TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY, SERIES B
Volume 9, Pages 288–321 (April 15, 2022)
https://doi.org/10.1090/btran/92

NEW BOUNDS FOR RYSER'S CONJECTURE AND RELATED PROBLEMS

PETER KEEVASH, ALEXEY POKROVSKIY, BENNY SUDAKOV, AND LIANA YEPREMYAN

Abstract. A Latin square of order n is an $n \times n$ array filled with n symbols such that each symbol appears only once in every row or column and a transversal is a collection of cells which do not share the same row, column or symbol. The study of Latin squares goes back more than 200 years to the work of Euler. One of the most famous open problems in this area is a conjecture of Ryser-Brualdi-Stein from 60s which says that every Latin square of order $n \times n$ contains a transversal of order $n - 1$. In this paper we prove the existence of a transversal of order $n - O(\log n/\log\log n)$, improving the celebrated bound of $n - O(\log^2 n)$ by Hatami and Shor. Our approach (different from that of Hatami-Shor) is quite general and gives several other applications as well. We obtain a new lower bound on a 40-year-old conjecture of Brouwer on the maximum matching in Steiner triple systems, showing that every such system of order n is guaranteed to have a matching of size $n/3 - O(\log n/\log\log n)$. This substantially improves the current best result of Alon, Kim and Spencer which has the error term of order $n^{1/2+o(1)}$. Finally, we also show that $O(n \log n/\log\log n)$ many symbols in Latin arrays suffice to guarantee a full transversal, improving on a previously known bound of $n^{2-\varepsilon}$. The proofs combine in a novel way the semi-random method together with the robust expansion properties of edge-coloured pseudorandom graphs to show the existence of a rainbow matching covering all but $O(\log n/\log\log n)$ vertices. All previous results, based on the semi-random method, left uncovered at least $\Omega(n^\alpha)$ (for some constant $\alpha > 0$) vertices.

1. Introduction

A Latin square of order n is an $n \times n$ array filled with n symbols so that every symbol appears only once in each row and in each column. A transversal is a collection of cells of the Latin square which do not share the same row, column or symbol. A full transversal is a transversal of order n. The study of Latin squares goes back to the work of Euler [13] in the 18th century, who asked a question equivalent to "for which n is there an $n \times n$ Latin square which can be decomposed into n disjoint full transversals?". Well-known examples of Latin squares are multiplication tables of finite groups. Latin squares have connections to 2-dimensional permutations, design theory, finite projective planes and error correcting codes.

Received by the editors February 25, 2021, and, in revised form, July 22, 2021.

2020 *Mathematics Subject Classification.* Primary 05D40, 05C65, 05B15, 05D15.

The research of the first author was supported in part by ERC Consolidator Grant 647678. The research of the third author was supported in part by SNSF grant 200021_196965. The research of the fourth author was supported by Marie Sklodowska Curie Global Fellowship, H2020-MSCA-IF-2018:846304.

It is easy to see that there are many Latin squares without full transversals (for example the addition table of the group $\mathbb{Z}_4$), and it is a hard problem to determine when full transversals exist. This question is very difficult even in the case of multiplication tables of finite groups. In 1955 Hall and Paige [15] conjectured that the multiplication table of a group G has a full transversal exactly if the 2-Sylow subgroups of G are trivial or non-cyclic. It took 50 years to establish this conjecture and its proof is based on the classification of finite simple groups (see [29] and the references therein). Very recently an alternative proof of this conjecture was found for large groups using tools from analytic number theory [12]. The most famous open problem on transversals in general Latin squares is the following conjecture of Ryser, Brualdi and Stein [10, 25, 27].

Conjecture 1.1. *Every $n \times n$ Latin square has a transversal of order $n-1$. Moreover, if n is odd it has a full transversal.*

Most research towards the Ryser-Brualdi-Stein conjecture has focused on proving that all $n \times n$ Latin squares have large transversals (trying to get as close to $n-1$ as possible). Here Koksma [19] found transversals of size $2n/3 + O(1)$ and Drake [11] improved this to $3n/4 + O(1)$. The first asymptotic proof of the conjecture was obtained by Brouwer, de Vries, and Wieringa [9] and independently by Woolbright [30] who found transversals of size $n - \sqrt{n}$. This was improved in 1982 by Shor [26] to $n - O(\log^2 n)$. His paper had a mistake which was later rectified, using the original approach, by Hatami and Shor [16]. For the last, nearly forty years, this was the best known bound for the Ryser-Brualdi-Stein conjecture. Our first theorem improves this result as follows.

Theorem 1.2. *There exists a constant k such that every $n \times n$ Latin square contains a transversal of order $n - k\frac{\log n}{\log\log n}$.*

A Latin array is an $n \times n$ square filled with an arbitrary number of symbols such that no symbol appears twice in the same row or column. Latin arrays are natural extensions of Latin squares, and also have been extensively studied. A familiar example of such an array is a multiplication table between elements of two subsets of equal size in some group. It is generally believed that extra symbols in a Latin array should help to find full transversals. Motivated by this Akbari and Alipour [2] conjectured that any Latin array of order n with at least $n^2/2$ different symbols contains a full transversal. Progress towards this conjecture was independently obtained by Best, Hendrey, Wanless, Wilson and Wood [7] (who showed that $(2-\sqrt{2})n^2$ symbols suffice) and Barát and Nagy [6] (who showed that $3n^2/4$ symbols suffice). Very recently Montgomery, Pokrovskiy, Sudakov [22] and Keevash and Yepremyan [17] independently showed that $n^{2-\varepsilon}$ many symbols suffice to guarantee a full transversal. Here we substantially improve these results.

Theorem 1.3. *There exists a constant k such that every $n \times n$ Latin array filled with $kn\log n/\log\log n$ many symbols contains a full transversal.*

It is worth pointing out that the problem of Akbari and Alipour is closely related to finding transversals in Latin squares, namely Conjecture 1.1. In particular, the last theorem implies Theorem 1.2. Indeed, start with an $n \times n$ Latin square and then substitute distinct new symbols in the first $k\log n/\log\log n$ rows, such that every symbol is used only once. Then Theorem 1.3 guarantees us a full transversal. Since this transversal can use at most $k\log n/\log\log n$ cells from the first $k\log n/\log\log n$

rows, upon removing these we are left with a transversal of the original Latin square which has size $n - k \log n / \log \log n$.

All the above results and problems can be rephrased as statements about matchings in hypergraphs. To see this, we construct from an $n \times n$ Latin square L the following 3-uniform hypergraph $\mathcal{H}$ on $3n$ vertices. The vertices of $\mathcal{H}$ are $V(\mathcal{H}) = R \cup C \cup S$, where R is the rows of L, C is the columns of L, and S is the symbols of L. There is an edge in $\mathcal{H}$ for every entry of L. If the (i, j)-th entry of L has symbol s, then $\{i, j, s\}$ is a hyperedge of $\mathcal{H}$. It is easy to check that under this transformation, the hypergraph we obtain is n-regular, there is exactly one edge containing a given pair of vertices, and that transversals in L correspond to matchings in $\mathcal{H}$.

The problem of finding nearly perfect matchings in regular hypergraphs has a long history in discrete mathematics and such results have many applications to other problems as well. For example Rödl [24] proved the Erdős-Hanani conjecture on existence of approximate designs by essentially showing that regular hypergraphs with bounded codegrees have nearly-perfect matchings. This paper introduced the celebrated technique of "Rödl's nibble" which is a versatile approach for finding large matchings in hypergraphs in semi-random manner. One famous example of a regular hypergraph with bounded codegress is a Steiner triple system, which is a 3-uniform hypergraph on n vertices in which every pair of vertices is in a unique edge. The existence of such triple systems was established by Kirkman in 1847. By definition, this hypergraph is $(n-1)/2$-regular and has all codegrees equal to one. The problem of existence of large matchings in Steiner triple systems was posed about forty years ago by Brouwer [8].

Conjecture 1.4. *Every Steiner triple system of order n contains a matching of size $(n-4)/3$.*

Over the years this conjecture attracted a lot of attention. Wang [28] showed that every Steiner triple system has a matching of size $2n/9 - O(1)$. Lindner and Phelps [20] found a matching of size $4n/15 - O(1)$. Brouwer [8] obtained the first asymptotic result by finding matchings of size $n/3 - O(n^{2/3})$. Using a clever refinement of Rödl's nibble combined with large deviation inequalities, Alon, Kim, and Spencer [3] obtained the best current bound. They show the existence of a matching covering all but $O(n^{1/2} \log^{3/2} n)$ vertices. Here we improve this twenty year old result and obtain the first sub-polynomial upper bound on the number of vertices uncovered by the maximum matching.

Theorem 1.5. *There is a constant k such that every Steiner triple system S on n vertices has a matching of size at least $n/3 - k \log n / \log \log n$.*

Our methods combine in a novel way the Rödl's nibble together with the robust expansion properties of edge-coloured pseudorandom graphs and apply in far more general settings than any of the above theorems and conjectures. The main technical theorem we prove can be used to show that 3-uniform hypergraphs satisfying certain pseudorandomness properties have a matching covering all but $O(\log n / \log \log n)$ vertices. All previous comparable theorems left n^α vertices uncovered.

1.1. **Proof ideas.**

Coloured graphs and rainbow matchings. Although our main results are about transversals in Latin arrays/squares and matchings in hypergraphs, all our proofs will take place in a different setting. This will be the setting of finding *rainbow matchings* in *properly edge-coloured* complete bipartite graphs. Recall that a *proper* edge-colouring of a graph is one where all edges incident to the same vertex have different colours. A matching in a coloured graph is rainbow if all its edges have different colours. A *linear* hypergraph is a hypergraph in which every pair of vertices lies in at most one edge. In this paper we will use extensively that the following three kinds of objects are equivalent:

- An $n \times n$ Latin array filled with m symbols.
- A linear 3-partite, 3-uniform hypergraph with partition sizes (n, n, m).
- A properly edge-coloured complete bipartite graph $K_{n,n}$ with m colours.

The connection between Latin arrays and linear hypergraphs was already described in the introduction. To see the reduction to coloured graphs consider an $n \times n$ Latin array L filled with m symbols. Using it we can construct the following proper edge-colouring of $K_{n,n}$. Label the vertices of $K_{n,n}$ by $\{x_1, \dots, x_n, y_1, \dots, y_n\}$, and join x_i to y_j with a colour ℓ edge whenever the ij-th entry of L is ℓ. This is a proper edge-colouring with m colours due to the properties of Latin arrays. A size t transversal in the Latin array corresponds to a rainbow matching with t edges in $K_{n,n}$. Note that in case of Latin squares we have $m = n$ in the above statement. Thus under this transformation, Theorem 1.2 is equivalent to the following.

Theorem 1.6. *There exists a constant k such that every properly n-edge-coloured $K_{n,n}$ has a rainbow matching of size $n - k\frac{\log n}{\log\log n}$.*

Similarly Theorem 1.3 has the following equivalent form.

Theorem 1.7. *There exists a constant k such that every properly edge-coloured $K_{n,n}$ with $kn\frac{\log n}{\log\log n}$ colours has a rainbow perfect matching.*

Although, as we already mentioned in the previous section, this theorem can be used to prove Theorem 1.6, we deduce both of them from a more general result about matchings in properly edge-coloured "typical" (i.e., both edges and colours have some pseudorandom properties) graphs which we obtain in Section 4.

The reduction of finding large matchings in a Steiner triple system S to a graph problem is slightly more subtle, and the details can be found in Section 6. The main idea is to randomly select a tripartition (A, B, C) of S and consider only the edges that respect this tripartition. The first two parts induce a properly edge-coloured graph G where we think of the colour of an edge ab with $a \in A, b \in B$ to be $c \in C$ if $abc \in S$. Note that any rainbow matching in G induces a matching of the same size in S. The graph G turns out to be typical in this coloured setting we mentioned above, and we can also guarantee $|A| = |B| = |C| = n/3$. Thus from our general result (more precisely, Corollary 4.6) it follows that G contains a rainbow matching of size $n/3 - O(\log n/\log\log n)$, and therefore, S has a matching of the same size.

Rödl Nibble and expansion. Rödl introduced a method called "Rödl's nibble", which can be used to find matchings in a wide variety of settings. In particular it applies in the setting of Theorem 1.6 to give a rainbow matching of size $n - O(n^{1-\alpha})$ (for some small constant $\alpha > 0$). Our ideas very much build on this result. At a high

level, our proof consists of starting with a matching produced by Rödl's nibble and then modifying it to get a matching of size $n - O(\log n/\log\log n)$. Although our methods apply for all coloured pseudorandom graphs let us demonstrate its main ideas for the simplest case, $K_{n,n}$.

The basic idea of Rödl's nibble is to construct a matching in several steps, each time taking a collection of random edges. To imitate this idea in our setting, given a properly edge-coloured $K_{n,n}$, we fix $q \in (0,1)$ and select every edge of $K_{n,n}$ with probability q/n. Then we delete all edges which share vertices or colours with other selected edges. This will certainly produce a rainbow matching. The matching produced like this is often called a "bite". How big will it be? Unfortunately not very big. The expected number of edges in the bite will be $qn(1-q/n)^{3(n-1)}$ which is roughly qne^{-3q} for large n. So the maximum size of the matching would be $n/3e$ achieved by $q = 1/3$. Rödl's brilliant idea was to perform several small bites one after another, deleting the vertices/colours used in each bite from the rest of the graph. Although after the first bite, the remaining graph will no longer be complete; it still turns out to be possible to repeatedly bite until the remainder has size $< O(n^{1-\alpha})$. This is based on the phenomenon that edges/vertices not used on each bite have pseudorandomness properties.

Our key new idea is to show that this matching has nice "expansion" properties. In fact, we only need to analyse these properties for the first bite. The structure of our proof is the following:

(S1) Obtain a rainbow matching M_0 via the first bite and show it satisfies certain expansion properties.

(S2) Delete vertices and colours of M_0 from $K_{n,n}$. The remaining graph will still have pseudorandomness properties with respect to both colours and vertices; therefore we can extend M_0 to a larger rainbow matching M of size $n - n^{1-\alpha}$. This step is done via using Rödl's nibble as a black box on coloured pseudorandom graphs.

(S3) The expansion properties that M_0 had can be transferred to M which will allow us to do switching-type arguments to increase M as long as we have $\Omega(\log n/\log\log n)$ unused colours outside of M. We do this iteratively, starting from M and obtaining a new matching of size one bigger at every step. The simplest "switching-type argument" we would have liked to employ here is to increase the matching using rainbow augmenting paths, that is, a rainbow path which starts and ends at an unmatched vertex and whose edges alternate between edges of M and edges of colours outside M. If such a path P exists, then the symmetric difference $M\Delta P$ is a larger rainbow matching. Unfortunately, in general we cannot guarantee to find such paths, so instead we employ a more complicated argument using a combination of alternating paths & cycles (see the discussion at the start of Section 4).

(S4) We can guarantee that the switching paths and cycles mentioned in (S3) are of length $O(\log n/\log\log n)$. This guarantees that the matchings at steps i and $i+1$ are not too "far" from each other in edit distance, more precisely, $|M_i \triangle M_{i+1}| = O(\log n/\log\log n)$. Because of this after $O(n^{1-\alpha})$ steps, $|M_i \triangle M| \le O(n^{1-\alpha}\log n/\log\log n) \ll |M|$; thus M_i will still have the expansion properties which guarantees we can find the alternating paths/cycles to do switchings along. Because of this as long as there are

$\Omega(\log n/\log\log n)$ many unused colours we may repeat the steps and after at most $O(n^{1-\alpha})$ times we get a rainbow matching of size at least $n - O(\log n/\log\log n)$.

The major part of this paper is devoted to establishing (S1) in Section 3. Next we discuss the notion of expansion we study. Our approach is heavily inspired by the idea that a "randomly chosen matching will satisfy pseudorandomness properties". The pseudorandomness property that we use is very different from the ones previously used in nibble-type proofs. It can be summarised as "the union of a random matching together with an arbitrary nearly regular graph D will have strong expansion properties". Here is a simplified version of what we prove:

Lemma 1.8. *Let $0 < q \ll 1$, $1 \ll d$, $q^{-1} \ll d \leq \sqrt{n}$. Given $K_{n,n}$ properly edge-coloured by n colours, let H be its subgraph formed by choosing every edge independently with probability q/n. Delete all edges of H which share vertices or colours with other edges of H and let M_0 be the resulting rainbow matching. Then with high probability*

(E1) *every collection D of d colours in $K_{n,n}$, and every set S of n/q^4d vertices there are at least $(1-q)n$ vertices that can be reached from S by a D-M_0 alternating path of length three, i.e., a path whose first and last edge is in D and the middle edge is in M_0.*

Notice that (E1) only provides expansion for large sets S but for our purposes we need it to hold for all sets. After we extend the matching M_0 to a larger rainbow matching M of size roughly $n - n^{1-\alpha}$ as described in (S2) we are able to iterate (E1) if we restrict to larger collections of colours and longer paths. In particular, we obtain the following refinement of (E1) with respect to M.

