work we need to assume that G^* and $G\backslash G^*$ and (G, G^*) are "nice" (namely, quasi-random in various ways), as well as various other conditions that allow to implement the initial nibble and to allow the entire argument to go through. All these conditions hold with high probability. We need also that $G\backslash G^*\backslash(\cup N)$ is sparse (having only small degrees), this also holds with high probability.

5.5. Step 2: cover

After packing most of $G\backslash G^*$ with triangles of N, we cover (again, by a random process) the left over part L of $G\backslash G^*$ by a collection M^c of edge-disjoint triangles each having two edges from G^*. This defines a set S of edges from G^*. See Fig. 4.

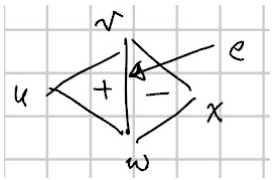

FIGURE 6. The rhombus.

Of course, we need to modify our collection of triangles. The strategy is not to try to modify the collection of triangles as to reduce S but, in a sense go in the other direction: We create collections of positive and negative triangles and enlarge them by a certain octahedral process (described below) which repeatedly replace two adjacent triangles by six others (or replace one triangle by seven others). As we go we improve the structure of these collections allowing us to use them to obtain a perfect cover.

5.6. Step 3: hole

Now we create two sets M^{out} and M^{in} of edge disjoint triangles in G^*, having the property that $\cup M^{\text{out}} = S \uplus M^{\text{in}}$ ($\uplus$ denotes disjoint union). This construction (both here and for the general case) heavily relies on the "integral designs" we considered in Section 3.3. Two clarifications: First, the triangles in these sets need not be from the template T. Second, when I say "we create" this accounts for using the nibble method and at times further massaging our initial steps.

5.7. The octahedron

Here is a detail of the argument taken in separation which we will use in Step 4. Suppose we start with the complete graph K_n (or with a smaller quasi random graph containing G^*), with a pair of edge disjoint collections of triangles. We want to modify these collections of edge-disjoint triangles to new such collections so that the edge $e = \{v, w\}$ (not in G^*) supported by triangles in the original collections will no longer be supported by triangles in the modified collections. We look at the two triangles containing e (Fig. 6) and embed them into an octahedron (Fig. 7). We give the triangles of the octahedron alternating $\pm$ signs. We can now replace the plus triangle containing e by the three minus triangles not containing e, and the minus triangle containing e by the three plus triangle not containing e. We also use the operation of replacing the four positive triangles by the four negative triangles in order to eliminate a participating triangle without changing the supported edges.

5.8. How to proceed: dream and reality

Consider $M_1 = M^c \cup M^{\text{in}}$, $M_2 = M^{\text{out}}$ and then $\cup M_1 = L \uplus \cup M^{\text{out}}$.

Dream plan: suppose that $M^{\text{out}} \subset T$ (in words, all triangles in M^{out} are in the template). Then we are in good shape: We consider the decomposition $N \cup (T \backslash M^{\text{out}}) \cup M_1$!

However, there is no reason to think that this can be achieved.

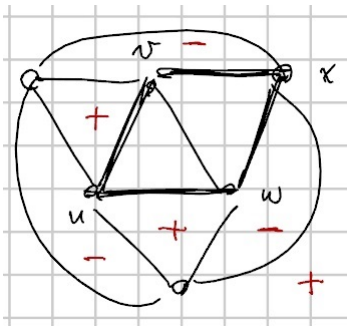

FIGURE 7. The octahedron.

Reality plan: what eventually works is to find two sets of edge-disjoint triangles M_3 and M_4 with edges in G^* with the following properties:

- $\cup M_3 = \cup M_4$,
- $M_3 \subset T$
- $M_2 \subset M_4$,

and use the decomposition

$$N \cup (T \backslash M_3) \cup (M_4 \backslash M_2) \cup M_1. \tag{4}$$

5.9. Two remarks

Remark 5.3. — One reason allowing to use the nibble probabilistic method to get a perfect design is that the greedy random process is taking place in a large hypergraph and we study what is happening on a smaller subhypergraph on a tiny constant fraction of the vertices.

Remark 5.4. — Some arguments about "canceling boundaries" are of importance in the proof. Unlike with usual homology here the boundaries are not signed and the octahedron is a "cycle" w.r.t. such an unsigned "boundary" operator. We note that these type of boundaries and the important role of octahedra can be seen also in the hypergraph regularity theorems [RS04, Gow07], and in the early work [CG89] on quasirandom hypergraphs.

