SUBSTRUCTURES IN LATIN SQUARES

MATTHEW KWAN, ASHWIN SAH, MEHTAAB SAWHNEY, AND MICHAEL SIMKIN

Abstract. We prove several results about substructures in Latin squares. First, we explain how to adapt our recent work on high-girth Steiner triple systems to the setting of Latin squares, resolving a conjecture of Linial that there exist Latin squares with arbitrarily high girth. As a consequence, we see that the number of order-n Latin squares with no *intercalate* (i.e., no 2×2 Latin subsquare) is at least $(e^{-9/4}n-o(n))^{n^2}$. Equivalently, $\Pr[\mathbf{N}=0]\ge e^{-n^2/4-o(n^2)}=e^{-(1+o(1))\mathbb{E}\mathbf{N}}$, where $\mathbf{N}$ is the number of intercalates in a uniformly random order-n Latin square.

In fact, extending recent work of Kwan, Sah, and Sawhney, we resolve the general large-deviation problem for intercalates in random Latin squares, up to constant factors in the exponent: for any constant $0<\delta\le 1$ we have $\Pr[\mathbf{N}\le(1-\delta)\mathbb{E}\mathbf{N}]=\exp(-\Theta(n^2))$ and for any constant $\delta>0$ we have $\Pr[\mathbf{N}\ge(1+\delta)\mathbb{E}\mathbf{N}]=\exp(-\Theta(n^{4/3}(\log n)^{2/3}))$.

Finally, we show that in almost all order-n Latin squares, the number of *cuboctahedra* (i.e., the number of pairs of possibly degenerate 2×2 subsquares with the same arrangement of symbols) is of order n^4, which is the minimum possible. As observed by Gowers and Long, this number can be interpreted as measuring "how associative" the quasigroup associated with the Latin square is.

1. Introduction

A *Latin square* (of order n) is an $n\times n$ array filled with the numbers 1 through n (we call these *symbols*), such that every symbol appears exactly once in each row and column. Latin squares are a fundamental type of combinatorial design, and in their various guises they play an important role in many contexts (ranging, for example, from group theory, to experimental design, to the theory of error-correcting codes). In particular, the multiplication table of any group forms a Latin square. A classical introduction to the subject of Latin squares can be found in [23], though recently Latin squares have also played a role in the "high-dimensional combinatorics" program spearheaded by Linial, where they can be viewed as the first nontrivial case of a "high-dimensional permutation"[1] (see for example [35–38]).

There are a number of surprisingly basic questions about Latin squares that remain unanswered, especially with regard to statistical aspects. For example, there is still a big gap between the best known upper and lower bounds on the number of order-n Latin squares (see for example [44, Chapter 17]), and there is no known algorithm that (provably) efficiently generates a uniformly random order-n Latin square[2]. Perhaps the main difficulty is that Latin squares are extremely "rigid" objects: in general there is very little freedom to make local perturbations to change one Latin square into another.

Some of the most fundamental questions in this area concern existence and enumeration of various types of substructures. We collect a few different results of this type.

Sah and Sawhney were supported by NSF Graduate Research Fellowship Program DGE-1745302. Sah was supported by the PD Soros Fellowship. Simkin was supported by the Center of Mathematical Sciences and Applications at Harvard University.

[1] To see the analogy to permutation matrices, note that a Latin square can equivalently, and more symmetrically, be viewed as an $n\times n\times n$ zero-one array such that every axis-aligned line sums to exactly 1.

[2] Jacobson and Matthews [20] and Pittenger [43] designed Markov chains that converge to the uniform distribution, but it is not known whether these Markov chains mix rapidly.

1.1. **Intercalates.** Perhaps the simplest substructures one may wish to consider are *intercalates*, which are order-2 Latin (combinatorial) subsquares. That is, an intercalate in a Latin square L is a pair of rows $i < j$ and a pair of columns $x < y$ such that $L_{i,x} = L_{j,y}$ and $L_{i,y} = L_{j,x}$. It is a classical fact that for all orders except 2 and 4 there exist Latin squares with no intercalates [28,29,41] (such Latin squares are said to have property "N_2"). As our first result, we obtain the first nontrivial lower bound on the *number* of order-n Latin squares with this property (upper bounds have previously been proved in [31,40]).

Theorem 1.1. *The number of order-n Latin squares with no intercalates is at least*

$$\left(e^{-9/4}n - o(n)\right)^{n^2}.$$

Theorem 1.1 is proved by adapting our recent work on *high-girth Steiner triple systems*; we discuss this further in Section 1.2.

For comparison, the total number of order-n Latin squares is well known[3] to be $(e^{-2}n - o(n))^{n^2}$. So, Theorem 1.1 can be interpreted as the fact that a *random* order-n Latin square is intercalate-free with probability at least $e^{-n^2/4-o(n^2)}$. Resolving a conjecture of McKay and Wanless [40] (see also [9,34]), Kwan, Sah, and Sawhney [31] recently proved[4] that the *expected* number of intercalates in a random order-n Latin square is $n^2/4 + o(n^2)$, so writing $\mathbf{N}$ for the number of intercalates in a random order-n Latin square, Theorem 1.1 corresponds to the Poisson-type inequality

$$\Pr[\mathbf{N} = 0] \geq \exp(-(1 + o(1))\mathbb{E}\mathbf{N}).$$

Combining this with [31, Theorem 1.2(a)], we obtain an optimal lower-tail large deviation estimate, up to a constant factor in the exponent.

Theorem 1.2. *Let $\mathbf{N}$ be the number of intercalates in a random order-n Latin square, and fix a constant $0 < \delta \leq 1$. Then*

$$\Pr[\mathbf{N} \leq (1 - \delta)\mathbb{E}\mathbf{N}] = \exp(-\Theta(n^2)).$$

We expect that Theorem 1.1 is best-possible (related conjectures appeared in [4,16]). In Conjecture 1.6, we make a specific prediction for the constant factor in the exponent, in the setting of Theorem 1.2.

The upper tails of $\mathbf{N}$ behave rather differently, due to the "infamous upper tail" phenomenon elucidated by Janson [22]. Specifically, it seems that the "most likely way" for $\mathbf{N}$ to be much larger than its expected value is for $\mathbf{N}$ to contain a "tightly clustered" set of intercalates (for example, if L contains the multiplication table of an abelian 2-group $(\mathbb{Z}/2\mathbb{Z})^k$, then it contains about 2^{3k} intercalates).

We are able to obtain a similarly optimal estimate for the upper tail, using very different methods from Theorem 1.2 (specifically, we adapt the machinery of Harel, Mousset, and Samotij [19], studying upper tails using high moments and entropic stability).

Theorem 1.3. *Let $\mathbf{N}$ be the number of intercalates in a random order-n Latin square, and fix a constant $\delta > 0$. Then*

$$\Pr[\mathbf{N} \geq (1 + \delta)\mathbb{E}\mathbf{N}] = \exp(-\Theta(n^{4/3}\log n)).$$

Theorem 1.3 closes the gap between lower and upper bounds recently proved by Kwan, Sah, and Sawhney in [31]. Actually, we are able to prove an even sharper large deviation inequality for intercalates in random *Latin rectangles*, in terms of a certain extremal function; see Theorem 2.2.

[3]This is classical but nontrivial; it follows from celebrated permanent estimates due to Bregman [7] and Egorychev–Falikman [11,12]. See for example [44, Chapter 17].

[4]This is easy to guess heuristically but surprisingly difficult to prove; we are not aware of a way to estimate the expectation $\mathbb{E}\mathbf{N}$ without going through large deviation estimates.

1.2. **Cycles and girth.** Recall that the girth of a graph is the length of its shortest cycle. In [38], Linial defines a *cycle*[5] in a Latin square L to be a set of rows A, a set of columns B, and a set of symbols C, with $|A|+|B|+|C|>3$, such that the $A\times B$ subarray of L contains at least $|A|+|B|+|C|-2$ symbols in the set C. He defines the *girth* of a Latin square L to be the minimum of $|A|+|B|+|C|$ over all such cycles in L. These definitions are motivated by the *Brown–Erdős–Sós problem* in extremal hypergraph theory, and in particular by an old conjecture of Erdős on the existence of high-girth *Steiner triple systems*, which we recently proved in [33] (see [33] for definitions and motivation). Answering a conjecture of Linial [38], we show that there exist Latin squares with arbitrarily high girth.

Theorem 1.4. *Given $g\in\mathbb{N}$, there is $N_{1.4}(g)\in\mathbb{N}$ such that if $N\ge N_{1.4}(g)$, then there exists a Latin square with girth greater than g.*

Theorem 1.4 can be proved in essentially the same way as the analogous theorem for Steiner triple systems in [33]. The only real complication concerns a "triangle-regularisation" lemma, which is much simpler in the setting of [33] than in the setting of Theorem 1.4. Basically, we need a fractional triangle-decomposition result for quasirandom tripartite graphs. Suitable techniques in the dense tripartite setting have already been developed by Montgomery [42] and Bowditch and Dukes [5]; we give a somewhat different proof suitable for our application (combining ideas from both these papers, and introducing some new ones).

Theorem 1.1 is closely related to Theorem 1.4: it is not hard to see that a Latin square has girth greater than 6 if and only if it has no intercalate. The methods in [33] (which we explain in this paper how to adapt to prove Theorem 1.4) easily yield a lower bound on the *number* of Latin squares or Steiner triple systems with girth greater than a given constant g (see [33, Theorem 1.3]), so with a simple calculation one can deduce Theorem 1.1 from the proof of Theorem 1.4

1.3. **Cuboctahedra.** A *cuboctahedron*[6] in a Latin square is a pair of pairs of rows $(r_1,r_2),(r_1',r_2')$ and a pair of pairs of columns $(c_1,c_2),(c_1',c_2')$ such that $L_{r_i,c_j}=L_{r_i',c_j'}$ for all $i,j\in\{1,2\}$. Essentially, this is a pair of 2×2 subsquares (not necessarily Latin) with the same pattern of entries, though degeneracies are allowed (e.g., a pair of 1×2 subrectangles with the same entries also counts, as does a pair of cells with the same entries). When discussing cuboctahedra, we will always refer to labeled cuboctahedra (unlike the case with intercalates).

A *loop* is an algebraic structure satisfying all the group axioms except for associativity. It is easy to see that the multiplication table of a loop is a Latin square, and that one can uniquely label the rows and columns of a Latin square to obtain the multiplication table of a loop[7]. In fact, a loop is a group if and only if its multiplication table has n^5 cuboctahedra (which is the maximum possible). This is sometimes called the *quadrangle condition* (due to Brandt [6]). As explored by Gowers and Long [18], the number of cuboctahedra in a Latin square is a measure of "how associative" its corresponding loop is.

We show that a random Latin square typically has $(4+o(1))n^4$ cuboctahedra.

Theorem 1.5. *A random order-n Latin square $\mathbf{L}$ has $(4+o(1))n^4$ cuboctahedra whp.*

[5]We remark that the word "cycle" in the context of Latin squares sometimes refers to a different object: every pair of rows, columns, or symbols defines a permutation which decomposes into "row-cycles", "column-cycles", or "symbol-cycles". We will not need this notion in the present paper.

[6]The reason for the name is that (as we will see in Figure 7.1), one can interpret these objects in such a way that they resemble geometric cuboctahedra. In [18], the authors simply call these objects "octahedra".

[7]A *quasigroup* is a generalisation of a loop where we moreover do not demand the existence of an identity element. From a combinatorial perspective, there is not much difference between loops and quasigroups: every multiplication table of a quasigroup can be transformed into the multiplication table of a loop by a unique relabeling of the rows and columns.

It is not hard to see that *every* Latin square has at least $(3-o(1))n^4$ cuboctahedra[8], so up to constant factors random Latin squares are typically "as non-associative as possible". It remains an interesting question whether there exist any Latin squares with $(4-\Omega(1))n^4$ cuboctahedra.

The upper bound in Theorem 1.5 (i.e., that almost every order-n Latin square has at most $(4+o(1))n^4$ cuboctahedra) may be proved with similar methods to Theorem 1.3. In fact, some aspects of the proof can be simplified substantially because we are not attempting to prove an optimal upper tail bound. We also state a general-purpose result that provides upper bounds on general configuration counts in random Latin squares, as long as a certain "stability" criterion is satisfied (Theorem 7.2).

To prove the lower bound in Theorem 1.5, we make use of some ideas developed in [13, 30, 31], via which a random Latin square can be approximated by the so-called *triangle removal process*. These ideas are subject to quantitative limitations of completion theorems due to Keevash [25, 26], and are therefore only suitable for controlling events which occur with probability extremely close to 1 (specifically, they must hold with probability $1-\exp(-\Omega(n^{2-b}))$ for a very small constant b). We therefore require some non-standard arguments to prove very high probability bounds (see Theorem 7.3 for a precise statement).

1.4. **Further directions.** There are a number of fascinating further directions of study in this area. Concerning the constant factors in the exponents in Theorems 1.2 and 1.3, we make some fairly precise conjectures. Again let $\mathbf{N}$ be the number of intercalates in a random order-n Latin square.

Conjecture 1.6. *For every constant* $0<\delta\le 1$*, we have*

$$\Pr[\mathbf{N}\le(1-\delta)\mathbb{E}\mathbf{N}]=\exp(-(\delta+(1-\delta)\log(1-\delta)+o(1))\mathbb{E}\mathbf{N}).$$

Conjecture 1.7. *For every constant* $\delta>0$*, we have*

$$\Pr\big[\mathbf{N}\ge(1+\delta)n^2/4\big]=\exp(-(1+o(1))\Phi(\delta\mathbb{E}\mathbf{N})\log n),$$

where $\Phi(N)$ *is the minimum number of nonempty cells in a partial Latin square with at least* N *intercalates.*

We do not know the asymptotic value of $\Phi(N)$ in general (though we do know this asymptotic value is $(4N)^{2/3}$ when $(4N)^{1/3}$ is a power of two; see Theorem 2.3), and this would also be interesting to investigate further. Relatedly, it would also be interesting to find the asymptotics of the maximum possible number of intercalates in an order-n Latin square (see [3, 8] for the current best bounds on this problem).

Regarding Latin subsquares of order greater than 2: it was conjectured by Hilton (see [10, Problem 1.7]) that for sufficiently large n there exist order-n Latin squares with no proper Latin subsquare at all (such Latin squares are said to have property "N_∞"). This conjecture remains open when $n=2^a3^b$ for $a\ge 1$ and $b\ge 0$ (see [39] and the references therein). Although a random order-n Latin square typically has about $n^2/4$ intercalates, Kwan, Sah, and Sawhney [31] conjectured that the number of 3×3 subsquares has an asymptotic Poisson distribution with mean $O(1)$, and that almost all Latin squares have no Latin subsquares of order 4 or higher (see also [40, Section 10]). That is to say, intercalates are the "hardest" type of Latin subsquare to avoid. In fact, it seems plausible that the random construction used to prove Theorem 1.4 may have property N_∞ with probability $\Omega(1)$, but an attempt to prove this would require deconstructing the proof of [33, Theorem 1.1] to a far greater extent than we do in this paper.

[8] In fact there are always at least this many *degenerate* cuboctahedra: there are always $(1-o(1))n^4$ cuboctahedra obtained by taking the same 2×2 subsquare twice, there are $(1-o(1))n^4$ cuboctahedra obtained by taking pairs of 1×2 subrectangles with the same pair of symbols, and there are $(1-o(1))n^4$ cuboctahedra obtained by taking pairs of 2×1 subrectangles with the same pair of symbols.

Finally, we reiterate that it would be interesting to determine the asymptotic minimum number of cuboctahedra in an order-n Latin square (i.e., the "least associative" it is possible for a Latin square to be).

1.5. **Outline.** In Section 2 we reduce Theorem 1.3 to a large deviation problem for random Latin rectangles. In Section 3 we prove an upper bound on upper tail probabilities for intercalates in random Latin rectangles, and in Section 4 we sketch how to prove a corresponding lower bound (this is not necessary for the proof of Theorem 1.3, but may be of independent interest). In Section 5 we prove some basic facts about a certain extremal function Φ, which features in Sections 2 to 4. In Section 6 we make some observations about intercalates in random Latin rectangles with very few rows; here we observe quite different behavior.

In Section 7 we explain how the ideas used to prove Theorem 1.3 can be generalized to a wide range of different configurations other than intercalates. Based on this, we explain how to prove the upper bound in Theorem 1.5 (though there is some tedious casework that we omit). We also discuss the machinery for probabilistic transference between Latin squares and the triangle removal process, and use this machinery to prove the lower bound in Theorem 1.5.

Finally, in Section 8 we explain how to adapt our work in [33] to prove Theorems 1.1 and 1.4. This section is mostly targeted towards readers who have read [33] or at least its proof outline, though we do provide a high-level summary of the overall approach.

1.6. **Notation.** We use standard asymptotic notation throughout, as follows. For functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$ or $f \lesssim g$ to mean that there is a constant C such that $|f| \le C|g|$, $f = \Omega(g)$ to mean that there is a constant $c > 0$ such that $f(n) \ge c|g(n)|$ for sufficiently large n, and $f = o(g)$ to mean that $f/g \to 0$ as $n \to \infty$. Subscripts on asymptotic notation indicate quantities that should be treated as constants. Also, following [24], the notation $f = 1 \pm \varepsilon$ means $1 - \varepsilon \le f \le 1 + \varepsilon$.

For a real number x, the floor and ceiling functions are denoted $\lfloor x \rfloor = \max\{i \in \mathbb{Z} : i \le x\}$ and $\lceil x \rceil = \min\{i \in \mathbb{Z} : i \ge x\}$. We will however mostly omit floor and ceiling symbols and assume large numbers are integers, wherever divisibility considerations are not important. All logarithms in this paper are in base e, unless specified otherwise.