(E2) *For $d = \log n/\log\log n$ and any collection D of d colours there exists a set of vertices V_0 of size at most qn such that the following holds. Every vertex not in V_0 can reach all but qn vertices via $D - M$ alternating rainbow paths of length $O(\log n/\log\log n)$.*

Note that (E2) also implies that between any two vertices of $K_{n,n}$ lying in different sides of the bipartition there is a $D - M$ alternating rainbow path of length $O(\log n/\log\log n)$. This property is enough to perform the modifications described in (S3). We find an alternating rainbow path to extend the matching M by one edge at a time much like in standard proofs of say Hall's matching theorem. In the applications of (E2) we let D to be the set of unused colours on M. The condition $|D| = \Omega(\log n/\log\log n)$ allows us to do iterations and also tells us when the process must stop and the enlargement of the matching is no longer possible. The reason why we need this many unused colours is because the length of the alternating paths, used to perform the switching-type arguments, can be as large as $\Omega(\log n/\log\log n)$, and we need to guarantee that these paths are rainbow. So we can repeat (S3) until the number of unused colours on M is $O(\log n/\log\log n)$.

2. Preliminaries

We will use asymptotic "$\ll$" notation to state our intermediate lemmas. When we write "$\delta \ll \varepsilon$" in the statement of a result, it means "for all $\varepsilon > 0$ and sufficiently small $\delta > 0$, the following statement is true". In particular "$n^{-1} \ll \varepsilon$" means "for all $\varepsilon > 0$ and sufficiently large n, the following is true". When we chain several inequalities like this, the quantity on the left is small relative to all constants on

the right. For example "$n^{-1} \ll \delta \ll \varepsilon$" means "for all $\varepsilon > 0$, there is a δ_0 such that for positive $\delta < \delta_0$ and sufficiently large n, the following is true". Sometimes we will abuse this notation and write "$n \gg \varepsilon^{-1}$" to mean "$n^{-1} \ll \varepsilon$".

For any positive reals $a, b \in \mathbb{R}$, we use "$x = a \pm b$" to mean "$a - b \leq x \leq a + b$". We also use the same notation with more than one instance of "$\pm$". We will write expressions of the form "$f = g$", where f and g are functions involving, one or more instances of "$\pm$". To interpret such an expression, first define $\max_\pm f$ to be the maximum value of f taken over all possible assignments of $+/-$ to each "$\pm$" symbol. Similarly define $\min_\pm f$. Then we say that "$f = g$" if $\max_\pm f \leq \max_\pm g$ and $\min_\pm f \geq \min_\pm g$ are both true. A useful example of this is that for $q \gg n^{-1}$ we have $(1 - q/n)^{\pm n} = 1 \pm 2q$. To see this notice that $1 - 2q \leq (1 - q/n)^n \leq (1 - q/n)^{-n} \leq 1 + 2q$. Note that this "=" relation is not symmetric in general — for example "$1 \pm 0.1 = 1 \pm 0.2$" is true while "$1 \pm 0.2 = 1 \pm 0.1$" is false.

For a graph G, the set of edges of G is denoted by $E(G)$ and the set of vertices of G is denoted by $V(G)$. The set of neighbours of v is denoted by $N_G(v)$, and $d_G(v) = |N_G(v)|$. For a coloured graph G and a colour c, denote by $E_G(c)$ the set of edges of colour c in G, and denote $V_G(c)$ for the set of vertices touching colour c edges. In all of these, we omit the "G" subscript when the graph G is clear from context. For a properly edge-coloured graph G, let $C(G)$ be the set of colours appearing on the edges of G, $C \subseteq C(G)$ and $v \in V(G)$ we denote by $N_C(v)$ the set of vertices w such that $vw \in E(G)$ and $c(vw) \in C$. For a graph G, a set of vertices A, let $G[A]$ denote the induced subgraph of G on A. For a coloured graph G, a set of colours $C \subseteq C(G)$, we let $G[C]$ to be the subgraph of G induced by edges of colours in C.

Let G, H be graphs on the same vertex set V. We say that a path $x_1 x_2 \dots x_t$ is G-H alternating if, for odd i, $x_i x_{i+1} \in E(G)$ and for even i, $x_i x_{i+1} \in E(H)$ (or in other words, the first edge is in G, and thereafter the edges alternate between G and H). For a set $S \subseteq V(G \cup H)$, we use $N^t_{G,H}(S)$ to denote the set of vertices v to which there is a G-H alternating path of length t from some $s \in S$.

2.1. **Probabilistic tools.** Here we gather basic probabilistic tools that we use. We use the Chernoff bounds. Most of these can be found in textbooks on the probabilistic method such as [21].

Lemma 2.1 (Chernoff bounds, [21])**.** *Given a binomially distributed variable $X \in Bin(n, p)$ for all $0 < a \leq 3/2$ we have*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq a\mathbb{E}[X]] \leq 2e^{-\frac{a^2}{3}\mathbb{E}[X]}.$$

Given a product space $\Omega = \prod_{i=1}^n \Omega_i$ and a random variable $X : \Omega \to \mathbb{R}$ we make the following definitions.

- Suppose that there is a constant c such that changing $\omega \in \Omega$ in any one coordinate changes $X(\omega)$ by at most c. Then we say that X is c-Lipschitz.
- Suppose that for any $s \in \mathbb{N}$ and $\omega \in \Omega$ with $X(\omega) \geq s$ there is a set $I \subseteq \{1, \dots, n\}$ with $|I| \leq rs$ such that every ω' which agrees with ω on coordinates in I also has $X(\omega') \geq s$. Then we say that X is r-certifiable.

We'll use the following two versions of Azuma's inequality.

Lemma 2.2 (Azuma's inequality, [5,21]). *For a product space $\Omega = \prod_{i=1}^{n} \Omega_i$ and a c-Lipschitz random variable $X : \Omega \to \mathbb{R}$, we have*

$$\mathbb{P}(|X - \mathbb{E}(X)| > t) \leq 2e^{\frac{-t^2}{nc^2}}.$$

Lemma 2.3 (Azuma's inequality for 0/1 product spaces, [3,18]). *Let $\Omega = \{0,1\}^n$ with the ith coordinate of an element of Ω equal to 1 with probability p_i. Let X be a c-Lipschitz random variable on Ω. Set $\sigma^2 = c^2 \sum_{i=1}^{n} p_i(1-p_i)$. For all $t \leq 2\sigma/c$, we have*

$$\mathbb{P}(|X - \mathbb{E}(X)| > t\sigma) \leq 2e^{\frac{-t^2}{4}}.$$

We'll also use the following version of Talagrand's inequality.

Lemma 2.4 (Talagrand inequality, [21]). *For a product space $\Omega = \prod_{i=1}^{n} \Omega_i$ and a c-Lipschitz, r-certifiable random variable $X : \Omega \to \mathbb{R}$, we have*

$$\mathbb{P}\left(|X - \mathbb{E}(X)| > t + 60c\sqrt{r\mathbb{E}(X)}\right) \leq 4e^{\frac{-t^2}{8c^2 r\mathbb{E}(X)}}.$$

We say a bipartite graph G with parts X, Y is (ε, p, n)-*regular* if

(P1) $|X| = |Y| = n(1 \pm n^{-\varepsilon})$,
(P2) for every v, $d(v) = pn(1 \pm n^{-\varepsilon})$.

Furthermore, G is (ε, p, n)-*typical* if

(P3) for every $u, v \in X$ or $u, v \in Y$ we have $|N(u) \cap N(v)| = p^2 n(1 \pm n^{-\varepsilon})$.

A bipartite graph G with bipartition (X, Y) and colour set C is called *coloured* (ε, p, n)-*regular*/ *coloured* (ε, p, n)-*typical* if it is properly edge-coloured and the following hold:

(P4) G is (uncoloured) (ε, p, n)-regular/(ε, p, n)-typical.
(P5) Define $G_{X,C}$ to be the bipartite graph with vertex bipartition (X, C), where xc is an edge for $x \in X, c \in C$ if there exists some $y \in Y$ such that $xy \in E(G)$ and $c(xy) = c$. Define $G_{Y,C}$ analogously. We require both $G_{X,C}$ and $G_{Y,C}$ to be a coloured (ε, p, n)-regular/(ε, p, n)-typical.

Note that a coloured (ε, p, n)-regular graph G is *coloured* (ε, p, n)-regular if additionally $|C(G)| = (1 \pm n^{-\varepsilon})n$ and every colour $c \in C(G)$ has $|E_G(c)| = (1 \pm n^{-\varepsilon})pn$. Similarly, a properly edge-coloured (ε, p, n)-typical graph is *coloured* (ε, p, n)-typical if these happen and additionally every pair of colours c, c' have $|V_G(c) \cap V_G(c') \cap X| = (1 \pm n^{-\varepsilon})p^2 n$ and $|V_G(c) \cap V_G(c') \cap Y| = (1 \pm n^{-\varepsilon})p^2 n$.

Frankl and Rödl [14] (also Pippenger, unpublished) showed that for all $\gamma \gg \epsilon \gg n^{-1}$ every n-vertex hypergraph with $(1 \pm \epsilon)pn$ degrees and codegrees at most one has a matching of order $(1-\gamma)n$. A corollary of this is that every coloured (γ, p, n)-regular graph has a rainbow matching of order $(1 - \gamma)n$ (to see this, associate a hypergraph with the (γ, δ, n)-regular graph as explained in the introduction and apply their theorem). We'll need the following standard version (which appeared in the literature before) of this result where the error term γn is polynomially related with n.

Lemma 2.5. *Let $n^{-1} \ll \gamma \ll \varepsilon$ and $n^{-1} \ll p \leq 1$. Every coloured (ε, p, n)-regular bipartite graph G has a rainbow matching of size $n - n^{1-\gamma}$.*

Proof. Notice that G is balanced bipartite with parts of size $(1 \pm n^{-\varepsilon})n$, every vertex has degree $(1 \pm n^{-\varepsilon})pn$, and every colour occurs at most $(1 + n^{-\varepsilon})pn$ times.

Now the lemma is strictly weaker than Lemma 4.6 from [22] (applied with $n = n, \gamma = n^{-\varepsilon}, \delta = p, p = n^{-\gamma}, \ell = 1$). □

Lemma 2.6 shows that a random subgraph of a typical bipartite graph is typical. There are two notions of what "random subgraph" means here. The most important one is to consider the subgraph formed by deleting every vertex/colour independently with fixed probability (case (a) below). We use the second case in Section 6 to reduce the problem of finding large matchings in Steiner systems to finding large matchings in special typical graphs.

Lemma 2.6. *Let $n^{-1} \ll p, q, \varepsilon \leq 1$. Let G be a coloured (ε, p, n)-typical bipartite graph with bipartition X, Y and colour set C such that $|X| = |Y| = |C| = n$. Let $X' \subseteq X, Y' \subseteq Y, C' \subseteq C$ be random sets obtained as follows:*

(a) *Every vertex/colour ends up in X', Y', C' independently with probability q.*
(b) *Suppose we can label $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_n\}$, $C = \{c_1, \ldots, c_n\}$ such that if $x_i y_j$ is an edge of G of colour c_k then all i, j, k must be distinct. Form X', Y', C' by choosing disjoint set of indices $I_X, I_Y, I_C \subseteq [n]$ such that independently every $i \in [n]$ is placed in I_X, I_Y and I_C with probability q and in none of them with probability $1 - 3q$. Set $X' = \{x_i : i \in I_X\}$, $Y' = \{y_i : i \in I_Y\}$, $C' = \{c_i : i \in I_C\}$*

Let H be formed by colour C' edges going from X' to Y'. Then with probability at least $1 - e^{-n^{1-\varepsilon/2}}$, H is coloured $(\varepsilon/8, qp, qn)$-typical.

Proof. We will show that with probability $1 - \frac{1}{3}e^{-n^{1-\varepsilon/2}}$, H is (uncoloured) $(\varepsilon/8, qp, qn)$-typical. By symmetry between X, Y, C, the same proof shows that $H_{X,C}$ and $H_{Y,C}$ are $(\varepsilon/8, qp, qn)$-typical. Thus we will have that with probability $1 - e^{-n^{1-\varepsilon/2}}$ all of H, $H_{X,C}$, and $H_{Y,C}$ are $(\varepsilon/8, qp, qn)$-typical, or equivalently H is coloured $(\varepsilon/8, qp, qn)$-typical. To give a unified proof of both statements $(a), (b)$ we will use Azuma's inequality.

Let u, v be two vertices on the same side of G, and $y \in N_G(u) \cap N_G(v)$. Without loss of generality, we may suppose that $u, v \in X, y \in Y$. Notice that in both (a) and (b) we have $\mathbb{P}(u \in X') = q$, $\mathbb{P}(c(uy) \in C', y \in Y') = q^2$, and $\mathbb{P}(c(uy), c(vy) \in C', y \in Y') = q^3$. Indeed, here we use that $c(uy) \in C'$, $c(vy) \in C'$, and $y \in Y'$ are independent events which is true in case (a) trivially. For case (b), suppose $y = y_i$, $c(uy) = c_j$, $c(vy) = c_k$, it is enough to show that all three indices i, j, k are distinct. Indeed, since $uy, vy \in E(G)$, it follows that i and k, and j and k are distinct. Finally, j and k are distinct since the edge-colouring is proper. Since G is (ε, p, n)-typical, we have $|X|, |Y| = (1 \pm n^{-\varepsilon})n$, $|N_G(u)| = (1 \pm n^{-\varepsilon})pn$ and $|N_G(u) \cap N_G(v)| = (1 \pm n^{-\varepsilon})p^2 n$. Thus we have

$$\mathbb{E}[|X'|], \mathbb{E}[|Y'|] = (1 \pm n^{-\varepsilon})qn$$
$$\mathbb{E}[|N_{C'}(u) \cap Y'|] = (1 \pm n^{-\varepsilon})q^2 pn, \text{ for all } u \in X$$
$$\mathbb{E}[|N_{C'}(u) \cap N_{C'}(v) \cap Y'|] = (1 \pm n^{-\varepsilon})q^3 p^2 n, \text{ for all } u, v \in X.$$

Notice that these random variables are all 2-Lipschitz and are each affected by at most $3(1 + n^{-\varepsilon})n \leq 4n$ coordinates. By Azuma's inequality we get that for $t = q^3 p^2 n^{1-\varepsilon/8}/2$ with probability $1 - 2e^{-t^2/16n} \geq 1 - e^{-n^{1-3\varepsilon/8}}$, each one of them are within t of their expectations. By taking union bound over all vertices and colours we obtain that with probability at least $1 - (n + n^{-\varepsilon})^3 e^{-n^{1-3\varepsilon/8}} \geq 1 - \frac{1}{3}e^{-n^{1-\varepsilon/2}}$ they

are all simultaneously within t of their expectations. So now the result follows from the definition of $(\varepsilon/8, qp, qn)$-typicality and the fact that $t+n^{1-\varepsilon} < q^3p^2n^{1-\varepsilon/8}$. □

We will need the following result about typical graphs. It is a bipartite variation of Lemma 5.5 from [22] (see also Lemma 2.1 in [4]), whose proof is straightforward from the original version.

Lemma 2.7. *Let $n \in \mathbb{N}$, $\varepsilon, p, \gamma \in (0,1]$ with $8n^{-\varepsilon} \leq \gamma$. Then every (ε, p, n)-typical bipartite graph H with sides X, Y satisfies the following. For every pair of subsets $A \subseteq X$, $B \subseteq Y$ with $|B| \geq \gamma^{-1}p^{-2}$:*

$$|e(A,B) - p|A||B|| \leq 2|A|^{\frac{1}{2}}|B|\gamma^{\frac{1}{2}}n^{\frac{1}{2}}p.$$

Proof. Let Adj_H be the adjacency matrix of H, and let $M = \mathrm{Adj}_H - pJ$, where J is the appropriately-sized all-ones matrix. Notice that for every pair of distinct vertices $y, y' \in Y$, we have

$$\begin{aligned}\sum_{v\in X} M_{y,v}M_{y',v} &= d_H(y,y') - p(d(y)+d(y')) + p^2|X| \\ &\leq (1+n^{-\varepsilon})p^2n - 2(1-n^{-\varepsilon})p^2n + p^2(1+n^{-\varepsilon})n \\ &\leq \gamma p^2 n/2.\end{aligned} \tag{1}$$

Next notice that we have

$$\begin{aligned}\big|e(A,B) - p|A||B|\big|^2 &= \left(\sum_{x\in A}\sum_{y\in B} M_{x,y}\right)^2 \\ &\leq |A|\sum_{x\in A}\left(\sum_{y\in B} M_{x,y}\right)^2 \\ &\leq |A|\sum_{x\in X}\left(\sum_{y\in B} M_{x,y}\right)^2 \\ &= |A|\sum_{x\in X}\left(\sum_{y\in B} M_{x,y}^2\right) + |A|\sum_{x\in X}\left(\sum_{y\neq y'\in B} M_{x,y}M_{x,y'}\right) \\ &\leq |X||A||B| + |A|\sum_{y\neq y'\in B}\left(\sum_{x\in X} M_{x,y}M_{x,y'}\right) \\ &\overset{(1)}{\leq} (1+n^{-\varepsilon})n|A||B| + |A|\sum_{y\neq y'\in B}\frac{\gamma p^2 n}{2} \\ &\leq (1+n^{-\varepsilon})n|A||B| + |A||B|^2\gamma p^2 n/2 \\ &\leq 2|A||B|^2\gamma p^2 n.\end{aligned}$$

Here the first inequality comes from the Cauchy-Schwarz inequality and the last inequality comes from $|B| \geq \gamma^{-1}p^{-2}$. Taking square roots gives the result. □

Next we show that the above result implies that for a coloured typical graph G and any set of d many colours in G, the subgraph of G induced by the edges of colours in D can have at most $O(n/d)$ many vertices of small degree, for $d = O(n^{\varepsilon})$.