5.10. Step 4: The shuffle

Recall that at this point we have assumed that we know how to build M^{in}, M^c, and M^{out} using the edges of G^*, but not necessarily triangles from T. It is "left" to find M_3 and M_4. Here, we need to explain how to implement our plan, and to indicate how the precise definition of the template triangles enters the picture.

Consider five elements $x_1, x_2, x_3, t_1, t_2 \in F^*$ with x_1, x_2, x_3 linearly independent over $Z/2Z$, $t_1 \neq t_2$. Write $t_3 = t_1 + t_2$, and $X = \text{span} < x_1, x_2, x_3 >$ (over $Z/2Z$). (Reminder: F was a field with 2^a elements and F^* are its non-zero elements.)

The *shuffle* $S_{x,t}$ is a complete tripartite 3-uniform hypergraph with 24 vertices. We have three sets $X + t_1$, $X + t_2$, and $X + t_3$ of eight vertices each, and consider all

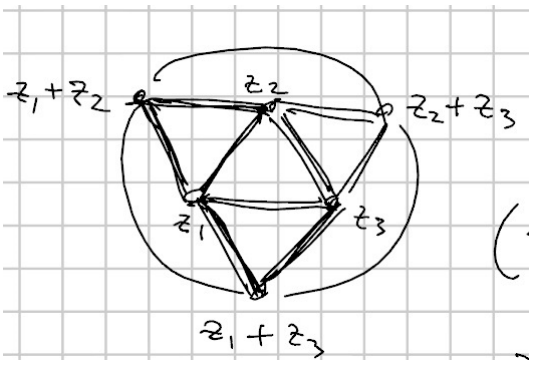

FIGURE 8. The shuffle.

8^3 triangles with one vertex taken from each set. (There is a positive probability that all triangles of the shuffle are supported by our graph.)

We will consider two decompositions of the shuffle into edge-disjoint triangles.

Decomposition I: All triangles of the form $x + t_1, y + t_2, x + y + t_3$.

Decomposition II: Translate of decomposition I by (x_1, x_2, x_3).

Note that all triangles in decomposition I are in the template. But since $x_3 \neq x_1 + x_2$ no triangle in decomposition II is in the template.

Now we can describe the random greedy construction of M_3, M_4. We consider a random available shuffle and add decomposition II to M_4 and decomposition I to M_3. Using the nice quasirandom properties we are careful to maintain throughout our construction (this requires also "massaging" M_1 and M_2 in the process), there are many available choices for the shuffle, and repeated use of it allows to achieve the desired M_3 and M_4. "□"

5.11. Quoting Calegari's reflection on Keevash's proof

"Random construction which 'nearly' solves the combinatorial problem, and then adjusting the result around the margins by formally expressing the error as a linear combination of formal 'differences' of designs of uniformly bounded size, and then treating 'negative' quantities of these small designs as 'holes' in the big uniform quantity.

Exactly the same idea (at this abstract level) is the key to the recent Kahn-Markovic proof of the Ehrenpreis conjecture [KM11], where one first uses a probabilistic (i.e. ergodic theoretic) argument to cover a hyperbolic surface with an almost equidistributed collection of pairs of pants with an almost prescribed geometry, almost all of which can be glued up, and then shows that the error can be formally glued up if one uses 'negative' pieces, which one then interprets as holes in big uniform collection (I like to think of these 'negative' pants as 'holes in the Dirac pants sea'...).

Exactly the same idea again was used by Alden Walker and I recently to show that random groups contain fundamental groups of closed surfaces [CW13] ; we first build 'most' of the surface by a random matching argument, then glue up the error formally

using 'negative' pieces (of bounded size), which can then be pulled out of the collection that was already matched.

No doubt the details of the constructions diverge considerably beyond this 'family resemblance' (this is already true in the latter two examples, where I understand the details of what is going on), but this resemblance at the abstract level seems to me to be much more than a triviality."