Acknowledgements. We thank Freddie Manners for suggesting the enumeration of cuboctahedra in random Latin squares.

2. Upper tails for intercalates: reducing to Latin rectangles

In the setting of Theorem 1.3, the lower bound $\Pr[\mathbf{N} \ge (1+\delta)\mathbb{E}\mathbf{N}] \ge \exp(-O(n^{4/3}\log n))$ already appears as [31, Theorem 1.1(d)]. So, to prove Theorem 1.3, it suffices to prove the following upper bound.

Theorem 2.1. *Fix a constant $\delta > 0$. Let $N(\mathbf{L})$ be the number of intercalates in a uniformly random order-n Latin square $\mathbf{L}$. Then*

$$\Pr[N(\mathbf{L}) \ge (1+\delta)n^2/4] \le \exp\Big(-\Omega_\delta(n^{4/3}\log n)\Big).$$

We will deduce Theorem 2.1 from the following related theorem, which gives a *sharp* upper tail estimate for random *Latin rectangles*. A $k \times n$ Latin rectangle is a $k \times n$ array containing the symbols $\{1, \ldots, n\}$, such that every symbol appears *at most* once in each row and column.

Theorem 2.2. *Fix a constant $\delta > 0$. Let $N(\mathbf{L})$ be the number of intercalates in a uniformly random $k \times n$ Latin rectangle $\mathbf{L}$, where $k = o(n)$ and $k = \omega((\log n)^{3/2})$. Let $\Phi(N)$ be the minimum number of nonempty cells in a partial Latin square with at least N intercalates. Then*

$$\Pr\left[N(\mathbf{L}) \ge (1+\delta)\frac{k^2}{4}\right] = \exp\left(-(1+o(1))\Phi\left(\delta\frac{k^2}{4}\right)\log n\right).$$

We remark that the assumption $k = \omega((\log n)^{3/2})$ is in fact necessary; qualitatively different behavior occurs for smaller k (we show this in Theorem 6.1). We believe that the assumption $k = o(n)$ is unnecessary (and that the conclusion holds even for $k = n$, as in Conjecture 1.7) but we are unable to prove this.

The estimate in Theorem 2.2 (approximately $n^{-\Phi(\delta k^2/4)}$) should be interpreted as being essentially the probability that a specific partial Latin square with $\Phi(\delta k^2/4)$ nonempty cells and $\delta k^2/4$ intercalates is present in $\mathbf{L}$. If we condition on this event, then the conditional expected number of intercalates becomes about $(1+\delta)k^2/4$.

Regarding the value of $\Phi(N)$, its order of magnitude is always $N^{2/3}$, and we are able to determine $\liminf_{N\to\infty} \Phi(N)/N^{2/3}$. However, it seems plausible that the asymptotic value of $\Phi(N)$ may in general depend on number-theoretic properties of N. It would be interesting to investigate this further.

Theorem 2.3. *Let $\Phi(N)$ be the minimum number of nonempty cells in a partial Latin square with at least N intercalates.*

(1) $\Phi(N) \geq (1+o(1))(4N)^{2/3}$ *for all N.*
(2) $\Phi(N) = (1+o(1))(4N)^{2/3}$ *when $(4N)^{1/3}$ is a power of two.*
(3) $\Phi(N) \leq (4+o(1))(4N)^{2/3}$ *for all N.*
(4) $\Phi(N+\varepsilon N) - \Phi(N) = o(N^{2/3})$ *if $\varepsilon = o(1)$.*

We will prove Theorem 2.3 in Section 5.

Theorem 2.2 can be separated into an upper bound and a lower bound on $\Pr[N(\mathbf{L}) \geq (1+\delta)k^2/4]$. Only the upper bound is necessary to prove Theorem 2.1; we conclude this section with the deduction.

Proof of Theorem 2.1, given the upper bound in Theorem 2.2. Let $k = \varepsilon n$, and let $\mathbf{L}_{\mathrm{r}}$ be a uniformly random $k \times n$ Latin rectangle, for $\varepsilon > 0$ sufficiently small such that

$$\Pr[N(\mathbf{L}_{\mathrm{r}}) \geq (1+\delta/2)k^2/4] \leq \exp(-\Omega_\delta(k^{4/3}\log n))$$

(such an ε exists by the upper bound in Theorem 2.2, and Theorem 2.3(1)). Now, if $N(\mathbf{L}) \geq (1+\delta)n^2/4$ then, by averaging, $\mathbf{L}$ has a set of k rows containing at least

$$(1+\delta)\frac{n^2}{4}\left(\binom{k}{2}\Big/\binom{n}{2}\right)$$

intercalates. Let $\mathbf{L}_{\leq k}$ be the Latin rectangle consisting of the first k rows of $\mathbf{L}$; by symmetry and a union bound we deduce

$$\Pr[N(\mathbf{L}) \geq (1+\delta)n^2/4] \leq \binom{n}{k}\Pr[N(\mathbf{L}_{\leq k}) \geq (1+\delta/2)k^2/4].$$

Then, the desired result follows from [40, Proposition 4] (which is a simple application of Bregman's inequality and the Egorychev–Falikman inequality for permanents). Specifically, comparing between $\mathbf{L}_{\leq k}$ and $\mathbf{L}_{\mathrm{r}}$, there is a multiplicative change of measure of at most $\exp\big(O(n(\log n)^2)\big)$ for any event. $\square$

3. Upper-bounding the upper tail in a random Latin rectangle

In this section we prove the upper bound in Theorem 2.1, building on techniques developed by Harel, Mousset and Samotij [19]. We first need some basic definitions and estimates.

Definition 3.1. A *partial Latin array* is an array of dimensions $k \times n$ with some cells filled with one of n symbols, in which no symbol appears more than once in any row or column. For partial Latin arrays S, S' with the same dimensions, we write $S \subseteq S'$ if S and S' agree on all cells where S is nonempty. Let $|S|$ be the number of nonempty cells in S, and let $N(S)$ be the number of intercalates in S. We say that a partial Latin array is *extendable* if there is a complete Latin rectangle of the same dimensions containing it.

Lemma 3.2. *Suppose $k = o(n)$. Let $\mathbf{L}$ be a uniformly random $k \times n$ Latin rectangle, and let Q be a $k \times n$ extendable partial Latin array with $|Q| = o(k\sqrt{n})$. Then*

$$\mathbb{E}[N(\mathbf{L}) \,|\, Q \subseteq \mathbf{L}] \le \frac{k^2}{4} + N(Q) + o(k^2).$$

To prove Lemma 3.2 we need a few auxiliary lemmas.

Lemma 3.3. *Let Q be an extendable $k \times n$ partial Latin array and let $\mathbf{L}$ be a uniformly random $k \times n$ Latin rectangle. Consider a row-column pair (r, c), and write $\mathbf{L}_{r,c}$ for the symbol in the corresponding cell of $\mathbf{L}$. Write $|Q_r|$ for the number of nonempty cells in row r of Q and suppose that $2k + |Q_r| \le n/2$. For any $s \in \{1, \ldots, n\}$, and any cell (r, c) which is empty in Q, we have*

$$\Pr[\mathbf{L}_{r,c} = s \,|\, Q \subseteq \mathbf{L}] \le \frac{1}{n} + O\left(\frac{k + |Q_r|}{n^2}\right).$$

Proof. Let $\mathcal{L}$ be the set of all $k \times n$ Latin rectangles $\mathbf{L}$ for which $Q \subseteq \mathbf{L}$, and let $\mathcal{L}_s \subseteq \mathcal{L}$ be the set of all such Latin rectangles L for which $L_{r,c} = s$. Consider the auxiliary bipartite graph $\mathcal{H}$ with parts $A = \mathcal{L}_s$ and $B = \mathcal{L}$, where we put an edge between $L \in A$ and $L' \in B$ if L' can be obtained from L by swapping the contents of $L_{r,c}$ with some cell in row r. (The trivial swap is allowed; note that elements in $\mathcal{L}_s$ appear on both sides of the bipartition.)

In this auxiliary bipartite graph:

(1) every $L \in B$ has degree at most 1, and
(2) every $L' \in A$ has degree at least $n - 2(k-1) - |Q_r|$.

To see (1), note that there is exactly one cell in row r with symbol s (swapping cell (r, c) with this cell may or may not produce a Latin rectangle). To see (2), we consider all $n - |Q_r|$ possible swaps. Not all of these produce Latin rectangles: at most $k - 1$ of the swaps bring a symbol into (r, c) which already appears in column c, and at most $k - 1$ of the swaps bring the symbol x into a column which already contains it.

It follows from (1) and (2) that

$$\Pr[\mathbf{L}_{r,c} \,|\, Q \subseteq \mathbf{L}] = \frac{|\mathcal{L}_s|}{|\mathcal{L}|} \le \frac{1}{n - 2(k-1) - |Q_r|},$$

which implies the desired result since $2k + |Q_r| \le n/2$. □

Lemma 3.4. *Let Q be a $k \times n$ partial Latin array with $|Q| = o(k\sqrt{n})$. Let $\mathbf{A}(Q)$ be the random array obtained from Q by independently putting a uniformly random symbol from $\{1, \ldots, n\}$ in each empty cell. Then the number of intercalates $N(\mathbf{A}(Q))$ in it satisfies*

$$\mathbb{E}[N(\mathbf{A}(Q))] \le \frac{k^2}{4} + N(Q) + o(k^2).$$

Remark. We assume that 2×2 subsquares composed all of the same element in $\mathbf{A}(Q)$ are counted as intercalates, though this will not matter to us.

Proof. For an intercalate in $\mathbf{A}(Q)$, say that one of its four entries was *forced* if it appeared in Q.

- The contribution to $\mathbb{E}[N(\mathbf{A}(Q))]$ from intercalates with four forced entries is $N(Q)$.
- The contribution to $\mathbb{E}[N(\mathbf{A}(Q))]$ from intercalates with zero forced entries is at most $\binom{k}{2}\binom{n}{2}/n^2 \le \frac{k^2}{4}$. Indeed, for any 2×2 subarray of empty cells in Q, the probability they form an intercalate in $\mathbf{A}(Q)$ is n^2/n^4.

- The contribution to $\mathbb{E}[N(\mathbf{A}(Q))]$ from intercalates with one forced entry is at most $|Q|(k-1)(n-1)/n^2 = o(k^2)$. Indeed, for a given nonempty entry in $|Q|$, there are at most $(k-1)(n-1)$ ways to choose an additional row and column yielding a 2×2 subarray of Q with 3 empty cells. For each such choice, the probability an intercalate is present in $\mathbf{A}(Q)$ is n/n^3.
- The contribution to $\mathbb{E}[N(\mathbf{A}(Q))]$ from intercalates with precisely two forced entries that are in the same row or column is at most $2|Q|^2(n-1)/n^2 \le |Q|^2/n = o(k^2)$. Indeed, for a given pair of nonempty entries in the same row or column of Q, there are at most $n-1$ ways to extend to a 2×2 subarray with 2 empty cells. For each such choice, an intercalate is present in $\mathbf{A}(Q)$ with probability $1/n^2$.
- The contribution to $\mathbb{E}[N(\mathbf{A}(Q))]$ from intercalates with at least two forced entries in different rows and columns, and at least one unforced entry, is at most $|Q|^2/n = o(k^2)$. Indeed, for a given pair of nonempty entries in different rows and columns of Q the probability the corresponding 2×2 subarray forms an intercalate in $\mathbf{A}(Q)$ is at most $1/n$.

The desired result follows. $\square$

Now we are ready to prove Lemma 3.2.

Proof of Lemma 3.2. Choose $\varepsilon = o(1)$ such that $|Q|^2/(k^2 n) = o(\varepsilon^2)$ (which is possible since $|Q| = o(k\sqrt{n})$). There are at most $|Q|/(\varepsilon n)$ "bad" rows of Q which have more than εn nonempty cells. In the conditional probability space given $Q \subseteq \mathbf{L}$, let $\mathbf{Q}'$ be obtained from Q by adding all entries of $\mathbf{L}$ in bad rows, and let $\mathcal{S}$ be the support of $\mathbf{Q}'$. Then

$$\begin{aligned}
\mathbb{E}[N(\mathbf{L}) \,|\, Q \subseteq \mathbf{L}] &\le \sum_{Q'\in\mathcal{S}} \Pr[\mathbf{Q}' = Q']\mathbb{E}[N(\mathbf{L}) \,|\, Q' \subseteq \mathbf{L}] \\
&= \sum_{Q'\in\mathcal{S}} \Pr[\mathbf{Q}' = Q']\big(N(Q') + \mathbb{E}[N(\mathbf{L}) - N(Q') | Q' \subseteq \mathbf{L}]\big) \\
&\le \sum_{Q'\in\mathcal{S}} \Pr[\mathbf{Q}' = Q']\big(N(Q') + (1 + O(\varepsilon + k/n))\mathbb{E}[N(\mathbf{A}(Q')) - N(Q')]\big) \\
&\le \sum_{Q'\in\mathcal{S}} \Pr[\mathbf{Q}' = Q']N(Q') + (1 + O(\varepsilon + k/n))\left(\frac{k^2}{4} + o(k^2)\right).
\end{aligned}$$

Here, in order to establish the first inequality we have used Lemma 3.3 and linearity of expectation over the possible intercalates in $\mathbf{A}(Q')$ which do not already appear in Q'. In order to establish the second inequality we used Lemma 3.4.

So, it suffices to prove that for any outcome $Q' \in \mathcal{S}$ we have $N(Q') = N(Q) + o(k^2)$. To see this, first note that we always have $|Q'| \le |Q|/\varepsilon$. An intercalate appearing in Q' but not in Q is determined by a bad row and an entry of Q' (namely, consider a row containing an element of $Q' \setminus Q$ in the intercalate and consider the opposite corner of the 2×2 subarray), so there are at most $(|Q|/(\varepsilon n)) \cdot (|Q|/\varepsilon) = o(k^2)$ such. $\square$

We now begin to adapt some of the ideas of Harel, Mousset and Samotij [19].

Definition 3.5. Say a $k \times n$ partial Latin array Q using the symbols $\{1, \ldots, n\}$ is a (δ, ε, C)-*seed* if

- Q is extendable,
- $|Q| \le Ck^{4/3}\log n$,
- $\mathbb{E}[N(\mathbf{L}) \,|\, Q \subseteq \mathbf{L}] \ge (1 + \delta - \varepsilon)k^2/4$.

Roughly speaking, these conditions say that Q has few nonempty cells, but its appearance in $\mathbf{L}$ is likely to dramatically increase the expected number of intercalates.

From now on we assume $k = o(n)$ and $k = \omega((\log n)^{3/2})$. Fix some $\varepsilon > 0$ which is small in terms of δ, and let C be large in terms of δ, ε (small and large enough to satisfy certain inequalities later in the proof; then at the end of the proof we will take $\varepsilon \to 0^+$ very slowly).

For a partial Latin array L, let $Z(L)$ be the indicator for the event that there is no (δ, ε, C)-seed $Q \subseteq L$. The rest of the proof of the upper bound in Theorem 2.2 boils down to the following two claims.

Claim 3.6. $\Pr[N(\mathbf{L}) \geq (1+\delta)k^2/4 \text{ and } Z(\mathbf{L}) = 1] \leq \exp\bigl(-\Omega_{\delta,\varepsilon}(Ck^{4/3}\log n)\bigr)$.

Claim 3.7. $\Pr[Z(\mathbf{L}) = 0] \leq \exp\bigl(-(1+o(1))\Phi((\delta - 3\varepsilon)k^2/4)\log n\bigr)$.

Once these two are proven, taking C sufficiently large in terms of δ, ε implies

$$\Pr[N(\mathbf{L}) \geq (1+\delta)k^2/4] \leq \exp(-(1+o(1))\Phi((\delta - 3\varepsilon)k^2/4)).$$

Then, sending $\varepsilon \to 0^+$ slowly and using Theorem 2.3(1) and Theorem 2.3(4), we obtain the desired upper bound in Theorem 2.2.

Proof of Claim 3.6. Say a *potential intercalate* is a $k \times n$ partial Latin array in which exactly four entries are nonempty, forming an intercalate. Say that a set of potential intercalates $\{I_1, \ldots, I_\ell\}$ are *compatible* if (a) every pair agrees on all cells where they are both nonempty, in which case it makes sense to consider their *union* $I_1 \cup \cdots \cup I_\ell$, and (b) the union is extendable.

