Thesis submitted in fulfillment of the degree of

Doctor of Philosophy in Mathematics

UNIVERSITY COLLEGE LONDON

Department of Mathematics

Large-scale structures in groups

Alp Müyesser

Supervised by Dr Alexey Pokrovskiy

This thesis is dedicated to the memory of my dear friend, Emily Zhu (1997-2023).

Declaration

I, Alp Müyesser, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Alp Müyesser

Abstract

This thesis is broadly concerned with finding perfect matchings in hypergraphs whose vertices represent group elements and edges represent solutions to systems of linear equations. For example, given a subset of a group, when is it possible to partition the subset into triples whose products are the identity element?

A well-known problem in this direction is the Hall-Paige conjecture from 1955 which asks for a characterisation of all groups whose multiplication table (viewed as a Latin square) contains a transversal. More precisely, a complete mapping of a group G is a bijection $\phi\colon G \to G$ such that $x \mapsto x\phi(x)$ is also bijective. Hall and Paige conjectured in 1955 that a finite group G has a complete mapping whenever $\prod_{x\in G} x$ is the identity in the abelianization of G. This was confirmed in 2009 by Wilcox, Evans, and Bray with a proof using the classification of finite simple groups.

Many problems in the area have a similar flavour, yet until recently they have been approached in completely different ways, using mostly algebraic tools ranging from the combinatorial Nullstellensatz to Fourier analysis.

The first result in this thesis, based on joint work with Alexey Pokrovskiy, gives a unified approach to attack these problems, using tools from probabilistic combinatorics. We show that for random-like and equal-sized subsets A, B, C of a group G, there exists a bijection $\phi\colon A \to B$ such that $x \mapsto x\phi(x)$ is a bijection from A to C whenever $\prod_{a\in A} a \prod_{b\in B} b = \prod_{c\in C} c$ in the abelianization of G. We show that this randomised version of the Hall-Paige conjecture can be used as a black-box to settle many old problems in the area for sufficiently large groups, including a conjecture of Snevily from 1999. As a by-product, we obtain the first combinatorial proof of the Hall-Paige conjecture.

The second result in this thesis, which is the author's own work, refines these tools further to solve a problem concerning the existence of complete mappings with a prescribed cycle type, confirming a conjecture of Friedlander, Gordon, and Tannenbaum from 1981.

Impact statement

The main contribution of this work is to academia. The particular area of research of the author is extremal and probabilistic combinatorics. Specifically, the author studies large-scale substructures in networks (or in mathematical terms, graphs/hypergraphs). A network consists of a collection of nodes and connections between certain pairs of nodes. Given an arbitrary network, can we determine if it is possible to travel through every single node whilst not visiting any node more than once? Such a travel itinerary is called a Hamilton path, the prototypical large-scale substructure a network could have. Computational complexity theory shows that it is virtually impossible to design an efficient algorithm to test whether a given network contains a Hamilton path. However, in practice, networks of interest come equipped with many additional properties that make it easier to find Hamilton paths. For example, (1) networks can be dense, meaning each node has many connections, or (2) networks can be symmetric, meaning the connections encode certain relations in some algebraic structure, or (3) networks can be pseudo- random, meaning the statistics of the connection patterns resemble that of a network generated randomly. The research of the author develops tools to find Hamilton paths in networks with additional properties such as those listed. This is a rich and vibrant field which has recently witnessed breakthroughs for several decades old conjectures. Moreover, concepts developed to tackle problems of this sort, such as the Regularity Lemma, expanders, the semi-random method, and absorption, have proved to be flexible enough to facilitate progress in several adjacent fields. Examples include the study of distribution of primes (number theory) and property testing (theoretical computer science).

The thesis itself is concerned with the questions of type (2) as indicated above. In particular, the thesis finds a bridge between Probabilistic Combinatorics and Combinatorial Group Theory by showing that the toolkit from the former field can be used to solve old problems in the latter field. This novel connection has allowed for the resolution of several old problems in the area, and confirmed conjectures made by several prominent mathematicians, including Snevily.

It seems likely that the new mathematical techniques developed in this thesis may have applications beyond those considered in the current manuscript. There remain several major open problems in the area for which our methods could provide a feasible angle of attack, including the famous Ryser's conjecture which asserts that all odd order Latin squares contain a full transversal. The work in the current thesis could be interpreted as asserting that this conjecture is true in a strong form whenever the underlying Latin square has sufficient algebraic structure.

Acknowledgments

I would like to begin by thanking Alexey Pokrovskiy, who has simply been a phenomenal PhD supervisor. He has invested an enormous chunk of his time to my growth as a mathematician, and he simultaneously gave me a lot of space and encouragement to be an independent researcher. I will feel forever indebted for all his support.

Next, I would like to thank Richard Montgomery, who, despite our lack of formal ties, has given me more inspiration and guidance than can be expected from even a PhD supervisor.

I have been lucky to cross paths with many brilliant mathematicians and friends, more than I can reasonably name here, who have all helped me find the determination I needed to write this thesis. However, this thesis is dedicated to one such brilliant mathematician and friend, Emily Zhu, who tragically passed away at a young age. Emily was one of the kindest and smartest people I knew. Thank you, Emily, for giving me the confidence to follow this through. To quote Uncle Iroh, "while it is always best to believe in oneself, a little help from others can be a great blessing."

Finally, I would like to thank my family for allowing me to choose my own path, and Émilie Guermeur for being by my side the whole way through.

UCL Research Paper Declaration Form

Referencing the doctoral candidate's own published work(s)

1. For a research manuscript that has already been published (if not yet published please skip to item 2):

   (a) What is the title of the manuscript?

   (b) Please include a link to or doi for the work:

   (c) Where was the work published?

   (d) Who published the work?

   (e) When was the work published?

   (f) List the manuscript's authors in the order they appear on the publication:

   (g) Was the work peer reviewed?

   (h) Have you retained the copyright?

   (i) Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)?

   If 'Yes' please give a link or doi:

If ‘No’ please seek permission from the relevant publisher and check the box next to the below statement:

□ I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.

2. For a research manuscript prepared for publication but that has not yet been published (if already published please skip to item 3):

   (a) What is the current title of the manuscript?

   A random Hall-Paige conjecture

   (b) Has the manuscript been uploaded to a preprint server ’e.g. medRxiv’?

   X

   If ‘Yes’ please please give a link or doi:

   https://doi.org/10.48550/arXiv.2204.09666

   (c) Where is the work intended to be published?

   (d) List the manuscript’s authors in the intended authorship order:

   Alp Müyesser, Alexey Pokrovskiy

   (e) Stage of publication:

   In review

3. For multi-authored work please give a statement of contribution covering all authors (if single-author please skip to item 4):

   Both authors contributed equally to the conception and writing of the project.

4. In which chapter(s) of your thesis can this material be found? Chapter 2.5.2

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/senior author unless this is not appropriate e.g. if the paper was a single-author work):

Candidate:

Alp Müyesser

Date:

7 May 2024

Supervisor/Senior Author signature (where appropriate):

Date:

UCL Research Paper Declaration Form

Referencing the doctoral candidate's own published work(s)

1. For a research manuscript that has already been published (if not yet published please skip to item 2):

   (a) What is the title of the manuscript?

   (b) Please include a link to or doi for the work:

   (c) Where was the work published?

   (d) Who published the work?

   (e) When was the work published?

   (f) List the manuscript's authors in the order they appear on the publication:

   (g) Was the work peer reviewed?

   (h) Have you retained the copyright?

   (i) Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)?

   If 'Yes' please give a link or doi:

If ‘No’ please seek permission from the relevant publisher and check the box next to the below statement:

□ I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.

2. For a research manuscript prepared for publication but that has not yet been published (if already published please skip to item 3):

   (a) What is the current title of the manuscript?

   Cycle type in Hall-Paige: A proof of the Friedlander-Gordon-Tannenbaum conjecture

   (b) Has the manuscript been uploaded to a preprint server ’e.g. medRxiv’?

   X

   If ‘Yes’ please please give a link or doi:

   https://doi.org/10.48550/arXiv.2303.16157

   (c) Where is the work intended to be published?

   (d) List the manuscript’s authors in the intended authorship order:

   Alp Müyesser

   (e) Stage of publication:

   In review

3. For multi-authored work please give a statement of contribution covering all authors (if single-author please skip to item 4):

   Both authors contributed equally to the conception and writing of the project.

4. In which chapter(s) of your thesis can this material be found? Chapter 3

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/senior author unless this is not appropriate e.g. if the paper was a single-author work):

Candidate:

Alp Müyesser

Date:

7 May 2024

Supervisor/Senior Author signature (where appropriate):

Date:

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

The results in this thesis provide ways of constructing mathematical objects that are in perfect equilibrium. Here, we think of an object as being in equilibrium if it does not contain any sub-components which are, in some sense, overloaded. We will momentarily give a concrete example, but first we remark that such mathematical objects find use in various interdisciplinary settings. For example, they can be used to design secure communication protocols as well as experiments that avoid unwanted correlations. The branch of mathematics concerned with the construction of these objects in equilibrium is often called "design theory". As we will explore further, design theory itself is a highly interdisciplinary branch of mathematics, requiring a expansive toolkit ranging from abstract algebra to probability theory. Moreover, even though the origins of design theory can be traced back to Euler's work in the 18th century, there are many fascinating unsolved conjectures in the area. This thesis addresses several such outstanding conjectures in a unified manner, resolving problems that had remained unsolved for decades.

As a concrete example, let us discuss the following puzzle, popularised by Euler in 1782. Suppose we have 5 different chess pieces, say a king, queen, rook, bishop, and knight, and we have 5 versions of each piece in different colours, say white, grey, red, blue, and green. Can we arrange the 25 pieces we have in a 5 by 5 grid so that no colour, nor piece, repeats in any of the rows or columns? See Figure 1.2 for a solution.

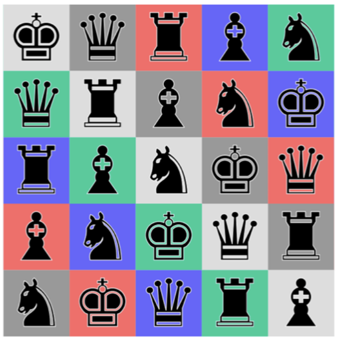


Figure 1.1: A solution to Euler's puzzle for a 5 by 5 grid. In mathematical terms, the object depicted in Figure 1.2 shows the existence of a mutually orthogonal Latin square of order 5.

This solution to Euler's puzzle could, for example, represent the optimal way to design an experiment to test the efficacy of five different medicines, each with five versions, and five different time slots to administer the medicine. The solution described implies that we can find a scheduling for this experiment in a "balanced" way, using only five test subjects. We refer the reader to [39] for more information about connections between solutions to Euler's puzzle and design of experiments.

Due to such practical applications, it is natural to wonder if there are solutions to Euler's puzzle when the number five is replaced with any larger number. Astonishingly, the analogous problem when five is replaced with six is much harder. This problem, known as Euler's 36 officers problem, stood open for more than a century before Tarry demonstrated in 1901 that in fact there are no solutions. How about seven, or larger? This is where the field of group theory comes in. A *group* is a mathematical object which describes all the possible symmetries a certain object has. For example, considering all possible configurations of a Rubik's cube, and all the possible ways to get from one configuration to another via performing rotations, naturally corresponds to a group.

Some groups, but not all, can be used to produce solutions to Euler's puzzle for large dimensions. As we will explain more precisely in the next section, here, the groups of interest are those whose multiplication tables contain transversals (see Section 1.2 for definitions). However, the structure of groups can be very intricate, and thus it is hard to determine which multiplication tables contain transversals. Despite this, in 1955, Hall and Paige conjectured that the multiplication table of a group contains a transversal if and only if the product of all group elements (in some arbitrary order) belongs in the commutator subgroup of that group. This in particular implies that given the multiplication table of a group, we can efficiently (in polynomial time) determine if there exists a transversal or not.

The Hall-Paige conjecture stood open for more than half a century. It was finally confirmed in 2009 with a long and difficult proof by Wilcox [60], Evans [22], and Bray [13]. In particular, this proof relies on the infamous Classification of Finite Simple Groups (whose proof spans thousands of pages) as well as extensive computer assistance.

A central contribution of this dissertation is a completely different and self-contained proof of a far-reaching generalisation of the Hall-Paige conjecture, obtained jointly with Pokrovskiy. Not only is this solution much shorter, but the generalised framework allows us to resolve several old conjectures adjacent to the Hall-Paige conjecture, including a conjecture of Gordon from 1961, Ringel from 1974, and Snevily from 1999. Although many of these conjectures have a similar flavour, up until now they were approached in completely different ways, using tools ranging from the Nullstellensatz to Fourier analysis. Our approach, combining tools from group theory, probability, and combinatorics, gives a unified method to address these conjectures simultaneously.

To illustrate just one application of our framework, let us consider the following 1961 conjecture of Gordon. This conjecture, which we will shortly state, is motivated by the fact that for many practical applications, it is desirable to find solutions to Euler's puzzle with an additional degree of symmetry. Going back to the example of the medical experiment, let us suppose that the outcome of the trial could potentially be influenced by which medicines are administered back to back. Then, in the solution to the puzzle described earlier, it becomes undesirable that king and queen appear consecutively both in the first row and the third row (as well as the fourth and the fifth rows), as for our medical experiment this means that some medicines are administered back to back more often than others, therefore potentially causing unwanted correlations.

Motivated by this, Gordon asked if there are solutions to Euler's puzzle where any two pieces appear consecutively in a row exactly once in the whole grid. Such solutions are called row-complete. Gordon himself discovered that the theory of groups is once again very useful in this setting. Therefore, he asked to characterise all groups that can be used to produce row-complete solutions to Euler's puzzle. In the literature, such groups are called sequenceable groups. Formally, a group G is sequenceable if and only if there exists a permutation $a_1, a_2, \ldots, a_n$ of its elements such that the partial products $b_i = a_1 \cdot a_2 \cdot \cdots \cdot a_i$ are all distinct. It turns out that the previous techniques used to address the Hall-Paige conjecture are not sufficient to characterise those groups which are sequenceable. On the other hand, the generalisation of the Hall-Paige conjecture we obtained with Pokrovskiy can be used to give a characterisation, thereby resolving the problem of Gordon [1].

As it turns out, the characterisation implies that the group corresponding to the Rubik's cube is sequenceable. We now explain what it means for the Rubik's cube to be sequenceable. A "rotation" in a Rubik's cube is just an instruction to perform a single rotation on a Rubik's cube. For example, "turn left face clockwise 90 degrees" is such an instruction. A "move" is just a finite sequence of rotations that can be followed one after the other, for example "turn top face counter-clockwise 90 degrees then turn back face clockwise 180 degrees" is a valid move. Formally, valid moves correspond to elements of the group associated with the Rubik's cube. There are as many moves as there are distinct configurations of a Rubik's cube, of which there are about 43 quintillion many. To check that the group associated to the Rubik's cube is sequenceable, we need to specify an order (or permutation) in which to apply

[1] A formal proof of this result is omitted from the thesis for space considerations; however, a proof can be found in [49].

all 43 quintillion moves to an initially solved Rubik's cube, so that after every move, the configuration of the Rubik's cube is one that we never saw after any of the previous moves. My dissertation shows that there is an algorithm that finds such a permutation of the moves, thereby certifying that the Rubik's cube is sequenceable. This means that the Rubik's cube is among the class of groups that can produce row-complete solutions to Euler's puzzle.

Unlike previous approaches in the area, our algorithm actually uses probability theory to produce a large fraction of the permutation. This means that up until we perform 42 quintillion moves or so, our algorithm gives a very simple instruction: "among the set of moves we can currently do which brings the Rubik's cube to a configuration we haven't already seen before, pick one randomly, and keep going". This probabilistic algorithm is simple to state, but rather difficult to analyse. For example, what if we just unexpectedly get stuck after performing 21 quintillion moves, meaning what if all the remaining moves we are yet to do bring the Rubik's cube back to a state we have already seen before? Our analysis shows that up until we do 42 quintillion moves, this is very unlikely to happen, which implies that there must be several ways to permute the first 42 quintillion moves so that we never get stuck.

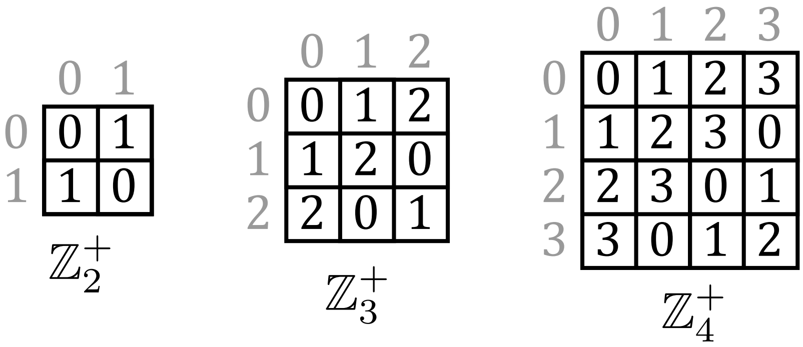
Although the probabilistic analysis does a majority of the work, completing the permutation is a very delicate task, and for this final stage our algorithm is significantly more complicated as we use algebraic tools. It is worth mentioning that the interplay between the probabilistic nature of the algorithm in the first stage and the algebraic nature in the final stage is critical. Prior to our approach, researchers had tried to resolve Gordon's conjecture using exclusively algebraic tools with little success. On the other hand, a purely probabilistic algorithm that does not take into account any of the algebraic symmetry is also destined to fail. The combination of the algebra and probability is therefore crucial to our algorithm. As described by Sir Timothy Gowers, problems of this sort, where there is too much structure for a purely randomised approach to work, and too little structure for a purely algebraic approach to work, are central in combinatorics. The celebrated resolution of the "existence of designs" conjecture by Prof Peter Keevash, for example, also falls into this category.

In the next section, we give a more mathematical introduction to the subject, and state the main result of the thesis.

1.2 Transversals in Latin squares

This thesis is about finding large-scale structures in finite groups using techniques from probabilistic combinatorics. The prototypical example of a “large-scale structure” in a group is a complete mapping, or equivalently, a transversal in the Latin square corresponding to the multiplication table of the group. We now define each of these terms. A *complete mapping* of a group G is a bijection $\phi\colon G \to G$ such that the function $x \mapsto x\phi(x)$ is also a bijection. A *Latin array* is an $n \times n$ array filled in with arbitrary symbols such that each symbol appears at most once in each row and column. A *Latin square* is an $n \times n$ Latin array with exactly n symbols. A *transversal* of a Latin array is a collection of n cells which do not share a row, a column, or a symbol. Denote by $M(G)$ the *multiplication table* of the group G whose rows and columns are labelled by the elements of the group, and the entry in row g_i and column g_j is the group element $g_i \cdot g_j$. Observe that $M(G)$ is a Latin square, and that $M(G)$ has a transversal if and only if G has a complete mapping.