6. PACKING, COVERING, AND DESIGNS - A FEW OTHER ADVANCES AND OPEN PROBLEMS

Keevash's achievement is extraordinary in solving a major open problem in combinatorics, and in developing and implementing with great difficulty and ingenuity existing and novel machineries towards a completely unexpected goal. I will mention in this section other important advances regarding packing, covering, designs, and highly regular combinatorial objects, and mention a few open problems. Let me mention that while this paper is an introduction to Keevash's work, it is by no mean an introduction to the reach theory of combinatorial designs. We do not mention here Hadamard matrices, room squares, strongly regular graphs, difference sets, and many other highly regular combinatorial structures. We refer the readers, e.g., to [DS92, CvL80, CvL91].

Finite projective planes. — We mainly discussed the situation when q and r are fixed and n goes to infinity. Understanding other regimes of parameters is also of great interest. The famous problem on the existence of projective planes of non-prime power order can be formulated as:

(1) Are there any designs of parameters $(q+1, 2, q^2+q+1, 1)$ when q is not a prime-power?

The central question regarding uniqueness of projective planes of prime order can be formulated as:

(2) Are there any designs of parameters $(q+1, 2, q^2+q+1, 1)$ when q is a prime, except those coming from a finite field?

It is hard to point the finger on where and why the "probabilistic heuristic" will fail for $(q, 2, n)$ designs when both q and n tend to infinity. It will certainly be interesting to understand this matter.

Further decompositions and stronger regularity. — Can we find a design admitting a decomposition into disjoint perfect matchings and, more generally, into disjoint designs with other parameters? We can seek structures with various recursive decompositions, and, ask also if we can decompose the complete hypergraph into such designs. In particular we can ask if "Kirkman systems" and "Sylvester systems" exist for constant q and r and large n when the corresponding divisibility conditions hold.

This is a good time to mention the classical result by Baranyai [BAR75] asserting that K_r^n can be decomposed into perfect matchings whenever r divides n. Baranyai's proof used ideas from linear programming.

We can also ask about stronger regularity conditions: Given a design S of parameters (n, q, r, λ) when is there a design of parameters $(n+1, q+1, r+1, \lambda)$ all whose links are isomorphic to S?

Pseudomanifolds, manifolds, and buildings. — Block designs with parameters $(n, q, q-1, 2)$ are *pseudomanifolds*. Considered as topological spaces they represent spaces with singularities of codimension at least two. The Heawood Conjecture proved by Ringel and Young asserts that for $q = 3$ such objects exist even if you require them to represent a prescribed triangulated surface. In higher dimensions you can impose various regularity conditions related to the conditions for block designs. For example, Altshuler [Alt78] constructed pseudomanifolds with a common prescribed vertex-link. Triangulations of $2d$-dimensional manifolds can have the property that every set of d vertices belongs to some face of dimension $2d$. When $d > 1$ there are only a handful of such triangulations known. A remarkable example is the 9-vertex triangulation of CP^2 by Kuhnel and Lassman [KL81]. Like for designs we can expect that infinite families for every d exist. Tits' Buildings are, of course, very regular combinatorial structures. Buildings, like projective geometries of dimension greater than two, represent a regime where regularity has strong algebraic consequences.

Designs, t-wise uniform permutations, conjugacy tables, and local central limit theorems. — Kuperberg, Lovett, and Peled [KLP12] used a novel probabilistic technique to show the existence of regular combinatorial objects and to approximately count them. They constructed and estimated the number of block designs B with parameters (n, q, r, λ) when q, r are arbitrary and λ is large, satisfying $\lambda = (n/r)^{\Omega(r)}$, and the corresponding condition for the complementary design (namely the design whose blocks are all q-blocks not in B). This yields new existence and counting results in many asymptotic regimes, e.g., when q grows as a power of n.

They developed local central limit theorems which enabled them to analyze the problem reformulated via a random walk on a lattice with a prescribed set of allowed steps. We will say a little more on their technique below.

A family of permutations on n elements is t-wise uniform if it acts uniformly on tuples of t elements. In another application of the method, Kuperberg, Lovett and Peled showed that there exist families of t-wise uniform permutations for all t, whose size is $n^{O(t)}$. (Before their work constructions of small families of t-wise uniform permutations were known only for $t = 1, 2, 3$.)

An independent body of works with a related method of developing and using local central limit theorems is by Barvinok and Hartigan [BH12] for approximately counting matrices with prescribed rows and column sums (contingency tables), graphs with prescribed degree sequences, and for approximately computing volumes of certain polytopes.

For related techniques for enumerating regular structures, see also [CM05, CGG+10]. When it comes to counting designs and regular structures, a good place to begin is in counting regular graphs. What is the number of d-regular graphs on n vertices? This is a fascinating problem on which a lot is known [MW90] and more is left to be explored.