Let $\ell = \lfloor Ck^{4/3}\log n/4 \rfloor$, and note that if $L' \subseteq L$ and $Z(L) = 1$ then $Z(L') = 1$ as well. We have

$$\begin{aligned}
\mathbb{E}[N(\mathbf{L})^\ell \cdot Z(\mathbf{L})] &\leq \sum_{\substack{I_1,\ldots,I_\ell \\ \text{compatible,} \\ Z(I_1 \cup \cdots \cup I_{\ell-1}) = 1}} \Pr[I_1 \cup \cdots \cup I_\ell \subseteq \mathbf{L}] \\
&= \sum_{\substack{I_1,\ldots,I_{\ell-1} \\ \text{compatible,} \\ Z(I_1 \cup \cdots \cup I_{\ell-1}) = 1}} \Pr[I_1 \cup \cdots \cup I_{\ell-1} \subseteq \mathbf{L}] \cdot \mathbb{E}[N(\mathbf{L}) | I_1 \cup \cdots \cup I_{\ell-1} \subseteq \mathbf{L}] \\
&\leq \sum_{\substack{I_1,\ldots,I_{\ell-1} \\ \text{compatible,} \\ Z(I_1 \cup \cdots \cup I_{\ell-2}) = 1}} \Pr[I_1 \cup \cdots \cup I_{\ell-1} \subseteq \mathbf{L}] \cdot \bigl((1+\delta-\varepsilon)k^2/4\bigr).
\end{aligned}$$

In the last step we use the definition of a seed: since $Z(I_1 \cup \cdots \cup I_{\ell-1}) = 1$ we see $I_1 \cup \cdots \cup I_{\ell-1}$ is not a (δ, ε, C)-seed (despite satisfying the first two properties in the definition of a seed). We can iterate the above inequality ℓ times to deduce that

$$\mathbb{E}[N(\mathbf{L})^\ell \cdot Z(\mathbf{L})] \leq \bigl((1+\delta-\varepsilon)k^2/4\bigr)^\ell.$$

Recalling that $\ell = Ck^{4/3}\log n/4$, Markov's inequality gives

$$\Pr[N(\mathbf{L}) \geq (1+\delta)k^2/4 \ \text{ and } \ Z(\mathbf{L}) = 1] \leq \frac{\mathbb{E}[N(\mathbf{L})^\ell \cdot Z(\mathbf{L})]}{\bigl((1+\delta)k^2/4\bigr)^\ell},$$

from which the claimed bound follows. ☐

Next, we turn to Claim 3.7. It would be too lossy to upper-bound $\Pr[Z(\mathbf{L}) = 0]$ by simply taking a union bound over all possible seeds Q. The next idea is to instead consider "minimal" subsets of seeds.

Definition 3.8. Say a $k \times n$ partial Latin array Q using the symbols $\{1, \ldots, n\}$ is a (δ, ε, C)-*core* if

- Q is extendable,
- $|Q| \leq Ck^{4/3}\log n$,
- $N(Q) \geq (\delta - 3\varepsilon)k^2/4$ (therefore $|Q| \geq \Phi((\delta - 3\varepsilon)k^2/4)$),
- For any $Q_- \subseteq Q$ obtained from Q by emptying a single nonempty cell, we have

$$N(Q) - N(Q_-) \geq \frac{\varepsilon k^2/4}{Ck^{4/3}\log n}.$$

Note that for every (δ, ε, C)-seed Q, we have $|Q| = o(k\sqrt{n})$ and hence Lemma 3.2 shows that $N(Q) \geq (\delta - 2\varepsilon)k^2/4$ (note this puts a bound on how fast ε decays at the end of the argument). By iteratively emptying cells that violate the fourth condition, we can always obtain a (δ, ε, C)-core $Q' \subseteq Q$. So, to prove Claim 3.7 it suffices to upper-bound the probability of containing a core. To this end, we collect a few observations about cores.

From now on, it will be sometimes convenient to think of a $k \times n$ partial Latin array with symbols in $\{1, \dots, n\}$ as a $k \times n \times n$ zero-one array (with "shafts" in the third dimension being identified with symbols), such that every row, column, and symbol has at most one "1". That is to say, symbols really have basically the same role as rows and columns.

Claim 3.9. *If $k = \omega((\log n)^{3/2})$ the number of (δ, ε, C)-cores with m nonempty cells is at most $n^{o(m)}$.*

To prove Claim 3.9 we need a few auxiliary observations.

Lemma 3.10. *There is $u_m \lesssim_{\delta,\varepsilon,C} mk^{-2/3} \log n$ such that every (δ, ε, C)-core with m nonempty cells contains at most u_m nonempty rows, at most u_m nonempty columns, and at most u_m distinct symbols.*

Proof. For any core Q, the fourth condition in the definition of a core says that every nonempty cell in Q participates in at least $w := (\varepsilon k^2/4)/(Ck^{4/3} \log n)$ intercalates.

This implies in particular that every nonempty row of Q contains at least w nonempty entries, so Q has at most $u_{|Q|} := |Q|/w$ nonempty rows. A similar argument applies to columns and symbols. □

Now, say an *ordered core* $\vec{Q}$ is a core Q equipped with an ordering $(r_1, c_1) \prec \cdots \prec (r_{|Q|}, c_{|Q|})$ of its nonempty cells. We write $\vec{Q}_i \subseteq \vec{Q}$ for the partial Latin array obtained from $\vec{Q}$ by emptying all nonempty cells except the first i of them. We say that $\vec{Q}$ is *good* if for every $|Q|/(\log n)^2 \leq i \leq |Q|$, there are at least $(i/|Q|)^3 k^{2/3}/(\log n)^2$ intercalates in $\vec{Q}_i$ containing the cell (r_i, c_i).

Claim 3.11. *If $k \geq (\log n)^{15}$, then a $1 - o(1)$ fraction of orderings of a (δ, ε, C)-core Q are good.*

Proof. Given a (δ, ε, C)-core Q, define the random ordered core $\vec{\mathbf{Q}}$ by taking a uniformly random ordering of the nonempty cells of Q. We wish to show that $\vec{\mathbf{Q}}$ is good whp.

We will take a union bound over all $|Q|/(\log n)^2 \leq i \leq |Q|$. So, fix such an i. Note that, given a choice of (r_i, c_i), the remaining nonempty cells in $\vec{\mathbf{Q}}_i$ comprise a uniformly random subset of $i - 1$ other nonempty cells of Q. It will be convenient to work with a closely related "binomial" random array: let $\mathbf{Q}^{\mathrm{bin}}_{i,r,c}$ be obtained by starting with Q, and emptying each cell other than (r, c) with probability $1 - (i-1)/(|Q| - 1)$. Any property that holds with probability $1 - n^{-\omega(1)}$ for $\mathbf{Q}^{\mathrm{bin}}_{i,r,c}$ also holds with probability $1 - n^{-\omega(1)}$ for $\vec{\mathbf{Q}}_i$ conditioned on the event $(r_i, c_i) = (r, c)$ (this follows from the so-called "Pittel inequality"; see [21, p. 17]). It will suffice to study $Q^{\mathrm{bin}}_{i,r,c}$.

Now, recall that in Q there are at least $\Omega_{\delta,\varepsilon,C}(k^{2/3}/\log n)$ intercalates containing (r, c). Each of these intercalates involves a disjoint set of three nonempty cells other than (r, c), and is therefore present in $\vec{\mathbf{Q}}^{\mathrm{bin}}_{i,r,c}$ with probability $((i-1)/(|Q| - 1))^3$. These intercalates are disjoint from each other aside from the shared cell (r, c). By $k \geq (\log n)^{15}$ and a Chernoff bound, the number of such intercalates in $\vec{Q}_i$ is at least $(i/|Q|)^3 k^{2/3}/(\log n)^2$ with probability $1 - n^{-\omega(1)}$. □

Proof of Claim 3.9. Let $u = u_m \lesssim_{\delta,\varepsilon,C} mk^{-2/3} \log n$ be as in Lemma 3.10. We have $u = o(m)$ since $k = \omega((\log n)^{3/2})$. There are $\binom{k}{u}\binom{n}{u}^2 \leq n^{3u} = n^{o(m)}$ ways to choose sets of u rows, columns, and symbols; we fix such a choice and count the good ordered (δ, ε, C)-cores involving only those rows, columns and symbols (this is valid by Lemma 3.10).

First, there are $\binom{u^2}{m} \le (eu^2/m)^m \le (O(m^2k^{-4/3})/m)^m \le (\log n)^{5m} = n^{o(m)}$ ways to choose a set of nonempty cells, and there are $m!$ ways to choose an ordering on these cells $(r_1, c_1), \dots, (r_m, c_m)$. Therefore it suffices to show there are $n^{o(m)}$ ways to choose the symbols in these ordered cells to produce an ordered core.

If $k \le (\log n)^{15}$ we use the trivial bound that there are at most $u \le (\log n)^{O(1)}$ choices at each step. Overall, there are at most $(\log n)^{O(m)} = n^{o(m)}$ ordered cores on this ordered list of cells.

Otherwise, we bound the number of good ordered cores on this list of cells (r_i, c_i). Then Claim 3.11 will imply a bound on the number of ordered cores as desired. For the first $m/(\log n)^2$ cells, we still use the trivial bound that there are at most $u \le n$ choices. For $m \ge i \ge m/(\log n)^2$, given any choices for the previous cells, we observe that the number of ways to choose a symbol for the ith cell is at most

$$\frac{u}{(i/m)^3 k^{2/3}/(\log n)^2} \le (\log n)^{11}.$$

Indeed, say a "potential intercalate at step i for symbol s" is a set of three cells other than (r_i, c_i) which have been filled in the previous $i-1$ steps and which would form an intercalate with (r_i, c_i) if it were filled with the symbol s. There are at most u potential intercalates at step i in total (corresponding to the at most u supported columns in our list of cells, say), and we must choose a symbol s such that there are at least $(i/m)^3 k^{2/3}/(\log n)^2$ potential intercalates for s.

It therefore follows that the number of good ordered cores on this list of ordered cells is bounded by $n^{m/(\log n)^2}(\log n)^{O(m)} = n^{o(m)}$ and we are done. □

Finally we prove Claim 3.7, which completes the proof of the upper bound in Theorem 2.2 as discussed earlier.

Proof of Claim 3.7. By [17, Theorem 4.7], for any $k \times n$ partial Latin array Q with at most $u = u_{Ck^{4/3}\log n}$ (from Lemma 3.10) nonempty columns, we have $\Pr[Q \subseteq \mathbf{L}] \le (O(1/n))^{|Q|}$. For $m_0 = \Phi((\delta - 3\varepsilon)k^2/4)$, by Claim 3.9 and the definition of Φ and cores, along with the fact that seeds contain cores, we have

$$\Pr[Z(\mathbf{L}) = 0] \le \sum_{m=m_0}^{Ck^{4/3}\log n} n^{o(m)}(O(1/n))^m \le \exp\bigl(-(1+o(1))\Phi((\delta - 3\varepsilon)k^2/4)\log n\bigr),$$

as desired. □

4. Lower-bounding the upper tail in a random Latin rectangle

In this section we sketch how to prove the lower bound in Theorem 2.2. This is not necessary for the proof of Theorem 1.3 but may be of independent interest. We refer to some of the ideas in Section 3, which should be read first.

Proof sketch of the lower bound in Theorem 2.2. Fix $\varepsilon > 0$ and some $C > 0$, and let Q be a partial Latin square with $(\delta + 2\varepsilon)k^2/4$ intercalates and $|Q| = \Phi((\delta + 2\varepsilon)k^2/4)$. As in the discussion directly after Definition 3.8 in Section 3, we can find a (δ, ε, C)-core $Q' \subseteq Q$ with at least $(\delta + \varepsilon)k^2/4$ intercalates by iteratively removing elements that violate the fourth condition. Then by Lemma 3.10, we see that Q' has at most $O_{\delta,\varepsilon,C}(k^{2/3}\log n)$ nonempty columns.

Now, [17, Theorem 4.7] says that $\Pr[M \subseteq \mathbf{L}]$ is very close to $n^{-|M|}$ for any "reasonably small" partial Latin square M. It implies that

$$\Pr[Q' \subseteq \mathbf{L}] \ge \left(\frac{1+o(1)}{n}\right)^{\Phi((\delta+2\varepsilon)k^2/4)}.$$

It also implies, with an easy second-moment calculation, that

$$\Pr\left[N(\mathbf{L}) \geq (1+\delta)\frac{k^2}{4} \,\middle|\, Q' \subseteq \mathbf{L}\right] = 1 - o(1).$$

The desired result follows, taking $\varepsilon \to 0^+$ and using Theorem 2.3(1,4). $\square$

5. Maximizing the number of intercalates

In this section we prove Theorem 2.3. To prove Theorem 2.3(1) we need the following extremal theorem, which follows directly from the "colored" version of the Kruskal–Katona theorem due to Frankl, Füredi, and Kalai [14].

Theorem 5.1. *Let $\mathcal{F}$ be a 3-partite graph with at least n^3 triangles. Then $\mathcal{F}$ has at least $3n^2$ edges.*

Proof of Theorem 2.3(1). Fix three disjoint sets R, C, S with $|R| = |C| = |S|$, such that the rows, columns, and symbols of Q lie in R, C, S respectively. Consider the 3-uniform tripartite graph G with vertex set $R \cup C \cup S$, obtained by adding a triangle between a row $r \in R$, column $c \in C$ and symbol $s \in S$ whenever the (r, c)-entry of Q contains s. Note that apart from the $|Q|$ triangles we directly added to form G (which are all edge-disjoint), there are at least $4N(Q)$ other triangles in G (the four triangles corresponding to an intercalate form four of the eight faces of an octahedron, and the four additional triangles forming the other four faces are unique to that intercalate). Note that G has $3|Q|$ edges, so the desired result follows from Theorem 5.1 with $n = \lfloor (4N(Q))^{1/3} \rfloor$. $\square$

Proof of Theorem 2.3(2–4). For (2–3) we simply consider the Latin square corresponding to the multiplication table of an abelian 2-group $(\mathbb{Z}/2\mathbb{Z})^k$, where 2^k is the smallest power of 2 bigger than $(4N+N^{3/4})^{1/3}$. It is easy to show (see for example [8]) that this Latin square has $(2^k)^2(2^k-1)/4 \geq N$ intercalates.

For (4) we observe that given a partial Latin square with $\Phi(N)$ nonempty cells and N intercalates, it is always possible to increase the number of intercalates by at least εN by adding a disjoint copy of the Latin square corresponding to the multiplication table $(\mathbb{Z}/2\mathbb{Z})^k$, where $2^k = \Theta((\varepsilon N)^{1/3})$ is appropriately chosen. $\square$

6. Latin rectangles with very few rows

Now, we show that the assumption $k = \omega((\log n)^{3/2})$ in Theorem 2.2 is in fact necessary. Note that $k^2 \leq k^{4/3} \log n$ precisely when $k \leq (\log n)^{3/2}$.

Theorem 6.1. *Fix a constant $\delta > 0$. Let $N(\mathbf{L})$ be the number of intercalates in a uniformly random $k \times n$ Latin rectangle $\mathbf{L}$, where $k = n^{o(1)}$ and $k = \omega(1)$. Then*

$$\Pr\left[N(\mathbf{L}) \geq (1+\delta)\frac{k^2}{4}\right] \geq \exp\left((\delta - (1+\delta)\log(1+\delta) - o(1))\frac{k^2}{4}\right).$$

The bound in Theorem 6.1 is essentially the probability that a Poisson distribution with mean $k^2/4$ is at least $(1+\delta)k^2/4$. We suspect that a matching upper bound holds whenever $k = o((\log n)^{3/2})$ and $k = \omega(1)$. For very slowly growing k this follows from [40, Theorem 3], which shows that for fixed k the number of intercalates in a uniformly random $k \times n$ Latin rectangle has distribution limiting to $\mathrm{Poi}(k(k-1)/4)$ as $n \to \infty$.

Proof sketch of Theorem 6.1. Fix $\varepsilon > 0$. Say a collection of $m := \lceil (\delta + 2\varepsilon)k^2/4 \rceil$ potential intercalates is *good* if no pair of these intercalates shares a column, or symbol. An easy calculation shows that the number of good collections of intercalates is $(1 - o(1))\binom{n^2(n-1)^2k(k-1)/4}{m}$ since $k = n^{o(1)}$.

Fix a $k \times n$ partial Latin array Q obtained by taking the union of m intercalates in a good collection. Let $N^*(\mathbf{L}, Q)$ be the number of intercalates in $\mathbf{L}$ using no entry of Q. An argument similar to the proof of Lemma 3.2 ignoring intercalates which have no entries in Q shows

$$\mathbb{E}[N(\mathbf{L}) - N(Q) - N^*(\mathbf{L}, Q) | Q \subseteq \mathbf{L}] = o(k^2).$$

Markov's inequality then shows

$$\begin{aligned}
\Pr\Big[(1+\delta)\frac{k^2}{4} \le N(\mathbf{L}) \le (1+\delta+4\varepsilon)\frac{k^2}{4} \,\Big|\, Q \subseteq \mathbf{L}\Big] \\
\ge \Pr\Big[(1-\varepsilon)\frac{k^2}{4} \le N^*(\mathbf{L}, Q) \le (1+\varepsilon)\frac{k^2}{4} \,\Big|\, Q \subseteq \mathbf{L}\Big] - o(1).
\end{aligned}$$

An easy second-moment calculation using [17, Theorem 4.7] (which says that $\Pr[M \subseteq \mathbf{L}]$ is very close to $n^{-|M|}$ for any "reasonably small" partial Latin square M), and counting similar to the proof of Lemma 3.4, we find

$$\Pr\Big[(1-\varepsilon)\frac{k^2}{4} \le N^*(\mathbf{L}, Q) \le (1+\varepsilon)\frac{k^2}{4} \Big| Q \subseteq \mathbf{L}\Big] = 1 - o(1).$$

Then [17, Theorem 4.7] yields

$$\begin{aligned}
&\Pr\Big[Q \subseteq \mathbf{L} \text{ and } (1+\delta)\frac{k^2}{4} \le N(\mathbf{L}) \le (1+\delta+4\varepsilon)\frac{k^2}{4}\Big] \\
&\qquad = \Pr[Q \subseteq \mathbf{L}] \Pr\Big[(1+\delta)\frac{k^2}{4} \le N(\mathbf{L}) \le (1+\delta+4\varepsilon)\frac{k^2}{4} \,\Big|\, Q \subseteq \mathbf{L}\Big] \\
&\qquad \ge (1 - o(1))\left(\frac{1-o(1)}{n^4}\right)^m.
\end{aligned}$$

Let $\mathbf{X}$ be the number of size-m good collections of intercalates in $\mathbf{L}$, so

$$\begin{aligned}
\Pr[N(\mathbf{L}) \ge (1+\delta)k^2/4]\binom{(1+\delta+4\varepsilon)k^2/4}{m} &\ge \mathbb{E}[\mathbf{X}\mathbb{1}_{(1+\delta)k^2/4 \le N(\mathbf{L}) \le (1+\delta+4\varepsilon)k^2/4}] \\
&\ge (1-o(1))\binom{n^2(n-1)^2k(k-1)/4}{m}\left(\frac{1-o(1)}{n^4}\right)^m
\end{aligned}$$

by linearity of expectation.