The study of transversals in Latin squares began more than two hundred years ago, when Euler posed a problem equivalent to determining for which n there exists a $n \times n$ Latin square whose entries can be partitioned into transversals (Euler's 36 officers problem discussed in the earlier section asserts that there exists no 6×6 Latin square whose entries can be partitioned into transversals, see also Figure 1.1 for a 5×5 Latin square partitioned into transversals). This motivates the study of transversals in multiplication tables because if the multiplication table of a group has a transversal, then it can be partitioned into transversals. To see this, we can translate the columns of a transversal by a non-identity group element



	0	1
0	0	1
1	1	0

$\mathbb{Z}_2^+$

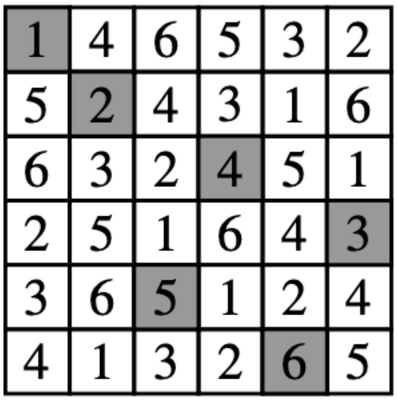
	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\mathbb{Z}_3^+$

	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\mathbb{Z}_4^+$

Figure 1.2: Multiplication tables of some small cyclic groups. For $\mathbb{Z}_3^+$, the identity function corresponds to a complete mapping (as well as a transversal of the multiplication table). It can also be observed that $\mathbb{Z}_2^+$ and $\mathbb{Z}_4^+$ does not contain a transversal/complete mapping.



1	4	6	5	3	2
5	**2**	4	3	1	6
6	3	2	**4**	5	1
2	5	1	6	4	**3**
3	6	**5**	1	2	4
4	1	3	2	**6**	5

Figure 1.3: A 6×6 Latin square with a transversal highlighted.

to produce another transversal, entirely disjoint from the first (we invite the reader to verify this for the displayed multiplication table of $\mathbb{Z}_3^+$ in Figure 1.2). However, some multiplication tables do not contain any transversals at all (such as $M(\mathbb{Z}_2^+)$ and $M(\mathbb{Z}_4^+)$), let alone partitions into transversals. Indeed, if we suppose that the multiplication table of a group has a transversal, equivalently, we know that the group has a complete mapping ϕ. Denote by π the permutation $x \mapsto x\phi(x)$, so we have that $\pi(x) = x\phi(x)$ for every $x \in G$. Let us assume for the moment that G is an abelian group, for simplicity, hence we will switch to additive notation (we use this convention for the rest of the thesis). Adding the n equations we have of the form $\pi(x) = x + \phi(x)$ (for each $x \in G$), we obtain the following.

$$\sum_{x \in G} \pi(x) = \sum_{x \in G} x + \phi(x) \tag{1.1}$$

As we assumed that the group is abelian, (1.1) rearranges into $\prod_{x \in G} x = e$ (where e is the identity element of G), giving a necessary condition for having a complete mapping in abelian groups. An immediate consequence is that even-order cyclic groups do not admit complete mappings (for example in $G = \mathbb{Z}_{2n}$ we have $\sum_{x \in G} x = n \neq e$).

For general (non-abelian) groups G, we can apply the same argument to derive that the product of all group elements (in any order) of G must be the identity when projected (using the standard quotient homomorphism) to any abelian subgroup of G. To get the most information possible out of this relation, we project these equations to the largest possible abelian subgroup of G, namely, the abelianization of G which we denote by G^{ab} (recall that $G^{ab} := G/G'$ where G' is the commutator subgroup of G, i.e. the subgroup generated by elements of the form $aba^{-1}b^{-1}$, $a, b \in G$).

Hence, we derive that if G is to have a complete mapping, $\prod_{x \in G} x \in G'$, where G' denotes the commutator subgroup of G. Equivalently, $\prod_{x \in G} \pi_{ab}(x) = e$ when projected to G^{ab} (where $\pi_{ab} : G \to G^{ab}$ is the quotient homomorphism). Since G^{ab} is abelian, we sometimes write this as "$\sum_{x \in G} x = 0$ in G^{ab}".

The condition that $\sum_{x \in G} x = 0$ in G^{ab} is known as the Hall-Paige condition [35]. We remark that the Hall-Paige condition is sometimes written as "all 2-Sylow subgroups of G are trivial or non-cyclic". This is equivalent to $\prod_{x \in G} x$ being trivial in G^{ab}, as shown by Hall and Paige themselves [35]. Perhaps astonishingly, the Hall-Paige condition is not only necessary, but also sufficient for the existence of a complete mapping. This was first conjectured by Hall and Paige [35]. The Hall-Paige Conjecture has a rich history, we refer the reader to the book of Evans [23] (Chapters 3-7) for a description of various approaches taken for this problem. The conjecture was finally shown to be true in a combination of papers by Wilcox [60], Evans [22], and Bray [13] in 2009. The original proof of Wilcox, Evans, and Bray uses an inductive argument which relies on the classification of finite simple groups. However, recently, a completely different proof for large groups was found by Eberhard, Manners, and Mrazović using tools from analytic number theory [19].

In this thesis, with the goal of giving a unified approach to many related conjectures in the area, we study complete mappings between subsets of groups. For example, given equal sized subsets A, B, C of a group G, is there a bijection $\phi\colon A \to B$ such that $\psi\colon A \to C$ defined via $\psi(a) := a\phi(a)$ is also a bijection? This corresponds to starting with the multiplication table of a group, and then deleting the rows corresponding to $G \setminus A$, columns corresponding to $G \setminus B$, and symbols corresponding to $G \setminus C$, and then searching for a transversal in the resulting structure, which is simply a Latin array with some missing entries. Generalising the Hall-Paige condition to this set-up, we see that

$$\sum A + \sum B = \sum C \text{ (in } G^{ab}) \tag{1.2}$$

is a necessary condition for the existence of such a map ϕ (we use $\sum S$ to denote $\sum_{s \in S} s$, $\prod S$ is defined analogously). Of course, we cannot expect (1.2) to be a sufficient condition for any triple of equal sized subsets[2]. However, our main theorem essentially states that for most subsets $A, B, C \subseteq G$, (1.2) is the only obstruction for finding the desired map ϕ. Recall that a p-**random subset** X of a finite set G is a subset sampled by including each element of G in X independently with probability p, and Δ denotes symmetric difference. When we say that an event holds with high probability, we mean that the probability of that event approaches 1 as n tends to infinity. The letter n throughout the thesis always denotes the order of the ambient group G.

Theorem 1.2.1 (Main result)**.** *Let G be a group of order n. Let $p \geq n^{-1/10^{100}}$. Let $R^1, R^2, R^3 \subseteq G$ be p-random subsets, sampled independently. Then, with high probability, the following holds.*

[2] For example, consider $G = \mathbb{Z}_{99}$, $A = \{98, 1\}$, $B = \{98, 1\}$, $C = \{49, 50\}$.

Let $X, Y, Z \subseteq G$ be equal sized subsets satisfying the following properties.

- $|X \Delta R^1| + |Y \Delta R^2| + |Z \Delta R^3| \le p^{10^{10}} n / \log(n)^{10^{15}}$
- $\sum X + \sum Y = \sum Z$ *in* G^{ab} *(or equivalently* $\prod X \prod Y (\prod Z)^{-1} \in G'$*)*

Then, there exists a bijection $\phi \colon X \to Y$ such that $x \mapsto x\phi(x)$ is a bijection from X to Z.

We remark that setting $p = 1$ and $X = Y = Z = G$, we see that the Hall-Paige conjecture holds for sufficiently large groups. However, Theorem 1.2.1 extends far beyond the setting of the Hall-Paige conjecture. Indeed, we can settle several longstanding conjectures at the interface of group theory and combinatorics using the full strength of Theorem 1.2.1. In the following section, we discuss several such problems. All of these problems are similar in spirit to the Hall-Paige conjecture in the sense that they concern finding large-scale structures in groups with certain desirable properties. However, all of these problems lack the level of symmetry present in the Hall-Paige problem. For example, some of these problems concern finding complete mappings in subsets of groups with limited structure, or they concern finding complete mappings permuting the group elements via a particular cycle type. Both of these constraints seem difficult to reason about using only algebraic techniques.

This perhaps explains why many of the commonly used tools, such as Alon's combinatorial Nullstellensatz, were only able to go so far in addressing these questions. On the other hand, one clearly needs use some of the group theoretic structure. Indeed, the key obstruction for problems of this type end up being some derivative of the Hall-Paige condition, which is inherently group theoretic. To cite Gowers [31], there are many problems in combinatorics where "there is too much choice for constructions to be easy to discover, and too little choice for simple probabilistic arguments to work" [31]. Our proof, similar in spirit to Keevash's celebrated construction of designs [40], uses probabilistic tools but also exploits the algebraic structure of the problem. We give a detailed overview of our strategy in Section 2.1.

From now on, additive notation will always imply that the corresponding operation is taking place in the abelianization of the ambient group (e.g. when for $A \subseteq G$ we write "$\sum A = e$" we mean that $\prod_{a\in A} \pi_{ab}(a) = e$ where $\pi_{ab} : G \to G^{ab}$ is the quotient map to the abelianization of G). Otherwise, operations within non-abelian groups will be denoted multiplicatively. The quantity n always denotes the size of the ambient group. We use e to denote the identity element of a group, and we sometimes use 0 to denote e in the special case of abelian groups.

1.3 Applications

As we already noted, the first application of our main result, Theorem 1.2.1, is an alternative proof that the Hall-Paige conjecture holds for all sufficiently large groups. This uses only the $p = 1$ case of Theorem 1.2.1, where we set the subsets X, Y, Z to be the entirety of the group G. We now mention another result we can recover easily, this time setting X, Y, Z to be sets of size $|G|-1$. Goddyn and Halasz recently proved that multiplication tables contain near transversals, that is, a collection of $n-1$ cells which do not share a row, column, or a symbol [29]. Perhaps surprisingly, there does not seem to be an easy way of deriving this from the Hall-Paige conjecture directly, and the proof in [29] is somewhat involved. However, we can derive this result from Theorem 1.2.1 as follows.

Proposition 1.3.1. *Let M be the multiplication table of a sufficiently large finite group G. Then, M contains a near transversal.*

Proof. Applying Theorem 1.2.1 with $p = 1$, we derive that for $|G|$ large enough, $R^1 = R^2 = R^3$ satisfies the conclusion of Theorem 1.2.1 with positive probability, and therefore, with probability exactly 1. Let $z \in G$ be some element equal to $\sum G$ in G^{ab}. Setting $X = Y = G \setminus \{e\}$ and $Z = G \setminus \{z\}$, we have that $\sum X + \sum Y = \sum Z$ in G^{ab}. This implies that there is a bijection $\phi\colon X \to Y$ such that $x \to x\phi(x)$ is injective. ϕ then corresponds to a near transversal in M, as desired. □

We remark that the above proposition is also a corollary of Montgomery's [46] recent breakthrough result asserting that the Ryser-Brualdi-Stein conjecture holds for large Latin squares.

The $p = 1$ case of Theorem 1.2.1 when applied with other subsets X, Y, Z has novel applications as well. We discuss such an application in the next section. The other application we discuss uses the full strength of Theorem 1.2.1. In fact, for these applications, we rely on an appropriate generalisation of Theorem 1.2.1 with a more complicated distribution on the sets R^1, R^2, R^3. In Section 2.3, we state this generalised version of Theorem 1.2.1.

1.3.1 Snevily's conjecture

Deleting k rows and k columns of a multiplication table, we obtain a natural Latin array which we call a *subsquare* of the multiplication table. In analogy with complete mappings, it is natural to ask which subsquares contain transversals. One may suspect that deleting rows and columns should only make it easier to find transversals, and Snevily's conjecture states that this indeed should be the case for abelian groups of odd order [55]. However, for even order abelian groups G, one may delete rows and columns so that the remaining Latin array is in fact the multiplication table of an even order subgroup $H \subseteq G$ (or a translate of such a multiplication table). Such subsquares cannot contain transversals as multiplication tables of even order cyclic groups do no admit transversals. In 1999, Snevily conjectured that this should be the only subtlety for cyclic even order groups. Below, we formally state both cases of Snevily's conjecture. Note that $A \times B$ denotes the subsquare of a group G obtained by keeping only the rows corresponding to A and columns corresponding to B in the multiplication table of G.

Conjecture 1.3.2 (Snevily, [55])**.** *Let $S = A \times B$ be a subsquare of the multiplication table of an abelian group G defined by two n-element sets $A, B \subseteq G$.*

1. *If $|G|$ is odd, then S has a transversal.*

2. *If $G \cong \mathbb{Z}_{2k}$ for some $k \in \mathbb{N}$, then S has a transversal unless there exists $g_1, g_2 \in G$ such that $g_1 + A \cong g_2 + B \cong \mathbb{Z}_{2j}$ for some $j \leq k$.*

The first case of the conjecture was verified by Alon in 2000 for prime order cyclic groups [4]. In 2001, Dasgupta, Károlyi, Serra, and Szegedy generalised Alon's result to arbitrary odd order cyclic groups [18]. Both of these results use the celebrated Combinatorial Nullstellensatz [3]. A decade later, Arsovski fully resolved the first case of Snevily's conjecture, using character theory [8]. On the other hand, as far as the authors are aware, no partial progress has been reported on the second case of the conjecture.

As observed by Wanless [59], the second part of Snevily's conjecture does not generalise straightforwardly to all even abelian groups due to the following construction of Akbari and Alireza [2]. Let $G = (\mathbb{Z}_2)^k$ for some $k \geq 1$, and let $a_1, a_2 \in G$ be distinct and let $b_1, b_2 \in G$ be distinct such that $a_1 + a_2 + b_1 + b_2 = 0$. Then, setting $A = G \setminus \{a_1, a_2\}$, $B = G \setminus \{b_1, b_2\}$, it is a simple exercise to check that the subsquare $A \times B$ does not contain a transversal. We show, perhaps surprisingly, that this is the only other barrier for an abelian subsquare to contain a transversal.

Theorem 1.3.3. *There exists an $n_0 \in \mathbb{N}$ such that the following holds for all $n \geq n_0$. Let G be an abelian group, and let $A, B \subseteq G$ with $|A| = |B| = n$. Then, $A \times B$ has a transversal, unless there exists some $k \geq 1$, $g_1, g_2 \in G$ and a subgroup $H \subseteq G$ such that one of the following holds.*

1. *$H \cong \mathbb{Z}_{2k} \times H_{odd}$ for some odd size group H_{odd}, and $H \cong g_1 A \cong g_2 B$*

2. *$H \cong (\mathbb{Z}_2)^k$, $g_1 A \cong H \setminus \{a_1, a_2\}$, $g_2 B \cong H \setminus \{b_1, b_2\}$ for some distinct $a_1, a_2 \in H$ and distinct $b_1, b_2 \in H$ such that $a_1 + a_2 + b_1 + b_2 = 0$.*

Note that this confirms both cases of Snevily's conjecture for sufficiently large subsquares. Note that $(\mathbb{Z}_2)^k$ contains translates of $\mathbb{Z}_2$, meaning that $(\mathbb{Z}_2)^k \times (\mathbb{Z}_2)^k$ contains subsquares of order 2 without transversals. However, this does not come up in the statement of Theorem 1.3.3 as we are only concerned with large subsquares. We discuss proving a stronger characterisation valid for all n in Section 3.6.

We take Snevily's conjecture further by proving a far more general theorem characterising subsquares without transversals of all groups.

Theorem 1.3.4. *There exists an $n_0 \in \mathbb{N}$ such that the following holds for all $n \geq n_0$. Let G be a group, and let $A, B \subseteq G$ with $|A| = |B| = n$. Then, $A \times B$ has a transversal, unless there exists some $k \geq 1$, $g_1, g_2 \in G$ and a subgroup $H \subseteq G$ such that one of the following holds.*

1. *H is a group that does not satisfy the Hall-Paige condition, and $A \cong g_1 H$ and $B \cong H g_2$.*

2. *$H \cong (\mathbb{Z}_2)^k$, $g_1 A \cong H \setminus \{a_1, a_2\}$, $g_2 B \cong H \setminus \{b_1, b_2\}$ for some distinct $a_1, a_2 \in H$ and distinct $b_1, b_2 \in H$ such that $a_1 + a_2 + b_1 + b_2 = 0$.*

Roughly speaking, Theorem 1.3.4 states that deleting rows and columns from a multiplication table only makes it easier to find transversals (supposing we do not end up with a translate of a multiplication table of a subgroup), except for a very specific scenario where we delete 2 rows and 2 columns summing to zero from the multiplication table of an elementary abelian 2-group. Theorem 1.3.4 is proved in Section 2.5.1.

1.3.2 The Friedlander-Gordon-Tannenbaum conjecture

Here, we are concerned with a strengthening of the Hall-Paige conjecture, regarding the existence of orthomorphisms with specific cycle types. Recall that the **cycle type** of a permutation π encodes how many cycles of each length are present when π is written as a product of disjoint cycles. Also recall that an **orthomorphism** of a finite group G is a bijection $\phi\colon G \to G$ such that $g \mapsto g^{-1}\phi(g)$ is also bijective. Orthomorphisms that consist of a single cycle come up naturally in Ringel's resolution of the Heawood map colouring conjecture, which motivated Ringel to ask for a classification of all groups with such orthomorphisms (see [53, 26, 6, 52]).

Several other problems of a similar flavour concerning “sequenceable groups” were raised by numerous authors with the motivation to construct Latin squares with additional properties (see [51] and Section 1.1.2 in [49]). There are also motivations to study orthomorphisms with other cycle types. For example, orthomorphisms that are products of disjoint 6-cycles give constructions such as “cyclic” Steiner triple systems [38].

A unifying conjecture in the area was given by Friedlander, Gordon, and Tannenbaum in 1981 [25].

Conjecture 1.3.5 (The Friedlander-Gordon-Tannenbaum (FGT) conjecture, 1981)**.** *Let G be an abelian group of order n satisfying the Hall-Paige condition. Suppose for some integer $k \geq 2$ that k divides $n-1$. Then, there exists an orthomorphism of G that fixes the identity element, and permutes the remaining elements as products of disjoint cycles of length k.*

The Hall-Paige conjecture is not very laborious to verify for abelian groups, and this was already achieved by Hall and Paige when they posed their conjecture. The FGT conjecture, on the other hand, has remained open for more than forty years.

Remark 1. A group admits a complete mapping if and only if it admits an orthomorphism, essentially because the map $g \to g^{-1}$ is a bijection. Therefore, the Hall-Paige conjecture is sometimes stated with respect to complete mappings instead of orthomorphisms. However, this equivalence does not hold when we make restrictions on the cycle type. For example, in an abelian group, there cannot be a complete mapping inducing any cycle of length 2, therefore the FGT conjecture does not hold when orthomorphisms are replaced with complete mappings (for a more detailed discussion of cycle types of complete mappings, see [10, 11]). However, some appropriate modification of the FGT conjecture likely holds for complete mappings as well, and we discuss this further in Section 3.6. We should also remark that, confusingly, orthomorphisms are called complete mappings in [25], but the convention in the thesis seems to be standard following the book of Evans [23].

There are several partial results towards the FGT conjecture in the literature. Friedlander, Gordon, and Tannenbaum themselves confirmed their conjecture for groups of order at most 15, and abelian p-groups where $p \geq 3$ [25]. We refer the reader to [23] for a more detailed overview (see also [10, 11, 58] for results about the very related concept of complete mappings). We just remark that the $k = 3$ and the cyclic group case of the FGT conjecture is open, signifying the difficulty of the problem. In this thesis, we resolve the FGT conjecture for sufficiently large groups.

