The existence of designs - after Peter Keevash

Gil Kalai

Bourbaki seminar, IHP, Paris, June 27, 2015
Introduction: Designs and Steiner systems

A set $S$ of $q$-subsets of an $n$-set $X$ is a design with parameters $(n, q, r, \lambda)$ if every $r$-subset of $X$ belongs to exactly $\lambda$ elements (“blocks”) of $S$.

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 18 months ago not a single Steiner system for $r > 5$ was known to exist.
Disigns exist!
Divisibility conditions

There are some necessary *divisibility conditions* for the existence of such a design, namely that

\[
\binom{q - i}{r - i} \text{ divides } \lambda \binom{n - i}{r - i}, \quad 0 \leq i \leq r - 1.
\]

(1)

To see that the divisibility conditions are necessary, fix any \( i \)-subset \( I \) of \( X \) and consider the sets in \( S \) that contain \( I \).
Keevash’s theorem

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

**Theorem:** For fixed $q$, $r$, and $\lambda$, there exist $n_0(q, r, \lambda)$ such that if $n > n_0(q, r, \lambda)$ satisfies the divisibility conditions then a design with parameters $(n, q, r, \lambda)$ exists.

In other words, for fixed $q$, $r$, and $\lambda$, the divisibility conditions are sufficient apart from a finite number of exceptional values of $n$. 
The proof has three ingredients:

The greedy random method. It was first used to get "approximate designs". Here it is interwind with the other ingredients. It is used first to get an approximate construction and second to patch it and get the exact construction.

A certain auxiliary algebraic construction (also with randomization) which is a crucial apparatus for the ability to "correct" an approximate decomposition toward an exact decomposition.

A certain "integral" relaxation of the notion of design (where blocks can have positive and negative integral weights) and the octahedral structure which is crucial for constructing those "integral designs".
Part 1: Regularity, symmetry and randomness
Regularity and symmetry

Designs are regular objects. You can get them from groups acting transitively on $r$-sets.

**Proposition:** Let $\Gamma$ be a $t$-transitive permutation group. Then the orbit of a set of size $k$ 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

**Proposition:** Let $\Gamma$ be a $t$-transitive permutation group, $t > 5$, then $G$ is $A_n$ or $S_n$. 
The probabilistic method

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 that kind, there is a positive probability for property $P$ to hold. The probabilistic method is of central importance in combinatorics (and other areas).
The probabilistic method: proving the existence of rare events (without enough statistical independence)

**Challenge:** Use a probabilistic argument to prove the existence of an injective map from a set $A$ of $n$ elements to a set $B$ with $m$ elements.

(1) Easy: when $m > n^2$

(2) Hard When $m > 10n$

(3) ??? when $m = 1.1n$; when $m = n$. 

Gil Kalai

Disigns exist!
Quasi randomness

Quasi randomness refers to structures with a similar behavior (for certain purposes) as random objects. For graphs: 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 independently by Thomasson, and by Chung, Graham, and Wilson. Other examples: primes; sets with diminishing Fourier coefficients; hypergraphs...
Part 2: Keevash’s results
Decomposition of quasi random hypergraphs: the language

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

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

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 $K_r^n$. 

Gil Kalai

Disigns exist!
Decomposition of quasi random hypergraphs: the setting

1) Dividibility.
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 $I \subseteq V(G)$, for all $1 \leq i \leq r$.

2) 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

$$(1 - c)p^{\lvert A \rvert}n \leq \lvert \cap_{S \in A} G(S) \rvert \leq (1 + c)p^{\lvert A \rvert}n.$$ (2)
Decomposition of quasi random hypergraphs: the result

Keevash’s main theorem is

**Theorem:** Let $1/n \ll c \ll d, 1/h \ll 1/q \leq 1/r$. 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.
The number of designs

**Theorem:** The number $S(n, q, r)$ of Steiner systems with parameters $(n, q, r)$ (where $n$ satisfies the divisibility conditions) satisfies

$$\log S(n, q, r) = (1 + o(1)) \left( \frac{n}{r} \right) \left( \frac{q}{r} \right)^{1} (q - r) \log n. \quad (3)$$
Part 2: History
Plucker, Kirkman and Steiner

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.)
Kirkman’s school girl problem

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.
The Kirkman’s schoolgirl problem
Kirkman’s school girl problem: extensions and solutions

Kirkman’s question can be asked in greater generality for every $n = 3(\text{mod} 6)$ and a partial solution was offered in 1852 by Spottiswoode. The general question was settled independently by Xi (a schoolteacher from Mongolia) 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.
Mathieu Groups, the Witt Design and the Golay codes

In the first half of the 20th century combinatorial designs played an important role in experimental designs in statistics, in group theory, and were closely related to error-correcting codes which are among the most important real-life applications of mathematics.