The desired result follows, taking $\varepsilon \to 0^+$ slowly. Here we use the assumption $k = \omega(1)$, the approximation $\binom{m}{q} = ((1+o(1))em/q)^q$ (which holds for $q = o(m)$ and $q = \omega(1)$) and the approximation $\binom{m}{q} = \exp((1+o(1))mH(q/m))$, for $H(t) = -t\log t - (1-t)\log(1-t)$ (which holds for $\min(q, m-q) = \Theta(m)$). $\square$

7. General configurations and cuboctahedra

In this section we prove Theorem 1.5. The upper bound and lower bound will be proved by quite different means, but for both we use the 3-uniform hypergraph formulation of a Latin square. A *colored triple system* is a properly 3-colored 3-uniform hypergraph, where the color classes are labeled "R", "C" and "S" (short for "rows", "columns" and "symbols"). A colored triple system is *Latin* if no pair of hyperedges intersect in more than one vertex. An order-n *partial Latin square* is a Latin colored triple system, where the color classes are

$$R = \{1, \ldots, n\}, \quad C = \{n+1, \ldots, 2n\} \quad S = \{2n+1, \ldots, 3n\}.$$

An order-n *Latin square* is a partial Latin square with exactly n^2 hyperedges.

7.1. **An upper bound on the number of cuboctahedra.** For the upper bound we adapt the ideas used to prove Theorem 2.1. In fact, we give a general high-probability upper bound for counts of configurations that satisfy a certain "stability" property. We make no attempt to obtain sharp tail estimates, so the proof of this upper bound is basically just a simpler version of the proof of Theorem 2.1. We will therefore be very brief with the details.

Definition 7.1. Fix a Latin colored triple system H, and let $X_H(Q)$ be the number of (labeled) copies of H in a colored triple system Q. Let $\mathbf{B}_{n,p}$ be a random colored triple system with color classes R, C, S, where each possible hyperedge is present with probability p, and let $\mathbf{B}_n = \mathbf{B}_{n,1/n}$. Note that $\mathbb{E}[X_H(\mathbf{B}_{n,p})] = (1+o(1))n^{v(H)}p^{e(H)}$, where $v(H)$ and $e(H)$ are, respectively, the numbers of vertices and hyperedges in H.

Say that H is *α-stable* if $\mathbb{E}X_H(\mathbf{B}_n) = O(n^\alpha)$ (i.e., $\alpha \geq v(H) - e(H)$), and there is $t = t(n) = \omega(n(\log n)^2)$ such that $\mathbb{E}[X_H(\mathbf{B}_n) \,|\, Q \subseteq \mathbf{B}_n] - \mathbb{E}[X_H(\mathbf{B}_n)] = o(n^\alpha)$ for any Latin colored triple system Q with at most t triples.

Theorem 7.2. *Fix an α-stable Latin colored triple system H , and let $\mathbf{L}$ be a uniformly random order-n Latin square. Then $X_H(\mathbf{L}) \leq \mathbb{E}X_H(\mathbf{B}_n) + o(n^\alpha)$ whp.*

Proof. Fix $\gamma > 0$ to be chosen later. Let $R' \subseteq R, C' \subseteq C, S' \subseteq S$ be the first γn rows, columns, and symbols of R, C, S, respectively. Let $\mathbf{L}_\mathrm{r}$ be a uniformly random Latin rectangle with rows indexed by R' and columns and symbols indexed by C, S, and let $\mathbf{L}'_\mathrm{r}$ be the Latin colored triple system obtained from $\mathbf{L}_\mathrm{r}$ by deleting all columns and symbols except those in C', S'.

Let $t(\cdot)$ be the function certifying α-stability (as in Definition 7.1). Consider a Latin colored triple system Q with color classes R', C', S' and at most $t(\gamma n)$ hyperedges. By Lemma 3.3, for any such Q we have

$$\mathbb{E}[X_H(\mathbf{L}'_\mathrm{r}) \,|\, Q \subseteq \mathbf{L}'_\mathrm{r}] \leq (1 + O(\gamma))\mathbb{E}[X_H(\mathbf{B}_{\gamma n,1/n}) \,|\, Q \subseteq \mathbf{B}_{\gamma n,1/n}].$$

By α-stability of H,

$$\begin{aligned}\mathbb{E}[X_H(\mathbf{B}_{\gamma n,1/n}) \,|\, Q \subseteq \mathbf{B}_{\gamma n,1/n}] - \mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) &= O_\gamma(\mathbb{E}[X_H(\mathbf{B}_{\gamma n}) \,|\, Q \subseteq \mathbf{B}_{\gamma n}] - \mathbb{E}X_H(\mathbf{B}_{\gamma n}))\\ &= o_\gamma(n^\alpha).\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}[X_H(\mathbf{L}'_\mathrm{r}) \,|\, Q \subseteq \mathbf{L}'_\mathrm{r}] &\leq (1 + O(\gamma))\mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + o_\gamma(n^\alpha)\\ &\leq \mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + O(\gamma) \cdot \gamma^{e(H)}\mathbb{E}X_H(\mathbf{B}_{\gamma n}) + o_\gamma(n^\alpha)\\ &\leq \mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + O(\gamma) \cdot \gamma^{e(H)} \cdot O((\gamma n)^\alpha) + o_\gamma(n^\alpha)\\ &\leq \mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + O(\gamma^{e(H)+\alpha+1}n^\alpha),\end{aligned}$$

using the first property of α-stability (that $\mathbb{E}X_H(\mathbf{B}_n) = O(n^\alpha)$). Let $\ell = \lfloor t(\gamma n)/e(H)\rfloor$; a similar calculation to that in the proof of Claim 3.6 shows that

$$\mathbb{E}X_H(\mathbf{L}'_\mathrm{r})^\ell \leq (\mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + O(\gamma^{e(H)+\alpha+1}n^\alpha))^\ell,$$

where we have noted that n is sufficiently large with respect to γ. By Markov's inequality and the fact that $\mathbb{E}X_H(\mathbf{B}_n) = O(n^\alpha)$, for a sufficiently large absolute constant M we have

$$\begin{aligned}\Pr[X_H(\mathbf{L}'_\mathrm{r}) \geq \mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + M\gamma^{e(H)+\alpha+1}n^\alpha] &\leq \frac{\mathbb{E}X_H(\mathbf{L}'_\mathrm{r})^\ell}{\left(\mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + M\gamma^{e(H)+\alpha+1}n^\alpha\right)^\ell}\\ &\leq \exp(-\omega_\gamma(n(\log n)^2)).\end{aligned}$$

Now, we finish similarly to the deduction of Theorem 2.1 from Theorem 2.2. First, using [40, Proposition 4] (i.e., Bregman's inequality and the Egorychev–Falikman inequality for permanents), any event that holds for $\mathbf{L}'_\mathrm{r}$ with probability $1-\exp(-\omega(n(\log n)^2))$ will hold with similar probability

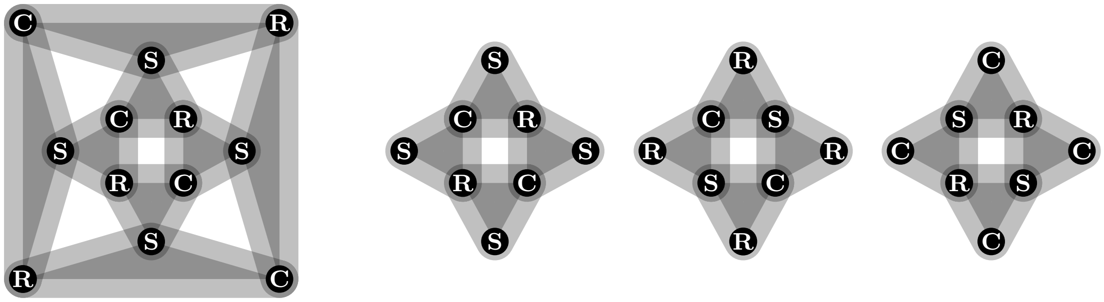

Figure 7.1. On the left, a nondegenerate cuboctahedron (as an $\{R, C, S\}$-colored triple system). On the right, the three dominant degenerate cuboctahedra. In the $n \times n$ array formulation of a Latin square, the first of these degenerate cuboctahedra corresponds to a 2×2 subsquare with distinct entries, taken twice. The second corresponds to a pair of distinct 2×1 subsquares with the same pair of symbols, and the third corresponds to a pair of distinct 1×2 subsquares with the pair of symbols.

for the restriction of $\mathbf{L}$ to the rows, columns, and symbols R', C', S'. Thus by a union bound and symmetry, we see that whp $\mathbf{L}$ has at most $\mathbb{E}X_H(\mathbf{B}_{\gamma n,1/n}) + M\gamma^{e(H)+\alpha+1}n^\alpha$ copies of H in any choice of γn rows, columns, and symbols. An averaging computation reveals that this property implies $X_H(\mathbf{L}) \le (1 + o(1))\big(\mathbb{E}X_H(\mathbf{B}_n) + M\gamma^{e(H)+\alpha+1-v(H)}n^\alpha\big)$. Note that $\alpha + 1 + e(H) - v(H) \ge 1$ and thus the desired result follows by taking $\gamma \to 0^+$ slowly. $\square$

We now prove the upper bound in Theorem 1.5 (i.e., that almost every order-n Latin square has at most $(4 + o(1))n^4$ cuboctahedra) using Theorem 7.2.

Proof of the upper bound in Theorem 1.5. We say that a cuboctahedron is *nondegenerate* if its defining rows r_1, r_2, r_1', r_2' are distinct, defining columns c_1, c_2, c_1', c_2' are distinct, and the four entries of the form L_{r_i,c_j} are distinct. The number of nondegenerate cuboctahedra in a Latin square L is $X_H(L)$, where H is a certain 8-hyperedge, 12-vertex colored triple system depicted on the left hand side of Figure 7.1. Note that $\mathbb{E}X_H(\mathbf{B}_n) = (1 - o(1))n^4$. We claim that H is 4-stable with $t(n) = n(\log n)^3$. Fix a Latin colored triple system Q with at most $t(n)$ hyperedges, and for a copy of H in $\mathbf{B}_n$, say one of its 8 hyperedges is *forced* if it appears in Q.

- The contribution to $\mathbb{E}[X_H(\mathbf{B}_n)\,|\,Q \subseteq \mathbf{B}_n]$ from copies of H with 1 forced hyperedge is $O(tn^9n^{-7}) = o(n^4)$, because there are t ways to choose the forced entry, and $O(n^9)$ ways to choose the other 9 vertices to specify a copy of H. Then, the probability that all 7 non-forced hyperedges are present is n^{-7}.
- The contribution from copies with 2 forced hyperedges is $O(t^2n^7n^{-6}) = o(n^4)$ for similar reasons.
- The contribution from copies with 3 forced hyperedges is $O(t^3n^{12-7}n^{-5}) = o(n^4)$, noting that every set of 3 hyperedges in a cuboctahedron spans at least 7 vertices.
- For the contribution from copies with 4 forced hyperedges:
  - If these 4 forced hyperedges are arranged in a "4-cycle", spanning 8 vertices, then the four forced hyperedges are determined by any three of them by the Latin property of Q, so the contribution is $O(t^3n^{12-8}n^{-4}) = o(n^4)$.
  - Otherwise, the four forced hyperedges span at least 9 vertices, and the contribution is $O(t^4n^{12-9}n^{-4}) = o(n^4)$.
- For the contribution from copies with 5 forced hyperedges we again distinguish cases: if the forced hyperedges contain a 4-cycle, then one can check that they are determined by some size-3 subset and the contribution is $O(t^3n^{12-10}n^{-3}) = o(n^4)$. Otherwise, at least 11 vertices are covered by Q so the contribution is $O(t^5n^{12-11}n^{-3}) = o(n^4)$.

- For the contribution from copies with 6 forced hyperedges, there are three "non-isomorphic" cases to consider: the non-forced hyperedges can share a vertex, be at distance 1, or be on "opposite sides" of the cuboctahedron. In all cases, one can check that the forced hyperedges can be determined by some size-3 subset, and therefore compute that the contribution is $O(t^3 n^{12-11} n^{-2}) = o(n^4)$.
- The contribution from copies with 7 forced hyperedges is $O(t^4 n^{-1}) = o(n^4)$, as the forced hyperedges are determined by some size-3 subset.
- The contribution from copies with 8 forced hyperedges is $O(t^3) = o(n^4)$, as a cuboctahedron is determined by some 3 of its hyperedges.

The difference between $\mathbb{E}[X_H(\mathbf{B}_n) \,|\, Q \subseteq \mathbf{B}_n]$ and $\mathbb{E}X_H(\mathbf{B}_n)$ arises from cuboctahedra with at least one forced hyperedge, so the above calculations show that H is 4-stable, as claimed. It follows that $X_H(\mathbf{L}) \le (1 + o(1))n^4$ whp, by Theorem 7.2.

The total number of cuboctahedra (including degenerate ones) in $\mathbf{L}$ can be expressed as a sum of quantities of the form $X_{H'}(\mathbf{L})$, where H' ranges over a variety of Latin colored triple systems obtained by identifying vertices of H in certain ways (respecting the Latin property).

For any such H' with eight hyperedges (i.e., some vertices are identified but no hyperedges coincide), we have $\mathbb{E}X_{H'}(\mathbf{B}_n) = O(n^{11}) \cdot n^{-8} = O(n^3)$ and the exact same case structure as above shows that H' is 4-stable (whether it is 3-stable is non-obvious but unnecessary for us). Theorem 7.2 then shows that $X_{H'}(\mathbf{L}) \le \mathbb{E}X_{H'}(\mathbf{B}_n) + o(n^4) = o(n^4)$ whp.

Now we consider H' with fewer than eight hyperedges. We will show that these H' deterministically contribute $(3 + o(1))n^4$ degenerate cuboctahedra. First, the dominant contribution comes from the three Latin colored triple systems depicted on the right of Figure 7.1. As discussed in Section 1.3, the contribution from these diagrams is $(1 + o(1))n^4$ each, for a total of $(3 + o(1))n^4$ (with probability 1). Starting from these "dominant" cases, there are four further H' that can be obtained by further identifying vertices (three with two hyperedges, and one with a single hyperedge). Each of these contribute only $O(n^3)$ to our count (again, with probability 1). As it turns out, one always obtains such a situation if any pair of faces is collapsed, other than "opposite faces".

It remains to consider degenerate cuboctahedra obtained by collapsing such "opposite faces". In the $n \times n$ array formulation of a Latin square, this corresponds to those cuboctahedra which consist of a pair of 2×2 subsquares with the same arrangement of symbols, intersecting in exactly one entry (this is only possible if the arrangement has two of the same symbol). It is easy to see that in all such configurations there are three vertices which, if known, determine the entire configuration, so the contribution from such configurations is $O(n^3)$ with probability 1.

The desired result follows by adding up all the contributions from degenerate H' (together with nondegenerate H). ∎

7.2. **The lower tail for cuboctahedra.** We now turn to the lower bound in Theorem 1.5. Recall the definition of a nondegenerate cuboctahedron from the proof of the upper bound in Theorem 1.5. The contribution from degenerate cuboctahedra is always at least $(3 - o(1))n^4$, so it suffices to prove the following strong lower tail bound for nondegenerate cuboctahedra.

Theorem 7.3. *Fix $\delta > 0$, and let $X_H(\mathbf{L})$ be the number of nondegenerate cuboctahedra in a uniformly random order-n Latin square $\mathbf{L}$. Then*

$$\Pr[X_H(\mathbf{L}) \le (1 - \delta)n^4] \le \exp(-\Omega_\delta(n^2)).$$

To prove Theorem 7.3, we use some machinery from [31] (building on ideas in [13, 30]), which allows one to approximate a random Latin square with the so-called *triangle-removal process*. We simply quote a number of statements which will be used in the proof; a more thorough discussion of the history of these techniques can be found in [31]. Let $\mathcal{L}_m$ be the set of order-n partial Latin squares with m hyperedges and let $\mathcal{L}$ be the set of order-n Latin squares.

Definition 7.4 (cf. [31, Definition 2.2]). The 3-partite *triangle removal process* is defined as follows. Start with the complete 3-partite graph $K_{n,n,n}$ on the vertex set $R \cup C \cup S$. At each step, consider the set of all triangles in the current graph, select one uniformly at random, and remove it. After m steps of this process, the set of removed triangles can be interpreted as a partial Latin square $L \in \mathcal{L}_m$ unless we run out of triangles before the mth step. Let $\mathbb{L}(n, m)$ be the distribution on $\mathcal{L}_m \cup \{*\}$ obtained from m steps of the triangle removal process, where "$*$" corresponds to the event that we run out of triangles.

Definition 7.5 ([31, Definition 2.3]). Let $\mathcal{T}_m \subseteq \mathcal{L}_m$ and $\mathcal{T} \subseteq \mathcal{L}$. We say that $\mathcal{T}_m$ is *ρ-inherited* from $\mathcal{T}$ if for any $L \in \mathcal{T}$, taking $\mathbf{L}_m \subseteq L$ as a uniformly random subset of m hyperedges of L, we have $\mathbf{L}_m \in \mathcal{T}_m$ with probability at least ρ.

We will need the following transference theorem for inherited properties, which compares a subset of a uniformly random Latin square to the outcome of the triangle removal process. This theorem builds on a similar theorem for Steiner triple systems proved by Kwan [30], using the work of Keevash [25, 26], and the tripartite case is similar; see [32].