Theorem 1.3.6. *The Friedlander-Gordon-Tannenbaum conjecture is true for all sufficiently large groups.*

We use methods from probabilistic combinatorics, so our proof needs large groups just to get concentration for some random variables with fairly simple distributions. We do not make this constant explicit to make the presentation neater. We also remark that even though the random Hall-Paige conjecture is a good starting point to prove Theorem 1.3.6, the proof of Theorem 1.3.6 is much more involved than (say) the proof of Snevily's conjecture discussed in the earlier section. Indeed, the entirety of Chapter 3 is devoted to the proof of Theorem 1.3.6. The additional challenges that arise in comparison to the setting of the Hall-Paige conjecture are discussed in detail in Chapter 3.

We verify the FGT conjecture by developing fairly general methods which can potentially be used to shed light on many adjacent embedding problems with an algebraic flavour. As we explore in Section 3.6.1, our methods seem adaptable for the study of graceful and harmonious graph labellings [28].

We make two further remarks.

Remark 2. At the time the FGT conjecture was posed, the Hall-Paige conjecture was known to be true for abelian groups, but not in general, which perhaps explains why Conjecture 1.3.5 is concerned only with abelian groups. Given the present work, it seems reasonable to suspect that the FGT conjecture can be extended to non-abelian groups, perhaps even quasi-groups/Latin squares, which would generalise the famous Ryser-Brualdi-Stein conjecture. We discuss this further in the concluding remarks, Section 3.6.

Remark 3. Our proof of Theorem 1.3.6 actually gives much more, and can be used to give many other cycle types that can be realised via orthomorphisms. We discuss this further in Section 3.6.

1.3.3 Further applications

Theorem 1.2.1 has several further applications that we have omitted in this thesis due to space considerations. For a more comprehensive overview of all the consequences of Theorem 1.2.1 (and its variants), we refer the reader to [49].

1.4 Organisation

The next chapter, Chapter 3 is devoted to proving the main result of the thesis, which is a stronger version of Theorem 1.2.1. This chapter also contains a proof of Snevily's conjecture (in a strong form). The following chapter, Chapter 3, contains a proof of the Friedlander-Gordon-Tannenbaum conjecture (for sufficiently large groups).

Chapter 2

A random Hall-Paige conjecture and Snevily's conjecture

2.1 Proof strategy for the main result

In this section, we attempt to give an accessible outline of a special case of our main result, concerning cyclic groups. We conclude in Section 2.1.1 by outlining some key difficulties we omit in the simplified discussion.

Firstly, instead of using the language of complete mappings, we will re-frame Theorem 1.2.1 as a hypergraph matching problem. Given a group G, we define the 3-uniform 3-partite multiplication hypergraph H_G as follows. Set $V(H_G) := G_A \sqcup G_B \sqcup G_C$ where G_* is a copy of G and $\sqcup$ indicates a disjoint union. Set $E(G) := \{(g_A, h_B, k_C) \in G_A \times G_B \times G_C \colon g_A h_B k_C = e\}$. Given $X, Y, Z \subseteq G$, we denote by $H_G[X, Y, Z]$ the induced subgraph of H_G given by the vertex subset $(G_A \cap X) \sqcup (G_B \cap Y) \sqcup (G_C \cap Z)$. Given equal sized subsets X, Y, Z, observe that finding a bijection $\phi \colon X \to Y$ such that $x \mapsto x\phi(x)$ is a bijection from X to Z is equivalent to finding a perfect matching in $H_G[X, Y, Z^{-1}]$ where $Z^{-1} = \{z^{-1} \colon z \in Z\}$. We will outline a proof for the following result. For the rest of the outline, fix $\varepsilon = 1/100$.

Proposition 2.1.1. *Let $X, Y, Z \subseteq \mathbb{Z}_n$ such that $|X| = |Y| = |Z| = n - O(n^{1-\varepsilon})$, and $\sum X + \sum Y + \sum Z = 0$. Then, $H_{\mathbb{Z}_n}[X, Y, Z]$ has a perfect matching.*

Proposition 2.1.1 is already novel, and would be sufficient, for example, to deduce Snevily's conjecture for large subsquares. We remark that Proposition 2.1.1 is a simple corollary of Theorem 1.2.1 which can be obtained by setting $p = 1$. We remark that throughout the rest of the paper, when we say that a subset $S \subseteq G$ is **zero-sum**, we mean that the product of all elements of S (in any order) is in G', the commutator subgroup. For abelian groups, this corresponds to $\sum S = 0$.

2.1.1 Absorption

Our main tool is the *absorption method*, a technique codified by Rödl, Ruciński, Szemerédi [54] (see also the earlier work of Erdős, Gyárfás, and Pyber [21]), adapted to the setting of hypergraphs defined by groups. Absorption is a general method that reduces the task of finding spanning structures to finding *almost* spanning structures. In most cases, the latter task is considerably simpler, as evidenced by the celebrated nibble method which roughly states that pseudorandom hypergraphs contain large matchings [5]. The main technical innovation in our paper is developing an absorption strategy for multiplication hypergraphs, which in the setting of Proposition 2.1.1, culminates in the following lemma.

Lemma 2.1.2 (Simpler version of Lemma 2.4.32)**.** *$H_{\mathbb{Z}_n}[X, Y, Z]$ contains a vertex subset $\mathcal{A}$ of size $o(n)$ such that for any $S \subseteq V(H_{\mathbb{Z}_n}[X, Y, Z]) \setminus \mathcal{A}$ of size $O(n^{1-\varepsilon})$ intersecting X, Y and Z in the same number of vertices, and satisfying $\sum S = 0$, $\mathcal{A} \cup S$ has a perfect matching.*

In Lemma 2.1.2, $\mathcal{A}$ functions as our *absorber*, in the sense that it can *absorb* small enough subsets by formingperfect matchings when combined with them. The key premise of the absorption method is that once a suitable absorber is found and set aside, the only remaining task is to find an *almost perfect matching* in the leftover set. Indeed, $\mathcal{A}$ can absorb whatever small subset S we fail to cover with the almost perfect matching.

We now explain in a bit more detail how Lemma 2.1.2 reduces the task of proving Proposition 2.1.1 to finding a matching of size $n - O(n^{1-\varepsilon})$ in $V(H_{\mathbb{Z}_n}[X,Y,Z]) \setminus \mathcal{A}$. First, note that setting $S = \emptyset$ in Lemma 2.1.2 implies that $\mathcal{A}$ has a perfect matching, and thus $\sum \mathcal{A} = 0$ (if a subset contains a perfect matching, by definition the subset can be partitioned into zero-sum sets, and hence is zero-sum itself). Now, suppose that having fixed the set $\mathcal{A}$, we were able to find a matching M_1 in $V(H_{\mathbb{Z}_n}[X,Y,Z]) \setminus \mathcal{A}$ covering all but $O(n^{1-\varepsilon})$ vertices. Let S denote the set of these leftover vertices. As $\mathcal{A}$ and $V(M)$ are disjoint zero-sum sets contained in $V(H_{\mathbb{Z}_n}[X,Y,Z])$, and $\sum X + \sum Y + \sum Z = 0$ by assumption, we have that $\sum S = 0$ as well. So by the property in Lemma 2.1.2, $\mathcal{A} \cup S$ spans another perfect matching, M_2 say. Then, $M_1 \cup M_2$ is the desired perfect matching of $H_{\mathbb{Z}_n}[X,Y,Z]$.

We remark that, in reality, deleting $\mathcal{A}$ from $H_{\mathbb{Z}_n}[X,Y,Z]$ would damage the pseudorandomness properties of the hypergraph too greatly to be able to find the desired M_1 using the Rödl nibble [5]. Therefore, here we actually need a slightly stronger version of Lemma 2.1.2 which can find $\mathcal{A}$ inside small random sets. This way, deleting $\mathcal{A}$ only spoils the pseudorandomness of a set R much smaller than the multiplication hypergraph itself. $R \setminus \mathcal{A}$ can then be dealt with using standard pseudorandomness arguments, see for example Lemma 2.2.8.

For Lemma 2.1.2 to hold, we remark that some condition on the value of $\sum S$ is necessary. Indeed, recall that if a subset admits a perfect matching, it has to be zero-sum. So, if S_1 and S_2 are two sets disjoint with $\mathcal{A}$ such that $\mathcal{A} \cup S_1$ and $\mathcal{A} \cup S_2$ both admit perfect matchings, it follows that

$$0 = \sum \mathcal{A} \cup S_1 = \sum \mathcal{A} \cup S_2$$

hence $\sum S_1 = \sum S_2$. Therefore, the set $\mathcal{A}$ can have the flexibility of combining with any member of a large family of sets $\mathcal{F}$ to produce perfect matchings only if $\sum S$ is fixed for all $S \in \mathcal{F}$. For convenience, we fix this sum to be 0, but Lemma 2.1.2 would remain true if we replaced 0 with any other fixed element.

Building the absorber from small subgraphs

Our starting point for building the absorber set $\mathcal{A}$, similar in spirit to most applications of the absorption method, is the existence of small subgraphs (gadgets) which give *local variability*. More precisely, we will rely on the existence of $O(1)$-sized gadgets, Q say, that can combine with $O(1)$ distinct sets, F_1 and F_2 say, each of size $O(1)$, such that $Q \cup F_1$ and $Q \cup F_2$ both induce perfect matchings in $H_{\mathbb{Z}_n}[X, Y, Z]$. We say that Q can *switch* between F_1 and F_2.

The power of the absorption method rests in the fact that small gadgets such as Q displaying rather limited variability can be combined in a way to build an absorber displaying *global variability*. By global variability, we mean the type of property that $\mathcal{A}$ has in the statement of Lemma 2.1.2. In particular, we are referring to how $\mathcal{A}$ can combine with essentially any subset of size $O(n^{1-\varepsilon})$, as opposed to just a few of size $O(1)$. To achieve this in our case, we will use a variant of the absorption technique called *distributive absorption*, initially developed by Montgomery [45]. The method has since been applied in numerous settings, notably in the proof of Ringel's conjecture by Montgomery, Pokrovskiy, and Sudakov [48]. For a detailed discussion of how gadgets such as Q can be combined to build an absorber, we refer the reader to the discussion in [48].

For readers who are familiar with the absorption method, we add a quick remark that common-place methods of building absorbers, such as those used in [54], do not work in our context for the following simple reason. Say we have a subset S and we are interested in subsets A of size k such that both A and $A \cup S$ span a perfect matching. In the usual absorption strategy, we would require that the number of subsets A with this property is $\Omega(n^k)$. However, if A spans a perfect matching, then $\sum A = 0$. The number of zero-sum subsets A of order k is $O(n^{k-1})$ – too little to appear in abundance when a positive fraction of k-subsets are randomly sampled. On the other hand, with the distributive absorption strategy, one may build absorbers with fewer gadgets, provided that one can show that the gadgets are well-distributed within the host structure.

Absorption for pairs

Typically, in applications of the distributive absorption method, one works with gadgets Q switching between F_1 and F_2 where F_1 and F_2 are both singletons. This would be impossible to implement in our context as if $Q \cup F_1$ and $Q \cup F_2$ both contain perfect matchings, then $\sum F_1 = \sum F_2$. Thus, if F_1 and F_2 were singletons, F_1 and F_2 would consist of the exact same vertex. Hence, if we want Q to be a gadget that actually gives us some flexibility, we can only hope to switch between sets of size at least 2. This motivates us to search for disjoint sets Q, F_1 and F_2 such that $Q \cup F_1$ and $Q \cup F_2$ both span perfect matchings, $|F_1| = |F_2| = 2$, and $\sum F_1 = \sum F_2 = 0$, where the final equality is chosen for convenience as in the statement of Lemma 2.1.2.

Finding Q with this property turns out to be rather easy. For example, fix disjoint sets $F_1 := \{a, -a\}$ and $F_2 := \{b, -b\}$ where neither a nor b is an involution. Let us view F_1 and F_2 as subsets of G_C. Consider some $x \in G_A$. Set $y := -x - a \in G_B$, set $z = x + a - b \in G_A$ and $w = -x + b$. Suppose that $x \neq z$ and $y \neq w$. Observe that $M_1 = \{(x, -x - a, a), (x + a - b, -x + b, -a)\}$ is a matching of $\{x, y, z, w\} \cup F_1$ and $M_2 = \{(x, -x+b, -b), (x+a-b, -x-a, b)\}$ is a matching of $\{x, y, z, w\} \cup F_2$. Hence, $Q_x := \{x, y, z, w\}$ is a gadget with the desirable property of switching between F_1 and F_2. See Figure 2.1 for an illustration. Moreover, it is not hard to see that there are many choices of x for which the corresponding sets Q_x are all disjoint, which is critical for the distributive absorption strategy.

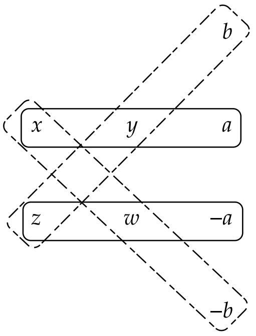


Figure 2.1: The gadget Q_x. Matchings M_1 and M_2 depicted in solid and dashed lines, respectively.

Building on the idea detailed in the previous paragraph, we can show that there exists gadgets like Q_x which can combine with any one of 100 (as opposed to just 2) pairs of inverses to produce a matching. This, combined with the usual distributive absorption strategy, is already sufficient to prove a version of Lemma 2.1.2 with additional hypotheses on the set S. Namely, one can show the following.

Lemma 2.1.3 (Simpler version of Lemma 2.4.29)**.** *Let S_1 be a $o(n)$-sized vertex subset of $H_{\mathbb{Z}_n}[X,Y,Z]$. Then, $H_{\mathbb{Z}_n}[X,Y,Z]$ contains a vertex subset $\mathcal{A}'$ of size $o(n)$ such that for any $S_2 \subseteq S_1$ of size $O(n^{1-\varepsilon})$ intersecting X, Y and Z in the same number of vertices, closed under the function $x \to -x$, and containing no involutions, we have that $\mathcal{A}' \cup (S_1 \setminus S_2)$ has a perfect matching.*

In the above lemma, it would arguably be more natural to insist that $\mathcal{A}' \cup S_2$ has a perfect matching, as opposed to $\mathcal{A}' \cup (S_1 \setminus S_2)$. Such a version of the lemma would also be correct, and be even easier to prove. We formulate the lemma in the form above for a reason that will become clear shortly.

From absorption for pairs to absorption for arbitrary zero-sum sets

Now, we discuss how we can derive Lemma 2.1.2 from Lemma 2.1.3. The key idea is encapsulated in the following lemma.

Lemma 2.1.4 (Simpler version of Lemma 2.4.31)**.** *$H_{\mathbb{Z}_n}[X,Y,Z]$ contains a vertex subset $\mathcal{T}$ of size $o(n)$ such that for any $S \subseteq V(H_{\mathbb{Z}_n}[X,Y,Z]) \setminus \mathcal{T}$ of size $O(n^{1-\varepsilon})$ intersecting X, Y and Z in the same number of vertices, and satisfying $\sum S = 0$, there exists a matching M of order $O(n^{1-\varepsilon})$ with $\mathcal{T} \cup S \supseteq V(M) \supseteq S$ satisfying also that $V(M) \setminus S$ is closed under $x \to -x$.*

Deriving Lemma 2.1.2 from Lemma 2.1.3 and Lemma 2.1.4 is a simple exercise. Indeed, let $\mathcal{T}$ be a vertex subset of $H_{\mathbb{Z}_n}[X,Y,Z])$ with the property in Lemma 2.1.4. Apply Lemma 2.1.3 with $S_1 = \mathcal{T}$ to obtain a vertex subset $\mathcal{A}'$. Set $\mathcal{A} := \mathcal{A}' \cup \mathcal{T}$. We invite the reader to check that $\mathcal{A}$ then satisfies the property that Lemma 2.1.2 requires.

To see how we prove a version of Lemma 2.1.4, we invite the reader to see Section 2.4.3. Similar in spirit to the proof of Lemma 2.1.3, our main trick here is to reduce (a version of) Lemma 2.1.4 to the existence of many small matchings with certain desirable properties (see Lemma 2.4.30).

Additional difficulties

We now point out several complications we omitted in the previous discussion, along with some technicalities that arise in the level of generality of our main theorem.

Ensuring distinctness. A fair portion of our arguments rely on the existence of constant sized matchings with specific properties, such as being closed under the map $x \to -x$. Often, the properties we require can be written as solutions to a particular system of linear equations. For example, consider Q_x defined in Section 2.1.1. The requirements from $\{x, y, z, w\}$ could be written as:

$$x + y + a = 0 \qquad z + w - a = 0 \qquad x + w + b = 0 \qquad z + y - b = 0$$

This is a system of equations with 4 variables and 4 constraints; however, any 3 of these equations imply the fourth, allowing us to easily deduce that there are many x, y, z, w with the desired properties. However, a solution to the system of equations is useful to us only when $x \neq z$ and $y \neq w$, as otherwise $\{(x, -x+b, -b), (x+a-b, -x-a, b)\}$ would not be a matching in $H_{\mathbb{Z}_n}$.

In the specific scenario outlined above, getting the relevant coordinates to be distinct is not particularly challenging. However, throughout the paper, we will require gadgets with properties significantly more complicated than those of Q_x. Furthermore, our main theorem works with subsets $X, Y, Z \subseteq G$ which are disjoint. Thus, we would have no chance of locating Q_x within $H_G[X, Y, Z]$ unless all coordinates x, y, z, w are distinct. Due to these technicalities, finding gadgets with the desired properties can become quite delicate.

Section 2.2.6 is entirely devoted to obtaining a sufficient set of conditions for a system of relations to yield solutions where each coordinate is distinct. The key result of that section, Lemma 2.2.32, is used in abundance throughout the paper.

The elementary abelian 2-group. The strategy of working with pairs of inverses $\{x, -x\}$ which are not involutions fails for obvious reasons in the elementary abelian 2-group. This turns out to be not a serious complication, as for general groups G, we will work with pairs $\{x, q_\phi x^{-1}\}$ for a carefully chosen q_ϕ so that $x \to q_\phi x^{-1}$ does not create too many fixed points.

Nonabelian groups. Although it turns out that the case of general abelian groups is not significantly more complicated than cyclic groups, there are serious issues to overcome with nonabelian groups. As just one example, suppose that we wish to find a gadget similar to Q_x from Section 2.1.1 in H_G, where G is a nonabelian group. Suppose that $F_1 = \{a, q_\phi a^{-1}\}$, $F_2 = \{b, q_\phi b^{-1}\}$. Our goal is then to find (many) Q such that $Q \cup F_1$ and $Q \cup F_2$ both can be perfectly matched. It is quite instructive to try to construct such Q, and we invite the reader to try to do so.

Some reflection shows that while it is difficult to construct Q switching between $\{a, q_\phi a^{-1}\}$ and $\{b, q_\phi b^{-1}\}$, it is considerably simpler to find *some* b' such that b' is in the same G'-coset as $q_\phi b^{-1}$ such that we can find a Q switching between $\{a, q_\phi a^{-1}\}$ and $\{b, b'\}$. This motivates us to search for gadgets not only switching between pairs, but also switching between elements of the same G'-coset. Achieving this latter task is considerably more technical. We accomplish this by first devising a strategy for switching between commutator elements (as opposed to arbitrary elements in the commutator subgroup). To switch between arbitrary elements of G', we have to rely on a non-trivial fact from representation theory (see Theorem 2.4.11). The statement we require is that arbitrary elements in the commutator subgroup can be written as products of just $O(\log n)$ commutators. Arguably, this is the only point in the proof of the main theorem where we use a group theoretic result beyond the undergraduate level. Due to bounds coming from Theorem 2.4.11 (which are tight), we are obliged to use gadgets of logarithmic size to switch between elements of the same G'-coset. This inflates the error rate in our main theorem by a polylogarithmic factor. We refer the reader to Section 2.4.1 for more details.