Kuperberg, Lovett, and Peled: a few more words on the setting and method. — Kuperberg, Lovett and Peled considered the following general questions: Let Φ be an integral matrix. Is there a small collection of rows of Φ whose average equals the average of all rows? How many such small collections are there? They develop a general method to show existence of such small collections and to estimate their number, when the matrix Φ satisfies certain arithmetic, boundedness and symmetry assumptions. These assumptions point to a possible relation of the problem with coding theory. In particular, the following surprising assumption, related to the notion of LDPC (low density parity check) codes, plays a significant role: the dual space to the space of columns of Φ has an integral basis whose members have low ℓ_1-norms.

The argument is based on choosing every row of Φ into the collection independently with probability p and considering the sum of the chosen rows as a random vector. A local central limit theorem is established for this random vector, showing that the probability for it to be at a point is approximately given by the suitably scaled density of a normal random variable at that point. In a large stroke this technique is related to general analytic methods in enumerative combinatorics [FS09, PW13].

q-analogs (Designs over finite fields). — One of the most important problems remaining is whether the existence theorems for designs hold with "sets" and "subsets" replaced by "vector spaces" and "subspaces" (over a finite field). Examples here are much fewer, The first nontrivial example did not appear until 1987 in a paper by Simon Thomas [Tho87].

Vertex-packing of small graphs in large random graphs. — An important problem in probabilistic combinatorics is the problem of vertex packing small graphs H into large random graphs G. The situation for perfect matching (when H is a single edge) is classic but even when H is a triangle this posed a difficult challenge for some decades until resolved by Johansson, Kahn and Vu [JKV08] again using a novel probabilistic technique. The surprising idea is to start with the complete graph and then randomly remove edges one at a time down to the threshold, making sure that at all times the (logarithm of the) number of surviving H-factors is not very far from the expectation. I don't know to what extent randomness can be replaced by quasirandomness (of some kind) for packing problems of this kind. A related recent development is a paper by Montgomery [Mon15] who proved that if a tree T has n vertices and maximum degree at most Δ, then a copy of T can almost surely be found in the random graph with average degree $\Delta \log^5 n$.

Random designs. — Now that we know that designs exist and even have good estimates for their number we can ask what are the properties of random designs and how to

generate random designs. These questions are interesting already for Steiner triple systems.

Two applications of Keevash's result and method. — Glebov and Luria [GL15+] used Keevash's method to estimate the number of 1-factorizations (packing by edge-disjoint perfect matching) of complete graphs with an even number of vertices. Lubotzky, Luria and Rosenthal [LLR15+] showed that the union of a constant number of designs given by Keevash's random construction is, with high probability, a good "coboundary expander," thus giving the first general construction for high-dimensional expanders (in the cohomological sense) with bounded degree.

A problem on packing trees. — Consider $n-1$ trees $T_1, T_2 \ldots, T_n$ where T_i has i vertices.

Conjecture (Gyárfás, 1963): There exists an edge-disjoint decomposition of K_n into $n-1$ parts so that the ith part is isomorphic to T_i.

Acknowledgment

I am thankful to Dan Calegari, Peter Cameron, Roman Glebov, Peter Keevash, Zur Luria, Ron Peled, Wojciech Samotij, and especially Jeff Kahn for very helpful comments and discussions.

REFERENCES

[AS00] N. Alon & J. Spencer, *The probabilistic method*, Wiley, New York, 2000.

[Alt78] A. Altshuler, 3-pseudomanifolds with preassigned links, *Trans. Amer. Math. Soc.* 241 (1978), 213–237.

[BC09] R. A. Bailey & P. J. Cameron, Combinatorics of optimal designs, in *Surveys in Combinatorics* 2009 (ed. S. Huczynska, J. D. Mitchell and C. M. Roney-Dougal), London Math. Soc. Lecture Notes 365, Cambridge Univ. Press 2009, pp. 19–73.

[BAR75] Zs. Baranyai, On the factorization of the complete uniform hypergraph, in A. Hajnal, R. Rado, and V. T. Sós, *Infinite and Finite Sets,* Proc. Coll. Keszthely, 1973, Colloquia Math. Soc. János Bolyai 10, North-Holland, 1975, pp. 91–107.

[Boh09] T. Bohman, The triangle-free process, *Adv. Math.* 221 (2009) 1653–1677.