Theorem 7.6 ([31, Theorem 2.4]). *Let $\alpha \in (0, 1/2)$. There is an absolute constant $\gamma > 0$ such that the following holds. Consider $\mathcal{T}_m \subseteq \mathcal{L}_m$ with $m = \alpha n^2$ and $\mathcal{T} \subseteq \mathcal{L}$ such that $\mathcal{T}_m$ is $1/2$-inherited from $\mathcal{T}$. Let $\mathbf{P} \sim \mathbb{L}(n, m)$ be a partial Latin square obtained by m steps of the triangle removal process, and let $\mathbf{L} \in \mathcal{L}$ be a uniformly random order-n Latin square. Then*

$$\Pr[\mathbf{L} \in \mathcal{T}] \leq \exp(O(n^{2-\gamma})) \Pr[\mathbf{P} \in \mathcal{T}_m].$$

The purpose of defining inherited properties is just that we can compare our property on a whole Latin square to a property on an initial segment of the triangle removal process (a direct comparison does not make sense since the triangle removal process is unlikely to complete a full Latin square).

Next, instead of analyzing the triangle removal process directly it is more convenient to compare an initial fraction of it to an independent model. Recall the definition of $\mathbf{B}_{n,p}$ from Definition 7.1.

Lemma 7.7 ([31, Lemma 5.2]). *Let $\mathcal{T}$ be a property of unordered partial Latin squares that is monotone decreasing, i.e., if $P \in \mathcal{T}$ and $P' \subseteq P$ then $P' \in \mathcal{T}$. Fix $\alpha \in (0, 1)$, let $\mathbf{P} \sim \mathbb{L}(n, \alpha n^2)$, and let $\mathbf{G}^*$ be the partial Latin square obtained from $\mathbf{B}_{n,\alpha/n}$ by simultaneously deleting every hyperedge which intersects another hyperedge in more than one vertex. Then*

$$\Pr[\mathbf{P} \in \mathcal{T}] \leq O(\Pr[\mathbf{G}^* \in \mathcal{T}]).$$

To apply this machinery we must first show that the property of having few nondegenerate cuboctahedra satisfies the inheritance property defined in Definition 7.5.

Lemma 7.8. *Fix $\alpha, \delta \in (0, 1)$. Let $\mathcal{T}^\delta \subseteq \mathcal{L}$ be the property that a Latin square $L \in \mathcal{L}$ has at most $(1 - \delta)n^4$ nondegenerate cuboctahedra, let $m = \alpha n^2$, and let $\mathcal{T}_m^\delta \subseteq \mathcal{L}_m$ be the property that a partial Latin square $P \in \mathcal{L}_m$ has at most $\alpha^8(1 - \delta/2)n^4$ nondegenerate cuboctahedra. Then $\mathcal{T}_m^\delta$ is $(1/2)$-inherited from $\mathcal{T}^\delta$.*

Proof. Fix $L \in \mathcal{T}^\delta$. Let $\mathbf{L}_m$ be obtained by taking m uniformly random hyperedges of L, and let $\mathcal{Q}$ be the set of nondegenerate cuboctahedra in L. For $Q \in \mathcal{Q}$, let $\mathbf{1}_Q$ be the indicator variable for the event that $Q \subseteq \mathbf{L}_m$, and let $\mathbf{X} = \sum_{Q \in \mathcal{Q}} \mathbf{1}_Q$ be the number of nondegenerate cuboctahedra in $\mathbf{L}_m$. For all $Q \in \mathcal{Q}$ we have $\mathbb{E}\mathbf{1}_Q = \alpha^8 + O(1/n)$, so $\mathbb{E}\mathbf{X} \leq \alpha^8(1 - \delta + o(1))n^4$. Also, for each pair of disjoint $Q, Q' \in \mathcal{Q}$ we have $\operatorname{Cov}(\mathbf{1}_Q, \mathbf{1}_{Q'}) = O(1/n)$. In every Latin square, every nondegenerate cuboctahedron shares a hyperedge with at most $8n^3$ others (choose which hyperedge overlaps, then choose the identities of one vertex adjacent to each of those three vertices in the cuboctahedron; repeatedly applying the Latin property, we see that there is at most one choice for the remaining vertices). Since $L \in \mathcal{T}^\delta$ has $|\mathcal{Q}| \leq n^4$, there are $O(n^7)$ intersecting pairs of nondegenerate cuboctahedra in $\mathcal{Q}$. Thus $\operatorname{Var}\mathbf{X} \le |\mathcal{Q}|^2 \cdot O(1/n) + O(n^7) = O(n^7)$. By Chebyshev's inequality, we conclude that

$$\Pr[\mathbf{L}_m \in \mathcal{T}_m^\delta] = \Pr[\mathbf{X} \le \alpha^8(1-\delta/2)n^4] = 1 - o(1) > 1/2.$$

That is to say, $\mathcal{T}_m^\delta$ is $(1/2)$-inherited from $\mathcal{T}^\delta$. □

Finally, we will need the following concentration inequality. The statement presented here appears for example in [30, Theorem 2.11], and follows from an inequality of Freedman [15].

Theorem 7.9. *Let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_N)$ be a sequence of independent, identically distributed random variables with $\Pr[\boldsymbol{\omega}_i = 1] = p$ and $\Pr[\boldsymbol{\omega}_i = 0] = 1-p$. Let $f : \{0,1\}^N \to \mathbb{R}$ satisfy $|f(\boldsymbol{\omega}) - f(\boldsymbol{\omega}')| \le K$ for all pairs $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \{0,1\}^N$ differing in exactly one coordinate. Then*

$$\Pr[|f(\boldsymbol{\omega}) - \mathbb{E}f(\boldsymbol{\omega})| > t] \le \exp\biggl(-\frac{t^2}{4K^2Np + 2Kt}\biggr).$$

We are ready to prove Theorem 7.3.

Proof of Theorem 7.3. Let $\alpha > 0$, which will later be chosen to be small with respect to δ. Let $\mathbf{G}^*$ be as in Lemma 7.7. Let $\mathcal{T}_m^\delta$ be the property that a partial Latin square, not necessarily having exactly m edges, has at most $(1-\delta/2)\alpha^8 n^4$ nondegenerate cuboctahedra. This property is clearly monotone decreasing. Let $\mathbf{P} \sim \mathbb{L}(n, \alpha n^2)$ and $\mathbf{L}$ be a uniformly random order-n Latin square. Let $X_H(\mathbf{L})$ be the number of nondegenerate cuboctahedra in $\mathbf{L}$. Then Lemma 7.8 and Theorem 7.6 along with Lemma 7.7 show

$$\Pr[X_H(\mathbf{L}) \le (1-\delta)n^4] \le \exp(n^{2-\gamma})\Pr[\mathbf{P} \in \mathcal{T}_m^\delta] \le \exp(O(n^{2-\gamma})) \cdot O(\Pr[\mathbf{G}^* \in \mathcal{T}_m^\delta]).$$

It now suffices to study $\mathbf{G}^*$ (via $\mathbf{B}_{n,\alpha/n}$, which determines it). Let $\mathbf{Q}'$ be the maximum size of a collection of nondegenerate cuboctahedra in $\mathbf{G}^*$ for which every hyperedge is in at most $2\alpha^7 n^2$ of the cuboctahedra in the collection. We claim that

$$\mathbb{E}\mathbf{Q}' \ge \alpha^8(1-\delta/4)n^4, \tag{7.1}$$

if α is sufficiently small with respect to δ (and n is sufficiently large). For the moment, assume this is true.

Note that we can view $\mathbf{Q}'$ as a function of n^3 different independent zero-one random variables (one for each possible hyperedge of $\mathbf{B}_{n,\alpha/n}$). If we add a hyperedge e to $\mathbf{B}_{n,\alpha/n}$, this can result in at most one hyperedge being added to $\mathbf{G}^*$ (e itself) and it can result in at most 3 hyperedges being removed from $\mathbf{G}^*$ (hyperedges which share more than one vertex with e). Similarly, removing a hyperedge from $\mathbf{B}_{n,\alpha/n}$ affects $\mathbf{G}^*$ by at most 3 hyperedges. Now, adding a hyperedge to $\mathbf{G}^*$ increases $\mathbf{Q}'$ by at most $2\alpha^7 n^2$ (and can never decrease $\mathbf{Q}'$). So, $\mathbf{Q}'$ is a $O(n^2)$-Lipschitz. Theorem 7.9 shows that

$$\begin{aligned}\Pr[\mathbf{G}^* \in \mathcal{T}_m^\delta] &\le \Pr[\mathbf{Q}' \le \alpha^8(1-\delta/2)n^4] \le \Pr[\mathbf{Q}' \le \mathbb{E}\mathbf{Q}' - \alpha^8\delta n^4/4] \\ &\le \exp\left(-\frac{(\alpha^8\delta n^4/4)^2}{4(O(n^2))^2 n^3(\alpha/n) + 2(O(n^2))(\alpha^8\delta n^4/4)}\right) = \exp(-\Omega_{\alpha,\delta}(n^2)),\end{aligned}$$

and the result follows.

Now it suffices to prove (7.1). Let $\mathbf{Q} = X_H(\mathbf{G}^*)$ be the number of nondegenerate cuboctahedra in $\mathbf{G}^*$. For a triple T let $\mathbf{Q}_T$ be the number of nondegenerate cuboctahedra in $\mathbf{G} \cup \{T\}$ which include the hyperedge T, and let $\mathbf{Q}_2$ be the sum of $\mathbf{Q}_T$ over all $T \in \mathbf{B}_{n,\alpha/n}$ for which $\mathbf{Q}_T \ge 2\alpha^7 n^2$. We have

$$\mathbf{Q}' \ge \mathbf{Q} - \mathbf{Q}_2$$

since we can consider the collection of nondegenerate cuboctahedra in $\mathbf{G}^*$ and simply remove all cuboctahedra which have an edge that is in more than $2\alpha^7 n^2$ cuboctahedra of $\mathbf{B}_{n,\alpha/n}$. Now,

$$\mathbb{E}\mathbf{Q} = (1+O(1/n))n^{12}\cdot\left(\frac{\alpha}{n}\right)^8\left(1-\frac{\alpha}{n}\right)^{24n+O(1)} = (e^{-24\alpha}+O(1/n))\alpha^8 n^4$$

by linearity of expectation: a specific nondegenerate cuboctahedron will be in $\mathbf{G}^*$ precisely if all its 8 hyperedges are in $\mathbf{G}$ and all triples sharing an edge with one of these 8 hyperedges (of which there are $24n+O(1)$) are not in $\mathbf{G}$.

We now fix a triple T and study $\mathbf{Q}_T$, which is a degree-7 polynomial of independent random variables. We see $\mu = \mathbb{E}\mathbf{Q}_T = (1+O(1/n))n^9(\alpha/n)^7 = \alpha^7 n^2 + O(n)$, and a straightforward second-moment calculation yields $\operatorname{Var}\mathbf{Q}_T = O(n^3)$. Therefore

$$\Pr[\mathbf{Q}_T \ge 2\alpha^7 n^2] \lesssim_\alpha n^{-1}.$$

In particular, for sufficiently large n we have

$$\begin{aligned}\mathbb{E}\mathbf{Q}_2 &= n^3\left(\frac{\alpha}{n}\right)\mathbb{E}\mathbf{Q}_T\mathbb{1}_{\mathbf{Q}_T\ge 2\alpha^7n^2} \le 3\alpha n^2\mathbb{E}(\mathbf{Q}_T-\mu)\mathbb{1}_{\mathbf{Q}_T\ge 2\alpha^7n^2}\\ &\le 3\alpha n^2(\mathbb{E}(\mathbf{Q}_T-\mu)^2)^{1/2}\Pr[\mathbf{Q}_T\ge 2\alpha^7n^2]^{1/2} \le 3\alpha n^2(O(n^3))^{1/2}(O_\alpha(n^{-1}))^{1/2} \lesssim_\alpha n^3.\end{aligned}$$

Thus as long as $e^{-24\alpha} \ge 1-\delta/8$ we find

$$\mathbb{E}\mathbf{Q}' \ge \mathbb{E}\mathbf{Q} - \mathbb{E}\mathbf{Q}_2 \ge (1-\delta/4)\alpha^8 n^4$$

for n sufficiently large. This establishes (7.1) and we are done. □

Remark. It appears that the method of proof of Theorem 7.3 (and similarly [31, Theorem 1.2(a)]) may apply to more general colored triple systems H, though we do not pursue this here.

8. High Girth Latin Squares

Recall that a *Steiner triple system* of order N is an N-vertex triple system (i.e., 3-uniform hypergraph) such that every pair of vertices is contained in exactly one triple. Equivalently, this is a *triangle-decomposition* of the complete graph K_N. A foundational theorem of Kirkman [27] states that order-N Steiner triple systems exist if and only if $N \equiv 1,3 \pmod 6$. The necessity of the arithmetic condition is easy to see: a graph can have a triangle-decomposition only if all its degrees are even and the total number of edges is a multiple of three.

As mentioned in the introduction, in [33] the authors proved the analogue of Theorem 1.4 for Steiner triple systems. We begin by recalling the relevant definition and theorem.

Definition 8.1. The *girth* of a triple system $\mathcal{S}$ is the smallest $g > 3$ such that there exists a set of g vertices spanning $g-2$ triangles of $\mathcal{S}$. If $\mathcal{S}$ contains no such vertex set, we say it has infinite girth.

Theorem 8.2 ([33, Theorem 1.1])**.** *Given $g \in \mathbb{N}$, there is $N_g \in \mathbb{N}$ such that if $N \ge N_g$ and N is congruent to 1 or 3 $\pmod 6$, then there exists a Steiner triple system of order N and with girth greater than g.*

Order-N Latin squares are naturally equivalent to triangle-decompositions of the complete tripartite graph $K_{N,N,N}$ (the three parts correspond to rows, columns, and symbols). In this language, the definition of girth in the introduction is equivalent to Definition 8.1 and Theorem 1.4 is equivalent to the statement that for every fixed g and sufficiently large $N \in \mathbb{N}$ there exists a triangle-decomposition of $K_{N,N,N}$ with girth greater than g.

In this section we will outline the proof of [33, Theorem 1.1] and explain the adaptations necessary to prove Theorem 1.4. We omit some details (even salient ones!) that are the same in the tripartite and non-partite cases. For a more detailed outline, addressing some of the subtleties and difficulties involved, we refer the reader to [33, Section 2] (the full proof of Theorem 8.2 is of course contained in [33] as well).

We will need the following notation and definition.

Notation. For a graph $G \subseteq K_{N,N,N}$, we write $G_{i,j} = G[V^i \cup V^j]$ for the graph of edges between V^i and V^j. Throughout, indices that naturally come in threes due to a tripartite structure are taken modulo 3 (so, for example, every edge lies in $G_{j-1,j}$ for some $j \in \{1,2,3\}$).

Definition 8.3. We say that $G \subseteq K_N$ is *triangle-divisible* if all the vertex degrees in G are even and $|E(G)|$ is a multiple of three. We call $G \subseteq K_{N,N,N}$ *triangle divisible* if for every $1 \leq j \leq 3$ and $v \in V^j$, we have $\deg_G(v, V^{j-1}) = \deg_G(v, V^{j+1})$.

We remark that the previous definition contains a certain abuse: $K_{N,N,N}$ can be viewed as a subgraph of K_M, for $M \geq 3N$. However, it will always be clear from context whether we are thinking of G as a tripartite graph.

Note that if G is obtained from a triangle divisible graph H by removing a triangle-divisible subgraph of H (so, in particular, by removing a collection of edge-disjoint triangles) then G itself is triangle divisible.

8.1. **High-girth iterative absorption in a nutshell.** Let $g \in \mathbb{N}$ and let K be either K_N or $K_{N,N,N}$. If $K = K_N$ we additionally assume that $N \equiv 1, 3 \pmod 6$. Denote by V^1, V^2, V^3 the three parts of $K_{N,N,N}$. We wish to prove that if N is sufficiently large then there exists a triangle-decomposition of K with girth greater than g. We do so by describing a probabilistic algorithm that whp constructs such a decomposition. We build on the method of *iterative absorption*, and especially its application to triangle-decompositions described in [1].

Besides iterative absorption, the other major ingredient in the proof is the *high girth triangle removal process*. This is a simple random greedy algorithm for constructing partial triangle-decompositions with girth greater than g, defined as follows. Beginning with an empty collection of triangles $\mathcal{C}(0)$, for as long as possible, choose a triangle T^* in K uniformly at random, subject to the constraint that $\mathcal{C}(t) \cup \{T^*\}$ is a partial triangle-decomposition of K with girth greater than g. Then, set $\mathcal{C}(t+1) = \mathcal{C}(t) \cup \{T^*\}$. This process, with $K = K_N$, was analyzed independently by Glock, Kühn, Lo, and Osthus [16] and by Bohman and Warnke [4]. They showed that whp the process constructs an approximate triangle-decomposition of K (i.e., covering all but a $o(1)$-fraction of $E(K)$). A generalization of this process [33, Section 9] was used by the authors as a crucial component in the proof of Theorem 8.2. In particular, the more general process is applicable to the tripartite setting of Theorem 1.4, and allows a more general family of constraints than just those corresponding to girth.

The high-girth iterative absorption procedure can be broken down as follows:

The vortex. We fix a small constant $\rho > 0$, and a *vortex* $V(K) = U_0 \supseteq U_1 \supseteq \ldots \supseteq U_\ell$, where $\ell = O(1)$ is a large constant and $|U_{i+1}| = (1+o(1))|U_i|^{1-\rho}$ for each $i < \ell$. We note that $|V(K)| = (1+o(1))|U_\ell|^{(1-\rho)^{-\ell}} = |U_\ell|^{O(1)}$. If $K = K_{N,N,N}$, we additionally require that for every i, the intersection of U_i with each of V^1, V^2, and V^3 has the same cardinality.