Lemma 2.8. *Let $n^{-1} \ll p, \varepsilon \leq 1$, $16 \leq 8p^2 d \leq n^{\varepsilon}$. Suppose G is a coloured (ε, p, n)-typical bipartite graph with bipartition (X, Y) and colour set C. Then for any set of d colours D the subgraph $G[D] \subseteq G$ induced by edges of colours in D has at most $\leq 32p^{-2}n/d$ many vertices of degree less than $pd/2$.*

Proof. Let J be the set of vertices in $G[D]$ of degree less than $pd/2$. We will show that $|J \cap X| \leq 16p^{-2}n/d$. Similarly one can prove that $|J \cap Y| \leq 16p^{-2}n/d$. Recall $G_{X,C}$ is defined on the vertex bipartition (X, C) where we put an edge xc if there is some $y \in Y$ such that $xy \in E(G)$ and $c(xy) = c$. By definition of coloured typical graphs, $G_{X,C}$ is (ε, p, n)-typical; hence we can apply Lemma 2.7 to the set $J \cap X \subseteq X$ and $D \subseteq C$ with $\gamma = p^{-2}d^{-1}$ (so that $|D| \geq \gamma^{-1}p^{-2}$). We obtain

$$p|J \cap X|d - 2p|J \cap X|^{1/2} d\gamma^{1/2} n^{1/2} \leq e_{G_{X,C}}(J \cap X, D) < |J \cap X|pd/2.$$

From here it follows that $|J \cap X| \leq 16\gamma n \leq 16p^{-2}n/d$. □

3. Expansion and its properties

The proof of our main technical theorem, which we present in this section, is based on finding a nearly spanning randomized rainbow matching M which "expands" in some sense. We then show that these expansion properties can be used to alter M via a series of switchings along alternating paths to obtain a new matching covering all but $O(\log n/\log\log n)$ vertices.

3.1. Typical coloured graphs are expanding.

In this subsection we prove that every typical graph has a large matching which is "expanding" with respect to any small collection of colours. First we define what we mean by expanding. Since by itself a matching is clearly not an expanding graph, we will always speak about expansion properties of a *union of two graphs*, one of which will always be a matching. Let G, H be two graphs. Recall that for a set $S \subseteq V(G) \cup V(H)$, we use $N^t_{G,H}(S)$ to denote the set of vertices v to which there is a length t path from some $s \in S$ whose edges alternate between G and H with the first edge belonging to G. Definition 3.1 is key in this paper.

Definition 3.1 (Expander)**.** For a matching M and a bipartite graph D we say that (D, M) is a (d, A, ε, n)-expander if every vertex set $S \subseteq X$ or $S \subseteq Y$ with $|S| \geq An/d$ has a subset S' with $|S'| = An/d^2$ and $|N^4_{D,M}(S')| \geq (1-\varepsilon)n$, where (X, Y) is the bipartition of $D \cup M$ with $|X|, |Y| \geq An/d$.

Note that, since in this definition the last edge on the 4-edge path is from M, it follows that if (D, M) is a (d, A, ε, n)-expander then $|M| \geq (1-\varepsilon)n$. We also want to point out few additional subtleties. First, it would be more natural to ask $|N^2_{D,M}(S)| \geq (1-\varepsilon)n$. However, it is not true that the second neighbourhoods expand (see details in the proof sketch of Lemma 3.2). Second, we have a stronger requirement that S has a subset of size $\Theta(n/d^2)$ which expands. This is done for the following two technical purposes.

We are able to show that every coloured pseudorandom graph G has a random rainbow matching such for any subgraph D induced by edges of any collection of $d = \log n/\log\log n$ many colours, every S of size roughly n/d expands (in the above sense) with probability at least $1 - e^{-|S|}$ (see Lemma 3.8). Then we would like to claim, by taking the union bound, that with high probability all sets S expand simultaneously. Unfortunately, when taking this approach, the probability that

there is some S which does not expand is at most roughly $\binom{n}{|S|}e^{-|S|} \gg 1$. Instead, for each non-expanding set S we find a smaller set S' of size $|S|/d$ which "captures" the expansion properties of S. Now the union bound gives us that the probability that some set S does not expand is at most $\binom{n}{|S|/d}e^{-|S|} \ll 1$. (This idea is similar to the applications of containers widely used in studying H-free graphs for fixed H, where one shows that there is a collection of containers of bounded size which contain all H-free graphs of certain size).

The second reason to have a smaller subset S' which captures the expansion of S is for finding rainbow $D-M$ alternating paths between almost all $x \in X$ and $y \in Y$ (see Section 3.2 for details). Let's assume $S \subseteq Y$. As the first step to achieve this, we need to show that almost all vertices $y \in Y$ have some rainbow alternating $D-M$ path of length four starting at some $s \in S$ and ending at y. Furthermore, we need that each of these paths avoids a prescribed set of colours and vertices of order εd. Since $|N^4_{D,M}(S')| = (1-o(1))n$ for each $y \in Y$ there is a $D-M$-alternating path that starts at some $s \in S'$ and ends at y. Some of these paths can be *bad* if either they are not rainbow or they do not avoid the prescribed set of vertices and colours. The number of such bad paths is at most roughly $\varepsilon d^2|S'|$, where $|S'|$ factor comes for the choice of starting vertex in S', εd comes from using a forbidden vertex or a colour and the second d factor is due to the fact that $\Delta(D) \leq d$ (see Lemma 3.14). So using that $|S'| \approx n/d^2$ we conclude that the number of bad paths is at most εn.

The main result of this section is to prove the following expansion properties of coloured typical bipartite graphs.

Lemma 3.2 (Main expansion lemma)**.** *Let $n^{-1} \ll q \ll p \leq 1$, $n^{-1} \ll \gamma \ll \varepsilon \ll 1$ and $n^{-\varepsilon/2} \leq d^{-1} \ll q$. Suppose G is a coloured (ε, p, n)-typical bipartite graph. Then there is a randomized rainbow matching M in G with the following property.*

For any bipartite graph D on the same biparition as G with $\Delta(D) \leq d$ and at most $96p^{-2}n/d$ vertices of degree less than $pd/6$, with probability at least $1-2e^{-n^{1-\varepsilon}}$ the following hold:

(i) $|M| \geq (1-n^{-\gamma})n$,
(ii) (D, M) *is a* (d, q^{-4}, q, n)*-expander.*

The proof of Lemma 3.2 is technical. Here we present a quick sketch. The randomized matching M in Lemma 3.2 will be composed of two bits — M_0 and M_1. First we choose M_0 by picking every edge in G with probability q/n and deleting all colour and vertex collisions, for some $0 < q \ll 1$. This matching will be of size roughly qn and will satisfy certain "expansion" properties in second and third neighbourhoods which we describe next. Fix some D as in the statement of Lemma 3.2 and suppose (X, Y) is the bipartition of $M_0 \cup D$. We first show that M_0 has the following property:

- **Second neighbourhood expansion:** For each set $S \subseteq X$ or $S \subseteq Y$ of size roughly n/d we have $|N^2_{M_0,D}(S)| = (1-o(1))n$ with probability $1-e^{-O(|S|)}$(see Lemma 3.8).

The above property simply means for say $S \subseteq X$ that if we follow edges coming out of S that belong to M_0 and then follow the edges of D we reach almost all of X. Notice that we cannot prove that the second neighbourhood is large for all sets S simultaneously. Indeed, let D be a disjoint union of complete bipartite graphs of

size d on X, Y with $|X| = |Y| = n$. Now let M_0 be any perfect matching on $K_{n,n}$. Now let H be any union of disjoint $K_{d,d}$'s in D. If we let $S = V(H) \cap X$, then $|N^2_{D,M}(S)| = |N_M(V(H) \cap Y)| = |S|$ that is S won't expand. Note that it doesn't matter whether we follow the edges in the order of M and then D or otherwise. Indeed, take $S' = N_M(S)$, then $|N^2_{M,D}(S')| = |N_D(S)| = |S| = |S'|$. However, if we look at the same example in the third neighbourhood, that is $N^3_{D,M}(S)$ and let's assume $|S| = n/d$ then if M was a randomly picked perfect matching, it is likely that $N^2_{D,M}(S)$ will hit a vertex from each $K_{d,d}$ outside of H thus resulting $N^3_{D,M}(S)$ being almost all of X. And this is what we prove; using the second neighbourhood expansion and that M_0 is picked randomly, we show that with high probability *all* large sets S have expansion in their third neighbourhood.

- **Third neighbourhood expansion:** With high probability, all sets $S \subseteq X$ or $S \subseteq Y$ of size roughly n/d will have subsets S' of size roughly n/d^2 and $|N^3_{D,M_0}(S')| = (1 - o(1))n$ (see Lemma 3.9).

Finally notice that to obtain the expansion in the fourth neighbourhood in the sense of Definition 3.1, M_0 is not enough, as it is only of size roughly qn. That is why we need to extend M_0 to a nearly spanning rainbow matching. Let H be obtained from G by deleting vertices and colours of M_0. Since G was coloured regular and M_0 was picked randomly, H will be coloured regular as well (Lemma 3.10). We find a nearly spanning rainbow matching M_1 in H, which by definition of H, will be edge and colour disjoint from M_0. This is done by applying Lemma 2.5 to H, which gives a rainbow matching M_1 of size roughly $n - |M_0| - n^{1-\gamma}$. Then, taking $M = M_0 \cup M_1$, for any nearly regular graph D on $V(G)$ with high probability we obtain that all large sets will expand as in the Definition 3.1.

We start with an easy lemma exhibiting a feature of nearly-regular graphs.

Lemma 3.3. *For $\kappa \leq 1 \leq d$, let D be a bipartite graph with bipartition (X, Y) and $\Delta(D) \leq d$. Let $S \subseteq X$ or $S \subseteq Y$ be a set with $|S| \geq 2d$ such that every $s \in S$ satisfies $d_D(s) \geq \kappa d$. Then there exists a set $S' \subseteq S$ such that $|S'| \leq |S|/d$ and $|N_D(S')| \geq \kappa|S|/4$.*

Proof. Take the maximal collection of vertex-disjoint stars of size $\kappa d/2$ in D whose centers are in S and let F be the vertex set of their union. We are done if $|F \cap X| \geq |S|/2d$. Indeed, in this case any set $S' \subseteq S$ containing $|S|/2d$ many of the centers of the stars of F satisfies the lemma. So, we may assume $|F \cap X| < |S|/2d$. Since $d_D(x) \geq \kappa d$ for every x, by maximality of F, we have $|N_D(x) \cap F| \geq \kappa d/2$ for all $x \in S \setminus F$. On the other hand, since $\Delta(D) \leq d$, for any $y \in Y \cap F$, $|N_D(y)| \leq d$. Thus,

$$\frac{\kappa d}{2}|S \setminus F| \leq e(S \setminus F, Y \cap F) \leq d|Y \cap F|.$$

This implies

$$|Y \cap F| \geq \frac{\kappa}{2}|S \setminus F| = \frac{\kappa}{2}(|S| - |X \cap F|) > \frac{\kappa|S|}{4},$$

where in the last inequality we used $|F \cap X| < |S|/2d$ and $d \geq 1$. Now $S' = F \cap X$ has $|S'| < |S|/2d$ and $|N_D(S')| \geq |Y \cap F| > \kappa|S|/4$ as required. □

Lemma 3.4. *Let $n^{-1} \ll q \ll p \leq 1$, $n^{-1} \ll \varepsilon < 1$ and $q^{-1} \ll d \leq qn/8$. Let G, D be two bipartite graphs on the same vertex set with bipartition (X, Y) such that G is coloured (ε, p, n)-typical, $\Delta(D) \leq d$ and all but at most $96p^{-2}n/d$ vertices have degrees less than $pd/6$ in D. Let H be derived from G by picking every edge*

independently with probability q^2/n. *Let* $S \subseteq X$ *or* $S \subseteq Y$ *with* $|S| = \frac{n}{dq^3}$. *Then with probability at least* $1 - e^{-q^7|S|}$ *we have* $|N^2_{H,D}(S)| \geq (1-q)n$.

Proof. Let J be the set of vertices $v \in X \cup Y$ such that $d_D(v) < pd/6$. Without loss of generality, let us assume $S \subseteq X$. Denote $B = \{x \in X \setminus J : e_G(N_D(x), S) \leq p^2 d|S|/30\}$.

Claim 3.5. $|B| \leq qn/4$.

Proof. Suppose, for contradiction, that $|B| > qn/4$. By Lemma 3.3 there is $B' \subseteq B$ with $|B'| \leq |B|/d$ such that $|N_D(B')| \geq p|B|/24 > pqn/96$. We apply Lemma 2.7 with $\gamma_{2.7} = pq/9600$ and obtain

$$\begin{aligned} e_G(N_D(B'), S) &\geq p|N_D(B')||S| - 2p|N_D(B')|^{\frac{1}{2}}|S|(\gamma n)^{\frac{1}{2}} \\ &> \frac{4}{5}p|N_D(B')||S| \geq \frac{1}{30}p^2|B||S|. \end{aligned}$$

On the other hand, by the definition of B' we have $e_G(N_D(B'), S) \leq p^2 d|S|/30 \cdot |B'| \leq p^2 \frac{|B||S|}{30}$, which is a contradiction. □

Claim 3.6. For every $x \in X \setminus (B \cup J)$, $\mathbb{P}[x \notin N^2_{H,D}(S)] \leq q/4$.

Proof. For each such x, define the set $S_x = \{s \in S : |N_G(s) \cap N_D(x)| \geq p^2 d/60\}$. Using $x \notin B$ we get

$$\frac{p^2 d|S|}{30} < e_G(N_D(x), S) \leq |S_x||N_D(x)| + |S \setminus S_x|\frac{p^2 d}{60} \leq d|S_x| + \frac{p^2 d|S|}{60},$$

implying that $|S_x| \geq p^2|S|/60$.

Now we compute the probability of the event $x \notin N^2_{H,D}(S)$.

$$\begin{aligned} &\mathbb{P}[x \notin N^2_{H,D}(S)] \\ &\quad = \mathbb{P}[\forall s \in S, \forall y \in N_D(x) \cap N_G(s) \text{ we have } sy \notin E(H)] \\ &\quad = \prod_{s \in S} \prod_{y \in N_D(x) \cap N_G(s)} \mathbb{P}[sy \notin E(H)] = \prod_{s \in S} \left(1 - \frac{q^2}{n}\right)^{|N_D(x) \cap N_G(s)|} \\ &\quad \leq \prod_{s \in S_x} \left(1 - \frac{q^2}{n}\right)^{|N_D(x) \cap N_G(s)|} \leq \prod_{s \in S_x} \left(1 - \frac{q^2}{n}\right)^{p^2 d/60} \\ &\quad \leq \prod_{s \in S_x} e^{-p^2 q^2 d/60n} \leq (e^{-p^2 q^2 d/60n})^{p^2|S|/60} = e^{-p^4 q^{-1}/3600} \leq q/4. \end{aligned}$$

Here the second equation comes from independence of the events "$sy \in E(H)$", the first inequality comes from $S_x \subseteq S$, the second one from the definition of S_x, the third one comes from $1 - x \leq e^{-x}$, the fourth one comes from $|S_x| \geq p^2|S|/60$, and the last one holds since $q \ll p \leq 1$. □

By linearity of expectation we have $\mathbb{E}[|N^2_{H,D}(S)|] \geq (1 - q/4)(|X| - |B| - |J|) \geq (1 - q/4)(n - n^{1-\varepsilon} - qn/4 - 96p^{-2}n/d) \geq (1 - q/2)n$. Notice that the random variable $|N^2_{H,D}(S)|$ is defined on the product space Ω consisting of all the edges in G from S to Y, where the probability of every coordinate being one is q^2/n. This product space has $e(S, Y) = |S|pn(1 \pm n^{-\varepsilon})$ coordinates. Notice that $|N^2_{H,D}(S)|$ is

d-Lipchitz, thus we can apply Lemma 2.3. Let $\sigma^2 = d^2 \sum_{i\in\Omega} \left(1 - \frac{q^2}{n}\right) \frac{q^2}{n}$. Note that

$$\frac{qd\sqrt{p|S|}}{2} \le \sigma = qd\sqrt{\sum_{i\in\Omega} \left(1 - \frac{q^2}{n}\right) \frac{1}{n}} \le 2qd\sqrt{p|S|}.$$

So let $t = q^3\sqrt{|S|}/4$. Note that $td \le 2\sigma$ and $\sigma t \le (2qd\sqrt{p|S|})(q^3\sqrt{|S|}/4) \le qn/2$, since $|S| = n/dq^3$. Thus by Azuma's inequality we have:

$$\begin{aligned}\mathbb{P}[|N^2_{H,D}(S)| \le (1-q)n] &\le \mathbb{P}[|N^2_{H,D}(S)| \le \mathbb{E}[|N^2_{H,D}(S)|] - \sigma t] \\ &\le 2e^{-t^2/4} = 2e^{-q^6|S|/64} \le e^{-q^7|S|}.\end{aligned}$$

□

The graph H produced by the previous lemma won't generally be a matching or a rainbow subgraph. Lemma 3.7 estimates how many of its edges conflict with other edges due to a vertex or a colour collision.