[BFL15] T. Bohman, A. Frieze & E. Lubetzky, Random triangle removal, *Adv. Math.* 280 (2015), 379–438.

[BK10] T. Bohman & P. Keevash, The early evolution of the H-free process, *Invent. Math.* 181 (2010), 291–336.

[BH12] A. Barvinok & J. A. Hartigan, An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums, *Trans. Amer. Math. Soc.* 364 (2012), 4323–4368.

[CW13] D. Calegari & A. Walker, Random groups contain surface subgroups, to appear in *Jour. Amer. Math. Soc.*, arXiv:1304.2188.

[CvL80] P. Cameron & J. H. van Lint, *Graphs, codes and designs,* Cambridge Univ. Press, 1980.

[CvL91] P. Cameron & J. H. van Lint, *Designs, graphs, codes and their links*, Cambridge Univ. Press, 1991.

[CM05] E. R. Canfield & B. D. McKay, Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums, *Electron. J. Combin.* 12 (2005), Research Paper 29, 31 pp. (electronic).

[CGG+10] E. R. Canfield, Z. Gao, C. Greenhill, B. D. McKay & R. W. Robinson, Asymptotic enumeration of correlation-immune Boolean functions, *Cryptogr. Commun.* 2 (2010), 111–126.

[CGW89] F. Chung, R. L. Graham & R. M. Wilson, Quasi-random graphs, *Combinatorica* 9 (1989), 345–362.

[CG89] F. Chung & R. L. Graham, Quasi-random hypergraphs, *Random Structures and Algorithms* 1 (1990), 105–124.

[CD06] C. J. Colbourn & J. H. Dinitz, *Handbook of combinatorial designs,* 2nd ed., Chapman & Hall / CRC, Boca Raton, 2006.

[DS92] J. H. Dinitz & D. R. Stinson, *Contemporary design theory,* Wiley, 1992.

[EH63] P. Erdős & H. Hanani, On a limit theorem in combinatorial analysis, *Publicationes Mathematicae Debrecen* 10 (1963), 10–13.

[FS09] P. Flajolet & R. Sedgewick, *Analytic combinatorics,* Cambridge Univ. Press, 2009.

[FR85] P. Frankl & V. Rödl, Near perfect coverings in graphs and hypergraphs, *European J. Combinatorics* 6 (1985), 317–326.

[GL15+] R. Glebov & Z. Luria, On the number of 1-factorizations, preprint.

[Gow07] W. T. Gowers, Hypergraph Regularity and the multidimensional Szemerédi Theorem, *Annals of Math.* 166 (2007), 897–946.

[GJ73] J. E. Graver & W. B. Jurkat, The module structure of integral designs, *J. Combinatorial Theory Ser. A* 15(1973), 75–90.

[Han61] H. Hanani, The existence and construction of balanced incomplete block designs, *Annals Mathematical Statistics* 32 (1961), 361–386.

[Han65] H. Hanani, A balanced incomplete block design, *Annals Mathematical Statistics* 36(1965), 7–11.

[JKV08] A. Johansson, J. Kahn & V. Vu, Factors in Random Graphs, *Random Structures and Algorithms* 33 (2008), 1-28.

[Kah96] J. Kahn, Asymptotically good list-colorings, *J. Combinatorial Theory Ser. A* 73 (1996) 1–59.

[Kah00] J. Kahn, Asymptotics of the list chromatic index for multigraphs, *Random Structures and Algorithms* 17 (2000), 117–156.

[KM11] J. Kahn & V. Markovic, The good pants homology and the Ehrenpreis conjecture, *Ann. of Math.* 182 (2015)1–72, arXiv:1101.1330

[Kee14] P. Keevash, The existence of designs, arXiv:1401.3665.

[KeeV] P. Keevash, Videotaped lectures, Jerusalem 2015. https://youtube/tN6oGXqS2Bs?list=PLTn74Qx5mPsSU6ysUXk-ssF6sZtvh-a-o

[Kee15] P. Keevash, Counting designs, arXiv:1504.02909.

[KR97] A. Kostochka & V. Rödl, Partial Steiner systems and matchings in hypergraphs, *Random Structures and Algorithms* 13 (1997), 335–347.

[KL81] W. Kühnel & G. Lassman, The unique 3-neighborly 4 manifold with 9 vertices, *J. Combinatorial Theory Ser. A* 35 (1983), 173–184.