2.2 Preliminaries

2.2.1 Probabilistic tools

Concentration inequalities

The below is a standard bound that can be found in many probability textbooks, for example see [5]. We will refer to it as Chernoff's bound.

Lemma 2.2.1 (Chernoff bound)**.** *Let $X := \sum_{i=1}^m X_i$ where $(X_i)_{i\in[m]}$ is a sequence of independent indicator random variables with $\mathbb{P}(X_i = 1) = p_i$. Let $\mathbb{E}[X] = \mu$. Then, for any $0 < \gamma < 1$, we have that $\mathbb{P}(|X - \mu| \geq \gamma\mu) \leq 2e^{-\mu\gamma^2/3}$.*

We use the following corollary of Chernoff's bound often: that if R is a p-random subset of an n-element set, then with high probability we have that $|pn - |R|| \leq \log n\sqrt{n}$.

In almost all instances, the Chernoff bound will be all we need. Otherwise, we will make use of Azuma's inequality, which we now state. Given a product probability space $\Omega = \prod_{i\in[n]} \Omega_i$, a random variable $X\colon \Omega \to \mathbb{R}$ is called C-Lipschitz if $|X(\omega) - X(\omega')| \leq C$ whenever ω and ω' differ in at most 1-coordinate. We will refer to the following standard bound as Azuma's inequality.

Lemma 2.2.2 (Azuma's inequality)**.** *Let X be C-Lipschitz random variable on a product probability space with n coordinates. Then, for any $t > 0$,*

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

Pseudorandom graphs and the Rödl nibble

Here, we give some tools to find matchings in pseudorandom hypergraphs covering all but a few vertices. Our approach here is complicated by the fact that we need to find large matchings in subsets of hypergraphs, and parts of the subsets we consider might have their elements adversarially chosen. This comes from the first step of the proof where we set aside an absorber, whose complement could potentially have poor pseudorandomness properties. The most important result from this section is Lemma 2.2.8, which tells us that subsets of multiplication tables contain large matchings whenever the subset is obtained by taking unions of two random sets, and one (potentially adversarially chosen) deterministic set. We now give the details.

For a hypergraph H, vertices u, v and a subset $U \subseteq V(H)$, we define the **pair degree** of (u, v) into U as the number of vertices in U which are in the neighbourhood of both u and v, i.e. the number of vertices z in U such that there exists $v, w \in V(H)$ such that $\{u, z, v\}$ and $\{v, z, w\}$ are both edges of H. We say that a r-partite r-uniform hypergraph H is (γ, p, n)-**regular** if every part has $(1 \pm \gamma)n$ vertices and every vertex has degree $(1 \pm \gamma)pn$. We say that H is (γ, p, n)-**typical** if, additionally, every pair of vertices x, y in the same part of H have pair degree $(1 \pm \gamma)p^2 n$ into every other part of H. We say that a hypergraph is **linear** if through every pair of vertices, there is at most one edge. Multiplication hypergraphs have all these properties.

Observation 2.2.3. *For a group G of order n, the multiplication hypergraph H_G is $(0, 1, n)$-typical and linear.*

Proof. It is immediate that all parts have size n. For any two vertices u, v in different parts, there is a unique edge through u and v (if $u \in A, v \in B$, then this edge is $(u, v, v^{-1}u^{-1})$. If $u \in B, v \in C$, then this edge is $(v^{-1}u^{-1}, u, v)$. If $u \in A, v \in C$, then this edge is $(u, u^{-1}v^{-1}, v)$), hence H_G is linear. This shows that all vertices have degree exactly n and pair degree exactly n to each part i.e. that the hypergraph is $(0, 1, n)$-typical. □

Frankl and Rödl [24] (also Pippenger, unpublished) showed that for all $\epsilon, p \gg \gamma \gg n^{-1}$ every (γ, p, n)-regular hypergraph has a matching of size $(1 - \epsilon)n$. We need a well known variant of this where ϵ, p, γ have polynomial dependencies on n.

Lemma 2.2.4. *Let n be sufficiently large. Every (γ, δ, n)-regular linear tripartite hypergraph has a matching covering all but at most $n^{1-1/500} + 3\gamma n$ vertices.*

Proof. There are various ways of proving this. We will deduce it from a result of Molloy-Reed — Theorem 1 from [44]. Applying that theorem with $k = 3$, $\Delta = (1+\gamma)\delta n$ gives us a decomposition of our (γ, δ, n)-regular hypergraph into $\Delta + c_k\Delta^{1-\frac{1}{k}} \log^4 \Delta \leq \delta n + \gamma\delta n + n^{6/7}$ matchings (for some constant c_k). By the pigeonhole principle one of these has at least $\frac{e(H)}{\delta n+\gamma\delta n+n^{6/7}} \geq \frac{(1-\gamma)^2\delta n^2/3}{\delta n+\gamma\delta n+n^{6/7}} \geq n/3 - (n^{1-1/500} + 3\gamma n)$ edges as required. □

Typical graphs have the following well-known pseudorandomness condition, which dates back to work of Thomason [57]. Note that for the rest of the section, we assume that bipartite graphs come with a partition of their vertex set as (A, B) and similarly tripartite hypergraphs come with a partition (A, B, C).

Lemma 2.2.5. *Let G be a (γ, δ, n)-typical bipartite graph. Then for every $A' \subseteq A, B' \subseteq B$, we have $e(A', B') = \delta|A'||B'| \pm 5\sqrt{(\delta+\gamma)n^3} + \gamma n^2$.*

Proof. This will be a consequence of Theorem 2 of [57]. First, delete at most γn vertices from one side of the graph to obtain a balanced bipartite graph. Now with parameters $p := \delta - \gamma$ and $\mu := 5\gamma n$ it is easy to see that the hypothesis of Theorem 2 are satisfied. From the conclusion of Theorem 2, for every $A' \subseteq A, B' \subseteq B$, we have $e(A', B') = p|A'||B'| \pm 5\sqrt{(\delta+\gamma)n^3}$. With another γn^2 term, we can account for the deleted vertices in the beginning, implying the desired bound. □

Typicality is preserved by taking random subsets, in the following sense.

Lemma 2.2.6. *Let $H = (A, B, C)$ be a tripartite linear hypergraph that is $(0, 1, n)$-typical. Let $p \geq n^{-1/600}$ and let $A' \subseteq A$ be p-random. Then, with probability at least $1 - 1/n^3$, the bipartite graph between B and C consisting of edges passing through A' is $(n^{-1/5}, p, n)$-typical.*

Proof. For some $c, c' \in C$, let $d_{A'}(c) = e(A', B, c)$ be the degree of c into A' and let $d_{A'}(c, c')$ denote the pair degree of (c, c') into A'. We have $d_A(c) = d_A(c, c') = n$ (by $(0, 1, n)$-typicality of H).

Note that $\mathbb{E}(d_{A'}(c)) = pd_A(c) = pn$ and $\mathbb{E}(d_{A'}(c, c')) = p^2 d_A(c, c') = p^2 n$ (using linearity of H) for all c, c'. Set $\gamma := n^{-1/5}$. By Chernoff's bound and a union bound, with probability at least $1 - 1/n^4$, for all c we have $d_{A'}(c) = \mathbb{E}(d_{A'}(c)) \pm \gamma n = pn \pm \gamma n$. Note that $d_{A'}(c, c')$ is 2-Lipschitz. This is because for each a, there is exactly one b with abc an edge, and one b with abc' an edge (using linearity of H). Hence by Azuma's inequality and a union bound, we have that with probability at least $1 - 1/n^4$, for each pair $c, c' \in C$, $d_{A'}(c, c') = \mathbb{E}(d_{A'}(c, c')) \pm \gamma n = p^2 n \pm \gamma n$. Corresponding bounds hold for $b, b' \in B$. With probability at least $1 - 3/n^4$ all these properties hold simultaneously. Whenever these properties all hold, we have that the bipartite graph (B, C) consisting of edges through A' is (γ, p, n)-typical as desired. □

Using the previous pseudorandomness property, we can derive the following lemma which states most vertices send approximately the expected number of edges through a random set and a deterministic set.

Lemma 2.2.7. *Let $H = (A, B, C)$ be a tripartite linear hypergraph that is $(0, 1, n)$-typical. Let $p \geq n^{-1/600}$ and let $A' \subseteq A$ be p-random. Then, with probability at least $1 - 1/n^3$, the following holds. For any $B' \subseteq B$, there are at most $n^{9/10}$ vertices $c \in C$ with $e_H(A', B', c) \neq p|B'| \pm n^{9/10}$.*

Proof. Set $\gamma = n^{-1/5}$. By Lemma 2.2.6, with probability at least $1 - 1/n^3$, we have that the bipartite graph between B and C consisting of edges passing through A' is $(n^{-1/5}, p, n)$-typical. Supposing this property holds, by Lemma 2.2.5, for any $B' \subseteq B, C' \subseteq C$, we have $e_H(A', B', C') = p|B'||C'| \pm 5\gamma^{1/2} n^2$. Let C^- be the set of vertices with $e_H(A', B', c) < p|B'| - \gamma^{1/4} n$. We have that $e_H(A', B', C^-) < p'|B'||C^-| - \gamma^{1/4} n|C^-|$ and $e_H(A', B', C^-) = p'|B'||C^-| \pm 5\gamma^{1/2} n^2$ implying $|C^-| \leq 10\gamma^{1/4} n$. Similarly letting C^+ be the set of vertices with $e_H(A', B', c) > p|B'| + \gamma^{1/2} n$, we get $|C^+| \leq 10\gamma^{1/4} n$. Plugging in the value of γ, this implies the lemma. □

The following lemma will allow us to find a large matching whenever we are given two random subsets and a deterministic subset.

Lemma 2.2.8. *Let $H = (A, B, C)$ be a tripartite linear hypergraph that is $(0, 1, n)$-typical. Let $p \geq n^{-1/600}$ and let $A' \subseteq A$ be p-random, and let B' a p-random subset* *of B, where A' and B' are not necessarily independent. Then, with probability at least $1 - n^{-2}$, the following holds.*

For any $C' \subseteq C$ of size $(1 \pm n^{-0.2})pn$, there is a matching covering all but $2n^{1-1/500}$ vertices in $A' \cup B' \cup C'$.

Proof. With probability at least $1 - n^{-2}$, A' and B' satisfy the conclusion of Lemma 2.2.7 and have size $(1 \pm n^{-0.2})pn$ (by Chernoff's bound). This means that $A' \cup B' \cup C'$ has $\leq n^{9/10}$ vertices with degree $\neq p^2 n \pm n^{19/20}$ in $H[A', B', C']$.

Deleting all such vertices gives a hypergraph satisfying the hypothesis of Lemma 2.2.4 with $\gamma := n^{-0.01}$, hence the desired matching exists. □

In some applications the following formulation which allows for three deterministic subsets as opposed to just one will be more convenient. The result follows simply by applying the previous result four times.

Lemma 2.2.9. *Let $H = H_G$ be a multiplication hypergraph. Let $p \geq n^{-1/650}$. Let A', B', C' be p-random subsets of A, B, C respectively, not necessarily independent. Set $R := A' \cup B' \cup C'$. With high probability the following holds. Let $q \leq 5p$. For any $X \subseteq V(H) \setminus R$ with $|X \cap A|, |X \cap B|, |X \cap C| = (1 \pm n^{-0.25})qn$, there is a matching in $R \cup X$ covering all but at most $n^{1-10^{-4}}$ vertices of $R \cup X$.*

Proof. Let q be a fixed rational number between 0 and 1 and denominator at most n. Suppose first that $q \leq n^{-1/600}$. We have that Lemma 2.2.8 holds for A' and B' with probability $\geq 1 - n^{1.5}$ and by Chernoff's bound $C' = (1 \pm n^{-0.2})pn$, with probability $\geq 1 - n^{-2}$. Both properties hold simultaneously with probability $\geq 1 - n^{-1.49}$. Then, using Lemma 2.2.8, $A' \cup B' \cup C'$ has a matching covering all but $n^{1-10^{-3}}$ vertices. Together with X, this gives $n^{1-10^{-3}} + 2n^{599/600} \leq n^{1-10^{-4}}$ vertices.

Suppose now that $q \geq n^{-1/600}$. For each $\diamond \in \{A, B, C\}$, partition $\diamond'$ into a q-random set $\diamond_1$, and a q-random set $\diamond_2$, and a $(p - 2q)$-random set $\diamond_3$. As $q, p - 2q \geq n^{-1/600}$ by assumption, with probability $\geq 1 - n^{-1.49}$ the pairs $(A_1, B_1), (A_2, C_1), (B_2, C_2)$ and (A_3, B_3) satisfy the property of Lemma 2.2.8, and $|\diamond_i| = (1 + n^{-0.22})\mathbb{E}[|\diamond_i|]$ for each $\diamond_i$.

Let $X_A = X \cap A$, $X_B = X \cap B$, $X_C = X \cap C$ to get sets of size $(1 \pm n^{-0.2})qn$. Using Lemma 2.2.8, we have matchings M_1, M_2, M_3, M_4 covering all, but at most $n^{1-10^{-3}}$ vertices of $A_1 \cup B_1 \cup X_C$, $A_2 \cup X_B \cup C_1$, $X_A \cup B_2 \cup C_2$, and $A_3 \cup B_3 \cup C_3$ respectively. In total, the number of uncovered vertices does not exceed $4n^{1-10^{-3}} \leq n^{1-10^{-4}}/10$.

Taking a union bound over all rational q with denominator at most n, we have shown that with high probability, there exists a matching covering all but $n^{1-10^{-4}}/10$ vertices. The statement for real values of q follows simply by using the property for the closest rational value q' to q with denominator at most n. Indeed, leaving out or deleting few elements, we can ensure that the set X has $|X \cap A|, |X \cap B|, |X \cap C| = (1 \pm n^{-0.25})q'n$, thereby obtaining a matching that covers all but $n^{1-10^{-4}}/10 + n^{1-10^{-4}}/10 \leq n^{1-10^{-4}}$ elements of the original sets $X \cup R$. □

2.2.2 The multiplication hypergraph

We make some clarifications regarding our notation with the multiplication hypergraph H_G of a group (recall that this was defined in Section 2.1). While referring to the parts of the multiplication hypergraph, we often omit the $A/B/C$ subscripts and think of A_G, B_G, C_G simply as copies of G. For example, whenever we have some vertices $v_1, v_2, \dots, v_k \in V(H_G)$, we write $v_1 v_2 \dots v_k$ to mean the product of the corresponding group elements in G (as opposed to in any of the copies A_G, B_G, C_G). Similarly, if $v \in V(H_G)$ and U is a subset of G, then we use $v \in U$ to mean that v is an element of G after dropping the $A/B/C$ subscript. For any subset $S \subseteq G$, we use $S_A/S_B/S_C$ to denote the corresponding subsets of $G_A/G_B/G_C$ respectively.

A subset S of $V(H_G)$ is called **balanced**, if $|S \cap G_\diamond|$ does not depend on the value of $\diamond \in \{A, B, C\}$.

2.2.3 Basic group theory definitions and results

Throughout the paper we use e to denote the identity element of a group G. Recall that we use G' to denote the commutator subgroup of a group G. That is, G' is the subgroup generated by elements of the form $[g,h] := ghg^{-1}h^{-1}$ where $g,h \in G$. We denote the abelianization of G (quotient of G by G') as G^{ab}. We sometimes call the elements of G^{ab} **G'-cosets**. For $g \in G$, we use $[g]$ to denote the unique G'-coset that g is a member of. When $g \in V(H_G)$, we think of $[g]$ as the G'-coset that resides in the same part $G_A/G_B/G_C$ that g resides in. As mentioned in the introduction, whenever we use additive notation together with elements of G, all operations take place in G^{ab}. We do this so that we don't have to use the $[g]$ notation excessively.

Lemma 2.2.10. *$g \in G'$ if, and only if, g can be written as $g = g_1 \dots g_t$ such there's a permutation σ of $[t]$ with $g_{\sigma(1)} \dots g_{\sigma(t)} = e$.*

Proof. To see the "only if" direction, write g as a product of commutators as

$$g = [a_1, b_1] \dots [a_t, b_t] = a_1 b_1 a_1^{-1} b_1^{-1} \dots a_t b_t a_t^{-1} b_t^{-1}.$$

Clearly, the latter product can be permuted as $a_1 a_1^{-1} b_1 b_1^{-1} \dots a_t a_t^{-1} b_t b_t^{-1} = e$, as required. For the "if" direction, consider some $g_1 \dots g_t$ which rearranges into $g_{\sigma(1)} \dots g_{\sigma(t)} = e$. Consider the quotient homomorphism $\phi : G \to G/G'$. Then since G/G' is abelian, we have $\phi(g_1 \dots g_t) = \phi(g_1) \dots \phi(g_t) = \phi(g_{\sigma(1)}) \dots \phi(g_{\sigma(t)}) = \phi(g_{\sigma(1)} \dots g_{\sigma(t)}) = \phi(e) = e$ i.e. $g_1 \dots g_t \in ker(\phi) = G'$ as required. □

The following is a well-known property of finite abelian groups.

Theorem 2.2.11 (Fundamental theorem of finite abelian groups)**.** *Let G be an abelian group. Then, G is isomorphic to a product of cyclic prime-power order groups.*

Given $g \in G$, $s(g)$ denotes the size of the set $\{x \in G \colon x^2 = g\}$.

Proposition 2.2.12 ([56]). *Let G be group such that there exists some $g \in G$ with $s(g) > (3/4)|G|$. Then, G is an elementary abelian 2-group and $g = 0$.*

Proof. We express our gratitude for all participants of the active discussion that took place on Math Overflow including Emil Jeřábek, Derek Holt, Saúl Rodríguez Martín, and Terry Tao. Here, we reproduce the argument of GH from MO [56]. Let $g \in G$, and suppose that $|s(g)| > (3/4)|G|$. Fix some $y \in G$, and define $S = \{x \in G \colon x^2 = g\}$ and $T = \{x \in S \colon xy \in S\}$. Observe that $|G \setminus T| \leq 2|G \setminus S|$, since $x \notin T$ only if $x \notin S$ or $xy \notin S$, and there are at most $|G \setminus S|$ many x of either type. By assumption $|G \setminus S| < |G|/4$, so $|G \setminus T| < |G|/2$. Consequently, $|T| > |G|/2$.

Observe that for any $x \in T$, we have that $(xy)^2 = g = x^2$, and so we have

$$xyx^{-1} = (xy)(xy)^{-2}(xy)^2x^{-1} = (xy)^{-1}x^2x^{-1} = (xy)^{-1}x = y^{-1}x^{-1}x = y^{-1}.$$

We claim that $C = \{x \in G \colon xyx^{-1} = y^{-1}\}$ is a coset of $C(y)$, the centralizer subgroup of y. To see this, take $x, w \in C$ and so $xyx^{-1} = y^{-1} = wyw^{-1}$, and rearranging we have that $(w^{-1}x)y(x^{-1}w) = y$, which implies that $w^{-1}x$ belongs to $C(y)$. Hence, w and x belong to the same coset of $C(y)$. Then, we can fix C' to be a coset of $C(y)$ with $C \subseteq C'$. Now, take some $w \in C'$. $|C| > |G|/2$ since $T \subseteq C$, so in particular, C is non-empty, so we can fix some $x \in C$. As $w, x \in C'$, $w^{-1}x \in C(y)$, so $(w^{-1}x)y(x^{-1}w) = y$, hence $xyx^{-1} = wyw^{-1}$. The left-hand side of the last equation is y^{-1} by definition of $x \in C$, so $wyw^{-1} = y^{-1}$, implying $w \in C$. So $C' = C$, and C is a coset of the desired form.