Lemma 3.7. *Let $n^{-1} \ll q \ll p \le 1$, $n^{-1} \ll \varepsilon < 1$. Let G be a bipartite graph with bipartition (X, Y) such that G is coloured (ε, p, n)-regular. Let H be derived from G by picking every edge independently with probability q/n, let $M \subseteq H$ be consisting of edges which don't share any vertices or colours with other edges in H, define $H' := H - M$. Then for any set $S \subseteq X$ with $|S| \ge q^{-3}$,*

$$\mathbb{P}(|N_{H'}(S)| \ge 5q^2|S|) \le e^{-q^3|S|}.$$

Proof. Let $xy \in E(G)$. For xy to be in M we need $xy \in H$ and also $e \notin H$ for all edges e sharing a vertex or a colour with xy. Thus

$$\begin{aligned}\mathbb{P}[xy \in M] &= \frac{q}{n}\left(1 - \frac{q}{n}\right)^{d_G(x)+d_G(y)+|E_G(c)|-3} \\ &= \frac{q}{n}\left(1 - \frac{q}{n}\right)^{3pn(1\pm n^{-\varepsilon})-3} = \frac{q}{n}\left(1 - \frac{q}{n}\right)^{3pn}(1 \pm 4qpn^{-\varepsilon}).\end{aligned}$$

Here the second equation uses coloured (ε, p, n)-regularity, and the third equation comes from $(1 - q/n)^{\pm 3pn^{1-\varepsilon}-3} = (1 \pm 4qpn^{-\varepsilon})$. This gives

$$\begin{aligned}\mathbb{P}[xy \in E(H')] &= \mathbb{P}[xy \in E(H)] - \mathbb{P}[xy \in E(M)] \\ &\le \frac{q}{n}\left(1 - \left(1 - \frac{q}{n}\right)^{3pn}(1 - 4qpn^{-\varepsilon})\right) \le \frac{3q^2p}{n} + 4q^2pn^{-1-\varepsilon} \le \frac{4q^2}{n},\end{aligned}$$

which implies $\mathbb{E}[|N_{H'}(S)|] \le 4q^2|S|$. Notice $|N_{H'}(S)|$ is 3-Lipschitz since adding or removing an edge e from H can affect at most two neighbouring edges or one edge of the same colour to be in H or not. $|N_{H'}(S)|$ is also 2-certifiable. Our product space is $\Omega = (x_1, x_2, \ldots, x_{|e(G)|})$, where each $x_i = 1$ if the ith edge is in H. Suppose the current outcome of H is described by $\omega \in \Omega$. Thus if $|N_{H'}(S)| \ge s$ then we can take I to be as follows. Note that for each edge e appearing in $|N_{H'}(S)|$ there exists an edge e' of the same colour or sharing a vertex with e which appears in H. We let I to be the coordinate of all edges e in $N_{H'}(S)$ and coordinates of corresponding e''s. This will guarantee that with respect to any ω' that agrees with ω on I must have $|N_{H'}(S)| \ge s$. Thus we can apply Talagrand's inequality with $t = q^2|S|/2$, $r = 2$, $c = 3$. We use that $60 \cdot 3\sqrt{2 \cdot 4q^2|S|} \le q^2|S|/2$ since $q \ll 1$ and $|S| \ge q^{-3}$ to get

$$\mathbb{P}[|N_{H'}(S)| > 5q^2|S|] \le 4e^{-\frac{(q^2|S|/2)^2}{8\cdot 9\cdot 2\cdot 4q^2|S|}} \le e^{-q^3|S|}.$$ □

Finally we are ready to prove our first lemma guaranteeing expansion in the second neighbourhood $N^2_{M,D}(S)$ of large sets S for a randomized rainbow matching M and a nearly regular graph D.

Lemma 3.8. *Let $n^{-1} \ll q \ll p \leq 1$, $n^{-1} \ll \varepsilon < 1$ and $q^{-1} \ll d \leq qn/8$. Let G, D be two bipartite graphs on the same vertex set with bipartition (X, Y) such that G is coloured (ε, p, n)-typical, $\Delta(D) \leq d$ and all but at most $96p^{-2}n/d$ vertices have degrees less than $pd/6$ in D. Let H be obtained from G by picking every edge with probability q^2/n, let $M \subseteq H$ be consisting of edges which don't share any vertices or colours with other edges in H. If $S \subseteq X$ with $|S| = \frac{n}{dq^3}$ then with probability at least $1 - 2e^{-q^6 n/d}$ we have $|N^2_{M,D}(S)| \geq (1-6q)n$.*

Proof. By Lemma 3.4 we have that $|N^2_{H,D}(S)| \leq (1-q)n$ with probability at most $e^{-q^7|S|}$. By Lemma 3.7 applied with $q_{3.7} = q^2$ we have $|N_{H\setminus M}(S)| \geq 5q^4|S|$ with probability at most $e^{-q^6|S|}$. By the union bound, we get that with probability at least $1 - 2e^{-q^6|S|}$ both of these events don't happen.

Since $|N_{H\setminus M}(S)| < 5q^4|S|$ and $\Delta(D) \leq d$, we have that at most $5dq^4|S|$ many edges of D touch $N_{H\setminus M}(S)$. Thus,

$$|N^2_{M,D}(S)| \geq |N^2_{H,D}(S)| - |N^2_{H\setminus M,D}(S)| \geq (1-q)n - 5dq^4|S| = (1-6q)n.$$

□

The next lemma builds on Lemma 3.8 and guarantees expansion in the third neighbourhood for subsets of size roughly n/d^2.

Lemma 3.9. *Let $n^{-1} \ll q \ll p \leq 1$, $n^{-1} \ll \varepsilon < 1$ and $q^{-1} \ll d \leq \sqrt{n}$. Let G, D be two bipartite graphs on the same vertex set with bipartition (X, Y) such that G is coloured (ε, p, n)-typical, $\Delta(D) \leq d$ and all but at most $96p^{-2}n/d$ vertices have degrees less than $pd/6$ in D. Let H be obtained from G by picking every edge with probability q^2/n and define $M \subseteq H$ to be consisting of edges which don't share any vertices or colours with other edges in H. Then with probability $\geq 1 - e^{-q^7 n/d}$ the following holds.*

For any $S \subseteq X$ or $S \subseteq Y$ with $|S| \geq \frac{25n}{pq^3d}$ there exists $S' \subseteq S$ such that $|S'| = \frac{24n}{pq^3d^2}$ such that $|N^3_{D,M}(S')| \geq (1-6q)n$.

Proof. Let J be the set of vertices $v \in X \cup Y$ with $d_D(v) \leq pd/6$. By assumption $|J| \leq 96p^{-2}n/d \leq \frac{n}{dpq^3}$ since $q \ll p$. Without loss of generality let us assume $S \subseteq X$. Since $|S| \geq \frac{25n}{pq^3d}$ by throwing away at most $\frac{n}{pq^3d}$ vertices we may assume $S \cap J = \emptyset$ and $|S| \geq \frac{24n}{pq^3d}$.

By Lemma 3.3, there is a set $S' \subseteq S$ with $|S'| \leq 24n/pq^3d^2$ such that $|N_D(S')| \geq n/dq^3$. By adding extra vertices from S to S' we may assume $|S'| = 24n/pq^3d^2$. Fix one such set S' for each S. We say that S' is *bad* if it has $|N^3_{D,M}(S')| < (1-6q)n$. To prove the lemma it is sufficient to show that with probability $\geq 1 - e^{-q^7 n/d}$, there are no bad sets S'.

Let $S' \subseteq X$ with $|S'| = \frac{24n}{pq^3d^2}$ and $|N_D(S')| \geq n/dq^3$. By Lemma 3.8 (applied to a subset of $N_D(S')$ of order exactly n/dq^3), with probability at least $1 - 2e^{-q^6 n/d}$, we have $|N^2_{M,D}(N_D(S'))| \geq (1-6q)n$. Recall that "$N^3_{D,M}(S')$" means we are looking

at edges going out of S' to be in the order of D, M and D, thus $N^2_{M,D}(N_D(S')) = N^3_{D,M}(S')$. So we have shown that $\mathbb{P}(S' \text{ is bad}) \leq 2e^{-q^6 n/d}$. By taking a union bound over all $S' \subseteq X$, with $|S'| = \frac{24n}{pq^3d^2}$ and using $d \gg q^{-1}$, we obtain

$$\mathbb{P}[\exists \text{ bad } S'] \leq \binom{n+n^{-\varepsilon}}{24n/pq^3d^2} \cdot 2e^{-q^6 n/d} \leq 2\left(\frac{2en}{24n/pq^3d^2}\right)^{\frac{24n}{pq^3d^2}} e^{-q^6 n/d}$$
$$\leq 2e^{-q^6 n/d + \frac{24n}{pq^3d^2}\log pq^3d^2} \leq 2e^{-q^6 n/2d}.$$

Similarly, the probability that there exists a bad $S' \subseteq Y$ is at most $2e^{-q^6 n/2d}$. Thus with probability at least $1 - e^{-q^7 n/d}$ there are no bad $S' \subseteq X \cup Y$. □

In the next lemma we show that if we have a coloured regular graph G then if we pick a random rainbow matching and delete all of its edges and colours from the graph G then the remaining graph is still a coloured regular graph.

Lemma 3.10. *Let $n^{-1} \ll p, q, \varepsilon$ with $\varepsilon \ll 1$, $q \leq 1/2$, and $p \leq 1$. Let G be coloured (ε, p, n)-regular bipartite graph with bipartition (X, Y). Let M be a random rainbow matching obtained from G by picking every edge with probability q/n and deleting all colour and vertex collisions. Let H be G with vertices and colours of M deleted. Then there are numbers $m > n/2, p' > p/2$ such that with probability at least $1 - e^{-n^{1-\varepsilon}}$, the graph H is $(\varepsilon/10, p', m)$-regular.*

Proof. Let $d = pn$ and $\alpha = \left(1 - \frac{q}{n}\right)^{3d} q$. We will see that every edge of G ends up in M with probability roughly α/n. Denote $x_H = |X \cap V(H)|$, $y_H = |Y \cap V(H)|$, and $c_H = |C(H)|$. For any vertex $v \in G$, let $d_H(v) = |N_{C(H)}(v) \cap V(H)|$, and note that for vertices $v \in H$ this is just their degree in H. Similarly, for any colour $c \in G$, let $e_H(c) = |E_G(c) \cap E(H)|$, and note that for colours $c \in C(H)$ this is just the number of edges they have in H. We need to show that with probability at least $1 - e^{-n^{1-\varepsilon}}$ the following hold for appropriately chosen p' and m:

(P1) $x_H, y_H = m(1 \pm m^{-\varepsilon/10})$,
(P2) $c_H = m(1 \pm m^{-\varepsilon/10})$,
(P3) $d_H(v) = p'm(1 \pm m^{-\varepsilon/10})$, for every $v \in V(H)$,
(P4) $e_H(c) = p'm(1 \pm m^{-\varepsilon/10})$, for every $c \in C(H)$.

Claim 3.11.

- $\mathbb{E}[x_H], \mathbb{E}[y_H] = n(1 - p\alpha)(1 \pm n^{-\varepsilon/5})$,
- $\mathbb{E}[d_H(v)] = pn(1 - p\alpha)^2(1 \pm n^{-\varepsilon/6})$ for every vertex $v \in V(G)$,
- $\mathbb{E}[c_H] = n(1 - p\alpha)(1 \pm n^{-\varepsilon/5})$,
- $\mathbb{E}[e_H(c)] = pn(1 - p\alpha)^2(1 \pm n^{-\varepsilon/6})$ for every colour $c \in C(G)$.

Proof. By the symmetry between vertices and colours, it is enough to show that the first two hold. We estimate several probabilities. At various points, to bound errors we use that for any positive constant k, $(1 - q/n)^{\pm kpn^{1-\varepsilon}} = (1 \pm n^{-\varepsilon/2})$ and $(1 \pm n^{-\varepsilon})^k = 1 \pm n^{-\varepsilon/2}$ (which holds as long as $n^{-1} \ll q, \varepsilon, p$). For a colour c edge xy, let $F(xy)$ be the set of edges of $G \setminus xy$ sharing a colour or vertex with xy. Note that since G is coloured (ε, p, n)-regular, we have

$$|F(xy)| = d_G(x) + d_G(y) + |E_G(c)| - 3 = 3d(1 \pm 2n^{-\varepsilon}).$$

- **The probability that an edge is in M:**
  We say an edge $e \in G$ is **chosen** if it was picked at the first step when generating M (with probability q/n). By definition of M, $e \in M$ exactly when e is chosen and none of the edges from $F(e)$ are chosen. This has probability

$$\mathbb{P}[e \in M] = \frac{q}{n}\left(1-\frac{q}{n}\right)^{|F(e)|} = \frac{q}{n}\left(1-\frac{q}{n}\right)^{3d(1\pm 2n^{-\varepsilon})} = \frac{\alpha}{n}(1\pm n^{-\varepsilon/2}).$$

- **The probability that a pair of edges are both in M:** Let f, e be a pair of edges which don't share any vertices or a colour. Notice that

$$|F(e)\cup F(f)| = |F(e)| + |F(f)| \pm 2 = 6d(1\pm 3n^{-\varepsilon}).$$

  By definition of M, we have $e, f \in M$ exactly when e, f are chosen and none of the edges of $F(e)\cup F(f)$ are chosen which happens with probability

$$\begin{aligned}\mathbb{P}[e,f\in M] &= \left(\frac{q}{n}\right)^2\left(1-\frac{q}{n}\right)^{|F(e)\cup F(f)|}\\ &= \left(\frac{q}{n}\right)^2\left(1-\frac{q}{n}\right)^{6d(1\pm 3n^{-\varepsilon})} = \frac{\alpha^2}{n^2}(1\pm n^{-\varepsilon/2}).\end{aligned}$$

- **The probability that a vertex/colour is in M:** For any v, we have

$$\begin{aligned}\mathbb{P}[v\in M] &= \sum_{y\in N_G(v)}\mathbb{P}[vy\in M] = d_G(v)\frac{\alpha}{n}(1\pm n^{-\varepsilon/2})\\ &= p\alpha(1\pm n^{-\varepsilon/2})^2 = p\alpha(1\pm n^{-\varepsilon/4}).\end{aligned}$$

  By the symmetry between vertices and colours, we also have $\mathbb{P}[c\in C(M)] = p\alpha(1\pm n^{-\varepsilon/4})$ for every colour c.
- **For an edge uv, the probability that u or $c(uv)$ is in M:** Fix an edge uv. We first estimate the probability that "$u\in M$ and $c(uv)\in M$". Notice that there are two ways this can happen — either $uv\in M$ or there are three distinct vertices, w, x, y such that uw, $xy\in M$ with $c(xy) = c(uv)$. We will see that the probability of the first event is negligible compared to the second. For an edge uv, let $J(uv)\subseteq E(G)\times E(G)$ be the set of pairs (uw, xy) as described above. We have

$$(d_G(u)-1)(|E_G(c(uv))|-2) \le |J(uv)| \le d_G(u)|E_G(c(uv))|.$$

  This implies $|J(uv)| = p^2n^2(1\pm n^{-\varepsilon})^3$ which implies

$$\begin{aligned}\mathbb{P}[u\in M, c(uv)\in M] &= \mathbb{P}[vu\in M] + \sum_{(e,f)\in J(uv)}\mathbb{P}[e,f\in M]\\ &= \frac{\alpha}{n}(1\pm n^{-\varepsilon/2}) + (pn)^2\frac{\alpha^2}{n^2}(1\pm n^{-\varepsilon/2})^4 = p^2\alpha^2(1\pm n^{-\varepsilon/5}),\end{aligned}$$

  where in the last equality we used that $\alpha\approx e^{-3pq}q$ and so $\alpha/n \ll p^2\alpha^2n^{-\varepsilon/2}$, as long as n is sufficiently large. Thus,

$$\begin{aligned}\mathbb{P}[u\in M \text{ or } c(uv)\in M] &= \mathbb{P}[u\in M] + \mathbb{P}[c(uv)\in M] - \mathbb{P}[u\in M, c(uv)\in M]\\ &= \left(2p\alpha - p^2\alpha^2\right)(1\pm n^{-\varepsilon/5}).\end{aligned}$$

Finally we are ready to estimate the expectations in the claim. By linearity of expectation,

$$\mathbb{E}[x_H] = \sum_{v\in X} \mathbb{P}(v \notin M) = n(1\pm n^{-\varepsilon})(1-p\alpha(1\pm n^{-\varepsilon/4})) = n(1-p\alpha)(1\pm n^{-\varepsilon/5}).$$

$$\begin{aligned}\mathbb{E}[d_H(v)] &= \sum_{u\in N_G(v)} (1-\mathbb{P}[u\in M \text{ or } c(vu)\in M]) \\ &= pn(1\pm n^{-\varepsilon})(1-(2p\alpha+p^2\alpha^2)(1\pm n^{-\varepsilon/5})) = pn(1-p\alpha)^2(1\pm n^{-\varepsilon/6}).\end{aligned}$$

□

Fix $m=(1-p\alpha)n$ and $p'=p(1-p\alpha)$. Notice that we may assume $m>n/2$ and $p'>p/2$ as we can guarantee $p\alpha<pq\le 1/2$, since n is sufficiently large.