The absorber graph. We set aside an *absorber* $H \subseteq K$. This is a graph, containing U_ℓ as an independent set, with the property that for every triangle-divisible graph $L \subseteq K[U_\ell]$, the graph $H \cup L$ admits a triangle-decomposition with girth greater than g. The construction for $K = K_N$ is given in [33, Theorem 4.1]. However, this construction is not tripartite, as would be required when $K = K_{N,N,N}$. Theorem 8.5, which we prove below, is the necessary tripartite analogue of [33, Theorem 4.1].

Initial sparsification. After setting aside the absorber, we perform the high-girth triangle removal process to obtain a partial Steiner triple system $\mathcal{I}$ covering all but a $o(1)$-fraction of the edges in $K \setminus H$. We obtain the lower bounds in [33, Theorem 1.3] and Theorem 1.1 by essentially multiplying the number of choices at each step of this process (we detail the calculation necessary to obtain Theorem 1.1 in Section 8.4). We remark that the initial sparsification has an important role beyond its usefulness in enumeration; see [33] for more details.

Cover down. The heart of the iterative absorption machinery is a "cover down" procedure. At each step k, given a set of edge-disjoint triangles $\mathcal{M}_0 \cup \cdots \cup \mathcal{M}_{k-1}$ in $K \setminus H$ covering all possible edges that are not contained in U_k, we find an augmenting set of edge-disjoint triangles $\mathcal{M}_k$ which covers all possible edges except ones contained in U_{k+1}. We do this in such a way that $\mathcal{M}_0 \cup \cdots \cup \mathcal{M}_k$ has girth greater than g.

After ℓ steps of this procedure, we will have found a set of edge-disjoint triangles $\mathcal{M}_0 \cup \cdots \cup \mathcal{M}_{k-1}$ in $E(K) \setminus E(H)$ covering all edges that are not contained in U_ℓ. We then use the defining property of the absorber to transform this into a triangle-decomposition of K. (Technically, beyond the fact that $\mathcal{M}_0 \cup \cdots \cup \mathcal{M}_{k-1}$ has high girth, we need to maintain the property that $\mathcal{M}_0 \cup \cdots \cup \mathcal{M}_{k-1} \cup \mathcal{B}$ has high girth, for each of the possible triangle-decompositions $\mathcal{B}$ associated with the absorber H).

Each iteration of the cover down procedure is proved using a three-stage randomized algorithm. The task is to find a set of edge-disjoint triangles in a particular graph G, covering all edges except those in a particular set U, and avoiding "obstructions to having high girth" with a particular collection of previously selected triangles.

- First, we run a generalized version of the high-girth triangle removal process to find an appropriate set of triangles $\mathcal{M}^*$ covering *almost* all of the edges which do not lie in U (and none of the edges inside U). Most of the leftover edges will be contained in a quasirandom *reserve graph* R, which is set aside before starting the process (this is a convenient way to control the approximate structure of the graph of leftover edges).

  In order for the process to succeed, it is necessary for its initial conditions to be quite regular. Specifically, for every edge, the number of available triangles including that edge (i.e., those that we permit ourselves to use) must vary by at most a multiplicative factor of $1 \pm |U|^{-c}$, with c an absolute constant.

  Naïvely, one might think to simply take our set of available triangles to be the set of triangles whose addition would not violate our girth condition. However, this set of triangles is not regular enough, and it is necessary to perform a *regularity boosting* step to obtain an appropriately regular subset of these triangles. Specifically, in [33, Lemma 5.1] (which is based on [1, Lemma 4.2]), we prove that given a set $\mathcal{T}$ of triangles satisfying certain weak regularity and "extendability" conditions, one can find a subset $\mathcal{T}' \subseteq T$ satisfying much stronger regularity conditions. This is accomplished by fixing an appropriate weight for each triangle in $\mathcal{T}$, and randomly subsampling with these weights as probabilities. The weights (which essentially correspond to a *fractional* triangle-decomposition of an appropriate graph) are constructed using weight-shifting "gadgets" defined in terms of copies of the complete graph K_5. Suitable gadgets are not available in the tripartite setting. Hence, we build on the more sophisticated techniques of Bowditch and Dukes [5] and Montgomery [42] to prove Lemma 8.11, which is the necessary tripartite regularity-boosting lemma. As stated in the introduction, this regularity-boosting lemma is the most substantial difference between the partite and non-partite cases.
- The remaining leftover edges (i.e., the edges in $G \setminus G[U]$ which are not covered by $\mathcal{M}^*$) can be classified into two types: "internal" edges which lie completely outside U, and "crossing" edges which have a single vertex in U. To handle the remaining internal edges, we use a random greedy algorithm to choose, for each leftover internal edge e, a suitable covering triangle T_e.

- At this point, the only leftover edges each have one vertex outside U and one vertex inside U. This allows us to reduce the problem to a simultaneous perfect matching problem: For every $v \in V(G) \setminus U$, let $W_v \subseteq U$ be the set of vertices u such that uv is still uncovered. Suppose that M_v is a perfect matching of W_v. Then $\{v \cup e : e \in M_v\}$ is a set of edge-disjoint triangles covering all the leftover crossing edges incident to v. Thus, it suffices to find a family of perfect matchings $\{M_v\}_{v \in V(G)\setminus U}$ such that the corresponding sets of triangles are edge-disjoint and satisfy appropriate girth properties.

  In [33] (i.e., the non-partite case $K = K_N$), the matchings M_v are found using a robust version of Hall's matching condition (cf. [33, Section 6]), applied to a uniformly random balanced bipartition of U. In the tripartite case we consider here the situation is simpler: the graph induced by each W_v is already bipartite due to the structure of $K_{N,N,N}$. Moreover, this bipartition of W_v is balanced, as W_v is obtained by removing a triangle-divisible graph from $K_{N,N,N}$. In particular we note that although the majority of the link graph is comprised of edges from the reserve graph, one does not need to maintain "divisibility" constraints for the reserve graph and instead divisibility follows simply by construction.

The final difference between the tripartite and non-partite cases concerns conditions regarding the graph of uncovered edges and the set of available triangles that must be verified before each stage of the iteration. One such set of conditions, called "iteration-typicality", is given in [33, Definition 10.1]. These must be replaced by a tripartite analogue, which we now define.

Definition 8.4 (Tripartite iteration-typicality)**.** Let $n \le N$. Fix a descending sequence of subsets $V(K_{n,n,n}) = U_k \supseteq \cdots \supseteq U_\ell$. Consider a graph $G \subseteq K_{n,n,n}$ and a set of triangles $\mathcal{A}$ in G. We say that $(G, \mathcal{A})$ is (p, q, ξ, h)-*tripartite-iteration-typical* (with respect to our sequence of subsets) if for every $1 \le j \le 3$:

- for every $k \le i < \ell$, every set $W \subseteq (V^j \cup V^{j+1}) \cap U_i$ of at most h vertices is adjacent (with respect to G) to a $(1 \pm \xi)p^h$ fraction of the vertices in $U_i \cap V^{j-1}$ and $U_{i+1} \cap V^{j-1}$, and
- for any i, i^* with $k \le i < \ell$, and $i^* \in \{i, i+1\}$, and any edge subset $Q \subseteq G[U_i] \cap G_{j,j+1}$ spanning $|V(Q)| \le h$ vertices, a $(1 \pm \xi)p^{|V(Q)|}q^{|Q|}$-fraction of the vertices $u \in U_{i^*} \cap V^{j-1}$ are such that $uvw \in \mathcal{A}$ for all $vw \in Q$.

To summarize, here are the changes required to the proof of Theorem 8.2 in order to prove Theorem 1.4:

- The vortex $V(K_{N,N,N}) = U_0 \supseteq \cdots \supseteq U_\ell$ should be chosen such that each U_k has the same number of vertices in each V^j.
- In the final step of the cover down procedure, in the non-partite case the problem was first reduced to a bipartite matching problem by taking a random bipartition of U_{i+1}. In the tripartite setting this is not necessary as the bipartite structure is already induced by $K_{N,N,N}$.
- During the iterations, replace iteration-typicality ([33, Definition 10.1]) with tripartite iteration-typicality (Definition 8.4) with $h = 6$. The analysis in [33, Section 9.1], where iteration-typicality is first established, requires only very minimal changes. Similarly, the verification that iteration-typicality is maintained throughout the iterations (which is done in [33, Section 10.4.2]) requires only small changes.
- In [33], the regularity boosting step that precedes the high-girth triangle removal process relies on "gadgets" which are not tripartite. We prove a tripartite regularity boosting lemma in Section 8.3.
- The absorbing structure must be tripartite. We give such a construction in Section 8.2.

The remainder of this section is devoted to constructing tripartite high-girth absorbers (Section 8.2), proving a tripartite regularity boosting lemma (Section 8.3), and providing the calculations that yield the lower bound in Theorem 1.1 (Section 8.4).

8.2. **Efficient tripartite high-girth absorbers.** Recall that a tripartite graph is *triangle-divisible* if for every vertex, its degree to both other parts is the same. This is a necessary but insufficient condition for triangle-decomposability. In this section we explicitly define a high-girth "absorbing structure" that will allow us to find a triangle-decomposition extending any triangle-divisible tripartite graph on a specific triple of vertex sets. Importantly, the size of this structure is only polynomial in the size of the distinguished vertex sets, which is needed for our proof strategy. For a set of triangles $\mathcal{R}$, let $V(\mathcal{R})$ be the set of all vertices in these triangles.

The following theorem encapsulates the properties of our absorbing structure, and is basically the same as [33, Theorem 4.1] (we just need to be a bit careful to ensure that everything is tripartite). Throughout, every tripartite graph will have a *fixed* tripartition (i.e., each vertex has a color in $\{1, 2, 3\}$). When we refer to a subgraph of a graph G, this subgraph inherits the tripartition G.

Theorem 8.5. *There is $C_{8.5} \in \mathbb{N}$ so that for $g \in \mathbb{N}$ there exists $M_{8.5}(g) \in \mathbb{N}$ such that the following holds. For any $m \geq 1$, there is a tripartite graph H with at most $M_{8.5}(g)m^{C_{8.5}}$ vertices containing a distinguished independent set $X^1 \cup X^2 \cup X^3$ (where each X^j contains exactly m vertices of color j), satisfying the following properties.*

- ***Ab1*** *For any triangle-divisible tripartite graph L on $X := X^1 \cup X^2 \cup X^3$ (with a consistent tripartition) there exists a triangle-decomposition $\mathcal{S}_L$ of $L \cup H$ which has girth greater than g.*
- ***Ab2*** *Let $\mathcal{B} = \bigcup_L \mathcal{S}_L$ (where the union is over all triangle-divisible tripartite graphs L on the vertex set X) and consider any tripartite graph K containing H as a subgraph. Say that a triangle in K is* nontrivially H-intersecting *if it is not one of the triangles in $\mathcal{B}$, but contains a vertex in $V(H) \setminus X$.*

  Then, for every set of at most g triangles $\mathcal{R}$ in K, there is a subset $\mathcal{L}_{\mathcal{R}} \subseteq \mathcal{B}$ of at most $M_{8.5}(g)$ triangles such that every Erdős configuration $\mathcal{E}$ on at most g vertices which includes $\mathcal{R}$ must either satisfy $\mathcal{E} \cap \mathcal{B} \subseteq \mathcal{L}_{\mathcal{R}}$ or must contain a nontrivially H-intersecting triangle $T \notin \mathcal{R}$.

Remark. Note that **Ab1**, with L as the empty graph on the vertex set X, implies that H itself is triangle-decomposable (hence triangle-divisible).

We will prove Theorem 8.5 by chaining together some special-purpose graph operations. Call a cycle in a tripartite graph a *tripartite cycle* if its length is divisible by 3 and every third vertex is in the same part.

Definition 8.6 (Path-cover)**.** Let the *path-cover* $\wedge X$ of a vertex set $X = X^1 \sqcup X^2 \sqcup X^3$ be the graph obtained as follows. Start with the empty graph on X. Then, for every unordered pair of u, v of distinct vertices in each part X^j, add $12|X|^2$ new paths of length 3 between u and v, introducing 2 new vertices for each (so in total, we introduce $\sum_j 24|X|^2\binom{|X^j|}{2}$ new vertices). We call these new length-3 cycles *augmenting paths*. Note that there are two ways to properly color a length-3 path between u and v; we include exactly $6|X|^2$ of each type, so the augmenting paths between u and v can be decomposed into $6|X|^2$ tripartite 6-cycles.

The key point is that for any triangle-divisible graph L on the vertex set X, the edges of $L \cup \wedge X$ can be decomposed into short cycles.

Lemma 8.7. *If a tripartite graph L on X is triangle-divisible, then $L \cup \wedge X$ can be decomposed into tripartite cycles of length at most 9. Additionally, if L is triangle-divisible, then so is $L \cup \wedge X$.*

The proof is analogous to that of [33, Lemma 4.3]. First, we note that triangle-divisible graphs have a decomposition into tripartite cycles (start at a vertex and move around the graph cyclically until a collision occurs, then remove this cycle and continue). We then "shorten" these cycles with our augmenting paths.

Next, it is a consequence of [2, Lemma 6.5] that for any triangle-divisible tripartite graph H, there is a tripartite graph $A(H)$ containing the vertex set of H as an independent set such that $A(H)$ and $A(H) \cup H$ are both triangle-decomposable. (In our case, we only care about the case $H \in \{C_3, C_6, C_9\}$, in which case it is not hard to construct $A(H)$ by hand.)

Definition 8.8 (Cycle-cover)**.** Let the *cycle-cover* of a tripartite vertex set Y be the graph $\triangle Y$ obtained as follows. Beginning with Y, for every $H \in \{C_3, C_6, C_9\}$ and every color-preserving injection $f : V(H) \to Y$ we add a copy of $A(H)$, such that each $v \in V(H) \subseteq V(A(H))$ in the copy of $A(H)$ coincides with $f(v)$ in Y. We do this in a vertex-disjoint way, introducing $|V(A(H))\backslash V(H)|$ new vertices each time. (Think of the copy of $A(H)$ as being "rooted" on a specific set of vertices in Y, and otherwise being disjoint from everything else.)

Lemma 8.9. *Let $Y = V(\wedge X)$ be the vertex set of the graph $\wedge X$. If a graph L on X is triangle-divisible, then $L \cup \wedge X \cup \triangle Y$ admits a triangle-decomposition.*

The proof of Lemma 8.9 is the same as that of [33, Lemma 4.5], so we omit the details.

If we consider any $X = X_1 \sqcup X_2 \sqcup X_3$ with $|X_i| = m$, then Lemma 8.9 implies that $\wedge X \cup \triangle(V(\wedge X))$ can "absorb" any triangle-divisible tripartite graph on X, though not necessarily in a high-girth manner. We use a tripartite modification of the *g-sphere-cover* of [33, Definition 4.6] to provide a girth guarantee (the definition is basically the same, but we glue a g-sphere only onto those triples which are tripartite).

Definition 8.10 (Sphere-cover)**.** Let the *g-sphere-cover* of a tripartite vertex set Z be the graph $\bigcirc_g Z$ obtained by the following procedure. For every tripartite triple T of vertices of Z, arbitrarily label these vertices as a, b_1, b_2. Then, append a "g-sphere" to the triple. Namely, first add $2g-1$ new vertices $b_3, \ldots, b_{2g}, c$. Then add the edges ab_j for $3 \le j \le 2g$, the edges cb_j for $1 \le j \le 2g$, the edges $b_j b_{j+1}$ for $2 \le j \le 2g-1$, and the edge $b_{2g}b_1$. This can be seen to preserve the tripartite property.

Note that every such g-sphere Q itself has a triangle-decomposition: specifically, we define the *out-decomposition* to consist of the triangles

$$cb_2b_3,\ ab_3b_4,\ cb_4b_5,\ ab_5b_6, \ldots, cb_{2g}b_1.$$

We also identify a particular triangle-decomposition of the edges $Q \cup T$: the *in-decomposition* consists of the triangles

$$cb_1b_2,\ ab_2b_3,\ cb_3b_4,\ ab_4b_5, \ldots, ab_{2g}b_1.$$

For a triple T in Z, let $\mathcal{B}^{\bigcirc}(T)$ be the set of all triangles in the in- and out-decompositions of the g-sphere associated to T. We emphasize that $T \notin \mathcal{B}^{\bigcirc}(T)$.

Finally, the proof of Theorem 8.5 is exactly like the proof of [33, Theorem 4.1]: let $Y = V(\wedge X)$, $Z = V(\triangle Y)$, and take $H = \wedge X \cup \triangle Y \cup \bigcirc_g Z$. The girth properties guaranteed by analogues of [33, Lemmas 4.7, 4.8] are enough to show the desired girth properties here.

8.3. **Triangle-regularization of tripartite graphs.** Given a tripartite set of triangles with suitable regularity and "extendability" properties, the following lemma finds a subset which is substantially more regular. This lemma can be used in place of [33, Lemma 5.1] (which itself is an adaptation of [1, Lemma 4.2]) for the proof of Theorem 1.4.