As $|C| > |G|/2$, and C is a coset, $C = G$ by Lagrange's theorem. This implies that $y = y^{-1}$ for each y, hence G is an elementary abelian 2-group. It follows that $g = 0$, as $s(g) = 0$ for some $g \neq 0$ in an elementary abelian 2-group. □

2.2.4 Generic elements, and choice of of a_ϕ, b_ϕ, c_ϕ

The following definition is critical.

Definition 2.2.13. A group element $g \in G$ is **generic** if $g \neq e$ and there are at most $n/10^{9000}$ solutions to $x^2 = g$ in G.

Let $N(G)$ denote the set of non-generic elements and note that $|N(G)| \leq 10^{9000}$. Similarly, we call vertices of H_G generic if the corresponding group element is generic.

As described in Section 2.1, it is critical to our method to pair up group elements with a fixed sum, where the fixed sum has some desirable properties. The following lemma serves to show that the pairing we desire exists.

Lemma 2.2.14. *For group G with $|G| \geq 10^{9900}$, there exist $a_\phi, b_\phi, c_\phi \in G$ with*

(a) $a_\phi b_\phi c_\phi = e$.

(b) There are at most 30 solutions to $x^2 = \{a_\phi, b_\phi, c_\phi\}$. In particular a_ϕ, b_ϕ, c_ϕ are generic.

(c) There are at most $30|G'|$ solutions to $x^2 \in [a_\phi] \cup [b_\phi] \cup [c_\phi]$. In particular, there are at most 30 self-paired cosets (i.e. cosets $[x]$ with $[x^2] = [a_\phi]/[b_\phi]/[c_\phi]$).

(d) If $|G'| \leq 10^{-9}n$, then $a_\phi, b_\phi, c_\phi \notin G'$.

Proof. For each $g \in G$, let $S(g) = \{x : x^2 = g\}$, and note that since every x is in precisely one $S(g)$, we have $\sum_{g\in G} |S(g)| = |\bigcup_{g\in G} S(g)| = |G| = n$. Let Y_1 be the set of elements y_ϕ with more than 10 solutions to $x^2 = y_\phi$, noting that $10|Y_1| \leq \sum_{y\in Y_1} |S(y)| \leq \sum_{y\in G} |S(y)| = n$, we get $|Y_1| \leq n/10$. Let Y_2 be the set of elements y_ϕ with $> 10|G'|$ solutions to $x^2 \in [y_\phi]$. Then $10|G'||Y_2| = \sum_{y\in Y_2} \sum_{z\in[y]} |S(z)| = \sum_{y\in Y_2} \sum_{k\in G'} |S(yk)| = \sum_{k\in G'} \sum_{y\in Y_2} |S(yk)| \leq \sum_{k\in G'} \sum_{y\in G} |S(yk)| = |G'|n$, which implies $|Y_2| \leq n/10$. If $|G'| \leq 10^{-9}n$, set $Y_3 = G'$, and if $|G'| > 10^{-9}n$. Let $X = X \setminus (Y_1 \cup Y_2)$ to get a set of size $|X| \geq n - n/10 - n/10 - 10^{-9}n \geq 0.79n$.

Letting $XX = \{xy : x, y \in X\}$ and $X^{-1} = \{z^{-1} : z \in X\}$, notice that $|XX|$, $|X^{-1}| \geq |X| \geq 0.79n$. Therefore there must exist some $v \in XX \cap X^{-1}$ i.e. $v = xy =$

z^{-1} for some $x, y, z \in X$. But then $xyz = e$, and so we can set $a_\phi = x, b_\phi = y, c_\phi = z$. □

Definition 2.2.15. We say that an element x is ϕ-**generic** if x, $a_\phi^{\pm 1}x$, $b_\phi^{\pm 1}x$, $c_\phi^{\pm 1}x$, $a_\phi^{\pm 1}b_\phi^{\pm 1}x$, $b_\phi^{\pm 1}c_\phi^{\pm 1}x$, $c_\phi^{\pm 1}a_\phi^{\pm 1}x$ are all generic.

Notice that the number of elements which aren't ϕ-generic is $\leq 30|N(G)| \leq 10^{9010}$. We call a subset ϕ-generic if all elements of that subset are ϕ-generic.

For each group G, we fix a triple a_ϕ, b_ϕ, c_ϕ with the properties as in Lemma 2.2.14. We call two vertices v and w of H_G coming from the same part a **pair** if $v \cdot w \in [x_\phi]$ where $x = a, b, c$ depending on whether $v, w \in A, B, C$.

We call a subset S of $V(H_G)$ **coset-paired** if S can be partitioned into a disjoint union of pairs. This is equivalent to the following statement. $|S \cap [g]| = |S \cap [x_\phi g^{-1}]|$ for every non-self-paired coset $[g]$ and $|S \cap [g]|$ is even for every self-paired coset $[g]$.

2.2.5 Symmetric sets

Recall that a p-**random** subset of set S is one obtained by sampling each element of S independently with probability p. Similarly, we say a collection of random sets $R_1, \dots, R_k \subseteq S$ is **disjoint** p-**random** if each element of S belongs to each R_i with probability p, and to none of the R_i with probability $1 - pk$, and these decisions are made independently for each element of S. Considering such disjoint distributions complicates our approach as it makes various gadgets significantly more difficult to find. The reason we are interested in such distributions is the applications we give later on in the paper. Indeed, all applications we give other than the alternative proof of Hall-Paige conjecture and Snevily's conjecture require that we work with such disjoint distributions.

In fact, we need to generalise the concept of a disjoint distribution even further so that we can work with random sets X, Y, Z where X, Y and $Z^{-1} = \{z^{-1} \colon z \in Z\}$ are sampled disjointly. This need comes from the applications to sequenceability and R-sequenceability (which are omitted in the thesis, but are present in [49]). Thankfully, this generalisation does not create many additional combinatorial difficulties. However, we still need the following definitions to state a single theorem that covers all of the applications we want to give.

For $g \in G$, define $\hat{g} := \{g, g^{-1}\}$, noting that $\hat{g}$ has size 1 or 2 (depending on whether g is an involution or not). For a subset $T \subseteq G$, let $\hat{T} = \{\hat{t} : t \in T\}$ and $\bigcup \hat{T} = \bigcup_{t \in T} \hat{t} = T \cup T^{-1}$. We say that a subset $T \subseteq G$ is **symmetric** if $T^{-1} = T$ (or equivalently if $T = \bigcup \hat{T}$). We call a subset $S \subseteq V(H_G)$ symmetric if $S \cap A, S \cap B, S \cap C$ are all symmetric. We say that R is a **symmetric p-random subset** of G if R is always symmetric and $\hat{R}$ is a p-random subset of $\hat{G}$ (or equivalently if R is formed by flipping an independent coin for each $\hat{g} \in \hat{G}$ and taking the union of all group elements for which heads comes up). We say that R^1, R^2, R^3 are **disjoint symmetric p-random subsets** of G if additionally the joint distribution of $\hat{R}^1, \hat{R}^2, \hat{R}^3$ is that of disjoint p-random subsets of $\hat{G}$. The following two lemmas are useful as they allow us to jump between these definitions.

Lemma 2.2.16. *Let R_1, R_2, R_3 be disjoint p-random sets. Then there are $S_1 \subseteq R_1$, $S_2 \subseteq R_2$, $S_3 \subseteq R_3$ so that the joint distribution on S_1, S_2, S_3 is that of disjoint symmetric p^2-random sets.*

Proof. Let T be a p- random subset of G, independent of R_1, R_2, R_3. For non-involutions g, place g, g^{-1} into S_i whenever $\{g, g^{-1}\} \subseteq R_i$. For involutions g, place g into S_i whenever $g \in R_i \cap T$. Notice that for each $g \in G$, this gives $\mathbb{P}(\hat{g} \subseteq S) = p^2$. Also, these events are mutually independent for different $\hat{g}, \hat{h}$ since they depend on different coordinates. □

Lemma 2.2.17. *Let Q_1, Q_2, Q_3, R be random sets with Q_1, Q_2, Q_3 being disjoint symmetric q-random and R being p-random and independent of Q_1, Q_2, Q_3. Then, there are $S_1 \subseteq Q_1 \cap R$, $S_2 \subseteq Q_2 \cap R$, $S_3 \subseteq Q_3 \cap R$ so that the joint distribution on S_1, S_2, S_3 is that of disjoint symmetric p^2q-random sets.*

Proof. Let T be a p-random subset of G, independent of R, Q_1, Q_2, Q_3. For non-involutions g, place g, g^{-1} into S_i whenever $\{g, g^{-1}\} \subseteq Q_i \cap R$. For involutions g, place g into S_i whenever $g \in Q_i \cap R \cap T$. Notice that for each $\hat{g} \in \hat{G}$, this gives

$\mathbb{P}(\hat{g} \subseteq S_i) = p^2 q$. Also, these events are mutually independent for different $\hat{g}, \hat{h}$ since they depend on different coordinates. □

2.2.6 Free products

The goal of this section is to prove Lemma 2.2.32, which is our main tool to find vertex-disjoint copies of constant sized substructures in multiplication hypergraphs. We first remind the reader of some standard group theoretic terminology. We use F_k to denote the free group on k generators $v_1, \dots, v_k$.

For a group G, we use $G * F_k$ to denote the free product of G and F_k. We call $v_1, \dots, v_k$ the **free variables** of the free product $G * F_k$. A **word** is simply an element of $G * F_k$. For $w \in G * F_k$, we define the length of w to be the minimum number ℓ needed to write $w = x_1 x_2 \dots x_\ell$ for $x_i \in \{v_i, v_i^{-1} : i = 1, \dots, k\} \cup G$. A word $w \in G * F_k$ is called **reduced** if it cannot be made shorter using the group operations i.e. if it doesn't contain consecutive elements of G, doesn't contain e, and doesn't contain consecutive v_i, v_i^{-1} or v_i^{-1}, v_i for any i. It is a standard property of free products that every $g \in G * F_k$ is uniquely expressible as a reduced word.

Lemma 2.2.18. *For $w \in G * F_k$, there's a unique way of writing $w = g_0 x_1 g_1 \dots x_t g_t$ with $x_i \in \{v_1, \dots, v_k, v_1^{-1}, \dots, v_k^{-1}\}$ and $g_i \in G$ such that we don't have "$x_i = x_{i+1}^{-1}$ and $g_i = e$" for any $i \in \{0, 1, \dots, t-1\}$.*

Proof. Let W be the set of words which can be written of the form $g_0 x_1 g_1 \dots x_t g_t$ with $x_i \in \{v_1, \dots, v_k, v_1^{-1}, \dots, v_k^{-1}\}$ and $g_i \in G$ such that we don't have "$x_i = x_{i+1}^{-1}$ and $g_i = e$" for any $i \in \{0, 1, \dots, t-1\}$. Let R be the set of reduced words of $G * F_k$. Let $f : W \to R$ be defined by mapping $w = g_0 x_1 g_1 \dots x_t g_t$ to the word formed by removing all copies of e from w. Let $g : R \to W$ be defined by mapping $w' = y_1 y_2 \dots y_k$ to the word formed by inserting e between any y_i, y_{i+1} which are both in $\{v_1, \dots, v_k, v_1^{-1}, \dots, v_k^{-1}\}$. It is easy to see that $f(g(w')) = w'$ and $g(f(w)) = w$ for any $w \in W$ and $w' \in R$ i.e. both functions are bijections. Thus, since every $g \in G * F_k$ is uniquely expressible as a reduced word, it is also uniquely expressible as a word in W. □

From the above, we get that every $w \in G*F_k$ can be written as $w = g_0x_1g_1 \ldots x_tg_t$ with $x_i \in \{v_1, \ldots, v_k, v_1^{-1}, \ldots, v_k^{-1}\}$ and $g_i \in G$. We say that $w \in G * F_k$ is **linear in** v_i there's a way of writing w like this with precisely one occurrence of v_i (meaning one occurrence v_i or v_i^{-1}, but not both). We say that w is **linear** if all variables occur at most once in w, and some variable occurs exactly once in w. A useful fact is that for a linear w, there's a unique way of writing it as $w = g_0x_1g_1 \ldots x_tg_t$ (with $x_i \in \{v_1, \ldots, v_k, v_1^{-1}, \ldots, v_k^{-1}\}$ and $g_i \in G$) such that there's at most one occurrence of each variable. This comes from Lemma 2.2.18, because in a linear $w = g_0x_1g_1 \ldots x_tg_t$ it is impossible to have $x_i = x_{i+1}^{-1}, g_i = e$ (since this would create two occurrences of the free variable x_i).

A homomorphism $\pi : G * F_k \to G$ is a **projection** if $\pi(g) = g$ for all $g \in G$.

Lemma 2.2.19. *For each function $f : \{v_1, \ldots, v_k\} \to G$, there is precisely one projection $\pi_f : G * F_k \to G$ which agrees with f on $\{v_1, \ldots, v_k\}$. In particular, there are precisely n^k projections $G * F_k \to G$.*

Proof. Let $f : \{v_1, \ldots, v_k\} \to G$ be a function. By the universal property of free groups, there is a unique homomorphism $\phi_f : F_k \to G$ which agrees with f on $\{v_1, \ldots, v_k\}$. Let $e_G : G \to G$ be the identity homomorphism. By the universal property of free products there is a unique homomorphism π_f which agrees with ϕ_f on F_k and agrees with e_G on G. Such a homomorphism is exactly a projection from $G * F_k$ to G.

For the "in particular" part, note that the number of functions $\{v_1, \ldots, v_k\} \to G$ is exactly n^k, and so there are this many projections $G*F_k \to G$ by the first part. □

We use $\pi_0 : G * F_k \to G$ to denote the projection which maps all $w \in F_k$ to e (i.e. the map coming from Lemma 2.2.19 via the function f mapping all v_i to e).

Observation 2.2.20. *For all $g, h \in G$ there are the same number of solutions to $x^2 \in [g]$ and $x^2 \in [h^2g]$.*

In particular, for any $w, w' \in G * F_k$, there are the same number of solutions to $x^2 \in [\pi_0(w^{-1}w')]$ an $x^2 \in [\pi_0(ww')]$ in G).

Proof. We have that $x^2 \in [g]$ if, and only if $(hx)^2 \in [h^2 g]$. Therefore, if S is the set of solutions to $x^2 \in [g]$, then hS is the set of solutions to $(hx)^2 \in [h^2 g]$. Since S and hS always have the same size in a group, this gives what we want.

The "in particular" part follows from the above by taking $g = \pi_0(w^{-1}w'), h = \pi_0(w)$, and noting that, since π_0 is a homomorphism, we have $\pi_0(ww') = h^2 g$ □

Definition 2.2.21. Let $w, w' \in G * F_k$. We say that w and w' are **strongly separable** if any of the following hold.

(a) $w^{-1}w'$ is linear in some free variable v_i. Note that this is equivalent to asking that a free variable v_i appears once in one of w/w', and never in the other.

(b) w, w' are linear, contain at least one free variable, and there is a $g \in G$ with g generic so that $w' \in \{gw, g^{-1}w, gw^{-1}, g^{-1}w^{-1}, wg, wg^{-1}, w^{-1}g, w^{-1}g^{-1}\}$.

(c) All of the following hold.

- $|G'| \leq 10^{-9}n$
- w and w' are linear and have the same free variables (potentially with different signs).
- We either have $\pi_0(ww') \notin G'$ or some free variable occurs with the same sign in w, w'.
- We either have $\pi_0(w^{-1}w') \notin G'$ or some free variable occurs with the opposite sign in w, w'.
- There are $\leq 90|G'|$ solutions to $x^2 \in [\pi_0(ww')]$ (or equivalently by Observation 2.2.20, there are $\leq 90|G'|$ solutions to $x^2 \in [\pi_0(w^{-1}w')]$ in G).

Definition 2.2.22. We say that two sets $S, T \subseteq G$ are **strongly separable** if every pair of elements $s \in S, t \in T$ are strongly separable.

The following observation is quite critical, and justifies why considering symmetric disjoint random sets doesn't create additional complications when compared to disjoint random sets.

Observation 2.2.23. *w, w' are strongly separable $\iff$ w^{-1}, w' are strongly separable $\iff$ $\hat{w}, \hat{w}'$ are strongly separable.*

Proof. Within this proof, we say separable to mean strongly separable. For "w, w' are separable $\iff$ w^{-1}, w' are separable": Consider the possible cases of Definition 2.2.21. If w, w' are separable by (a), then w^{-1}, w' are also separable by (a) (this is immediate when one considers the "a free variable v_i appears once in one of w/w', but not both" version of (a)). If w, w' are separable by (b), then w^{-1}, w' are also separable by (b) — this is true because the set $\{gw, g^{-1}w, gw^{-1}, g^{-1}w^{-1}, wg, wg^{-1}, w^{-1}g, w^{-1}g^{-1}\}$ doesn't change if you replace each "w" with "w^{-1}". If w, w' are separable by (c), then w^{-1}, w' are also separable by (c). (The first bullet point doesn't involve w, w'. The second bullet point doesn't change by replacing w with w^{-1} because w, w^{-1} are always linear in the same free variables. The 3rd and 4th bullet points get exchanged when replacing w by w^{-1}. The 5th bullet point doesn't change when replacing w by w^{-1} since there is the same number of solutions to $x^2 \in [\pi_0(ww')]$ and $x^2 \in [\pi_0(w^{-1}w')]$ by Observation 2.2.20).

The direction "$\hat{w}, \hat{w}'$ are separable $\implies$ w, w' are separable" is immediate from the definition of "S, T are separable". For " w, w' are separable $\implies$ $\hat{w}, \hat{w}'$ are separable", note that once we know that w, w' are separable, we also know that the pairs (w^{-1}, w'), $(w, (w')^{-1})$, $(w^{-1}, (w')^{-1})$ are separable (all coming from "w, w' are separable $\iff$ w^{-1}, w' are separable"). This gives that $\hat{w}, \hat{w}'$ are separable. □

Observation 2.2.24. *Let $S \subseteq G * F_k$ be a set of linear elements. For each $T \subseteq \{v_1, \ldots, v_k\}$, let S_T be the set of $w \in W$ such that the free variables in w are exactly the set T. Then the sets $S_T, S_{T'}$ are strongly separable for distinct T, T'.*

Proof. If T, T' are distinct, then there's some $v_i \in T \Delta T'$ say $v_i \in T \setminus T'$. For any $w \in S_T, w' \in S_{T'}$, we have that v_i appears in w (just once by linearity) but not w', and so part (a) of the definition of "strongly separable" applies. □

Definition 2.2.25. Let $w, w' \in G{*}F_k$. We say that w and w' are **weakly-separable** if either they are strongly-separable, or they satisfy the following property.

(b') For some non-identity element g, the equation $w = w'$ rearranges into $e = g$.

We remark that the key difference between strong and weak separability comes from the property (b') not necessarily holding when w is replaced with w^{-1}, i.e. Observation 2.2.23 fails for weak separability. This will not be an issue while we search for gadgets, as we rely on weak separation only to separate elements coming from the same set A_G, B_G, C_G.