Notice that the random variables x_H, y_H, c_H, $d_H(v)$, and $e_H(c)$ depend on the probability space $\Omega=\{0,1\}^{E(G)}$ with every coordinate being 1 with probability q/n. All these variables are 3-Lipshitz. Set $\sigma^2=3^2\sum_{e\in E(G)} q/n(1-q/n)$ and notice that

$$\frac{9pqn}{4} < \sigma^2 = 9pqn(1-q/n)(1\pm n^{-\varepsilon})^2 < 10pqn.$$

Set $t=n^{1/2-\varepsilon/3}$. Note that $t\le 2\sigma/3$ and $t\sigma<3n^{1-\varepsilon/3}$, since n is sufficiently large. By Lemma 2.3 we have

$$\mathbb{P}\Big[|x_H-\mathbb{E}[x_H]|>3n^{1-\varepsilon/3}\Big] < 2e^{-\frac{n^{1-\frac{2\varepsilon}{3}}}{4}} < \frac{e^{-n^{1-\varepsilon}}}{n^3}.$$

Similarly one can show that each of the random variables y_H, c_H, $d_H(v)$, and $e_H(c)$ are within $3n^{1-\varepsilon/3}$ of their expectations with probability at least $1-\frac{e^{-n^{1-\varepsilon}}}{n^3}$. If we take a union bound over all vertices and colours, we can guarantee that x_H, y_H, c_H, $d_H(v)$, and $e_H(c)$ are simultaneously all within $3n^{1-\varepsilon/3}$ of their expectations for all c and v. To conclude that (P1)–(P4) hold, it remains to check that $3n^{1-\varepsilon/3}+mn^{-\varepsilon/6}\le m^{1-\varepsilon/10}$. □

We now prove the main result of this section.

Proof of Lemma 3.2. Fix $\hat{q}=(25p^{-1}q^4)^{1/3}$. Note that $\hat{q}\ll p$ since $q\ll p$. Suppose G has bipartition (X,Y). Let M_0 be generated by picking every edge of G with probability $\hat{q}^2/n$ and deleting all vertex and colour collisions. Let H be obtained from G by removing the vertices and colours of M_0. By Lemma 3.10, with probability at least $1-e^{-n^{1-\varepsilon}}$ we have that H is $(\varepsilon/10,p',m)$-regular for some suitable p' and m. When H is $(\varepsilon/10,p',m)$-regular, by Lemma 2.5, there is a rainbow matching M_1 in H of size $\ge m-m^{1-2\gamma}$. Note that $|V(H)\cap X|=|X|-|M_0|$ and when H is $(\varepsilon/10,p',m)$-regular we have $|V(H)\cap X|=m(1\pm m^{-\varepsilon/10})$. Therefore, it follows that $m\ge|X|-|M_0|-m^{1-\varepsilon/10}\ge n-|M_0|-n^{1-\varepsilon}-n^{1-\varepsilon/10}$, which implies that $|M_1\cup M_0|\ge m-m^{1-2\gamma}+|M_0|\ge n-2n^{1-\varepsilon/10}-n^{1-2\gamma}\ge n-2n^{1-2\gamma}\ge n-n^{1-\gamma}$, since $\gamma\ll\varepsilon$. We will show that the conclusion of the lemma holds for the randomized rainbow matching $M=M_0\cup M_1$. Notice that with probability at least $1-e^{-n^{1-\varepsilon}}$ we have

$\mathcal{E}_1$: $|M|\ge(1-n^{-\gamma})n$.

Let D be a bipartite graph with the same bipartition as G having $\Delta(G)\le d$ and at most $\le 96p^{-1}n/d$ vertices of degrees less than $pd/6$. We can apply Lemma 3.9 to M_0, G, and D and obtain that with probability at least $1-e^{-\hat{q}^7n/d}$ we have

$\mathcal{E}_2$: For any $S \subseteq X$ or $S \subseteq Y$ with $|S| \geq \frac{25n}{dp\hat{q}^3}$ there exists $S' \subseteq S$ such that $|S'| = \frac{24n}{p\hat{q}^3d^2}$ such that $|N^3_{D,M_0}(S')| \geq (1-6\hat{q})n$.

So with probability at least $1 - e^{-n^{1-\varepsilon}} - e^{-\hat{q}^7 n/d} \geq 1 - 2e^{-n^{1-\varepsilon}}$ both events $\mathcal{E}_1$ and $\mathcal{E}_2$ happen. Clearly (i) holds then, let us show that (ii) holds as well. Let D be as earlier and without loss of generality assume $S \subseteq X$ with $|S| \geq \frac{n}{dq^4}$. Then since $q \ll p$ it follows that $|S| \geq \frac{25n}{dp\hat{q}^3}$. Since $\mathcal{E}_2$ holds there exists $S' \subseteq S$ such that $|S'| = \frac{24n}{p\hat{q}^3d^2} \geq 24n/25q^4d^2$ such that $|N^3_{D,M_0}(S')| \geq (1-6\hat{q})n$. Since $|X \setminus M| \leq n + n^{1-\varepsilon} - (n - n^{1-\gamma}) \leq 2n^{1-\gamma}$, it follows that

$$|N^4_{D,M}(S')| \geq |N^3_{D,M_0}(S')| - |X \setminus M| \geq (1-6\hat{q})n - 2n^{1-\gamma} \geq (1-q)n,$$

where the last inequality holds since $q \ll p$. Finally we can always add extra vertices of $S \setminus S'$ to S' to make it exactly of size n/q^4d^2. This finishes the proof. □

3.2. **Switchings via expansion.** The main result of this section is the lemma below which is our main tool for doing switchings. It says that if we have two expanders (D_1, M) and (D_2, M) such that D_1 and D_2 are two bipartite graphs on the same vertex set then almost all pairs of vertices lying in the opposite sides of the bipartition of the graph $D_1 \cup D_2 \cup M$ have a short rainbow $(D_1 \cup D_2) - M$-alternating path between them.

Lemma 3.12. *Let $d^{-1/2} \leq A^{-1} \leq \varepsilon/100 \ll 1$ and further, $d \log d \geq 8A^2 \log n$. Suppose we are given two bipartite graphs D_1, D_2 and a rainbow matching M on the bipartition (X, Y) with $\Delta(D_1), \Delta(D_2) \leq d$, $M \cup D_1 \cup D_2$ properly edge-coloured, $C(M)$, $C(D_1)$, $C(D_2)$ pairwise disjoint, and $|X|, |Y| < (2-4\varepsilon)n$. If for both $i = 1, 2$, (D_i, M) is a (d, A, ε, n)-expander then there is a set $B \subseteq X \cup Y$ of at most $4An/d$ vertices, such that for all $u, v \notin B$ lying in the opposite sides of the bipartition, there is a $(D_1 \cup D_2)$-M-alternating rainbow path from u to v of length at most $8\lceil \frac{\log n}{\log (d/4A)} \rceil$.*

The alternating paths found by the above lemma will be used to go from one rainbow matching to another. When proving the existence of large rainbow matchings in typical graphs, we will start from some rainbow matching and iteratively do such switchings, eventually enlarging the original rainbow matching to one of a desired size. We need a simple lemma which claims that if we have an expander (D, M) then any "small" perturbation of the matching M will keep the expansion properties. (In applications of this lemma "small" would mean sub-polynomial.)

Lemma 3.13. *Suppose we are given two matchings M_1, M_2 and a bipartite graph D with $\Delta(D) \leq d$, all on the same bipartition (X, Y). If (D, M_1) is a (d, A, ε, n)-expander and $|M_1 \triangle M_2| < \varepsilon n/10d^2$ then (D, M_2) is a $(d, A, 2\varepsilon, n)$-expander.*

Proof. Let $S \subseteq X$ or Y with $|S| \geq An/d$. Since (D, M_1) is a (d, A, ε, n)-expander, there is a subset $S' \subseteq S$ with $|S'| = An/d^2$ such that $|N^4_{D,M_1}(S')| \geq (1-\varepsilon)n$. By definition, for each $v \in N^4_{D,M}(S')$ there is a D-M_1-alternating path of length four from S' to v. We call such an alternating D-M_1 path of length four *bad* if it uses any vertex from $V(M_1) \setminus V(M_2)$ and *good* otherwise. The number of bad paths is at most $5\Delta(D)^2|V(M_1) \setminus V(M_2)| \leq 10|M_1 \triangle M_2|d^2 < \varepsilon n$. Indeed, there are 5 ways to choose which vertex of the path of length four is in $V(M_1) \setminus V(M_2)$ and then at most $\Delta(D)^2$ ways to choose D-M_1-alternating path which has this vertex in the

correct position. Hence there are at least $(1-2\varepsilon)n$ good paths. Every good path is an alternating D-M_2 path, and so we have $|N^4_{D,M_2}(S')| \geq (1-2\varepsilon)n$. □

We say a path P in an edge-coloured graph G *avoids* a vertex subset $V' \subseteq V(G)$ if it doesn't contain any vertex from V'. Similarly, P *avoids* a colour subset $C' \subseteq C(G)$ if it does not contain any edge of colours from C'. To prove Lemma 3.12 we first show that given an expander (D, M) such that D and M are colour disjoint, all large sets S "expand" in the following coloured fashion: almost every vertex in $v \in V(D) \cup V(M)$ can be reached from some $s \in S$ via a rainbow $D-M$-alternating path of length four and additionally, this path avoids some small set of forbidden colours and vertices prescribed to s a priori (Lemma 3.14). Then we apply this iteratively to obtain a similar expansion property for smaller sets (Lemma 3.15). Finally via applying this iteration multiple times we show that almost all vertices can reach almost all vertices via rainbow $D-M$-alternating paths of length $O(\log n/\log\log n)$ (Lemma 3.16).

Lemma 3.14. *Let $d^{-1/2} \leq A^{-1} \leq \varepsilon/100 \ll 1$. Suppose we are given a bipartite graph D with $\Delta(D) \leq d$, M a rainbow matching such that $M \cup D$ has bipartition (X, Y), is properly edge-coloured, and $C(M)$ and $C(D)$ are disjoint. Let $C = C(D) \cup C(M)$, $V = V(D) \cup V(M)$. If (D, M) is a (d, A, ε, n)-expander, then for any $S \subseteq X$ or Y with $|S| = An/d$ and any collections of "forbidden" colours and vertices $\{C(s) \subseteq C | s \in S\}$, $\{V(s) \subseteq V | s \in S\}$ with $|C(s)| \leq A^{-2}d$, $|V(s)| \leq A^{-2}d$, $s \notin V(s)$ for all $s \in S$ the following holds. There are at least $(1-2\varepsilon)n$ vertices $v \in V$ for which there is a D-M-alternating rainbow path P_v of length four going from some $s_v \in S$ to v and avoiding $C(s_v)$ and $V(s_v)$.*

Proof. Since (D, M) is a (d, A, ε, n)-expander there exists $S' \subseteq S$ of order An/d^2 such that $|N^4_{D,M}(S')| \geq (1-\varepsilon)n$. For every $v \in N^4_{D,M}(S')$, there is some $s \in S'$ and a D-M-alternating path P_v of length four going from s to v. We say that P_v is *bad* if either P_v is not rainbow, or P_v doesn't avoid $C(s), V(s)$. We say P_v is good otherwise. We will show that the total number of bad paths among all the paths $\{P_v\}_{v\in V}$ is at most εn. Note that this would be enough for the conclusion of the lemma as we can take the final vertex set to be $\{v \in N^4_{D,M}(S') | P_v \text{ is good}\}$. To count the total number of bad paths P_v, we count for all $s \in S'$ how many bad paths start at s.

Fix a vertex $s \in S'$. Notice that a bad D-M alternating path $sxyzw$ must satisfy at least one of the following:

- $sxyzw$ is not rainbow. Since $D \cup M$ is properly edge-coloured, and M is rainbow and colour disjoint from D, this happens only when $c(sx) = c(yz)$. There are at most $\Delta(D) \leq d$ such D-M alternating paths starting from s, since there are at most d choices for x and at most one for all the other vertices.
- Some vertex among x, y, z, w is from $V(s)$. The number of such D-M alternating paths is at most $4|V(s)|d$. Indeed, there are at most four choices for which some vertex among x, y, z, w is in $V(s)$. Then $|V(s)|$ choices to specify that vertex v. Now suppose $x = v$ then there is at most one choice for y depending if x is covered by the matching M or not, at most d choices for z since $\Delta(D) \leq d$ and finally at most one choice for w. So there are at most d such paths with $x = v$. Similar argument applies if $y = v$ or $z = v$ or $w = v$.

- Some edge among sx, xy, yz, zw has a colour appearing in $C(s)$. The number of such D-M alternating paths is at most $4|C(s)|d$. Again, there are $4|C(s)|$ choices to specify which one of the edges sx, xy, yz, zw has colour $c \in C(s)$. Suppose we are counting the number of paths $sxyzw$ with $c(sx) = c$. Since the colouring is proper there is at most one edge of colour c coming out of s therefore at most one choice for vertex x. Then there is at most one choice for y, depending if x is covered by the matching M or not, at most d choices for z since $\Delta(D) \leq d$ and finally at most one choice for w. A similar analysis with the same bound will apply if $c(xy) = c$, $c(yz) = c$ or $c(zw) = c$.

Thus the total number of bad paths starting at s is $\leq 8A^{-2}d^2 + d \leq 9A^{-2}d^2$ (using $d^{-1/2} \leq A^{-1}$). Summing over all $s \in S'$, we get that the total number of bad paths is at most $9A^{-2}d^2|S'| = 9A^{-1}n \leq \varepsilon n$ as desired. □

Recall that $N^t_{D,M}(S)$ denotes the set of vertices to which there is a D-M-alternating path of length t starting in S. We use $\hat{N}^t_{G,H}(S)$ to denote the set of vertices to which there is a D-M-alternating *rainbow* path of length t starting in S.

Lemma 3.15. *Let $d^{-1/2} \leq A^{-1} \leq \varepsilon/100 \ll 1$ and $t \leq A^{-2}d/4$. Suppose we are given a bipartite graph D with $\Delta(D) \leq d$, M a rainbow matching such that $M \cup D$ has bipartition (X, Y), is properly edge-coloured, and $C(M)$ and $C(D)$ are disjoint. If (D, M) is a (d, A, ε, n)-expander then for every set of vertices $S \subseteq X$ or Y with $|\hat{N}^{4t}_{D,M}(S)| \geq (1 - 2\varepsilon)n$ there is $S' \subseteq S$ with $|S'| = \lceil 2A|S|/d \rceil$ and $|\hat{N}^{4t+4}_{D,M}(S)| \geq (1 - 2\varepsilon)n$.*

Proof. For each $v \in \hat{N}^{4t}_{D,M}(S)$, by definition there exists $s_v \in S$ and a D-M-alternating rainbow path P_v of length $4t$ from s_v to v. For each v fix such s_v and P_v. For each $s \in S$, let $p(s)$ be the number of paths P_v starting at s. We have $\sum_{s\in S} p(s) = |\hat{N}^{4t}_{D,M}(S)| \geq (1 - 2\varepsilon)n$. Let $S' \subseteq S$ be a subset of size $\lceil 2A|S|/d \rceil$ with $\sum_{s\in S'} p(s)$ maximum. By averaging, $\frac{1}{|S'|}\sum_{s\in S'} p(s) \geq \frac{1}{|S|}\sum_{s\in S} p(s)$. Thus $|\hat{N}^{4t}_{D,M}(S')| \geq \sum_{s\in S'} p(s) \geq \frac{|S'|}{|S|}\sum_{s\in S} p(s) \geq (1 - 2\varepsilon)2An/d \geq An/d$.