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 $(24, 8, 5)$.

Mathieu groups and Witt designs are closely related to the Golay error-correcting codes, discovered in 1949, which have much practical use.
The existence conjecture for the case $r = 2$, namely when every pair of elements belong to $\lambda$ blocks, was solved by Wilson in 1972.

**Theorem:** If $n$ satisfies the divisibility conditions and is large enough then designs of parameters $(n, q, 2, \lambda)$ exist.
Tierlink’s theorem, 1987

In 1987 Teirling showed that for a large $\lambda$ depending on $q$ and $r$ (but not on $n$) designs exist!

**Theorem:** For every $q$ and $r$ there is $\lambda = \lambda(k, r)$ such that designs of parameters $(n, q, r, \lambda)$ exist.
The necessary conditions for designs are sufficient for something

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 $\alpha_S$ 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 in the early 1970s Wilson and independently Graver and Jurkat asked for integral solutions. They proved:

**Theorem:** For every $n, q, r, \lambda$, if $n$ satisfies the divisibility conditions then integral designs exists.

The proof of this result is not difficult and the existence of integral designs plays an important role in Keevash’s work. ”Octahedra” (and in general, ”signless boundaries”) plays a role in the proof.
Part 3: The greedy random method
The probabilistic heuristic

Consider a $q$-hypergraph with $n$ vertices and $b = \binom{n}{r}/\binom{q}{r}$ edges. Write $a = \frac{n}{q}$. 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*. 
The existence of approximate designs - Rödl1985

**Theorem:** For every fixed $q$ and $r$ there exist a *nearly* Steiner system of parameters $(n, q, r)$, namely a system of 
$(1 + o(1))(n)(q)^{-1}$ $q$-subsets of $[n]$ such that every $r$ set is included in at most one block in the system.
The Rödl nibble

The idea of the proof is this. You choose at random $\epsilon b$ 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 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.
The Pippinger-Spencer theorem

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 – 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. Keevash’s virtuously develop and apply the greedy random method.
Part 4: Keevash’s proof for triangle decompositions
Part 4: Triangle decompositions: the setting

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 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 $G$. 
Triangle decompositions: the result

**Theorem:** 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.
We choose \( a \) so that \( 2^{a-2} < n \leq 2^{a-1} \). We consider a random map from \( V \) into \( \mathbb{F}_{2^a}^* \) (the non-zero elements of a field with \( 2^a \) element). We let \( T \) to be those triangles \( \{x, y, z\} \) in \( G \) such that \( x + y + z = 0 \). We let \( G^* = \bigcup T \), namely the union of all edges in triangles in \( T \). Note that the number of triangles in the template is roughly \( K^3 d(G)^3 \binom{n}{3} \) where \( K \) is some constant between 1/2 and 1/4.
The octahedron

Figure: The rombus
The shuffle

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 $\mathbb{Z}/2\mathbb{Z}$, $t_1 \neq t_2$. Write $t_3 = t_1 + t_2$, and $X = \text{span} < x_1, x_2, x_3 >$ (over $\mathbb{Z}/2\mathbb{Z}$). (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 $8^3$ 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.

Disigns exist!
Step 1: The nibble

We want to use the nibble (greedy random) method to find a collection $N$ of edge-disjoint triangles whose union is most of $G \setminus G^*$. We will not modify $N$ any further. In order for the method to work we need to assume that $G^*$ and $G \setminus G^*$ 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 $G \setminus G^* \setminus (\bigcup N)$ is sparse (having only small degrees), this also holds with high probability.
Step 1+2: The nibble qnd the hole
Step 3:

Figure: Step 3 (picture out of scale)
Part 5: Discussion and open problems
Every design has at least as many blocks as the size of the ground state

The number \( b \) of blocks in a design is at least \( n \). Consider the \( n \times b \) incidence matrix of a design. **Claim:** the rows \( v_1, v_2, \ldots, v_b \) are always linearly independent. Indeed for some \( x > y \),

\[
\left\langle \sum \alpha_i v_i, \sum \alpha_i v_i \right\rangle = x \sum_i \alpha_i^2 + y \sum_{i \neq j} \alpha_i \alpha_j
\]

\[
= x \sum_i \alpha_i^2 + y \sum_{i \neq j} \alpha_i \alpha_j
\]

\[
= (x - y) \sum_i \alpha_i^2 + y \left( \sum_i \alpha_i \right)^2,
\]

which can vanish only when all \( \alpha_i \)'s vanish. This implies that a \( (q + 1, 2, n, 1) \) designs do not exist if \( n < q^2 + q + 1 \).
Thank you very much