Definition 2.2.26. Let $S \subseteq G * F_k$. We say that a homomorphism $\phi : G * F_k \to G$, **separates** S if for every weakly-separable $w, w' \in S$ we have $\phi(w) \neq \phi(w')$.

Recall that $(G * F_k)'$ denotes the commutator subgroup of $(G * F_k)$.

Lemma 2.2.27. *For a group G, let $w, w' \in G * F_k$ satisfy part (c) of the definition of "strongly separable". Then $w = w'$ rearranges to $u^2 = \pi_0(w^{-1}w')k$ for some $k \in (G * F_k)'$ and some $u \in F_k$ which is either linear or equals e. Additionally, $u = e$ only if the free variables occur with the same signs in w, w'.*

Proof. Let $w = g_0 x_1 g_1 \dots x_t g_t$ and $w' = g'_0 x'_1 g'_1 \dots x'_{t'} g'_{t'}$ (with $x_i, x'_i \in \{v_1, \dots, v_k, v_1^{-1}, \dots, v_k^{-1}\}$ and $g_i, g'_i \in G$). Note that the assumption "w and w' are linear and have the same free variables" implies that $t = t'$, and that there is a permutation σ of $[t]$ so that $x'_i \in \{x_{\sigma(i)}, x_{\sigma(i)}^{-1}\}$ for $i = 1, \dots, t$. Partition $[t] = I^+ \cup I^-$ with $I^+ = \{i : x'_i = x_{\sigma(i)}\}$ and $I^- = \{i : x'_i = x_{\sigma(i)}^{-1}\}$. Set $u = \prod_{i \in I^-} (x'_i)^{-1}$, noting that u is in F_k, and that $u = e \iff I^- = \emptyset \iff$ the free variables occur with the same signs in w, w'. This also implies that if $u \neq e$, then u is linear. Note that $w = w'$ rearranges to $u^2 = \pi_0(w^{-1}w')k$ where $k = \pi_0(w^{-1}w')^{-1}u^2 w' w^{-1}$. Notice that

$$\begin{aligned}
&\pi_0(w^{-1}w')^{-1}u^2 w' w^{-1} = \pi_0(w')^{-1}\pi_0(w) u^2 w' w^{-1} \\
&\quad = (g_t'^{-1} \dots g_1'^{-1} g_0'^{-1})(g_0 \dots g_{t-1} g_t)(\prod_{i \in I^-} (x'_i)^{-1})(\prod_{i \in I^-} (x'_i)^{-1})(g'_0 x'_1 g'_1 \dots x'_t g'_t)(g_t^{-1} x_t^{-1} g_{t-1}^{-1} \dots x_1^{-1} g_0^{-1})
\end{aligned}$$

Notice that the above product can be permuted into the identity. Indeed each $g_i'^{-1}$ in the 1st bracket cancels with g'_i in the 5th bracket, each g_i in the 2nd bracket cancels with g_i^{-1} 6th bracket, for $i \in I^-$ each $(x_i')^{-1}$ in the 3rd bracket cancels with x_i' in the 5th bracket, each $(x_i')^{-1}$ in the 4th bracket cancels with $(x_{\sigma(i)})^{-1} = x_i'$ in the 6th bracket, and for $i \in I^+$ each x_i' in the 5th bracket cancels with $x_{\sigma(i)}^{-1} = (x_i')^{-1}$ in the 6th bracket. Thus, by Lemma 2.2.10, $k \in (G * F_k)'$. □

Lemma 2.2.28. *Let $w \in G * F_k$ be linear in some free variable v_i and let $g \in G$. Then there are exactly n^{k-1} projections $\pi : G * F_k \to G$ having $\pi(w) = g$.*

Proof. Without loss of generality, suppose $i = k$. By Lemma 2.2.19 there are exactly n^{k-1} projections $\pi : G * F_{k-1} \to G$. We will show that for every such projection, there is a unique projection $\pi' : G * F_k \to G$ that agrees with π on $G * F_{k-1}$ and additionally has $\pi'(w) = g$. To see this, note that the equation $w = g$ in the group $G * F_k$ rearranges into $v_k = h$ where $h \in G * F_k$ such that the free variable v_k doesn't occur in h (this is possible because w is linear in v_k). This shows that for any projection π', the equation $\pi'(w) = g$ is equivalent to $\pi'(v_k) = \pi'(h)$ (using that for any $g \in G$ we have $\pi'(g) = g$ for any projection π'). Since $h \in G * F_{k-1}$, the image $\pi(h) \in G$ is defined. Thus a projection π' agrees with π on $F_{k-1} * G$ and also has $\pi'(w) = g$ if and only if $\pi'(v_1) = \pi(v_i), \dots, \pi'(v_{k-1}) = \pi(v_{k-1})$, and also $\pi'(v_k) = \pi(h)$. By Lemma 2.2.19, there is a unique projection satisfying this, as required. □

Lemma 2.2.29. *Let $S \subseteq G * F_k$ be a set of elements which are each linear in at least one variable, and let $U \subseteq G$. Then the number of projections $\pi : G * F_k \to G$ for which $\pi(S)$ intersects U is $\leq |S||U|n^{k-1}$.*

Proof. For each $w \in S$ and $u \in U$, by Lemma 2.2.28, there are n^{k-1} projections $\pi : G * F_k \to G$ with $\pi(w) = u$. Thus, summing over all w, u, there are at most $\leq |S||U|n^{k-1}$ projections with $\pi(S)$ intersecting U. □

Lemma 2.2.30. *Let n be sufficiently large. Let $S \subseteq G * F_k$ be a set of size ≤ 1000. Then there are at most $0.1n^k$ projections $\pi : G * F_k \to G$ which do not separate S. If $|G'| > 10^{-9}n$, then this can be improved to "at most $10^{-8900}n^k$ projections".*

Proof. The total number of projections $\pi : G * F_k \to G$ is n^k by Lemma 2.2.19. Consider two weakly separable words $w, w' \in S$. We will count the number of projections for which $\pi(w) = \pi(w')$. There are four cases, depending on which part of the definition of weakly separable applies to w, w'.

(a) Note that the equation $\pi(w) = \pi(w')$ can be rearranged into $\pi(w^{-1}w') = e$. Since in (a), $w^{-1}w'$ is linear in some variable, Lemma 2.2.28 tells us that there are $n^{k-1} \leq n^k/10^{9000}$ projections $\pi : G * F_k \to G$ for which $\pi(w^{-1}w') = e$.

(b') Since $w = w'$ rearranges into $e = g$, the equation $\pi(w) = \pi(w')$ rearranges into $\pi(e) = \pi(g)$. But this is impossible for a projection π (since the definition of "projection" gives that $\pi(e) = e$ and $\pi(g) = g$). Thus there are zero projections with $\pi(w) = \pi(w')$ in this case.

(b) If we are not in some case covered by the previous bullet point, then the equation $w = w'$ rearranges into $(w)^2 = g$ for a generic group element $g \in G$. Let T_g be the set of solutions to $x^2 = g$ to get a set of size $\leq n/10^{9000}$, using the definition of a generic group element. For a projection π to have $\pi(w) = \pi(w')$, it must have (using that π is a homomorphism) $\pi(w)^2 = \pi(g) = g$ and so $\pi(w) \in T_g$. By Lemma 2.2.29, there are $|T_g|n^{k-1} \leq n^k/10^{9000}$ projections with $\pi(w) \in T_g$.

(c) Using Lemma 2.2.27, $w = w'$ rearranges into $u^2 = \pi_0(w^{-1}w')k$ where $u \in F_k$ is either linear or $u = e$, and $k \in (G * F_k)'$. If $u = e$, then we know that free variables occur with the same signs in w, w', which tells us that $\pi_0(w^{-1}w') \notin G'$. This gives that $[\pi_0(w^{-1}w')k] \neq G'$, and so $e \neq \pi_0(w^{-1}w')k = u^2 = e$, a contradiction. Thus in this case, there are no projections with $\pi(w) = \pi(w')$. So we can assume that $u \neq e$ which implies by Lemma 2.2.27 that u is linear.

Let T be set of solutions to $x^2 \in [\pi_0(w^{-1}w')]$ in G. Since we are in (c), we have $|T| \leq 90|G'| \leq 90 \cdot 10^{-9}n \leq 10^{-7}n$. For $\pi(w) = \pi(w')$ to hold, we must have (using that π is a homomorphism) $\pi(u^2) = \pi(\pi_0(w^{-1}w')k)$, which implies (using that π is a projection) $\pi(u)^2 = \pi_0(w^{-1}w')\pi(k)$. Since

$\pi_0(w^{-1}w')\pi(k) \in [\pi_0(w^{-1}w')]$ for any projection π (we have $\phi(H') \subseteq G'$ for any group homomorphism $\phi : H \to G$, since if x is a commutator in H, then $\phi(x)$ is a commutator in G), this would imply that $\pi(w) \in T$. By Lemma 2.2.29, there are $|T|n^{k-1} \leq 10^{-7}n^k$ projections with $\pi(w) \in T$ as required.

There are at most $\binom{|S|}{2} \leq \binom{1000}{2} \leq 10^6$ pairs of weakly separable $w, w' \in S$, and for each of them there are $\leq n^k/10^7$ projections with $w = w'$. Thus in total there are $\leq 10^6 n^k/10^7 = 0.1n^k$ projections which don't separate S. This implies the lemma when $|G'| \leq 10^{-9}n$. When $|G'| > 10^{-9}n$, note that case (c) can't occur, so we actually have $\leq n^k/10^{9000}$ projections with $w = w'$ for each separable w, w'. This gives a total of $\leq 2 \cdot 10^6 n^k/10^{9000} \leq n^k/10^{8900}$ projections which don't separate S. □

Lemma 2.2.31. *Let n be sufficiently large, $k \leq 200$, and $S \subseteq G * F_k$ a set of ≤ 1000 elements which are linear in at least one variable. There are projections $\pi_1, \dots, \pi_{10^{-3000}n}$ which separate S and have $\pi_1(S), \dots, \pi_{10^{-3000}n}(S)$ disjoint. If $|G'| > 10^{-9}n$, then we can additionally ensure that $\pi_j(v_i) \in G'$ for all free variables v_i and projections π_j.*

Proof. For $|G'| \leq 10^{-9}n$ say that a projection π is good if it separates S. For $|G'| > 10^{-9}n$ say that a projection π is good if it separates S and has $\pi(v_i) \in G'$ for all $v_i \in F_k$. Our task is to find $10^{-3000}n$ good projections $\pi_i(S)$, which have $\pi_i(S)$ disjoint for different i. Consider a maximal family $\pi_1, \dots, \pi_t$ of good projections. which have $\pi_1(S), \dots, \pi_t(S)$ disjoint. Let $T = \pi_1(S) \cup \dots \cup \pi_t(S)$. By maximality, we have that all good projections π have $\pi(S) \cap T \neq \emptyset$. Lemma 2.2.29 tells us that the number of projections with $\pi(S) \cap T \neq \emptyset$ is $\leq |T|n^{k-1} \leq t|S|n^{k-1} \leq 1000tn^{k-1}$. Thus we established that there are $\leq 1000tn^{k-1}$ good projections.

Suppose $|G'| \leq 10^{-9}$. Lemma 2.2.30 tells us that there are $\leq 0.1n^k$ bad projections, and hence $\geq n^k - 0.1n^k = 0.9n^k$ good projections. Thus $t \geq 0.9n/1000 \geq 10^{-3000}n$.

Now suppose $|G'| > 10^{-9}n$: There are $|G'|^k$ projections with $\pi(v_i) \in G'$ for all $v_i \in F_k$ (using Lemma 2.2.19). By Lemma 2.2.30, there are $\leq n^k/10^{8900} \leq 0.1(10^{-9}n)^{200} \leq 0.1(10^{-9}n)^k \leq 0.1|G'|^k$ projections that don't separate S. Hence there remain $\geq 0.9|G'|^k \geq 0.9 \cdot 10^{-9k} n^k$ projections that both separate S and have all $\pi(v_i) \in G'$ i.e. there are $\geq 0.9 \cdot 10^{-9k} n^k$ good projections. Combining with "there are $\leq 1000tn^{k-1}$ good projections", this gives $t \geq 0.9 \cdot 10^{-9k} n/1000 \geq 10^{-3000} n$. □

Lemma 2.2.32. *Let $p \geq n^{-1/700}$. Let R_A, R_B, R_C be disjoint p-random symmetric subsets of G and set $R = R_A \cup R_B \cup R_C$. With high probability, the following holds:*

Let $k \leq 200$, $S \subseteq G * F_k$ a set of ≤ 600 elements of length ≤ 200 which are each linear in at least one variable, and $U \subseteq G$ with $|U| \leq p^{800} n/10^{4000}$. Then there is a projection $\pi : G * F_k \to G$ which separates S and has $\pi(S) \subseteq R \setminus U$. Moreover:

- *For any $S_A, S_B, S_C \subseteq S$, with S_A, S_B, S_C being pairwise strongly separable sets, we can ensure $S_A \subseteq R_A, S_B \subseteq R_B, S_C \subseteq R_C$.*
- *If $|G'| > 10^{-9}n$, then we can additionally ensure that $\pi(v_i) \in G'$ for all free variables v_i.*

Proof. We can assume that n is sufficiently large (otherwise the lemma is vacuous since "with high probability" wouldn't mean anything). First fix $k \leq 200$, and some set $S \subseteq G$, and $S_A, S_B, S_C \subseteq S$ as in the lemma. Note $|\bigcup \hat{S}| \leq 2|S| \leq 800$. Fixing $m := n/10^{3000}$, apply Lemma 2.2.31 to get projections $\pi_1, \dots, \pi_m$ which separate $\bigcup \hat{S}$ and have $\pi_1(\bigcup \hat{S}), \dots, \pi_m(\bigcup \hat{S})$ disjoint (and additionally, when $|G'| > 10^{-9}n$, having that $\pi_j(v_i) \in G'$ for all free variables v_i and projections π_j).

Note that for each i we have $\pi_i(\hat{S}_A), \pi_i(\hat{S}_B), \pi_i(\hat{S}_C)$ pairwise disjoint. Indeed if say $\pi_i(\hat{S}_A) \cap \pi_i(\hat{S}_B) \neq \emptyset$, then there would be some $a \in S_A, b \in S_B$ with $\pi_i(\hat{a}) = \pi_i(\hat{b})$. We know that a, b separable, which implies that $\hat{a}, \hat{b}$ are separable (by Observation 2.2.23). But π_i separates $\bigcup \hat{S}$, which shows that $\hat{a}, \hat{b}$ are disjoint.

For each i, $\hat{g} \in \pi_i(\hat{S})$, let

$$ABC(\hat{g}) = \begin{cases} A \text{ if } \hat{g} \in \pi_i(\hat{S}_A) \\ B \text{ if } \hat{g} \in \pi_i(\hat{S}_B) \\ C \text{ if } \hat{g} \in \pi_i(\hat{S}_C) \\ C \text{ otherwise} \end{cases} .$$

Note that this is well defined by the previous paragraph. For each $i, \hat{g} \in \pi_i(S)$, let $E_i^{\hat{g}}$ be the event"$\hat{g} \subseteq R_{ABC(\hat{g})}$". Note that we have $P(E_i^{\hat{g}}) = p$. Note that $E_i^{\hat{g}}, E_j^{\hat{h}}$ are independent for $\hat{g} \neq \hat{h}$ ($E_i^{\hat{g}}$ depends only on the coordinate $\hat{g}$, $E_j^{\hat{h}}$ depends only on $\hat{h}$). Let $E_i = \bigcap_{\hat{g} \in \pi_i(\hat{S})} E_i^{\hat{g}}$. We have $\mathbb{P}(E_i) = \prod_{\hat{g} \in \pi_i(\hat{S})} \mathbb{P}(E_i^{\hat{g}}) \geq p^{|S|}$. For all $i = 1, \dots, m$, the events E_i are independent (since E_i, E_j depend on the coordinates in $\pi_i(\bigcup \hat{S}), \pi_j(\bigcup \hat{S})$ respectively. These are disjoint for $i \neq j$). By linearity of expectation, the expected number of indices i for which E_i occurs is at least $p^{|S|}m$. By Chernoff's Bound, there are at least $p^{|S|}m/2$ indices for which E_i occurs with probability

$$\geq 1 - 2e^{\frac{1}{6}p^{|S|}m} \geq 1 - 2e^{\frac{1}{6}(n^{-1/700})^{600}n/10^{3000}} \geq 1 - 2e^{-n^{1/9}/10^{6000}}.$$

Since $|U| \leq p^{800}n/10^{4000} \leq p^{600}n/(2 \cdot 10^{3000}) = p^{|S|}m/2$, there is at least one such index with $\pi_i(S) \cap U = \emptyset$. This projection π_i satisfies the lemma.

To get the lemma for all possible families $\{k, S, S_A, S_B, S_C\}$, notice that there are $o(e^{n^{1/7}/36000})$ such families. Indeed, there are 200 choices for k and for each k, there are $\leq 201(400n)^{201}$ length ≤ 200 elements $w \in G * F_k$. There are $\leq (201(400n)^{201})^{600}$ sets of ≤ 600 such words. Hence, there are $\leq ((201(400n)^{201})^{600})^4 = o(e^{n^{1/37}/36000})$ families of 4-tuples of such subsets. So we can take a union bound over all such families. □

Corollary 2.2.33. *Let $p \geq 3n^{-1/1400}$. Let R be random set which is either p-random or symmetric p-random. With high probability, the following holds:*

Let $k \leq 200$, $S \subseteq G * F_k$ a set of ≤ 600 elements of length ≤ 200 which are each linear in at least one variable, and $U \subseteq G$ with $|U| \leq p^{1600}n/10^{4002}$. Then there is a projection $\pi : G * F_k \to G$ which separates S and has $\pi(S) \subseteq R \setminus U$.

Proof. Using Lemma 2.2.16, we can choose random sets $R_A, R_B, R_C \subseteq R$ such that the joint distribution on R_A, R_B, R_C is that of disjoint symmetric $p^2/9$-random sets. Now, the lemma follows from Lemma 2.2.32. □

We end with a simple application of the above lemma for later use.

Lemma 2.2.34. *Let $p \geq n^{-1/700}$. Let G be a group R a symmetric p-random subset of G. With high probability, the following holds.*

For any generic $x_\phi \in G$ and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4001}$, there are distinct and ϕ-generic $x, x' \in R \setminus U$ with $xx' = x_\phi$.

Proof. With high probability, Lemma 2.2.32 applies. Let x_ϕ, U be as in the lemma. Add all non-ϕ-generic elements to U in order to get a set U' with $|U'| \leq p^{800}n/10^{4000}$. Let $x_\phi \in G$ and consider the set $\{v_1, x_\phi v_1^{-1}\} \subseteq G * F_1$. Note that $v_1, x_\phi v_1^{-1}$ are separable (by part (b) of the definition), and so Lemma 2.2.32 gives a projection $\pi : G * F_1 \to G$ with $\pi(v_1), \pi(x_\phi v_1^{-1})$ distinct and contained in $R \setminus U'$ setting $x = \pi(x_\phi v_1^{-1}), x' = \pi(v_1)$ gives the lemma. □

2.3 The main theorem and its variants

This section is devoted to stating the main technical result of the thesis, and collecting various consequences thereof (including Theorem 1.2.1) which will be more convenient to use for various applications we give. For the applications, we need variants of Theorem 1.2.1, where the probability distributions on R^1, R^2, R^3 are different from the one given (e.g. it is sometimes useful to sample R^1, R^2, R^3 disjointly rather than independently). We begin by stating a more technical version of Theorem 1.2.1 which covers all the different distributions of these sets that we might need. The following definition is the most general case. We remark that for the main application of this strategy given in this thesis, namely the resolution of a generalisation of Snevily's conjecture, the preceding level of generality is not required, and Theorem 1.2.1 is in fact sufficient.