Let $T \subseteq \hat{N}^{4t}_{D,M}(S')$ be a subset of size exactly An/d. To each vertex $v \in T$, assign forbidden sets of colours and vertices $C(v) := C(P_v)$, $V(v) := V(P_v) \setminus \{v\}$, and note that $|C(v)|, |V(v)| \leq 4t \leq A^{-2}d$. By Lemma 3.14, we get $(1 - 2\varepsilon)n$ vertices u together with rainbow $D - M$-alternating paths Q_u of length four from some $v_u \in T$ such that Q_u avoids $C(v_u)$ and $V(v_u)$. It is easy to check that for each u, $P_{v_u} \cup Q_u$ is a rainbow $D - M$-alternating path of length $4t + 4$. This finishes the proof. □

Lemma 3.16. *Let $d^{-1/2} \leq A^{-1} \leq \varepsilon/100 \ll 1$ and further, $d \log d \geq 8A^2 \log n$. Suppose we are given a bipartite graph D with $\Delta(D) \leq d$, M a rainbow matching such that $M \cup D$ has bipartition (X, Y), is properly edge-coloured, and $C(M)$ and $C(D)$ are disjoint. If (D, M) is a (d, A, ε, n)-expander then for $t = \lceil \frac{\log n}{\log (d/4A)} \rceil$, all but possibly at most $2An/d$ vertices $v \in V(M) \cup V(D)$ satisfy*

$$|\hat{N}^{4t}_{D,M}(v)| \geq (1 - 2\varepsilon)n.$$

Proof. Suppose the lemma is false. Then without loss of generality, there are at least An/d vertices $v \in X$ such that $|\hat{N}^{4t}_{D,M}(v)| < (1-2\varepsilon)n$. Let. Let $S_0 \subseteq X$ be this set of vertices.

Choose $S_1 \subseteq S_0$ to be of size exactly $\lceil An/d \rceil$. Then by Lemma 3.14 it follows that $|\hat{N}^{4}_{D,M}(S_1)| \geq (1-2\varepsilon)n$ (we assign $C(s) = V(s) = \emptyset$ for all $s \in S_1$). Now we can iteratively apply Lemma 3.15 and obtain sets $S_1 \supseteq S_2 \supseteq \cdots \supseteq S_t$ with $|\hat{N}^{4i}_{D,M}(S_i)| \geq (1-2\varepsilon)n$ and $|S_i| = \lceil 2A|S_{i-1}|/d \rceil$ such that $|S_t| = 1$. Indeed, note that $|S_1| \leq 4An/d$ and for all $i \geq 2$, $|S_i| \leq 4A|S_{i-1}|/d$. Moreover, the iterative steps can be applied because $i \leq t \leq 2\log n/\log d \leq A^{-2}d/4$. Therefore for $t = \lceil \frac{\log n}{\log (d/4A)} \rceil$ we must have $|S_t| \leq 1$, but since S_i is always non-empty we have $|S_t| = 1$. Thus there is a vertex $s \in S_t$ with $|\hat{N}^{4t}_{M,D}(s)| \geq (1-2\varepsilon)n$, contradicting the definition of S_0. □

We now prove the main lemma of this section.

Proof of Lemma 3.12. Set $t = \lceil \frac{\log n}{\log d/4A} \rceil$. By applying Lemma 3.16 first with (D_1, M) and then with (D_2, M) we obtain a set B of order at most $4An/d$ such that all vertices outside B have $|\hat{N}^{4t}_{D_i,M}(v)| \geq (1-2\varepsilon)n$ for $i = 1, 2$. Now let $u \in X, v \in Y$ be two vertices outside B, then $|\hat{N}^{4t}_{D_1,M}(u)| \geq (1-2\varepsilon)n$ and $|\hat{N}^{4t}_{D_2,M}(v)| \geq (1-2\varepsilon)n$. Notice that $\hat{N}^{4t}_{D_1,M}(u) \subseteq X \cap V(M)$ and $\hat{N}^{4t}_{D_2,M}(v) \subseteq Y \cap V(M)$ (since these sets are defined by *even* length alternating paths from u and v). For any $x \in N^{4t}_{D_1,M}(u)$, by definition, there is a rainbow $D_1 - M$-alternating path going from u to x of length $4t$ whose last edge is in M, call this path P_{ux}. Similarly, for any $y \in N^{4t}_{D_1,M}(v)$, there is a rainbow $D_2 - M$-alternating path going from v to y of length $4t$ whose last edge is in M, call this P_{vy}. We claim that there is a pair $x \in X, y \in Y$ such that $xy \in M$ is the last edge of P_{ux} and P_{vy}. Indeed, otherwise we will have $(2-4\varepsilon)n \leq |\hat{N}^{4t}_{D_1,M}(u)| + |\hat{N}^{4t}_{D_2,M}(v)| \leq |M| \leq |X|, |Y| < (2-4\varepsilon)n$, a contradiction. So $P_{ux} \cup P_{vy}$ is a rainbow walk which must contain a rainbow $(D_1 \cup D_2)$-M-alternating path from u to v of length at most $8t$. (In fact, since u, v lie in the opposite sides of the bipartition, this path must be of odd length). □

4. Large matchings in coloured typical graphs

In this section, we combine previous ones to show that typical graphs have large rainbow matchings. We prove the following technical theorem which will imply all our other theorems.

Theorem 4.1. *Let $n^{-1} \ll k^{-1} \ll p \leq 1$, $n^{-1} \ll \varepsilon < 1$ and fix $d = \frac{k \log n}{\log \log n}$. Suppose that we have graphs $G \subseteq H$ with the following properties:*

- *H is properly edge-coloured, bipartite with bipartition (X, Y) such that $|X| = |Y| = n$, and every vertex $v \in V(H)$ has $|N_H(v) \cap V(G)| \geq 0.3pn$.*
- *G is coloured (ε, p, n)-typical with at least $n + 6d$ colours.*

Then H has a rainbow perfect matching.

The full power of the above theorem will only be used to prove our results about generalized Latin arrays. For our results about Latin squares and Steiner systems, a weakening of this result stated as Corollary 4.6 will be sufficient.

The above theorem is proved using the approach described in the Introduction (see Section 1.1, (S1)–(S4)). Note that there are two graphs in the assumption of

Theorem 4.1. Inside the typical graph G, we find a randomized rainbow matching M of size $n - n^{1-\gamma}$ with expansion properties with respect to any collection D of d colours of the graph G. By these expansion properties we know that almost all $x \in X$ and $y \in Y$ have short rainbow $D - M$-alternating paths between them in G (Lemma 3.12). Below think of D being some subset of unused colours on G, sometimes these colours can come from the graph H. We iteratively increase the size of the matching until we get a perfect rainbow matching in H. At each step we obtain a rainbow matching M^i of size $|M^{i-1}| + 1$, such that the edit distance between each M^i and M is still sufficiently small. This guarantees that the expansion properties that M originally had are still preserved for M^i. Note that since G has at least $n + 6d$ colours we always have at least $\Omega(\log n/\log\log n)$ unused colours outside of M.

For rather technical reasons, to perform switchings, we first randomly split G into three graphs G_1, G_2, G_3 and find randomized large matchings $M_1 \subseteq G_1, M_2 \subseteq G_2, M_3 \subseteq G_3$ in these subgraphs using Lemma 3.2. We set $M = M_1 \cup M_2 \cup M_3$ to get a matching of size $n - n^{1-\gamma}$. The advantage of splitting like this is that it now gives us three disjoint matchings with expansion properties which will be useful for finding *disjoint* alternating paths/cycles for switching purposes. Such alternating paths are found using Lemma 3.12. The way we use alternating paths/cycles for enlarging the matching is illustrated in Figure 1. The idea is to first fix two vertices x_0, y_0 which we want to add to the matching. Imagine x_0 and y_0 had edges x_0y_1' and y_0x_1' going to M_1 of colours still unused on $M = M_1 \cup M_2 \cup M_3$. Assume x_1' is matched to y_1 in M_1 and y_1' is matched to x_1 in M_1. Suppose between x_1 and y_1 we could find an alternating $D - M$ rainbow path P (this is true for *almost all* x_1 and y_1). Then we could switch the M-edges to non-M-edges and vice versa on the path $P \cup \{x_0y_1', x_1y_1', y_0x_1', x_1'y_1\}$. This would increase the size of the matching M by one immediately. However, we cannot guarantee that such x_1', y_1' will be present in G_1, that is, such that $c_2 := c(x_0y_1'), c_3 := c(y_0x_1') \notin C(M)$. But we can always guarantee that a choice of x_1' and y_1' will be present such that c_2 appears on M_2 and c_3 appears on M_3. So then if say x_2y_2 is the c_2-edge in M_2, by the expansion properties of M_2 we know there is a $D - M$-alternating rainbow path between x_2 and y_2 which together with the edge x_2y_2 induces an alternating cycle along which if we "switch", that is, we make all the non-edges of M_2 edges of M_2 and vice versa, M_2 remains to be rainbow. We apply the same argument to kick out the colour c_3 from G_3. Now colours c_2 and c_3 become available to use in the matching, and thus we can do the aforementioned switching along the edges of the path $P \cup \{x_0y_1'\} \cup \{y_0x_1'\}$, thus increasing the matching M by size one. Finally note that at each step $|M^i \triangle M^{i+1}| = O(\log n/\log\log n)$, since we switch along at most three paths/cycles of $O(\log n/\log\log n)$ length. Because of this after at most $O(n^{1-\gamma})$ steps, $|M^i \triangle M| \le O(n^{1-\gamma}\log n/\log\log n) \ll |M|$; thus M^i will still have the expansion properties, therefore we can iterate this approach by having M^i instead of M.

Proof of Theorem 4.1. We can assume that $\varepsilon \ll 1$ since any (ε, p, n)-typical graph is also (ε', p, n)-typical for all $\varepsilon' < \varepsilon$. Choose auxiliary constants $q = k^{-1/9}$, γ satisfying $n^{-1} \ll \gamma \ll \varepsilon$.

We call colours of G *large*. Notice that by coloured (ε, p, n)-typicality of G, large colours have $(1 \pm n^{-\varepsilon})n$ edges and so there are less than $n^{1+\varepsilon}$ large colours. Denote $C_0 = C(H) \setminus C(G)$.

Partition $V(G)$ and $C(G)$ into three sets V_1, V_2, V_3 and C_1, C_2, C_3 with each vertex/colour ending up in each V_i/C_i independently with probability 1/3. For $i, j = 1, 2, 3$ we let $G_{i,j}$ to be the subgraph induced by the vertex set V_i and by the colour set C_j. We also denote by $G_i = G_{i,i}$.

Claim 4.2. With positive probability

(i) For all $i, j \in \{1, 2, 3\}$, $G_{i,j}$ is $(\varepsilon/8, p/3, n/3)$-typical.
(ii) For all $i, j \in \{1, 2, 3\}$, every vertex $v \in V(H)$ satisfies $|N_{C_i \cup C_0}(v) \cap V_j| \geq pn/40$.

Proof. By Lemma 2.6 (a) applied with $q_{2.6} = 1/3$, with probability at least $1 - e^{-n^{1-\varepsilon/2}}$, G_i is $(\varepsilon/8, p/3, n/3)$-typical for each $i = 1, 2, 3$. Thus (i) holds with probability at least $1 - 9e^{-n^{1-\varepsilon/2}}$.

Property (ii) follows from Chernoff's bound. Fix a vertex $v \in V(H)$. Then every $y \in N_H(v) \cap V(G)$ is in $N_{C_i \cup C_0}(v) \cap V_j$ independently with probability $\geq 1/9$. Indeed, if $c(vy) \notin C(G)$ this happens when $y \in V_j$ which has probability 1/3. When $c(vy) \in C(G)$, then this happens when both $y \in V_j$ and $c(vy) \in C_i$ which has probability 1/9. So we get $\mathbb{E}[|N_{C_i \cup C_0}(v) \cap V_j|] \geq pn/30$. So by Chernoff's bound, the probability that $|N_{C_i \cup C_0}(v) \cap V_j|] < pn/40$ is less than $e^{-pn/960}$. Thus, with probability at least $1 - 9n^{-1}e^{-pn/960}$, (ii) holds. □

Claim 4.3. For each $i = 1, 2, 3$, there is a rainbow matching M_i in G_i such that the following hold:

(a) $|M_i| \geq n/3 - n^{1-\gamma}$.
(b) For every set of d large colours D, define D_i to be the subgraph induced by edges on V_i which have colours from D. Then the pair (D_i, M_i) is a $(d, q^{-4}, q, n/3)$-expander.

Proof. Fix $i = 1, 2, 3$. Let D be a set of d large colours. By the pigeonhole principle there exists some $j \in \{1, 2, 3\}$ such that $|D \cap C_j| \geq d/3$. Since $G_{i,j}$ is coloured $(\varepsilon/8, p/3, n/3)$-typical it follows from Lemma 2.8 that $G_{i,j}[D \cap C_j]$ has at most $32(p/3)^{-2}(n/3)/(d/3) = 96(p/3)^{-2}(n/3)/d$ vertices of degree $\leq (p/3)(d/3)/2 = pd/18$, which implies that so does D_i (note that D_i potentially has more colours but that can only increase the degrees of vertices). Therefore by Lemma 3.2 (applied to D and G_i) with probability at least $1 - 2e^{-n^{1-\varepsilon/8}}$ (i) and (ii) hold with respect to D_i. Thus, by the union bound, the probability that $|M_i| < n/3 - n^{1-\gamma}$ or there exists some D_i which is not $(d, q^{-4}, q, n/3)$-expander is at most $\binom{|C(G)|}{d} 2e^{-n^{1-\varepsilon/8}} \leq (n + n^{1-\varepsilon})^d \cdot 2e^{-n^{1-\varepsilon/8}} \ll 1$. □

Let $M = M_1 \cup M_2 \cup M_3$. We claim that M can be "extended" to a perfect rainbow matching M' of H using colours of $C_0 \cup (C(G) \setminus C(M))$ such that $|M' \triangle M| \leq 98n^{1-\gamma}\frac{\log n}{\log d}$. Indeed, pick r largest such that $|M'| = |M| + r$, M' is rainbow and $|M' \triangle M| \leq 49r\frac{\log n}{\log d}$. If M' is not a perfect matching, then there exist vertices $x_0 \in X, y_0 \in Y$ outside of M'. From Claim 4.3 (a), we have $|M| \geq 3(n/3 - n^{1-\gamma})$ which gives $r \leq 3n^{1-\gamma}$.

First of all, note that M must be missing at least $6d$ many large colours (since there are at least $n + 6d$ large colours in total). For $j = 1, 2, \ldots, 6$, let D^j be disjoint collections of such large colours each of size d. Note that since these colours

are large, Claim 4.3 tells us that for every $i = 1, 2, 3$ and $j = 1, 2, \ldots, 6$ the pair (D_i^j, M_i) is a $(d, q^{-4}, q, n/3)$-expander.

Denote $M_i' := M_i \cap M'$, for all $i = 1, 2, 3$. Note that $|M_i' \triangle M_i| \leq |M' \triangle M| \leq 147n^{1-\gamma}\frac{\log n}{\log d} \leq qn/30d^2$. It follows that by Lemma 3.13, for every $i = 1, 2, 3$ and $j = 1, 2, \ldots, 6$ the pair (D_i^j, M_i') is a $(d, q^{-4}, 2q, n/3)$-expander.

Setting $A = q^{-4}, \varepsilon' = 2q, \ell = 8\left\lceil \frac{\log(n/3)}{\log(d/4A)} \right\rceil \leq \frac{16\log n}{\log d}$ notice that we have $4d^{-1/2} \leq A^{-1} \leq \varepsilon'/100 \ll 1$ and $d\log d \geq 8A^2\log(n/3)$. So by Lemma 3.12, for each $i = 1, 2, 3$ there is a subset $J_i \subseteq V(G_i)$ of size at most $4n/q^4d$ such that for all $x, y \in V_i \setminus J_i$ lying in different parts of the bipartition of G_i, there is a rainbow $(D_i^{2i-1} \cup D_i^{2i})$-$(M_i')$-alternating path of length at most ℓ from x to y in $V(G_i)$. To finish the proof we need the following two simple claims, whose statements are illustrated by Figure 1.

Claim 4.4. There is an edge $x_1y_1' \in M_1'$ such that $x_1, y_1' \notin J_1$ and $x_0y_1' \in E(H)$ such that either $c(x_0y_1') \notin C(M')$ or there exists an edge $x_2y_2 \in M_2'$ such that $c(x_0y_1') = c(x_2y_2)$ with $x_2, y_2 \notin J_2$.

Proof. Recall that $|N_{C_2\cup C_0}(x_0) \cap V_1| \geq pn/40$ and so we have one of the following two options:

(i) $|N_{C_0}(x_0) \cap V_1| \geq pn/80$,
(ii) $|N_{C_2}(x_0) \cap V_1| \geq pn/80$.

Case 1. Let $F(x_0)$ be the set of vertices $y_1' \in V_1$ satisfying one of the "forbidden" properties below.

(F1) $y_1' \in N_{C_0\cup C_2}(x_0) \cap V_1 \setminus V(M_1')$. The number of these is at most $|V_1 \setminus V(M_1')| \leq |V_1 \setminus V(M_1)| + 2|M_1 \triangle M_1'| \leq \frac{n^{1-\varepsilon/8}}{3} + n^{1-\gamma} + 294n^{1-\gamma}\frac{\log n}{\log d} \ll pn/80$.

(F2) $y_1' \in J_1$ or the vertex that y_1 is matched to in M_1' is in J_1. The number of these is at most $|J_1| \leq 4n/q^4d \ll pn/80$.

(F3) $y_1' \in V_1$ such that $c(x_0y_1) \in C_0 \cap C(M')$. Notice that this is possible as when we extended M to M' we potentially used some of the colours in C_0. However, the number of these is at most $|M' \setminus M| \leq 147n^{1-\gamma}\frac{\log n}{\log d} \ll pn/80$.