Lemma 8.11. *There are $C_{8.11} \in \mathbb{N}$ and $n_{8.11} : \mathbb{N} \times \mathbb{R} \to \mathbb{N}$ such that the following holds. We are given $q \in (0,1)$ and $C \ge C_{8.11}$, and let $n \ge n_{8.11}(C,q)$. Suppose $q \in (0,1)$, $\xi = C^{-8}$, and $p \in (n^{-1/12}, 1)$, let G be a balanced tripartite graph on $3n$ vertices $V_1 \sqcup V_2 \sqcup V_3$, and let $\mathcal{T}$ be a collection of triangles of G, satisfying the following properties.*

(1) Every edge $e \in E(G)$ is in $(1 \pm \xi)p^2qn$ triangles of $\mathcal{T}$.
(2) For every $j \in \{1,2,3\}$ and every set $S \subseteq V^{j-1} \cup V^j$ with $|S| \le 6$ in G, there are $(1 \pm \xi)p^{|S|}n$ common neighbors of S in V_{j+1}.

(3) *For every $i \in [3]$ and every set Q of at most 6 edges between V^{j-1} and V^j with $|Q| \le 6$, there are between $C^{-1}p^{|V(E)|}n$ and $Cp^{|V(Q)|}n$ vertices $u \in V_{i+1}$ which form a triangle in $\mathcal{T}$ with every edge in Q.*

Then, there is a subcollection $\mathcal{T}' \subseteq \mathcal{T}$ such that every edge $e \in E(G)$ is in $(1 \pm n^{-1/4})p^2qn/4$ triangles of $\mathcal{T}'$.

As in [33, Lemma 5.1] and [1, Lemma 4.2], we will prove Lemma 8.11 by sampling from a fractional clique-decomposition. However, the construction of this clique decomposition will be substantially more involved. We will borrow ideas from general work of Montgomery [42] on fractional clique-decompositions of partite graphs (in particular, making use of "gadgets" to shift weight around), but certain simplifications are possible in the setting of triangle-decompositions. Also, instead of directly describing an *explicit* fractional triangle-decomposition (as in [33, Lemma 5.1], [1, Lemma 4.2], and [42]), we construct our fractional triangle-decomposition as the limit of an iterative "adjustment" process. This was inspired by some related ideas of Bowditch and Dukes [5].

Proof of Lemma 8.11. For $v \in V(G)$ let $\mathcal{T}_v$ be the set of triangles of $\mathcal{T}$ involving v and for $e \in E(G)$ let $\mathcal{T}(e)$ be the set of triangles of $\mathcal{T}$ containing e. Given $\phi : \mathcal{T} \to \mathbb{R}$, define the *vertex-weight* of $v \in V(G)$ and *edge-weight* of $e \in E(G)$ to be

$$\phi^{\mathrm{vtx}}(v) = \sum_{T \in \mathcal{T}_v} \phi(T), \qquad \phi^{\mathrm{edge}}(e) = \sum_{T \in \mathcal{T}(e)} \phi(T).$$

The total weight is $\phi^{\mathrm{sum}} = \sum_{T \in \mathcal{T}} \phi(T)$. We say ϕ is *vertex-balanced* or *edge-balanced*, respectively, if

$$\phi^{\mathrm{vtx}}(v) = \frac{3\deg_G(v)}{2|E(G)|}\phi^{\mathrm{sum}} \ \forall v \in V(G), \qquad \phi^{\mathrm{edge}}(e) = \frac{3}{|E(G)|}\phi^{\mathrm{sum}} \ \forall e \in E(G).$$

Summing over the edges incident to each vertex shows that edge-balancedness implies vertex-balancedness.

Our goal is to produce $\phi_* : \mathcal{T} \to [0,1]$ with $\phi_*^{\mathrm{edge}} \equiv p^2qn/4$ (so in particular ϕ_* is edge-balanced). This suffices to prove the lemma: if we take $\mathcal{T}'$ to be a random subcollection of $\mathcal{T}$ in which each $T \in \mathcal{T}$ included with probability $\phi_*(T)$ independently, then the Chernoff bound and union bound show that $\mathcal{T}'$ satisfies the desired properties with high probability.

So, for the rest of the proof we construct our desired edge-balanced weight function ϕ_*. We do this in several steps: first, we build a vertex-balanced function by averaging over certain simple functions $\chi_{u,v}$. Then, we iteratively adjust this function to make it edge-balanced (our final edge-balanced function will be a fixed point of a certain contraction map). Finally, we divide by an appropriate constant to obtain our desired function $\phi_* : \mathcal{T} \to [0,1]$ with $\phi_*^{\mathrm{edge}} \equiv p^2n/4$.

Step 1: Constructing a vertex-balanced function. For $j \in \{1,2,3\}$ and distinct $u, v \in V^j$, let $\mathcal{T}_{2,1,1}(u,v)$ be the set of copies of $K_{2,1,1}$ in G containing u, v, such that both triangles of the copy of $K_{2,1,1}$ are in $\mathcal{T}$ (this can only happen when u, v are in the size-2 part of the $K_{2,1,1}$). For $H \in \mathcal{T}_{2,1,1}(u,v)$ let $H(u), H(v)$ be the triangles involving u and v respectively.

For a vertex u, let

$$f_u = |\mathcal{T}_u| - \frac{3\deg_G(u)}{2|E(G)|}|\mathcal{T}|,$$

and note that for each $j \in \{1,2,3\}$, we have $\sum_{u \in V^j} f_u = 0$.

For $j \in \{1,2,3\}$ and distinct $u, v \in V^j$, define $\chi_{u,v} : \mathcal{T} \to \mathbb{R}$ by

$$\chi_{u,v}(T) = \frac{1}{|\mathcal{T}_{2,1,1}(u,v)|} \sum_{H \in \mathcal{T}_{2,1,1}(u,v)} (\mathbb{1}_{T=H(u)} - \mathbb{1}_{T=H(v)}),$$

and for $j \in \{1,2,3\}$, define $\phi_{V^j} : \mathcal{T} \to \mathbb{R}$ by

$$\chi_{V^j} = -\frac{1}{2n} \sum_{u,v \in V^j : u \neq v} (f_u - f_v)\chi_{u,v}.$$

We note some key facts about the vertex-weights of these functions: for any $w \in V(G)$ we have

$$\chi^{\mathrm{vtx}}_{u,v}(w) = \begin{cases} 1 & \text{if } w = u, \\ -1 & \text{if } w = v, \\ 0 & \text{otherwise,} \end{cases}$$

and thus, recalling that $\sum_{u \in V_i} f_u = 0$, we have

$$\chi^{\mathrm{vtx}}_{V^j}(w) = \begin{cases} -f_w & \text{if } w \in V^j, \\ 0 & \text{if } w \notin V^j. \end{cases}$$

Let $\mathbf{1} : \mathcal{T} \to \mathbb{R}$ be the all-1 function and define $\phi_0 = \mathbf{1} + \chi_{V^1} + \chi_{V^2} + \chi_{V^3}$. Note that ϕ_0 is vertex-balanced: for any $w \in V(G)$ we have

$$\phi_0^{\mathrm{vtx}}(w) = \sum_{T \in \mathcal{T}_w} (1 + \chi_{V^1}(T) + \chi_{V^2}(T) + \chi_{V^3}(T)) = |\mathcal{T}_w| + 0 + 0 - f_w = \frac{3\deg_G(w)}{2|E(G)|}|\mathcal{T}|.$$

Step 2: Defining "adjuster" functions. For a 6-cycle $J \subseteq G$ alternating between some V^{j-1} and V^j, let $\mathcal{T}_{3,3,1}(J)$ be the set of vertices $v \in V^{j+1}$ which complete a triangle in $\mathcal{T}$ with each edge of J. For an edge $e \in G_{j-1,j}$, let $G(e)$ be the set of 6-cycles $J \subseteq G$, alternating between V^{j-1} and V^j, which contain e.

For $e \in E(G)$, and a 6-cycle $J \in G(e)$, we define $\psi_{J,e} : \mathcal{T} \to \mathbb{R}$ by

$$\psi_{J,e}(T) = \frac{1}{|\mathcal{T}_{3,3,1}(J)|} \sum_{v \in \mathcal{T}_{3,3,1}(J)} \sum_{e' \in J} (-1)^{d_J(e',e)} \mathbb{1}_{T = e' \cup \{v\}},$$

where $d_J(e',e)$ is the minimum number of times e' must be rotated around the cycle in either direction to coincide with e. For an edge $e \in E(G)$ between V^{j-1} and V^j, define

$$\psi_e = \frac{1}{|G(e)|} \sum_{J \in G(e)} \psi_{J,e}.$$

We now claim that these two functions have zero vertex-weights (and can thus be used to modify edge-weights without modifying vertex-weights). Indeed, for $e \in E(G)$, $J \in G(e)$, $v \in \mathcal{T}_{3,3,1}(J)$ and any vertex $w \in V(G)$, we have

$$\sum_{T \in \mathcal{T}_w} \sum_{e' \in J} (-1)^{d_J(e',e)} \mathbb{1}_{e_j \cup \{v\}}(T) = 0.$$

(Basically, the idea is that this alternating sum around a 6-cycle contributes 0 to every vertex-weight.) It follows that $\psi^{\mathrm{vtx}}_{J,e}(w) = \psi^{\mathrm{vtx}}_e(w) = 0$, as claimed.

Step 3: One-step adjustment of a vertex-balanced function. For a function $\phi : \mathcal{T} \to \mathbb{R}$, define the edge-discrepancy function $\phi^{\mathrm{disc}} : E(G) \to \mathbb{R}$ by

$$\phi^{\mathrm{disc}}(e) = \phi^{\mathrm{edge}}(e) - \frac{3}{|E(G)|}\phi^{\mathrm{sum}}.$$

Note that ϕ is edge-balanced if and only if $\phi^{\mathrm{disc}} \equiv 0$, and note that for $v \in V^j$ and $i \neq j$, and any vertex-balanced ϕ, we have

$$\sum_{e \in G_{i,j} : v \in e} \phi^{\mathrm{disc}}(e) = \phi^{\mathrm{vtx}}(v) - \frac{3\deg_G(v)}{2|E(G)|}\phi^{\mathrm{sum}} = 0. \tag{8.1}$$

For a vertex-balanced function $\phi : \mathcal{T} \to \mathbb{R}$ we define its one-step adjustment $A(\phi) : \mathcal{T} \to \mathbb{R}$ by

$$A(\phi) = \phi - \sum_{e \in E(G)} \phi^{\mathrm{disc}}(e)\psi_e.$$

Note that $A(\phi)$ is again vertex-balanced, since the ψ_e have vertex-weight zero.

Step 4: Bounding discrepancy after adjustment. We now analyze how edge-discrepancy changes between ϕ and $A(\phi)$. First, we claim that for any for $e \in E(G)$ and $J \in G(e)$, and any $f \in E(G)$, we have

$$\psi_{J,e}^{\mathrm{edge}}(f) = \begin{cases} (-1)^{d_J(f,e)} & \text{if } f \in J, \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, this follows from the fact that for $e \in E(G)$ $J \in G(e)$, $v \in \mathcal{T}_{3,3,1}(J)$, and any $f \in E(G)$, we have

$$\sum_{T \in \mathcal{T}(f)} \sum_{e' \in J} (-1)^{d_J(e',e)} \mathbb{1}_{T = e' \cup \{v\}}(T) = \begin{cases} (-1)^{d_J(f,e)} & \text{if } f \in J, \\ 0 & \text{otherwise.} \end{cases}.$$

(Here we used that the edges between v and a vertex of J are zeroed out by the alternating subtraction.) It follows that for $e, f \in G_{j-1,j}$ we have

$$\psi_e^{\mathrm{edge}}(f) = \frac{c_0(e,f) - c_1(e,f) + c_2(e,f) - c_3(e,f)}{|G(e)|}$$

where $c_i(f,e)$ is the number of 6-cycles $J \subseteq G_{j-1,j}$ including f and e with $d_J(e,f) = i$. Note that if e, f are not between the same pair of parts then $\psi_f^{\mathrm{edge}}(e) = 0$, and that if $e = f$ then $\psi_e^{\mathrm{edge}}(f) = 1$.

Note that $\psi_e^{\mathrm{sum}} = 0$ for all e, so for any $f \in G_{j-1,j}$ we have

$$\begin{aligned} A(\phi)^{\mathrm{disc}}(f) = \phi^{\mathrm{disc}}(f) - \sum_{e \in E(G)} \phi^{\mathrm{disc}}(e)\psi_e^{\mathrm{edge}}(f) &= - \sum_{\substack{e \in G_{j-1,j} \\ e \neq f}} \phi^{\mathrm{disc}}(e)\psi_e^{\mathrm{edge}}(f) \\ &= - \sum_{\substack{e \in G_{j-1,j} \\ e \neq f}} \phi^{\mathrm{disc}}(e) \frac{-c_1(e,f) + c_2(e,f) - c_3(e,f)}{|G(e)|}. \end{aligned} \tag{8.2}$$

We simplify this expression by defining some further statistics. Suppose $e, f \in G_{j-1,j}$ are distinct. If e, f share a vertex (which we denote $e \sim f$), let $g_1(e,f)$ be the number of extensions of $e \cup f$ to a 6-cycle in $G_{j-1,j}$ (note $g_1(e,f) = c_1(e,f)$). If e, f do not share a vertex but there are vertices $u \in e$ and $v \in f$ that are adjacent to each other in G (which we denote $e \sim_{u,v} f$), then let $g_2(e,f,u,v)$ be the number of 6-cycles in $G_{j-1,j}$ containing the edges e, uv, f. If e, f share no vertex (which we denote $e \not\sim f$), let $g_3(e,f)$ be the number of 6-cycles $J \subseteq G_{j-1,j}$ in which $d_J(e,f) = 3$ (i.e., e and f are on opposite sides of the 6-cycle). It will be convenient to write $g_1(e,e) = g_2(e,e,u,v) = g_3(e,e) = 0$ as well as $g_3(e,f) = 0$ when $f \sim e$. Finally, given a vector $\vec{a} \in \mathbb{R}^t$ let $\mathrm{disc}(\vec{a}) = \sum_{j=1}^t |a_j - \mu|$ for $\mu = (a_1 + \cdots + a_t)/t$. We will use the key fact that if $b_1 + \cdots + b_t = 0$ then

$$|a_1 b_1 + \cdots + a_t b_t| \leq \sum_{i=1}^t |(a_i - \mu) b_i| \leq \|\vec{b}\|_\infty \,\mathrm{disc}(\vec{a}). \tag{8.3}$$

Now, continuing from (8.2), using (8.3) with (8.1), we obtain

$$|A(\phi)^{\mathrm{disc}}(f)| = \left| \sum_{e \sim f} \frac{g_1(e,f)}{|G(e)|} \phi^{\mathrm{disc}}(e) - \sum_{u,v,e : e \sim_{u,v} f} \frac{g_2(e,f,u,v)}{|G(e)|} \phi^{\mathrm{disc}}(e) + \sum_{e \not\sim f} \frac{g_3(e,f)}{|G(e)|} \phi^{\mathrm{disc}}(e) \right|$$

$$\le \|\phi^{\mathrm{disc}}\|_\infty \left(\sum_{v\in f} \mathrm{disc}\left(\left(\frac{g_1(e,f)}{|G(e)|}\right)_{\substack{e\in G_{j-1,j}:\\ v\in e}}\right) + \sum_{\substack{u,v:\, v\in f,\\ uv\in G_{j-1,j}}} \mathrm{disc}\left(\left(\frac{g_2(e,f,u,v)}{|G(e)|}\right)_{\substack{e\in G_{j-1,j}:\\ u\in e}}\right)\right.$$
$$\left. + \sum_{u\in V_i\setminus e} \mathrm{disc}\left(\left(\frac{g_3(e,f)}{|G(e)|}\right)_{\substack{e\in G_{j-1,j}:\\ u\in e}}\right)\right).$$

Now we use the quasirandomness hypotheses in the lemma statement, which we can apply repeatedly to count copies of any graph H in $G_{j-1,j}$ extending any given set of vertices and edges (we are assuming $p \ge n^{-1/12}$, so the contribution from "degenerate 6-cycles" with repeated vertices is negligible). For distinct edges $e \sim f$ we have

$$\frac{g_1(e,f)}{|G(e)|} = \frac{(1\pm O(\xi))p^4n^3}{(1\pm O(\xi))^4p^5n^4} = \frac{1\pm O(\xi)}{pn}.$$

For $e \sim_{u,v} f$ we have

$$\frac{g_2(e,f,u,v)}{|G(e)|} = \frac{(1\pm O(\xi))p^3n^2}{(1\pm O(\xi))p^5n^4} = \frac{1\pm O(\xi)}{p^2n^2},$$

and for $e \not\sim f$ we have

$$\frac{g_3(e,f)}{|G(e)|} = \frac{(1\pm O(\xi))p^4n^2}{(1\pm O(\xi))p^5n^4} = \frac{1\pm O(\xi)}{pn^2}.$$

We deduce

$$|A(\phi)^{\mathrm{disc}}(e)| \le \|\phi^{\mathrm{disc}}\|_\infty\left(O(pn)\cdot\frac{O(\xi)}{pn} + O(p^2n^2)\cdot\frac{O(\xi)}{p^2n^2} + O(pn^2)\cdot\frac{O(\xi)}{pn^2}\right) \lesssim \xi\|\phi^{\mathrm{disc}}\|_\infty.$$

That is to say, if ξ is sufficiently small, then repeated application of the "adjustment" map A reduces the discrepancy of ϕ.

Step 5: Final estimates. We need a few further bounds before completing the proof (we will use these to ensure that our final weight function ϕ_* has its values in $[0,1]$).