Definition 2.3.1. Let G be a group and let R_1, R_2, R_3 be random subsets of G. We say that R_1, R_2, R_3 are q-**slightly-independent**, if there random subsets $Q^1 \subseteq R^1$, $Q^2 \subseteq R^2$, $Q^3 \subseteq R^3$ such that the joint distribution on Q^1, Q^2, Q^3 is that of disjoint symmetric, q-random subsets of G.

The following observation about this definition is useful.

Observation 2.3.2. *If R_1, R_2, R_3 are q-slightly-independent, then so are R_1, R_2^{-1}, R_3.*

Proof. We have $Q^1 \subseteq R^1$, $Q^2 \subseteq R^2$, $Q^3 \subseteq R^3$ such that the joint distribution on Q^1, Q^2, Q^3 is that of disjoint, symmetric, q-random subsets of G. Note that since Q^2 is symmetric, $Q^2 = (Q^2)^{-1}$. Also, $(Q^2)^{-1} \subseteq (R^2)^{-1}$. Thus, Q^1, Q^2, Q^3 witness R_1, R_2^{-1}, R_3 being q-slightly-independent. □

The following is the strongest version of the main theorem that we prove in this paper. It is proved in Section 2.4.

Theorem 2.3.3. *Let $q \geq n^{-1/10^{101}}$. Let G be a group of order n. Let $R^1, R^2, R^3 \subseteq G$ be p-random, q-slightly-independent subsets. Then, with high probability, the following holds.*

Let X, Y, Z be equal-sized subsets of G_A, G_B, and G_C respectively, satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C)\Delta(X \cup Y \cup Z)| \leq q^{10^{17}} n/\log(n)^{10^{17}}$
- $\sum X + \sum Y + \sum Z = 0$ *(in G^{ab})*
- $e_G \notin X \cup Y \cup Z$

Then, $H_G[X, Y, Z]$ contains a perfect matching.

We now state and prove a number of consequences of this theorem, where the distributions on R^1, R^2, R^3 are more natural. Firstly, the following theorem is equivalent to Theorem 1.2.1 (we will prove this formally later on in the section).

Theorem 2.3.4. *Let $p \geq n^{-1/10^{102}}$. Let G be a group of order n. Let $R^1, R^2, R^3 \subseteq G$ be p-random subsets, sampled independently. Then, with high probability, the following holds.*

Let X, Y, Z be equal-sized subsets of G_A, G_B, and G_C respectively, satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C)\Delta(X \cup Y \cup Z)| \leq p^{10^{18}} n/\log(n)^{10^{18}}$

- $\sum X + \sum Y + \sum Z = 0$ *(in G^{ab})*

Then, $H_G[X, Y, Z]$ contains a perfect matching.

When R^1, R^2, R^3 are sampled disjointly, then the statement of the theorem needs to change slightly. In this case, when the group is $\mathbb{Z}_2^k$, then it is impossible to cover e with a hyperedge contained in $R^1 \cup R^2 \cup R^3$. Thus to get a matching, we need to additionally have the condition "$e_G \notin X \cup Y \cup Z$ if $G \cong \mathbb{Z}_2^k$" (Proposition 2.2.12 shows that this is the only local obstruction of this type).

Theorem 2.3.5. *Let $q \geq n^{-1/10^{102}}$. Let G be a group of order n. Let $R^1, R^2, R^3 \subseteq G$ be p-random disjoint subsets. Then, with high probability, the following holds.*

Let X, Y, Z be equal-sized subsets of G_A, G_B, and G_C respectively, satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C)\Delta(X \cup Y \cup Z)| \leq p^{10^{18}} n/\log(n)^{10^{18}}$
- $\sum X + \sum Y + \sum Z = 0$ *(in G^{ab})*
- $e_G \notin X \cup Y \cup Z$ *if $G \cong \mathbb{Z}_2^k$.*

Then, $H_G[X, Y, Z]$ contains a perfect matching.

Finally, we have two versions of the theorem which are intermediate between the previous two. First one which asks R^1, R^2 to be sampled disjointly, and R^3 to be sampled independently.

Theorem 2.3.6. *Let $p \geq n^{-1/10^{102}}$. Let G be a group of order n. Let $R^1, R^2 \subseteq G$ be disjoint p-random subsets, and let $R^3 \subseteq G$ be a p-random subset, sampled independently with R^1 and R^2. Then, with high probability, the following holds.*

Let X, Y, Z be subsets of G_A, G_B, and G_C be equal sized subsets satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C)\triangle(X \cup Y \cup Z)| \leq p^{10^{18}} n/\log(n)^{10^{18}}$
- $\sum X + \sum Y + \sum Z = 0$ *(in the abelianization of G)*

- *If $G = (\mathbb{Z}_2)^k$ for some k, suppose that $e \notin Z$.*

Then, $H_G[X, Y, Z]$ contains a perfect matching.

The next theorem is almost the same as the previous one, with the difference that we sample $(R^1)^{-1}, R^2$ disjointly (as opposed to R^1, R^2).

Theorem 2.3.7. *Let $p \geq n^{-1/10^{102}}$. Let G be a group of order n. Let $(R^1)^{-1}, R^2 \subseteq G$ be disjoint p-random subsets, and let $R^3 \subseteq G$ be a p-random subset, sampled independently with R^1 and R^2. Then, with high probability, the following holds.*

Let X, Y, Z be subsets of G_A, G_B, and G_C be equal sized subsets satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C) \triangle (X \cup Y \cup Z)| \leq p^{10^{18}} n / \log(n)^{10^{18}}$
- $\sum X + \sum Y + \sum Z = 0$ *(in the abelianization of G)*
- *If $G = (\mathbb{Z}_2)^k$ for some k, suppose that $e \notin Z$.*

Then, $H_G[X, Y, Z]$ contains a perfect matching.

These theorems all have almost the same proof. The idea is to first show that the distribution of R_1, R_2, R_3 is q-slightly-independent (and so Theorem 2.3.3) applies. When the sets X, Y, Z don't contain the identity, then this already gives what we want. When X, Y, Z do contain the identity, then we first find a small matching covering all copies of the identity in X, Y, Z, and then find another matching covering all remaining vertices using Theorem 2.3.3.

Proof of Theorem 2.3.5. Note that R_1, R_2, R_3 are q-slightly-independent for $q = p^2/9$. To see this, consider disjoint $1/3$-random sets $T_1, T_2, T_3 \subseteq G$, chosen independently of R_1, R_2, R_3. We have that $R_1 \cap T_1, R_2 \cap T_2, R_3 \cap T_3$ are disjoint $p/3$-random subsets of G. Indeed for every $g \in G$, we have $\mathbb{P}(g \in R_i \cap T_i) = \mathbb{P}(g \in R_i)\mathbb{P}(g \in T_i) = p/3$, and $\mathbb{P}(g \in R_i \cap T_i \cap R_j \cap T_j) \leq \mathbb{P}(g \in T_i \cap T_j) = 0$. These imply $\mathbb{P}(g \notin \bigcup_{i=1}^3 R_i \cap T_i) = 1 - 3\alpha p$. Also for different g, h, their locations are independent of each other (since there was no correlations between distinct g, h, in any of $R_1, R_2, R_3, T_1, T_2, T_3$), giving that $R_1 \cap T_1, R_2 \cap T_2, R_3 \cap T_3$ are disjoint $p/3$-random subsets of G. Use Lemma 2.2.16 to pick disjoint, symmetric, q-random subsets $Q_1 \subseteq R_1 \cap T_1, Q_2 \subseteq R_2 \cap T_2, Q_3 \subseteq R_3 \cap T_3$. Now Q_1, Q_2, Q_3 demonstrate R_1, R_2, R_3 being q-slightly-independent.

So Theorem 2.3.3 applies to R_1, R_2, R_3. We also have the following property if $G \neq \mathbb{Z}_2^k$:

P: For each $i, j \in \{1, 2, 3\}$, $i \neq j$, $R^i \times R^j$ contains $\geq p^2n/1000$ pairs of the form (x, x^{-1}) where for each such pair $\{x, x^{-1}\} \cap \{y, y^{-1}\} = \emptyset$ (this follows by Chernoff's bound).

Now consider sets X, Y, Z as in the theorems. Use (P) to pick distinct elements g_A, g_B, g_C with $g_A \in R^2$, $g_A^{-1} \in R^3$, $g_B \in R^1$, $g_B^{-1} \in R^3$, $g_C \in R^1$, $g_C^{-1} \in R^2$. Noting that we have $p^2n/1000$ choices for each g_A, g_B, g_C, we can choose them to have $\{g_A, g_A^{-1}\}, \{g_B, g_B^{-1}\}, \{g_C, g_C^{-1}\}$ disjoint from each other and from $R^1 \setminus X, R^2 \setminus Y, R^3 \setminus Z$. The result is that the three edges $f_A := (e_A, g_A, g_A^{-1}), f_B := (g_B, e_B, g_B^{-1}), f_C := (g_C, g_C^{-1}, e_C)\}$ form a matching with all vertices, other than possibly e_B, e_B, e_C contained in $X \cup Y \cup Z$. Let $N = \{f_i : e_i \in X \cup Y \cup Z\}$ to get a matching contained in $X \cup Y \cup Z$ covering all copies of the identity in $X \cup Y \cup Z$.

Let $X' = X \setminus N, Y' = Y \setminus N, Z' = Z \setminus N$, noting that these have the same size and have $\sum X' + \sum Y' + \sum Z' = 0$ in G^{ab} (due to N being a matching). Thus the property of Theorem 2.3.3 applies to give a perfect matching M in $H_G[X', Y', Z']$. Now $M \cup N$ satisfies the theorem. □

Next we show how to modify the above proof to obtain Theorems 2.3.4, 2.3.6, 2.3.7.

Proof of Theorems 2.3.4, 2.3.6, 2.3.7. First notice that in all three theorems, we have that R^1, R^2, R^3 are q-slightly-independent. In Theorems 2.3.4, 2.3.6 this is exactly the first paragraph of the proof of Theorem 2.3.5. For Theorem 2.3.7, that paragraph shows that $(R^1)^{-1}, R^2, R^3$ are q-slightly-independent. But, then by Observation 2.3.2, we have that R^1, R^2, R^3 are q-slightly-independent.

Next note that property (P) holds in the following cases:

- Theorem 2.3.4: here property (P) always holds.
- Theorem 2.3.6: here property (P) always holds for $(i, j) = (2, 3)$ and $(i, j) = (1, 3)$. For $(i, j) = (1, 2)$ property (P) holds when $G \neq \mathbb{Z}_2^k$.
- Theorem 2.3.7: here property (P) always holds for $(i, j) = (2, 3)$ and $(i, j) = (1, 3)$. For $(i, j) = (1, 2)$ property (P) holds when $G \neq \mathbb{Z}_2^k$.

The rest of the proofs are the same as in Theorem 2.3.5 — property (P) produces a matching of size ≤ 3 covering all copies of the identity in $X \cup Y \cup Z$, and then the property of Theorem 2.3.3 gives a matching covering the rest of $X \cup Y \cup Z$. □

Finally we show how to derive Theorem 1.2.1 as stated in the introduction.

Proof of Theorem 1.2.1 via Theorem 2.3.4. Let $R^1, R^2, R^3 \subseteq G$ be p-random subsets, independently sampled. Observe that $R^1, R^2, (R^3)^{-1} \subseteq G$ are also p-random subsets, independently sampled, hence with high probability Theorem 2.3.4 holds. Let $X, Y, Z \subseteq G$ be subsets with the properties as in the statement of Theorem 1.2.1. Then, the sets $X, Y, Z^{-1} \subseteq G$ clearly satisfy the two properties required by Theorem 2.3.4 with respect to $R^1, R^2, (R^3)^{-1} \subseteq G$. Thus, $H_G[X, Y, Z^{-1}]$ contains a perfect matching, say M. Define the bijection $\phi \colon X \to Y$ so that x maps to the unique element y of Y such that x and y are contained in an edge together in M. As for each edge $(x, \phi(x), z)$ of M, $x\phi(x)z = e$, $x \mapsto x\phi(x)$ is a bijection $X \to (Z^{-1})^{-1} = Z$, as desired. □

2.3.1 Complete mappings and orthomorphisms

We conclude this section with a version of the main theorem that fits better with the proof of Lemma 2.5.8. Given a triple of subsets of a group G as (X, Y, Z), a **complete mapping** is a bijection $\phi\colon X \to Y$ such that the induced map from X to Z via $x \to x\phi(x)$ is also a bijection, whereas an **orthomorphism** is a bijection $\phi\colon X \to Y$ such that the induced map from X to Z via $x \to x^{-1}\phi(x)$ is also a bijection.

Observation 2.3.8. *Let X, Y, Z be subsets of a group G. Then,*

- (X, Y, Z) *admits a complete mapping if and only if* $H_G[X, Y, Z^{-1}]$ *has a perfect matching.*
- (X, Y, Z) *admits an orthomorphism if and only if* $H_G[X^{-1}, Y, Z^{-1}]$ *has a perfect matching.*

Proof. Note that $x\phi(x) = z$ if and only if $x\phi(x)z^{-1} = e$. As $z \to z^{-1}$ is a bijection between Z and Z^{-1}, we have that whenever ϕ is a complete mapping of X, Y, Z, we have a perfect matching of $H_G[X, Y, Z^{-1}]$ given by the edges of the form $(x, \phi(x), \phi(x)^{-1}x^{-1})$. Given a perfect matching of $H_G[X, Y, Z^{-1}]$, we can similarly define ϕ so that ϕ maps the G_A coordinate of each matched edge to the G_B coordinate to get a complete mapping of $(X, Y, (Z^{-1})^{-1}) = (X, Y, Z)$. This proves the first bullet point.

Similarly, $x^{-1}\phi(x) = z$ if and only if $x^{-1}\phi(x)z^{-1} = e$. Since $x \to x^{-1}$ is also a bijection between X and X^{-1}, if ϕ is an orthomorphism of X, Y, Z, we have a perfect matching of $H_G[X^{-1}, Y, Z^{-1}]$ given by the edges of the form $(x^{-1}, \phi(x), z^{-1})$. And given a perfect matching of $H_G[X^{-1}, Y, Z^{-1}]$, taking ϕ to map the inverse of the G_A coordinate of each matched edge into the G_B coordinate of that same matched edge, we obtain an orthomorphism of $((X^{-1})^{-1}, Y, (Z^{-1})^{-1}) = (X, Y, Z)$. This proves the second bullet point. □

Theorem 2.3.9. *Let $p \geq n^{-1/10^{102}}$. Let G be a group of order n. Let $R^1, R^2 \subseteq G$ be disjoint p-random subsets, and let $R^3 \subseteq G$ be a p-random subset, sampled independently with R^1 and R^2. Then, with high probability, the following holds.*

Let X, Y, Z be equal-sized subsets of G_A, G_B, and G_C satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C) \triangle (X \cup Y \cup Z)| \leq p^{10^{18}} n / \log(n)^{10^{18}}$
- *One of the following identities holds in the abelianization of G.*

  C. $\sum X + \sum Y = \sum Z$

  O. $\sum Y - \sum X = \sum Z$
- *If $G = (\mathbb{Z}_2)^k$ for some k, then $e \notin Z$.*

*Then, if **C** holds, (X, Y, Z) admits a complete mapping, and if **O** holds, (X, Y, Z) admits an orthomorphism.*

Proof. Observe that $R^1, R^2, (R^3)^{-1} \subseteq G$ are each p-random subsets, and $(R^3)^{-1}$ is also a random subset sampled independently with R^1 and R^2. So Theorem 2.3.6 applies to $R^1, R^2, (R^3)^{-1}$ with high probability. Also, the random sets $(R^1)^{-1}, R^2, (R^3)^{-1}$ have the property that the inverse of the first random set is sampled disjointly with the second, and the third is sampled independently with the first two. So Theorem 2.3.7 applies to these random subsets.

Suppose both properties above hold, and fix X, Y, Z be given with the listed properties. Suppose **C** holds. Then, X, Y, Z^{-1} satisfy the properties required with respect to $R^1, R^2, (R^3)^{-1}$ so $H_G[X, Y, Z^{-1}]$ has a perfect matching by Theorem 2.3.6, implying that (X, Y, Z) has a complete mapping by Observation2.3.8. Suppose instead that **C** holds. Then, X^{-1}, Y, Z^{-1} satisfy the properties required with respect to $(R^1)^{-1}, R^2, (R^3)^{-1}$ so $H_G[X^{-1}, Y, Z^{-1}]$ has a perfect matching by Theorem 2.3.7, implying that (X, Y, Z) has an orthomorphism by Observation 2.3.8. □

2.4 Proof of the main theorem

In this section, we prove the main result of the thesis, Theorem 2.3.3. We begin by clarifying that for every (sufficiently large) group G, we fix $a_\phi, b_\phi, c_\phi \in G$ with properties as in Section 2.2.4. Definitions such as ϕ-generic, pair, and coset-paired are with respect to these fixed choices of $a_\phi, b_\phi, c_\phi \in G$ given by Lemma 2.2.14.

2.4.1 Absorbers

In this section we give constructions of absorbers i.e. subsets $R \subseteq V(H_G)$ which can be extended into a matching in several different ways. Lemma 2.4.24 is the main result of this section, and the only result we need for the rest of the paper. The following definition precisely describes what we will be looking for.

Definition 2.4.1. Let $\mathcal{F} = \{S_1, \dots, S_t\}$ be a family of subsets of $V(H)$ for a hypergraph H. We say that a set of vertices R **m-absorbs** $\mathcal{F}$ if for every subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size m, there is a hypergraph matching whose vertex set is exactly $R \cup \bigcup_{S_i \in \mathcal{F}'} S_i$.

It will be convenient to note that when $t = 2$ and $m = 1$, then the above definition is equivalent to "Let X, Y be sets of vertices in a hypergraph H. We say that a set of vertices R **1-absorbs** $\{X, Y\}$ if there are hypergraph matchings R^-, R^+ whose vertex sets are exactly $V(R^-) = R \cup X$ and $V(R^+) = R \cup Y$."

The following lemma shows how the parameter h changes when we pass to a subfamily of $\mathcal{F}$.

Lemma 2.4.2. *Let $\mathcal{F}$ be a family of disjoint subsets of $V(H)$ and $\mathcal{F}' \subseteq F$ a subfamily with $|\mathcal{F} \setminus \mathcal{F}'| = t$. If R h-absorbs $\mathcal{F}$ then*

1. *R h-absorbs $\mathcal{F}'$.*

2. *$R \cup \bigcup_{S \in \mathcal{F} \setminus \mathcal{F}'} S$ $(h-t)$-absorbs $\mathcal{F}'$.*

Proof. Part (1) is immediate from the definition of absorbing. For part (2), notice that for any subfamily $\mathcal{F}'' \subseteq \mathcal{F}'$ of size $h - t$, the subfamily $\mathcal{F}'' \cup (\mathcal{F} \setminus \mathcal{F}')$ has size h. Therefore there is a matching with vertex set $R \cup \bigcup_{S \in \mathcal{F}'' \cup (\mathcal{F} \setminus \mathcal{F}')} S = (R \cup \bigcup_{S \in \mathcal{F} \setminus \mathcal{F}'} S) \cup \bigcup_{S \in \mathcal{F}''} S$. □

We build larger absorbers from smaller ones. The following lemma allows us to take unions of 1-absorbers to get another 1-absorber.