Since $|F(x_0)| \ll |N_{C_0}(x_0) \cap V_1|$ we can pick a vertex $y_1 \in N_{C_0}(x_0) \cap V_1$ not satisfying (F1)–(F3). Let x_1 be the vertex that is matched to y_1' in M_1. It is easy to check that the following hold: $c(x_0y_1') \in C_0 \setminus C(M')$, $x_1, y_1' \notin J_1$.

Case 2. In this case $F(x_0)$ will include the vertices y_1' satisfying (F1) or (F2) and additionally the properties below.

(F4) $y_1' \in N_{C_2}(x_0) \cap V_1$ such that $c(x_0y_1') \in C_2 \setminus C(M_2')$. The number of these is at most $|C_2 \setminus C(M_2')| \leq |C_2 \setminus C(M_2)| + |M_2 \triangle M_2'| \leq \frac{n^{1-\varepsilon/8}}{3} + n^{1-\gamma} + 147n^{1-\gamma}\frac{\log n}{\log d} \ll pn/80$.

(F5) $y_1' \in N_{C_2}(x_0) \cap V_1$ such that $c(x_0y_1') \in C(M_2')$ but if we look at the edge x_2y_2 of M_2' which has colour $c(x_0y_1')$ either $x_2 \in J_2$ or $y_2 \in J_2$. The number of these is at most $|J_2| \leq 4n/q^4d \ll pn/80$.

Since $|F(x_0)| \ll |N_{C_2}(x_0) \cap V_1|$ we can pick a vertex $y_1' \in N_{C_2}(x_0) \cap V_1$ not satisfying (F1), (F2), (F4), (F5). Let x_1 be the vertex that is matched to y_1' in M_1 and let x_2y_2 be the edge in M_2' of colour $c(x_0y_1)$ (by the choice of y_1' such an

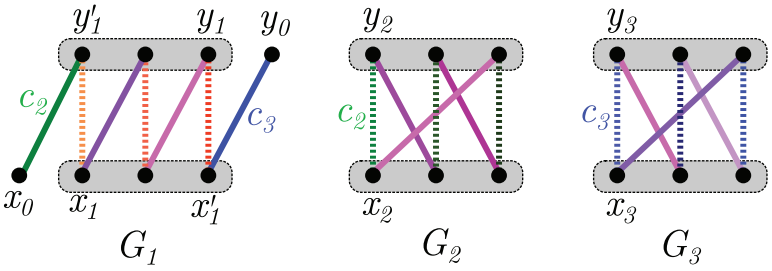

FIGURE 1. The alternating path $P_1 \cup \{x_1y_1', x_1'y_1, x_0y_1', y_0x_1'\}$ and alternating cycles $P_2 \cup \{x_2y_2\}$ and $P_3 \cup \{x_3y_3\}$. The dashed edges denote edges of M' which are removed from the matching. The solid edges denote edges of colours in $\cup_{j=1}^{6} D^j \cup C_0$ which are added to the matching.

edge exists). Additionally, it is easy to check that the following hold: $x_1, y_1' \notin J_1$, $x_2, y_2 \notin J_2$. □

Claim 4.5. There is an edge $x_1'y_1 \in M_1'$ such that $x_1' \neq x_1, y_1 \neq y_1'$ and $x_1', y_1 \notin J_1$ and $y_0x_1' \in E(H)$ such that $c(y_0x_1') \neq c(x_0y_1')$ and either $c(y_0x_1') \notin C(M')$ or there exists an edge $x_3y_3 \in M_3'$ such that $c(y_0x_1') = c(x_3y_3)$ with $x_3, y_3 \notin J_2$.

Proof. The proof is identical to the proof of Claim 4.4, with extra conditions $x_1' \neq x_1$,$y_1' \neq y_1$, $c(y_0x_1') \neq c(x_0y_1')$ which affect the calculations on $F(y_0)$ only by negligible amount. So we omit the proof. □

We may assume both Claim 4.4 and Claim 4.5 hold with the second outcome, as otherwise the proof is even simpler (in that case P_2 or P_3 can be taken as empty, in the proof below). Let $x_1, y_1, x_1', y_1', x_2, y_2, x_3, y_3$ be as in the claims. For each $i = 1, 2, 3$ there is a rainbow $(D_{2i-1} \cup D_{2i})$-M_i'-alternating path P_i of length at most ℓ from x_i to y_i in V_i. Note that the paths P_1, P_2, P_3 don't share any vertices or colours. Finally let M'' be obtained from M' by switching the matching edges along alternating cycles $P_2 \cup \{x_2y_2\}$ and $P_3 \cup \{x_3y_3\}$, and by switching along the alternating path $P_1 \cup \{x_0y_1', x_1y_1', x_1'y_0, x_1'y_1\}$ (see Figure 1). It is not hard to see that this is a rainbow matching with $|M''| = |M'| + 1 = |M| + r + 1$ and such that

$$|M'' \triangle M| \leq |M' \triangle M| + 3\ell + 6 \leq 49r\frac{\log n}{\log d} + 48\frac{\log n}{\log d} + 6 \leq 49(r+1)\frac{\log n}{\log d}.$$

This contradicts the maximality of M'; therefore M' must have been a perfect rainbow matching of H. □

As a corollary of the above theorem, coloured typical graphs have rainbow matchings covering all but at most $O(\log n / \log\log n)$ vertices.

Corollary 4.6. *Let $n^{-1} \ll k^{-1} \ll p \leq 1$, $n^{-1} \ll \varepsilon < 1$ and fix $d = \frac{k \log n}{\log\log n}$. Let G be coloured (ε, p, n)-typical bipartite graph with parts of size $\geq n$ and at least n colours. Then G has a rainbow matching of size $n - 6d$.*

Proof. Let X, Y be the parts of the bipartition of G. By the assumptions, we have $n \leq |X|, |Y| \leq n + n^{1-\varepsilon}$. Fix $n' = n - 6d$, and delete vertices from each part of G to get a balanced bipartite graph G' with parts of size n'.

Notice that the number of vertices deleted from each part of G is between $6d$ and $n^{1-\varepsilon} + 6d \ll n'^{1-\varepsilon/2}$. We claim that G' is $(\varepsilon/2, p, n')$-typical. Indeed, for every vertex $v \in V(G')$, $pn(1-n^{-\varepsilon}) - n^{1-\varepsilon} - 6d \leq d_{G'}(v) \leq pn(1+n^{-\varepsilon})$, thus $d_{G'}(v) = pn'(1 \pm n'^{-\varepsilon/2})$. Similarly, it can be shown that for any $u, v \in V(G')$, $d_{G'}(u,v) = p^2 n'(1 \pm n'^{-\varepsilon/2})$, and that for every colours c, c' we have $e_{G'}(c), e_{G'}(c') = pn'(1 \pm n'^{-\varepsilon/2})$ and $|V_{G'}(c) \cap V_{G'}(c') \cap X|, |V_{G'}(c) \cap V_{G'}(c') \cap Y| = p^2 n'(1 \pm n'^{-\varepsilon/2})$. This implies that G' is $(\varepsilon/2, p, n')$-typical and has at least $n = n' + 6d$ colours.

Thus the assumptions of Theorem 4.1 are satisfied with H and G being the same graph, G'. It follows that G' has a perfect rainbow matching, which induces a rainbow matching of size $n' = n - 6d$ in G, as required. □

We are now ready to reduce the proof of one of our main results from Corollary 4.6.

Proof of Theorems 1.2 *and* 1.6. As mentioned in Section 1.1, Theorems 1.2 and 1.6 are equivalent. So we will just prove Theorem 1.6.

We may assume $n \geq n_0$ for some implicit n_0 sufficiently large, as otherwise the theorem is vacuously true for $k = \frac{n_0 \log\log n_0}{\log n_0}$. So we choose k such that $1 \ll k \ll n_0$. Let $K_{n,n}$ be properly n-edge-coloured. Since every colour forms a perfect matching, we have that $K_{n,n}$ is coloured $(1,1,n)$-typical with $\varepsilon = 1 \gg 1/n_0$. Thus, we can apply Corollary 4.6 to this $K_{n,n}$ and obtain a rainbow matching of size $n - k\log n/\log\log n$, as desired. □

5. Transversals in generalized latin squares

In this section we prove Theorem 1.3. We will use the following result of Pokrovskiy, Montgomery and Sudakov [22].

Theorem 5.1. *There exists $\alpha > 0$ such that for all $n^{-\alpha}/\alpha < \varepsilon < 1$ the following holds. If $K_{n,n}$ is properly edge-coloured graph with at most $(1-\varepsilon)n$ colours having more than $(1-\varepsilon)n$ edges then $K_{n,n}$ has $(1-\varepsilon)n$ edge disjoint perfect rainbow matchings.*

We'll also use Lemma 5.2 giving small rainbow matchings in coloured bipartite graphs.

Lemma 5.2. *Let G be a properly edge-coloured balanced bipartite graph with $\delta(G) \geq d$, parts of size n, with every colour appearing at most $n/12$ times and such that $n \geq 3d + 12$. Then G has a rainbow matching of size at least $3d/2$.*

Proof. Let the parts of G be A, B with $|A| = |B| = n$. Let M_1 be a maximum rainbow matching in G. For contradiction, assume $|M_1| < 3d/2$. Let $A_1 := V(M_1) \cap A$, $B_1 := V(M_1) \cap B$, let $A_0 := A \setminus A_1$ and $B_0 := B \setminus B_1$.

Let B_1' be the set of vertices in B_1 which have at least two edges of unused colours going to A_0. Similarly define A_1'. Note that if there is any edge $ab \in M_1$ such that $a \in A_1'$ and $b \in B_1'$ then we can get a larger matching by replacing ab by different coloured edges from a to B_0 and b to A_0, thus contradicting the maximality of M_1. It follows that $|M_1| \geq |A_1'| + |B_1'|$. By the minimum degree condition we have $e_G(A_0, B) \geq d|A_0|$. Note also that all edges of unused colours must be adjacent to A_1 or B_1. Let $\tilde{e}_G(A_0, B_1)$ be the number of edges going from A_0 to B_1 using only unused colours. Using that every colour occurs $\leq n/12$ times and $|M_1| = |B_1|$, we have

$$\tilde{e}_G(A_0, B_1) \geq e_G(A_0, B) - |M_1|n/12 \geq d|A_0| - |B_1|n/12.$$

On the other hand, from the definition of B_1',

$$\tilde{e}_G(A_0, B_1) \leq |A_0||B_1'| + |B_1|.$$

Thus, we get that $|B_1'| \geq d - \frac{|B_1|}{|A_0|}(1 + n/12) \geq 3d/4$ (using $|B_1| \leq 3d/2$, and $|A_0| \geq n - 3d/2$ and $n \geq 3d + 12$). Similarly, $|A_1'| \geq 3d/4$. Therefore we get that $|M_1| \geq |A_1'| + |B_1'| \geq 3d/2$. □

Before proving Theorem 1.7 we explain main ideas. The basic idea is to employ Theorem 4.1. Given a proper edge-colouring of $K_{n,n}$ with roughly $n \log n / \log\log n$ many colours, by Theorem 5.1, we may assume that at least $n - o(n)$ colours appear $n - o(n)$ times. Call these colours *large*. The rest of the colours will be so-called *small* colours. Note that small colours might even appear only once in the entire graph $K_{n,n}$. However, since there are many of them, that is, roughly of order $n \log n / \log\log n$, we can greedily select a rainbow matching M_0 of size at least $O(\frac{\log n}{\log\log n}) + t$ containing only small colours (this is done in Claim 5.3). Now look at the graph obtained from $K_{n,n}$ by keeping only the edges of large colours and deleting the vertices of M_0 (note that in particular we exclude all the colours appearing on M_0). This graph might have some *bad* vertices of low degree, since they were adjacent to many edges of small colours in $K_{n,n}$. However, using the property that large colours appear $n - o(n)$ times and there are $n - o(n)$ many of them one can prove that there are only few such bad vertices. So we can delete them as well and call the remaining graph G. We show that G is coloured typical. Also note that G contains $n - t = n - |M_0| + O(\frac{\log n}{\log\log n})$ large colours. Finally let H to be the original $K_{n,n}$ minus the vertices and colours of M_0 removed. After checking that the graphs H and G satisfy all the conditions of Theorem 4.1 we obtain a rainbow matching M in H of size $|V(H)| = n - |M_0|$. Then $M_0 \cup M$ is a rainbow matching of size exactly n in $K_{n,n}$.

Proof of Theorem 1.7. We may assume $n \geq n_0$ for some sufficiently large n_0. As otherwise the theorem is true vacuously for $k_0 = \frac{n_0 \log\log n_0}{\log_{n_0}}$. Choose k such that $1 \ll k \ll n_0$. Let α be derived from Theorem 5.1. Fix $d = \frac{k \log n}{72 \log\log n}$ and $\varepsilon_0 := n^{-\alpha/2}$.

Let $K_{n,n}$ be properly coloured with at least $72nd$ colours. By Theorem 5.1, we may assume that more than $(1 - \varepsilon_0)n$ colours have more than $(1 - \varepsilon_0)n$ edges. We call such colours *large*. If a colour has less than $(1 - \varepsilon_0)n$ edges we call it *small*.

Choose t, so that the number of large colours is $n - t$. If the number of large colours is $> n$, then we instead fix $t = 0$. This way $0 \leq t \leq n^{1-\alpha/2}$ always holds.

Claim 5.3. There exists a rainbow matching of small colour edges of size $t + 6d$.

Proof. If a small colour appears less than $n/12$ times we call it *tiny*, otherwise *medium*. Let m be the number of medium colours.

If $m \geq t + 6d$, then, using that $n/12 \geq 3(t + 6d)$, we can greedily pick one edge per medium colour and obtain a rainbow matching of size $t+6d$. So we may assume the number of medium colours is less than $t + 6d$. As the total number of edges is n^2, the number of large colours must be less than $(1 + 2\varepsilon_0)n$. Thus there are at least $kn \log n / \log\log n - (1 + 2\varepsilon_0)n - t - 6d \geq kn \log n / 2 \log\log n$ tiny colours. Furthermore, we may assume $m < t - 12d$. Indeed, we can greedily pick a rainbow matching M_{tiny} of tiny colours of size $18d$. We can do this because each edge in M_{tiny} forbids $2n$ edges which intersect it, so it forbids at most $2n$ tiny colours, but

the number of available tiny colours is at least $kn \log n/2 \log\log n = 36dn$. Thus, if $m \geq t - 12d$, then we can find a rainbow matching M_{medium} of medium colours of size $t - 12d$ greedily in $K_{n,n} \setminus V(M_{tiny})$, since each medium colour appears at least $n/12 \geq 3(t + 6d)$ times. Then $M_{medium} \cup M_{tiny}$ is a matching of small colours of size $t + 6d$.

So we may assume $m < t - 12d$. Let G be the subgraph of $K_{n,n}$ induced by tiny colours. Note that, by definition of t, every vertex is incident to at least t edges of small colours. Since there are m medium colours, we obtain that $\delta(G) \geq t - m \geq 12d$. We can apply Lemma 5.2 to G and obtain a rainbow matching M_{tiny} of tiny colours of size at least $3(t - m)/2 > t - m + 6d$, since $m < t - 12d$. Again in $K_{n,n} \setminus V(M_{tiny})$ we can greedily pick a rainbow matching M_{medium} of size m of medium colours, since each one of these colours appears at least $n/12 \geq 3(t + 6d)$ times. Taking $M_{medium} \cup M_{tiny}$ finishes the proof. □

Let M_0 be a rainbow matching of size $t + 6d$ from the above claim. Let V_{small} be the set of vertices that are incident to more than $2\sqrt{\varepsilon_0}n$ small-coloured edges. Note that $|V_{small}| \leq 2\sqrt{\varepsilon_0}n$ (otherwise, we would get more that $2\varepsilon_0 n^2$ small colour edges in the graph, contradicting "more than $(1 - \varepsilon_0)n$ colours have more than $(1 - \varepsilon_0)n$ edges"). Let G be obtained from $K_{n,n}$ by removing all edges of small colours and all vertices from $V(M_0) \cup V_{small}$. Let H be $K_{n,n}$ with all colours and vertices of M_0 removed. It is easy to see that $G \subset H$. Next we check that we can apply Theorem 4.1 to G and H.

Notice that H is balanced bipartite with parts of size $n' = n - t - 6d$. Notice that the parts in G have size $\geq n' - |V_{small}| \geq n' - 2\sqrt{\varepsilon_0}n = n' - 2n^{1-\alpha/4} \geq n'(1 - n'^{-\alpha/5})$ and also that the number of colours in G is $\geq (1 - \varepsilon_0)n \geq n'(1 - n'^{-\alpha/5})$. Every vertex $v \in V(H)$ satisfies $|N_H(v) \cap V(G)| \geq n - 2|M_0| - |V_{small}| \geq n - 2\sqrt{\varepsilon_0}n - 2(t + 6d) \geq 0.3n'$.