First, we consider ϕ_0. Recalling the definition of χ_{V^j}, for each $j \in \{1,2,3\}$ we have

$$|\chi_{V^j}(T)| \lesssim \frac{1}{n}\sup_{u\in V(G)}|f_u| \cdot \sum_{u,v\in V^j}\frac{1}{|\mathcal{T}_{2,1,1}(u,v)|}\sum_{H\in\mathcal{T}_{2,1,1}(u,v)}\mathbb{1}_{T\in\{H(u),H(v)\}}.$$

By the assumptions of the lemma, for any u we have $|\mathcal{T}_u| = (1\pm O(\xi))(pn)(p^2qn)$ (considering the edges containing u, and then considering, for each such edge, the number of triangles including that edge), and therefore $|f_u| = O(\xi p^3qn^2)$. For any u,v we have $|\mathcal{T}_{2,1,1}(u,v)| \ge (1-\xi)(p^2n)(C^{-1}p^3n)$ (considering all length-2 paths between u,v, and then considering, for each such 2-path, the number of triangles which extend both the edges that 2-path to triangles). Also, $\sum_{u,v\in V^j}\sum_{H\in\mathcal{T}_{2,1,1}(u,v)}\mathbb{1}_{T\in\{H(u),H(v)\}}$ can be interpreted as a count of copies of $K_{2,1,1}$ which contain T; this is always at most Cp^2n.

So, recalling the definition of ϕ_0 and the all-1 function $\mathbf{1}:\mathcal{T}\to\mathbb{R}$, we have

$$\|\phi_0 - \mathbf{1}\|_\infty = \|\chi_{V^1}+\chi_{V^2}+\chi_{V^3}\|_\infty \lesssim C^2\xi. \tag{8.4}$$

We also upper-bound $\|\phi_0^{\mathrm{disc}}\|_\infty$: first recall that every edge is in $(1\pm\xi)p^2qn$ triangles, so it follows from (8.4) that $\|(\phi_0-\mathbf{1})^{\mathrm{disc}}\|_\infty \lesssim C^2\xi p^2n$. We then again use that every edge is in $(1\pm\xi)p^2n$ triangles to see that $\|\mathbf{1}^{\mathrm{disc}}\|_\infty \lesssim \xi p^2qn$, from which we deduce that $\|\phi_0^{\mathrm{disc}}\|_\infty \lesssim C^2\xi p^2n$.

Now, we prove a bound on $\|A(\phi)-\phi\|_\infty$ for vertex-balanced ϕ. Say a *6-pyramid* $\mathcal{J}$ is the set of all triangles containing v and a vertex of J, for some 6-cycle $J\in G_{j-1,j}$ and some vertex $v\in\mathcal{T}_{3,3,1}(J)$.

Let N_T be the number of 6-pyramids (with respect to the entire graph G and set of triangles $\mathcal{T}$) containing a particular triangle T. Note that for any $e \in E(G)$ we have

$$|\psi_e(T)| \le \frac{1}{|G(e)| \inf_{J \in G(e)} |\mathcal{T}_{3,3,1}(J)|} \sum_{\mathcal{J}} \mathbb{1}_{T \in \mathcal{J}} \le \frac{1}{C^{-1}p^6 n \cdot (1-O(\xi))^4 p^5 n^4} \sum_{\mathcal{J}} N_T \le \frac{2C}{p^{11}n^5} N_T,$$

where the sums are over all 6-pyramids $\mathcal{J}$ whose underlying 6-cycle contains e.

To specify a 6-pyramid containing T, whose underlying 6-cycle contains a particular edge $f \in T$, we need to specify four additional vertices. The first three of these vertices must each extend a particular edge to a triangle, and then the final vertex must extend two different edges to triangles. So, by the assumptions of the lemma (and accounting for the three choices of f) we have $N_T \le 3(Cp^2n)^3(Cp^3) \le C^4p^9n^4$.

Recalling the definition of $A(\phi)$, it follows that for any vertex-balanced ϕ we have

$$|A(\phi)(T) - \phi(T)| \le \|\phi^{\mathrm{disc}}\|_\infty \cdot 6 \cdot \frac{2C}{p^{11}n^5} N_T \lesssim \frac{C^5}{p^2 n}\|\phi^{\mathrm{disc}}\|_\infty.$$

(here we are using that each 6-pyramid contributes to ψ_e for 6 different e). We summarize

$$\|A(\phi)^{\mathrm{disc}}\|_\infty \lesssim \xi \|\phi^{\mathrm{disc}}\|_\infty, \quad \|A(\phi) - \phi\|_\infty \lesssim \frac{C^5}{p^2n}\|\phi^{\mathrm{disc}}\|_\infty$$

for any vertex-balanced ϕ, and

$$\|\phi_0 - \mathbf{1}\|_\infty \lesssim C^2\xi, \quad \|\phi_0^{\mathrm{disc}}\|_\infty \lesssim C^2\xi p^2 n.$$

Step 6: Iteration. Let $\phi_k = A^{(k)}(\phi_0)$ be the result of iterating the "adjustment map" k times. We claim that ϕ_k approaches a limiting weight function $\phi_\infty \in \mathbb{R}^{\mathcal{T}}$ as $k \to \infty$. Indeed, note that $\|\phi_k^{\mathrm{disc}}\|_\infty \le (O(\xi))^k \|\phi_0^{\mathrm{disc}}\|_\infty$ and hence $\|\phi_{k+1} - \phi_k\|_\infty \lesssim \frac{C^5}{p^2n}(O(\xi))^k\|\phi_0^{\mathrm{disc}}\|_\infty$. Hence, if ξ is sufficiently small (i.e., C is large), the sequence $(\phi_k)_{k\in\mathbb{N}}$ is Cauchy in $\ell^\infty(\mathcal{T})$, from which the claim follows.

Note that ϕ_∞ is edge-balanced, as $\|\phi_\infty^{\mathrm{disc}}\|_\infty = \lim_{k\to\infty}\|\phi_k^{\mathrm{disc}}\|_\infty = 0$. Also, summing our bound on $\|\phi_{k+1} - \phi_k\|_\infty$ over k, we obtain

$$\|\phi_\infty - \phi_0\|_\infty \lesssim \frac{C^5}{p^2n}\|\phi_0^{\mathrm{disc}}\|_\infty \lesssim C^7\xi.$$

Recalling that $\|\phi_0 - \mathbf{1}\|_\infty \lesssim C^2\xi$ and that $\xi = C^{-8}$, we deduce that ϕ_∞ takes values in $1 \pm C^8\xi/10 = 1 \pm 1/10$ for sufficiently large C. It follows that $\alpha := \phi_\infty^{\mathrm{edge}}(e)$ (which is the same for all e) is of the form $(1 \pm 1/10)(1 \pm \xi)p^2qn$ and that $\beta := 4\alpha/(p^2qn) = (1 \pm 1/10)(1 \pm \xi)4$. Therefore $\phi_* = \phi_\infty/\beta$ is a suitable function with $\phi_*^{\mathrm{edge}} = p^2qn/4$, as desired. □

8.4. **Counting intercalate-free Latin squares.** In this section we perform the necessary calculations to prove Theorem 1.1. An analogous result for high-girth Steiner triple systems was proved by the authors in [33].

We can interpret the random construction described in Section 8.1 as producing a triangle-decomposition of $K_{N,N,N}$ in which the set of triangles is *ordered*. Indeed, the triangles in $\mathcal{I}$ arising from the initial sparsification process come first in the ordering (in the order they are chosen in the process), and then the remaining triangles come afterwards (in some arbitrary order; this order will not be important for us). We will lower-bound the number of ordered triangle-decompositions that can arise from our random construction, and then divide by $(N^2)!$ to obtain a lower bound on the number of intercalate-free Latin squares. This type of argument was first used by Keevash [26] to count Steiner triple systems.

The analysis in [33, Section 9] implies that whp the initial sparsification process succeeds in constructing a partial Latin square $\mathcal{I}$ of girth greater than 6 (i.e., an intercalate-free partial Latin square) with $M := \lfloor (1-N^{-\nu})N^2 \rfloor$ triangles, for some absolute constant $\nu > 0$. Furthermore, the analysis shows that whp for every $0 \le t \le M$, the number of choices available to the process at step t is $(1 \pm N^{-\nu})A(t)$, with

$$A(t) = N^3\left(1 - \frac{t}{N^2}\right)^3 \exp\left(-\frac{t^3}{N^6}\right).$$

We say an outcome of the initial sparsification process is *good* if it satisfies the above properties.

Remark. There is a compelling heuristic explanation for the above formula. Denote by $\mathcal{C}(t)$ the set of triangles chosen in the first t steps of the process. Observe that a triangle T is available to be selected only if none of its edges has been covered by $\mathcal{C}(t)$ and there does not exist an intercalate containing T with the other three triangles in $\mathcal{C}(t)$. Note that the density of edges not covered by $\mathcal{C}(t)$ is approximately $(1 - t/N^2)$. Thus, if the graph of uncovered edges is pseudorandom, then there are approximately $N^3\big(1 - t/N^2\big)^3$ triangles with all edges uncovered. Additionally, if we fix an uncovered triangle T then the number of intercalates containing T (in $K_{N,N,N}$) is approximately N^3. If we think of $\mathcal{C}(t)$ as a random set of triangles of the same density $|\mathcal{C}(t)|/N^3 = t/N^3$ then the expected number of intercalates containing T whose other three triangles are in $\mathcal{C}(t)$ is approximately $N^3\big(t/N^3\big)^3 = t^3/N^6$. Thus, the factor $\exp\big(-t^3/N^6\big)$ captures the heuristic that the number of intercalates containing T, whose other three triangles are in $\mathcal{C}(t)$, has an approximate Poisson distribution with parameter t^3/N^6.

Now, in our initial sparsification process, we choose a uniformly random triangle at each step (among the valid choices at that step). The probability that this process produces a specific good ordered set of triangles is therefore $\prod_{t=0}^{M-1}((1 \pm N^{-\nu})A(t))^{-1}$. Since the process produces a good ordered set of triangles with probability $1 - o(1)$, it follows that the number of good outcomes that can arise is

$$\begin{aligned}(1-o(1))\prod_{t=0}^{M-1}(1 \pm N^{-\nu})A(t) &= \big((1 \pm N^{-\nu})N^3\big)^{N^2} \exp\left(\sum_{t=0}^{M-1} \log\left(\left(1 - \frac{t}{N^2}\right)^3 \exp\left(-\frac{t^3}{N^6}\right)\right)\right) \\ &= \Big((1 \pm N^{-\nu/2})N^3\Big)^{N^2} \exp\left(N^2 \int_0^1 \log\big((1-x)^3 \exp(-x^3)\big)\,dx\right) \\ &= \Big((1 \pm N^{-\nu/2})N^3\Big)^{N^2} \exp\left(-\frac{13N^2}{4}\right) = \left(\Big(1 \pm N^{-\nu/2}\Big)\frac{N^3}{e^{13/4}}\right)^{N^2}.\end{aligned}$$

Dividing this expression by $(N^2)! = ((1+o(1))N^2/e)^{N^2}$ yields the desired result, once we additionally note that whp such a choice of triangles can be completed to a full high-girth Latin square (due to the prior analysis).

References

[1] Ben Barber, Stefan Glock, Daniela Kühn, Allan Lo, Richard Montgomery, and Deryk Osthus, *Minimalist designs*, Random Structures Algorithms **57** (2020), 47–63.

[2] Ben Barber, Daniela Kühn, Allan Lo, Deryk Osthus, and Amelia Taylor, *Clique decompositions of multipartite graphs and completion of Latin squares*, J. Combin. Theory Ser. A **151** (2017), 146–201.

[3] Padraic Bartlett, *Completions of ϵ-dense partial Latin squares*, J. Combin. Des. **21** (2013), 447–463.

[4] Tom Bohman and Lutz Warnke, *Large girth approximate Steiner triple systems*, J. Lond. Math. Soc. (2) **100** (2019), 895–913.

[5] Flora C. Bowditch and Peter J. Dukes, *Fractional triangle decompositions of dense 3-partite graphs*, J. Comb. **10** (2019), 255–282.

[6] H. Brandt, *Über eine Verallgemeinerung des Gruppenbegriffes*, Math. Ann. **96** (1927), 360–366.

[7] L. M. Brègman, *Certain properties of nonnegative matrices and their permanents*, Dokl. Akad. Nauk SSSR **211** (1973), 27–30.

[8] Joshua M. Browning, Peter J. Cameron, and Ian M. Wanless, *Bounds on the number of small Latin subsquares*, J. Combin. Theory Ser. A **124** (2014), 41–56.

[9] Nicholas J. Cavenagh, Catherine Greenhill, and Ian M. Wanless, *The cycle structure of two rows in a random Latin square*, Random Structures Algorithms **33** (2008), 286–309.
[10] József Dénes and A Donald Keedwell, *Latin squares: New developments in the theory and applications*, Annals of Discrete Mathematics, vol. 46, North-Holland, Amsterdam, 1991.
[11] G. P. Egorychev, *The solution of van der Waerden's problem for permanents*, Adv. in Math. **42** (1981), 299–305.
[12] D. I. Falikman, *Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix*, Mat. Zametki **29** (1981), 931–938, 957.
[13] Asaf Ferber and Matthew Kwan, *Almost all Steiner triple systems are almost resolvable*, Forum Math. Sigma **8** (2020), Paper No. e39, 24.
[14] Peter Frankl, Zoltán Füredi, and Gil Kalai, *Shadows of colored complexes*, Math. Scand. **63** (1988), 169–178.
[15] David A. Freedman, *On tail probabilities for martingales*, Ann. Probability **3** (1975), 100–118.
[16] Stefan Glock, Daniela Kühn, Allan Lo, and Deryk Osthus, *On a conjecture of Erdős on locally sparse Steiner triple systems*, Combinatorica **40** (2020), 363–403.
[17] C. D. Godsil and B. D. McKay, *Asymptotic enumeration of Latin rectangles*, J. Combin. Theory Ser. B **48** (1990), 19–44.
[18] W. T. Gowers and J. Long, *Partial associativity and rough approximate groups*, Geom. Funct. Anal. **30** (2020), 1583–1647.
[19] Matan Harel, Frank Mousset, and Wojciech Samotij, *Upper tails via high moments and entropic stability*, arXiv:1904.08212.
[20] Mark T. Jacobson and Peter Matthews, *Generating uniformly distributed random Latin squares*, J. Combin. Des. **4** (1996), 405–437.
[21] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
[22] Svante Janson and Andrzej Ruciński, *The infamous upper tail*, vol. 20, 2002, Probabilistic methods in combinatorial optimization, pp. 317–342.
[23] A. Donald Keedwell and József Dénes, *Latin squares and their applications*, second ed., Elsevier/North-Holland, Amsterdam, 2015, With a foreword to the previous edition by Paul Erdős.
[24] Peter Keevash, *The existence of designs*, arXiv:1401.3665.
[25] Peter Keevash, *The existence of designs II*, arXiv:1802.05900.
[26] Peter Keevash, *Counting designs*, J. Eur. Math. Soc. (JEMS) **20** (2018), 903–927.
[27] Thomas P Kirkman, *On a problem in combinations*, Cambridge and Dublin Mathematical Journal **2** (1847), 191–204.
[28] A. Kotzig, C. C. Lindner, and A. Rosa, *Latin squares with no subsquares of order two and disjoint Steiner triple systems*, Utilitas Math. **7** (1975), 287–294.
[29] Anton Kotzig and Jean Turgeon, *On certain constructions for Latin squares with no Latin subsquares of order two*, Discrete Math. **16** (1976), 263–270.
[30] Matthew Kwan, *Almost all Steiner triple systems have perfect matchings*, Proc. Lond. Math. Soc. (3) **121** (2020), 1468–1495.
[31] Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *Large deviations in random Latin squares*, arXiv:2106.11932.
[32] Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *Note on random Latin squares and the triangle removal process*, arXiv:2109.15201.
[33] Matthew Kwan, Ashwin Sah, Mehtaab Sawhney, and Michael Simkin, *High-Girth Steiner Triple Systems*, arXiv:2201.04554v3.
[34] Matthew Kwan and Benny Sudakov, *Intercalates and discrepancy in random Latin squares*, Random Structures Algorithms **52** (2018), 181–196.
[35] Nathan Linial and Zur Luria, *An upper bound on the number of high-dimensional permutations*, Combinatorica **34** (2014), 471–486.
[36] Nathan Linial and Zur Luria, *Discrepancy of high-dimensional permutations*, Discrete Anal. (2016), Paper No. 11, 8.
[37] Nathan Linial and Michael Simkin, *Monotone subsequences in high-dimensional permutations*, Combin. Probab. Comput. **27** (2018), 69–83.
[38] Nati Linial, *Challenges of high-dimensional combinatorics*, László Lovász 70th birthday conference, Budapest, https://www.cs.huji.ac.il/~nati/PAPERS/challenges-hdc.pdf, 2018.
[39] Barbara Maenhaut, Ian M. Wanless, and Bridget S. Webb, *Subsquare-free Latin squares of odd order*, European J. Combin. **28** (2007), 322–336.
[40] B. D. McKay and I. M. Wanless, *Most Latin squares have many subsquares*, J. Combin. Theory Ser. A **86** (1999), 322–347.
[41] Mary McLeish, *On the existence of Latin squares with no subsquares of order two*, Utilitas Math. **8** (1975), 41–53.

[42] Richard Montgomery, *Fractional clique decompositions of dense partite graphs*, Combin. Probab. Comput. **26** (2017), 911–943.
[43] Arthur O. Pittenger, *Mappings of Latin squares*, Linear Algebra Appl. **261** (1997), 251–268.
[44] J. H. van Lint and R. M. Wilson, *A course in combinatorics*, second ed., Cambridge University Press, Cambridge, 2001.

Institute of Science and Technology Austria (IST Austria), 3400 Klosterneuburg, Austria
Email address: matthew.kwan@ist.ac.at

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Email address: {asah,msawhney}@mit.edu

Center of Mathematical Sciences and Applications, Harvard University, Cambridge, MA 02138, USA
Email address: msimkin@cmsa.fas.harvard.edu