[KLP12] G. Kuperberg, S. Lovett & R. Peled, Probabilistic existence of regular combinatorial objects, Proc. 44th ACM STOC, 2012.

[LL15+] N. Linial & Z. Luria, An upper bound on the number of high-dimensional permutations, *Combinatorica,* to appear.

[LLR15+] A. Lubotzky, Z. Luria & R. Risenthal, Construction of bounded degree coboundary expanders in high dimensions, preprint.

[MW90] B. D. McKay & N. C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, *European J. Combinatorics* 11 (1990), 565–580.

[Mon15] R. Montgomery, Embedding bounded degree spanning trees in random graphs, arXiv:1405.6559.

[N-W70] C. St. J. A. Nash-Williams, An unsolved problem concerning decomposition of graphs into triangles, *Combinatorial Theory and its Applications III,* North Holland (1970), 1179–1183.

[PW13] R. Pemantle & M. C.Wilson, *Analytic combinatorics in several variables*, Cambridge Univ. Press, 2013.

[PS89] N. Pippinger & J. H. Spencer, Asymptotic behaviour of the chromatic index for hypergraphs, *J. Combinatorial Theory Ser. A* 51 (1989), 24–42.

[Rod85] V. Rödl, On a packing and covering problem, *European J. Combinatorics* 6 (1985), 69–78.

[RS04] V. Rödl & J. Skokan, Regularity lemma for uniform hypergraphs, *Random Structures and Algorithms* 25 (2004), 1–42.

[RRS06] V. Rödl, A. Ruciński & E. Szemerédi, A Dirac-type theorem for 3-uniform hypergraphs, *Combin. Probab. Comput.* 15 (2006), 229–251.

[Spe95] J. Spencer, Asymptotic packing via a branching process, *Random Structures and Algorithms* 7 (1995), 167–172.

[Sze12] E. Szemerédi, Is laziness paying off? (Absorbing method), *Colloquium De Giorgi* 2010–2012 Volume 4 of the series Publications of the Scuola Normale Superiore pp. 17–34.

[Tei87] L. Teirlinck, Non-trivial t-designs without repeated blocks exist for all t, *Discrete Math.* 65 (1987), 301–311.

[Tho87] A. Thomason, Designs over finite fields. Geom. Dedicata 24 (1987), 237–242.

[Tho85] A. Thomason, Pseudo-random graphs, in: *Proceedings of Random Graphs,* Poznań 1985, M. Karoński, ed., *Annals of Discrete Math.* 33 (North Holland 1987), 307–331.

[Vu00] V. Vu, New bounds on nearly perfect matchings in hypergraphs: higher codegrees do help, *Random Structures and Algorithms* 17 (2000), 29–63.

[Wil72a] R. M. Wilson, An existence theory for pairwise balanced designs I. Composition theorems and morphisms, *J. Combinatorial Theory Ser. A* 13 (1972), 220–245.

[Wil72b] R. M. Wilson, An existence theory for pairwise balanced designs II. The structure of PBD-closed sets and the existence conjectures, *J. Combinatorial Theory Ser.* A 13 (1972), 246–273.

[Wil75] R. M. Wilson, An existence theory for pairwise balanced designs III. Proof of the existence conjectures, *J. Combinatorial Theory Ser. A* 18 (1975), 71–79.

[Wil72c] R. M. Wilson, Cyclotomy and difference families in elementary abelian groups, *J. Number Th.* 4 (1972) 17–42.

[Wil74b] R. M. Wilson, Constructions and uses for pairwise balanced designs, Part I, in *Combinatorics*, M. Hall and J. H. van lint (eds), Mathematisch Centrum Amsterdam, 1974, pp. 18–41.

[Wil73] R. M. Wilson, The necessary conditions for t-designs are sufficient for something, *Utilitas Math.* 4(1973), 207–215.

[Wil74] R. M. Wilson, Nonisomorphic Steiner triple systems, *Math. Z.* 135 (1974), 303–313.

[Wil03] R. Wilson, The early history of block designs, *Rend. del Sem. Mat. di Messina* 9 (2003), 267–276.

Gil KALAI

Hebrew University of Jerusalem
Institute of Mathematics
Givat-Ram
Jerusalem 91904, Israel
and
Department of Mathematics and Computer Science
Yale University
New Haven CT, USA
E-mail : kalai@math.huji.ac.il