Next we show that G is coloured $(\alpha/5, 1, n')$-typical. Using the fact that G consists of edges of only large colours and that vertices in G are adjacent to at most $2\sqrt{\varepsilon_0}n$ small coloured edges in the original graph $K_{n,n}$, we get that the following is true for any $u, v \in V(G)$ on the same side of G and colours $c, c' \in C(G)$.

$$\begin{aligned}
d_G(v) &\geq n - 2\sqrt{\varepsilon_0}n - |V_{small}| - 2|M_0| \geq (1 - n'^{-\alpha/5})n' \\
d_G(u, v) &\geq n - 4\sqrt{\varepsilon_0}n - |V_{small}| - 2|M_0| \geq (1 - n'^{-\alpha/5})n' \\
|E_G(c)| &\geq (1 - \varepsilon_0)n - |V_{small}| - 2|M_0| \geq (1 - n'^{-\alpha/5})n' \\
|V_G(c) \cap V_G(c') \cap X| &\geq (1 - 2\varepsilon_0)n - |V_{small}| - |M_0)| \geq (1 - n'^{-\alpha/5})n' \\
|V_G(c) \cap V_G(c') \cap Y| &\geq (1 - 2\varepsilon_0)n - |V_{small}| - |M_0| \geq (1 - n'^{-\alpha/5})n'
\end{aligned}$$

Note that this is enough to conclude that all three graphs G, $G_{X,C}$ and $G_{Y,C}$ are (uncoloured) $(\alpha/5, 1, n')$-typical. Finally G contains $n - t = n' + 6d$ large colours. So G and H satisfy all the assumptions of Theorem 4.1, therefore we obtain a perfect rainbow matching M in H (whose colours, by definition, are disjoint from M_0). Finally, $M \cup M_0$ is a perfect rainbow matching in $K_{n,n}$. □

6. Large matchings in Steiner systems

In this section we improve the bound on Brouwer's conjecture about matchings in Steiner triple systems.

Proof of Theorem 1.5. We may assume $n \geq n_0$ for n_0 sufficiently large. Choose ${n_0}^{-1} \ll k^{-1} \ll 1$ and fix $d = \frac{k \log n}{6 \log \log n}$. We assume $V(S) = \{x_1, x_2, \ldots, x_n\}$. In a Steiner triple system we have $n \equiv 1$ or $3 \pmod 6$. We'll first prove the result when $n \equiv 3 \pmod 6$. In the other case, the same proof works if we apply it to a subgraph of S formed by deleting a vertex. Let G be an auxiliary bipartite simple graph with bipartition $X = Y = V(S)$, colour set $V(S)$, and edge ab having colour c whenever $\{a, b, c\} \in E(S)$. Using that S is a Steiner triple system, notice that G is properly n-edge-coloured $K_{n,n}$ minus a perfect rainbow matching. It has codegrees $n - 2$ and hence it is in particular coloured $(1 - o(1), 1, n)$-typical.

We randomly construct a bipartite graph H as follows. Partition $[n]$ into three disjoint sets I_X, I_Y, I_C by putting independently every i in one of the sets I_X, I_Y, I_C with probability $1/3$. Let $A = \{x_i : i \in I_X\}$, $B = \{x_i : i \in I_Y\}$, $C = \{x_i : i \in I_C\}$. Let H be the subgraph of G consisting of edges from A to B having a colour in C. We claim that the following simultaneously hold with positive probability:

(P1) H is properly coloured $(1/8, 1/3, n/3)$-typical.
(P2) $|A| = |B| = |C| = n/3$.

We show that (P1) holds with high probability and (P2) holds with positive probability; thus the claim will follow.

(P1) holds with probability at least $1 - o(n^{-3})$.

We can apply Lemma 2.6 to H as it is easy to check that it satisfies the assumptions in (b) with $q_{2.6} = 1/3$, $\varepsilon_{2.6} = 1 - o(1), p_{2.6} = 1$. Indeed, since S is 3-uniform, we have that for every edge $e \in G$ going through $x_i \in X, x_j \in Y, x_\ell \in C(G)$ the indices i, j, ℓ are distinct. Thus Property (i) holds with probability at least $1 - e^{-n^{1-\varepsilon/2}} \geq 1 - o(n^{-3})$.

(P2) holds with probability at least n^{-3}.

Notice that out of all the possible outcomes of the random variables $|A|, |B|, |C|$, the outcome $|A| = |B| = |C| = n/3$ is the most likely one which happens with probability at least $1/n^3$. (The outcome $|A| = a, |B| = b, |C| = c$ has probability $\frac{1}{3^n}\binom{n}{a,b,c}$. For $n \equiv 0 \pmod 3$ the multinomial coefficient $\binom{n}{a,b,c}$ is maximized when $a = b = c$).

Now we are ready to apply Corollary 4.6 to H. We obtain a rainbow matching M in H of size at least $n/3 - 6d = n/3 - k \log n / \log \log n$. Now it easy to see that the triples $M_S = \{(a, b, c(ab)) | ab \in M\}$ induce a hypergraph matching in S of the same size. Indeed, the fact that $(a, b, c(ab))$ is an edge of S for all $ab \in M$ follows by definition of G. To see that M_S is a matching notice that for distinct edges $a_1b_1, a_2b_2 \in M$, all four endpoints a_1, a_2, b_1, b_2 correspond to four distinct vertices in S because (A, B, C) induce a partition of $V(S)$ and M is a matching. Finally $c(a_1b_1) \neq c(a_2b_2)$ since M is rainbow and $c(a_1b_1), c(a_2b_2)$ are distinct from a_1, a_2, b_1, b_2 since (A, B, C) induce a partition of $V(S)$. □

7. Concluding remarks

A far reaching generalisation of the Ryser-Brualdi-Stein conjecture was proposed in 1975 by Stein [27]. He defined an equi-n-square as an $n \times n$ array filled with n symbols such that every symbol appears *exactly n times*. Notice that Latin squares are equi-n-squares, but there are many equi-n-squares which are not Latin. Stein [27] conjectured that all equi-n-squares contain a transversal of size $n - 1$. If true, this would imply that Latin squares have size $n - 1$ transversals.

Recently, the second and third author [23] disproved Stein's conjecture by constructing equi-n-squares without transversals of size $n - \log n/42$. On the other hand, our Theorem 1.2 gives transversals in Latin squares of size at least $n - O(\log n/\log\log n)$. Thus, combining these two results, we obtain a full separation between Latin and equi-n-squares.

Despite being false, Stein's conjecture remains one of the outstanding problems in the area. In particular it would be very interesting to determine whether it is true asymptotically i.e. is it true that every equi-n-square has a transversal of size $n - o(n)$. Here, the best currently known result is due to Aharoni, Berger, Kotlar, and Ziv [1], who used topological methods, to show that equi-n-squares always have a transversal of size at least $2n/3$.

All our results can be rephrased as results about finding large matchings in 3-uniform hypergraphs. Here the results are about *linear* 3-uniform hypergraphs i.e. ones where every pair of vertices are contained in at most one edge. As mentioned in the introduction, $n \times n$ Latin squares correspond to linear n-regular, 3-partite, 3-uniform hypergraphs with n vertices in each part. Using the technique of Rödl's nibble, there have been general results proved about finding large matchings in linear regular (and nearly-regular) hypergraphs. In particular Alon, Kim, and Spencer [3] showed that linear 3-uniform, pn-regular hypergraphs of order n have matchings of size $n - O(n^{1/2}\log^{3/2} n)$. Our results show that if additionally a certain graph associated with the hypergraph is pseudorandom, then the matching can cover all but $O(\log / \log\log n)$ vertices. Specifically, for a 3-uniform hypergraph $\mathcal{H}$, define its shadow $\partial\mathcal{H}$ to be the graph formed by replacing every edge of $\mathcal{H}$ by a triangle. Then Corollary 4.6 is equivalent to the following.

Theorem 7.1. *Let $n^{-1} \ll k^{-1} \ll p \leq 1$, $n^{-1} \ll \varepsilon < 1$. Let $\mathcal{H}$ be a 3-uniform, tripartite linear hypergraph with partition (V_1, V_2, V_3), $|V_1| = |V_2| = |V_3| = n$. Suppose that for all $i \neq j$, the induced subgraph $\partial\mathcal{H}[V_i \cup V_j]$ of the shadow between V_i and V_j is (ε, p, n)-typical. Then $\mathcal{H}$ has a matching of size $n - \frac{k\log n}{\log\log n}$.*

If we only assume (ε, p, n)-regularity rather than (ε, p, n)-typicality then the hypergraph above is nearly pn-regular and is only known to have a matching of size $n - n^{1-\gamma}$ (e.g. from Lemma 2.5). With the added typicality condition we get a much larger matching. For non-tripartite hypergraphs we can prove the following analogue.

Theorem 7.2. *Let $n^{-1} \ll k^{-1} \ll p \leq 1$, $n^{-1} \ll \varepsilon < 1$. Let $\mathcal{H}$ be a 3-uniform linear hypergraph on n vertices. Suppose that for vertex v we have $|N_{\partial\mathcal{H}}(v)| = (1\pm n^{-\varepsilon})pn$ and for every pair of vertices u, v, $|N_{\partial\mathcal{H}}(v)| = (1\pm n^{-\varepsilon})pn$ and $|N_{\partial\mathcal{H}}(u)\cap N_{\partial\mathcal{H}}(v)| = (1 \pm n^{-\varepsilon})p^2 n$. Then $\mathcal{H}$ has a matching of size $n - \frac{k\log n}{\log\log n}$.*

This theorem is proved identically to Theorem 1.5. Indeed, the only change that needs to be made is to observe that the graph G constructed in that proof will be (ε, p, n)-typical (rather than $(1 - o(1), 1, n)$-typical as in Theorem 1.5). Due to applications to Latin squares and Steiner triple systems, it is worthwhile to study further the hypergraphs appearing in Theorems 7.1 and 7.2. In particular it would be interesting to determine if they always have matchings of size $n - O(1)$ or not.

Acknowledgments

We'd like to thank a referee for a careful reading of the paper and for many suggestions that improved the presentation. Part of this research was done when the third author visited London School of Economics. He wants to thank LSE for its hospitality and for creating a stimulating research environment.

References

[1] Ron Aharoni, Eli Berger, Dani Kotlar, and Ran Ziv, *On a conjecture of Stein*, Abh. Math. Semin. Univ. Hambg. **87** (2017), no. 2, 203–211, DOI 10.1007/s12188-016-0160-3. MR3696146

[2] Saieed Akbari and Alireza Alipour, *Transversals and multicolored matchings*, J. Combin. Des. **12** (2004), no. 5, 325–332, DOI 10.1002/jcd.20014. MR2079255

[3] Noga Alon, Jeong-Han Kim, and Joel Spencer, *Nearly perfect matchings in regular simple hypergraphs*, Israel J. Math. **100** (1997), 171–187, DOI 10.1007/BF02773639. MR1469109

[4] Noga Alon, Michael Krivelevich, and Benny Sudakov, *List coloring of random and pseudo-random graphs*, Combinatorica **19** (1999), no. 4, 453–472, DOI 10.1007/s004939970001. MR1773652

[5] Kazuoki Azuma, *Weighted sums of certain dependent random variables*, Tohoku Math. J. (2) **19** (1967), 357–367, DOI 10.2748/tmj/1178243286. MR221571

[6] János Barát and Zoltán Lóránt Nagy, *Transversals in generalized Latin squares*, Ars Math. Contemp. **16** (2019), no. 1, 39–47, DOI 10.26493/1855-3974.1316.2d2. MR3904714

[7] Darcy Best, Kevin Hendrey, Ian M. Wanless, Tim E. Wilson, and David R. Wood, *Transversals in Latin arrays with many distinct symbols*, J. Combin. Des. **26** (2018), no. 2, 84–96, DOI 10.1002/jcd.21566. MR3745158

[8] A. E. Brouwer, *On the size of a maximum transversal in a Steiner triple system*, Canadian J. Math. **33** (1981), no. 5, 1202–1204, DOI 10.4153/CJM-1981-090-7. MR638375

[9] A. E. Brouwer, A. J. de Vries, and R. M. A. Wieringa, *A lower bound for the length of partial transversals in a Latin square*, Nieuw Arch. Wisk. (3) **26** (1978), no. 2, 330–332. MR480083

[10] Richard A. Brualdi and Herbert J. Ryser, *Combinatorial matrix theory*, Encyclopedia of Mathematics and Its Applications, vol. 39, Cambridge University Press, Cambridge, 1991, DOI 10.1017/CBO9781107325708. MR1130611

[11] David A. Drake, *Maximal sets of Latin squares and partial transversals*, J. Statist. Plann. Inference **1** (1977), no. 2, 143–149, DOI 10.1016/0378-3758(77)90019-2. MR485434

[12] S. Eberhard, F. Manners, and R. Mrazović, *An asymptotic for the Hall–Paige conjecture*, Preprint `arXiv:2003.01798`, 2020.

[13] L. Euler, *Recherches sur un nouvelle espéce de quarrés magiques*, Verhandelingen uitgegeven door het zeeuwsch Genootschap der Wetenschappen te Vlissingen, pages 85–239, 1782.

[14] P. Frankl and V. Rödl, *Near perfect coverings in graphs and hypergraphs*, European J. Combin. **6** (1985), no. 4, 317–326, DOI 10.1016/S0195-6698(85)80045-7. MR829351

[15] Marshall Hall and L. J. Paige, *Complete mappings of finite groups*, Pacific J. Math. **5** (1955), 541–549. MR79589

[16] Pooya Hatami and Peter W. Shor, *A lower bound for the length of a partial transversal in a Latin square*, J. Combin. Theory Ser. A **115** (2008), no. 7, 1103–1113, DOI 10.1016/j.jcta.2008.01.002. MR2450332

[17] Peter Keevash and Liana Yepremyan, *On the number of symbols that forces a transversal*, Combin. Probab. Comput. **29** (2020), no. 2, 234–240, DOI 10.1017/s0963548319000282. MR4079635

[18] Jeong Han Kim, Benny Sudakov, and Van H. Vu, *On the asymmetry of random regular graphs and random graphs*, Random Structures Algorithms **21** (2002), no. 3-4, 216–224, DOI 10.1002/rsa.10054. Random structures and algorithms (Poznan, 2001). MR1945368

[19] Klaas K. Koksma, *A lower bound for the order of a partial transversal in a Latin square*, J. Combinatorial Theory **7** (1969), 94–95. MR239988

[20] C. C. Lindner and K. T. Phelps, *A note on partial parallel classes in Steiner systems*, Discrete Math. **24** (1978), no. 1, 109–112, DOI 10.1016/0012-365X(78)90179-6. MR522740

[21] Michael Molloy and Bruce Reed, *Graph colouring and the probabilistic method*, Algorithms and Combinatorics, vol. 23, Springer-Verlag, Berlin, 2002, DOI 10.1007/978-3-642-04016-0. MR1869439

[22] Richard Montgomery, Alexey Pokrovskiy, and Benjamin Sudakov, *Decompositions into spanning rainbow structures*, Proc. Lond. Math. Soc. (3) **119** (2019), no. 4, 899–959, DOI 10.1112/plms.12245. MR3964824

[23] Alexey Pokrovskiy and Benny Sudakov, *A counterexample to Stein's equi-n-square conjecture*, Proc. Amer. Math. Soc. **147** (2019), no. 6, 2281–2287, DOI 10.1090/proc/14220. MR3951411

[24] Vojtěch Rödl, *On a packing and covering problem*, European J. Combin. **6** (1985), no. 1, 69–78, DOI 10.1016/S0195-6698(85)80023-8. MR793489

[25] H. J. Ryser, *Neuere probleme der kombinatorik*, Vorträge über Kombinatorik, Oberwolfach, **69** (1967), 91.

[26] P. W. Shor, *A lower bound for the length of a partial transversal in a Latin square*, J. Combin. Theory Ser. A **33** (1982), no. 1, 1–8, DOI 10.1016/0097-3165(82)90074-7. MR665651

[27] S. K. Stein, *Transversals of Latin squares and their generalizations*, Pacific J. Math. **59** (1975), no. 2, 567–575. MR387083

[28] Shinmin Patrick Wang, *On self-orthogonal Latin squares and partial transversals of Latin squares*, ProQuest LLC, Ann Arbor, MI, 1978. Thesis (Ph.D.)–The Ohio State University. MR2627641

[29] Stewart Wilcox, *Reduction of the Hall-Paige conjecture to sporadic simple groups*, J. Algebra **321** (2009), no. 5, 1407–1428, DOI 10.1016/j.jalgebra.2008.11.033. MR2494397

[30] David E. Woolbright, *An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols*, J. Combinatorial Theory Ser. A **24** (1978), no. 2, 235–237, DOI 10.1016/0097-3165(78)90009-2. MR472562

Mathematical Institute, University of Oxford, Oxford, United Kingdom
Email address: keevash@maths.ox.ac.uk

Department of Mathematics, University College London, United Kingdom
Email address: dr.alexey.pokrovskiy@gmail.com

Department of Mathematics, ETH, 8092 Zürich, Switzerland
Email address: benny.sudakov@gmail.com

Department of Mathematics, Emory University, Atlanta
Email address: liana.yepremyan@emory.edu