Lemma 2.4.3. *Suppose that $\{R_1, \dots, R_t\}, \{X_1, \dots, X_t\}, \{Y_1, \dots, Y_t\}$ are three families of disjoint sets with $\bigcup R_i$ disjoint from $\bigcup (X_i \cup Y_i)$. Suppose that R_i 1-absorbs $\{X_i, Y_i\}$ for $i = 1, \dots, t$. Set $Z = (\bigcup X_i) \cap (\bigcup Y_i)$. Then, $\bigcup R_i \cup Z$ 1-absorbs $\{\bigcup X_i \setminus Z, \bigcup Y_i \setminus Z\}$.*

Proof. By the definition of 1-absorbs we have matchings R_i^- and R_i^+ with vertices $V(R_i^-) = R_i \cup X_i$ and $V(R_i^+) = R_i \cup Y_i$. Note that $\bigcup R_i^-$ and $\bigcup R_i^+$ are matchings (since $\{R_1, \dots, R_t\}, \{X_1, \dots, X_t\}, \{Y_1, \dots, Y_t\}$ are families of disjoint sets with $\bigcup R_i$ disjoint from $\bigcup (X_i \cup Y_i)$). Also $V(\bigcup R_i^-) = \bigcup R_i \cup \bigcup X_i = (\bigcup R_i \cup Z) \cup (\bigcup X_i \setminus Z)$ and $V(\bigcup R_i^+) = \bigcup R_i \cup \bigcup Y_i = (\bigcup R_i \cup Z) \cup (\bigcup Y_i \setminus Z)$. Thus $\bigcup R_i^-$ and $\bigcup R_i^+$ are matchings satisfying the definition of "$\bigcup R_i \cup Z$ 1-absorbs $\{\bigcup X_i \setminus Z, \bigcup Y_i \setminus Z\}$". □

We state a technical consequence of the above lemma for later use.

Lemma 2.4.4. *Suppose that we have distinct vertices $a, b, c, d \in V(H_G)$, vertices $w_1, \dots, w_k \in V(H_G) \setminus \{a, b, c, d\}$ and disjoint sets $R_0, R_1, \dots, R_{k-1} \subseteq V(H_G) \setminus \{w_1, \dots, w_k, a, b, c, d\}$ with R_0 1-absorbing $\{\{a, w_1, b\}, \{a, w_k, b\}\}$ and R_i 1-absorbing $\{w_i, w_{i+1}\}$ for $i = 1, \dots, k-1$. Then, there is a subset $R' \subseteq \bigcup_{i=0}^{k-1} R_i \cup \{w_1, \dots, w_k\}$ which 1-absorbs $\{\{a, b\}, \{c, d\}\}$.*

Proof. Without loss of generality, we can assume that $w_1, \dots, w_k$ are distinct (by passing to a subset of $\{w_1, \dots, w_k\}$ of distinct vertices, and a corresponding subfamily of $\{R_1, \dots, R_{k-1}\}$). Now, the lemma follows from Lemma 2.4.3 with $\{R_0, \dots, R_k\}$,

$$\{X_0, \dots, X_k\} := \{\{a, w_1, b\}, \{w_2\}, \{w_3\}, \dots, \{w_k\}\},$$

and

$$\{Y_1, \dots, Y_t\} := \{\{c, w_k, d\}, \{w_1\}, \{w_2\}, \dots, \{w_{k-1}\}\}.$$

□

Constructing 1-absorbing sets

The following lemma shows that there are many edges through generic vertices in H_G.

Lemma 2.4.5. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For any generic $v \in V(H_G)$ and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4000}$, there is an edge e of H_G passing through v and having the other two vertices in $R \setminus U$.

Proof. With high probability, Lemma 2.2.32 applies. We'll just look at the case $v \in G_A$, the other two cases are symmetric. Fix some generic $v \in G_A$. Thinking of b as a free variable, consider the set $S = \{b, b^{-1}v^{-1}\} \subseteq G * F_1$. Note that this is a set of linear elements with $w = b, w' = b^{-1}v^{-1}$ separable (by (b) since v^{-1} is generic). Lemma 2.2.32 now implies what we want. □

The following is the basic building block for the absorbers that we construct. It shows that we can 1-absorb sets of the form $\{[a, b]c, c\}$.

Lemma 2.4.6. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For any $a, b, c \in G_A$ with $c, c^{-1}bab^{-1}a^{-1}, c^{-1}bab^{-1}, c^{-1}ba, c^{-1}b$ generic and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4010}$, there is a set $R' \subseteq R \backslash U$ of size ≤ 14 which 1-absorbs $\{[a, b]c, c\}$.

Proof. With high probability, the properties of Lemmas 2.2.32 and 2.4.5 hold. Fix some $a, b, c \in G_A$ with

$$c, c^{-1}bab^{-1}a^{-1}, c^{-1}bab^{-1}, c^{-1}ba, c^{-1}b$$

all being generic and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4010}$. First suppose that $[a, b] = e$ i.e. that $[a, b]c = c$. From Lemma 2.4.5 we know that there is an edge $e = \{c, s, t\}$ of H_G with $s, t \in R \setminus U$. Then $R' = \{s, t\}$ satisfies the lemma (with both M^- and M^+ for the definition of "1-absorbs" equal to the single edge e).

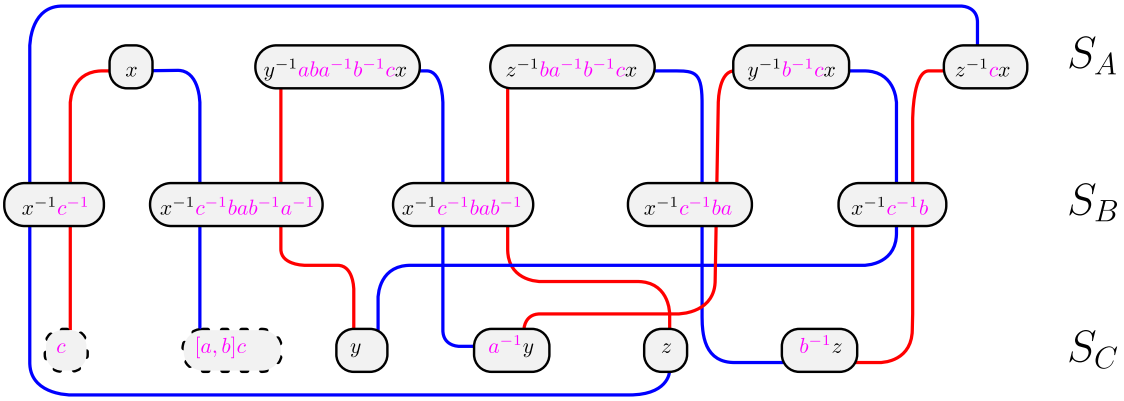


Figure 2.2: The set $S \subseteq G * F_3$ in Lemma 2.4.6. Black letters x, y, z are free variables, while pink letters are elements of G. The two elements $[a, b]c$, c are not part of S (and are just pictured to show how S 1-absorbs $\{[a, b]c, c\}$).

Now suppose $[a, b] \neq e$. Notice that this implies that $a, b, bab^{-1}a^{-1}, bab^{-1}, ba, b^{-1}a^{-1}, ab^{-1}a^{-1}, ab^{-1} \neq e$ also. Thinking of x, y, z as free variables in $G * F_3$, consider the set S given in Figure 2.2, with partition $S = S_A \cup S_B \cup S_C$. Notice that all words in S are linear, all pairs of words weakly-separable, and S_A, S_B, S_C are pairwise strongly separable (see Figure 2.3 for justification). Using Lemma 2.2.32, there is some projection $\pi : G * F_3 \to G$ which separates S and has $\pi(S) \subseteq R \setminus U$. Any such $\pi(S)$ 1-absorbs $\{[a, b]c, c\}$ (using the red/blue matchings in Figure 2.2). □

The following lemma is identical to the previous one, except that it weakens the assumption on what elements are generic.

S_x :

	x	$x^{-1}c^{-1}$	$x^{-1}c^{-1}bab^{-1}a^{-1}$	$x^{-1}c^{-1}bab^{-1}$	$x^{-1}c^{-1}ba$	$x^{-1}c^{-1}b$
x		c^{-1}	$c^{-1}bab^{-1}a^{-1}$	$c^{-1}bab^{-1}$	$c^{-1}ba$	$c^{-1}b$
$x^{-1}c^{-1}$	c^{-1}		$bab^{-1}a^{-1}$	bab^{-1}	ba	b
$x^{-1}c^{-1}bab^{-1}a^{-1}$	$c^{-1}bab^{-1}a^{-1}$	$bab^{-1}a^{-1}$		a^{-1}	$b^{-1}a^{-1}$	$ab^{-1}a^{-1}$
$x^{-1}c^{-1}bab^{-1}$	$c^{-1}bab^{-1}$	bab^{-1}	a^{-1}		b^{-1}	ab^{-1}
$x^{-1}c^{-1}ba$	$c^{-1}ba$	ba	$b^{-1}a^{-1}$	b^{-1}		a
$x^{-1}c^{-1}b$	$c^{-1}b$	b	$ab^{-1}a^{-1}$	ab^{-1}	a	

$S_{x,y}$:

	$y^{-1}aba^{-1}b^{-1}cx$	$y^{-1}b^{-1}cx$
$y^{-1}aba^{-1}b^{-1}cx$		aba^{-1}
$y^{-1}b^{-1}cx$	aba^{-1}	

S_y :

	y	$a^{-1}y$
y		a^{-1}
$a^{-1}y$	a^{-1}	

$S_{x,z}$:

	$z^{-1}ba^{-1}b^{-1}cx$	$z^{-1}cx$
$z^{-1}ba^{-1}b^{-1}cx$		$ba^{-1}b^{-1}$
$z^{-1}cx$	$ba^{-1}b^{-1}$	

S_z :

	z	$b^{-1}z$
z		b^{-1}
$b^{-1}z$	b^{-1}	

Figure 2.3: Proofs for weak/strong separability and linearity of all pairs $w, w' \in S$ in Lemma 2.4.6 to justify the application of Lemma 2.2.32. For separability, first we have partitioned S into five subsets $S = S_x \cup S_{x,y} \cup S_{x,z} \cup S_z \cup S_y$ based on which free variables appear in each $w \in S$ (as in Observation 2.2.24). By Observation 2.2.24, any w, w' in different subsets are strongly separable by part (a) of the definition. For each of the sets $S_x, S_{x,y}, S_{x,z}, S_z, S_y$ we give a table explaining why the w, w' in that set are strongly/weakly-separable. Note that for words coming from different $S_A/S_B/S_C$ we need to show strong separability, but for words coming from the same part, weak separability suffices. Blue cells represent w, w' being separable via part (b) of the definition of "separable", green cells represent w, w' being weakly separable via part (b'), and grey cells represent w, w' not being separable/weakly-separable. The group element inside each blue cell is a generic element g so that $w' \in \{gw, g^{-1}w, gw^{-1}, g^{-1}w^{-1}, wg, wg^{-1}, w^{-1}g, w^{-1}g^{-1}\}$ (thus checking (b) for w, w'). The group element inside the green cells is a non-identity element g so that $w = w'$ rearranges into $e = g$ (thus checking (b') for w, w'). Observe that green cells are used only between pairs of words coming from the same part of $S_A/S_B/S_C$, meaning that we have strong separation for pairs of words coming from different parts $S_A/S_B/S_C$, as needed. To see that every $s \in S$ is linear notice that every word pictured has no repetitions of black letters.

Lemma 2.4.7. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For any $a, b, x \in G_B$ with $x, x[a,b]$ generic and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4020}$, there is a set $R' \subseteq R \setminus U$ of size ≤ 16 which 1-absorbs $\{x[a,b], x\}$.

Proof. With high probability, the properties of Lemmas 2.2.32, 2.4.5 and 2.4.6 hold. Fix some $a, b, x \in G_B$ with $x, x[a,b]$ generic and $U \subseteq V(H_G)$ with

$|U| \leq p^{800}n/10^{4020}$. As in Lemma 2.4.6, the conclusion trivially follows from Lemma 2.4.5 if $[a,b]=e$, so assume that this doesn't happen. With v the free variable in $G * F_1$, consider the sets of words $T := \{v, [b,a]x^{-1}v^{-1}, x^{-1}v^{-1}\}$ and $S = \{x^{-1}v^{-1}, vxaba^{-1}b^{-1}, vxaba^{-1}, vxab, vxa\}$. Define $T_A = \{v\}, T_B = \{[b,a]x^{-1}v^{-1}, x^{-1}v^{-1}\}$, $T_C = \emptyset$ to get a partition of T. It is easy to check that all $w \in T \cup S$ are linear (since v appears precisely once in each $w \in T \cup S$), that T_A, T_B separable (by part (b) of the definition, using that $x, x[a,b]$ are generic), and that $[b,a]x^{-1}v^{-1}, x^{-1}v^{-1}$ are weakly separable (since $[b,a]x^{-1}v^{-1} = x^{-1}v^{-1}$ rearranges into $[b,a]=e$). By Lemma 2.2.32, there is a projection π which separates $T \cup S$ and has $\pi(T \cup S) \subseteq R \setminus (U \cup N(G))$. Since all pairs of words in T are separable, we have that the vertices in $\pi(T)$ are distinct (and so $e^- = (\pi(v), x[a,b], [b,a]x^{-1}\pi(v)^{-1})$ and $e^+ = (\pi(v), x, x^{-1}\pi(v)^{-1})$ are edges of H_G contained in $R \cup \{x, x[a,b]\}$).

Using Lemma 2.4.6 with $a' = b, b' = a, c = x^{-1}\pi(v^{-1})$ we find a set Q disjoint from $U \cup \pi(T)$ which 1-absorbs $\{[b,a]x^{-1}\pi(v)^{-1}, x^{-1}\pi(v)^{-1}\}$ (all the required elements are generic for that lemma as a consequence of $\pi(S) \cap N(G) = \emptyset$). Now $Q \cup \pi(T)$ 1-absorbs $\{x[a,b], x\}$. This can be seen directly or by first noticing that $\{\pi(v)\}$ 1-absorbs $\{\{x[a,b], [b,a]x^{-1}\pi(v)^{-1}\}, \{x, x^{-1}\pi(v)^{-1}\}\}$ (the single-edge matchings e^-, e^+ witness this). Then Lemma 2.4.3 shows that $Q \cup \pi(T)$ 1-absorbs $\{x[a,b], x\}$. □

For two sets $X, Y \subseteq G$ and $s \in G$, use XsY to denote $\{xsy : x \in X, y \in Y\}$. The following lemma lets us 1-absorb a pair of size 3 sets.

Lemma 2.4.8. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For any distinct, generic $a, b, c, d \in G_B$ and $X, Y, U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4000}, |X|, |Y| \leq 5$, there is a $R' \subseteq R \setminus U$ of size 6 and s with $XsY \subseteq R \setminus (U \cup R')$ such that R' 1-absorbs $\{\{a, dsc, b\}, \{c, bsa, d\}\}$.

Proof. With high probability, the property of Lemma 2.2.32 holds. Suppose we have generic $a, b, c, d \in G_B$ and $X, Y, U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4000}$, $|X|, |Y| \leq 5$. Let $X' = X \cup \{b, d\}, Y' = Y \cup \{a, c\}$. For free variables u, w, consider the sets of words $S = \{u, a^{-1}u^{-1}, c^{-1}u^{-1}, w, b^{-1}w^{-1}, d^{-1}w^{-1}\}$ with $S_A = \{u, b^{-1}w^{-1}, d^{-1}w^{-1}\}, S_B = \emptyset, S_C = \{w, a^{-1}u^{-1}, c^{-1}u^{-1}\}$. Notice that all words in $S \cup (X'wuY')$ are linear, all pairs of words in S are weakly separable (see Figure 2.5), and S_A, S_B, S_C are pairwise strongly separable. Also the pair of sets $(S, X'wuY')$ is strongly separable (using part (a) of the definition of separable). By Lemma 2.2.32, there is some projection π separating $S \cup (X'wuY')$ and having $S \cup (X'wuY') \subseteq R \setminus U$. For any such projection, $R' = \pi(S)$ and $s := \pi(wu)$ satisfy the lemma (with the red/blue matchings in Figure 2.4 satisfying the definition of R' 1-absorbing $\{\{a, dsc, b\}, \{c, bsa, d\}\}$. Note that this works regardless of whether $dwuc = bwua$ or not). □

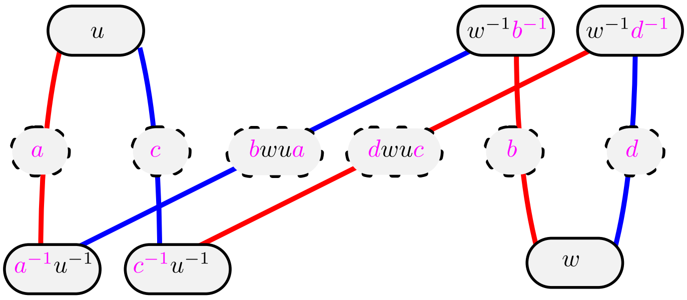


Figure 2.4: The set of words $S \subseteq G * F_2$ for Lemma 2.4.8. First, second, and third row of words represent S_A, S_B, and S_C, respectively. Black letters u, w are free variables of F_2, while pink letters are elements of G. The elements $a, b, c, d, bwua, dwuc$ are not part of S (and are pictured just to demonstrate how the absorption works).

The following technical lemma allows us to 1-absorb certain pairs of sets of size 2.

Lemma 2.4.9. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For any $g, x, y, a, b \in G_B$ with $yg, x, y, x[a, b]g$ distinct, generic and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4030}$, there is a set $R' \subseteq R \setminus U$ of size ≤ 60 which 1-absorbs $\{\{yg, x\}, \{y, x[a, b]g\}\}$.

S_u :

	u	$a^{-1}u^{-1}$	$c^{-1}u^{-1}$
u		a^{-1}	c^{-1}
$a^{-1}u^{-1}$	a^{-1}		ac^{-1}
$c^{-1}u^{-1}$	c^{-1}	ac^{-1}	

S_w :

	w	$w^{-1}b^{-1}$	$w^{-1}d^{-1}$
w		b^{-1}	d^{-1}
$w^{-1}b^{-1}$	b^{-1}		bd^{-1}
$w^{-1}d^{-1}$	d^{-1}	bd^{-1}	

Figure 2.5: Justification for linearity and separability of pairs (w, w') in S in Lemma 2.4.8. For separability, first split S into $S_u = \{u, a^{-1}u^{-1}, c^{-1}u^{-1}\}$, $S_w = \{w, b^{-1}w^{-1}, d^{-1}w^{-1}\}$ based on which free variables appear in the elements (as in Observation 2.2.24). By Observation 2.2.24, pairs (w, w') with w, w' in different sets S_u/S_w fall under part (a) of the definition of strongly separable, so it remains to check pairs inside S_u and S_w. Justification for this is given in the two tables above (with the same conventions as in Figure 2.3, in particular, green cells are used only between pairs of vertices coming from the same part $S_A/S_B/S_C$). Relevant elements are generic/non-identity as a consequence of a, b, c, d being distinct and generic). To see that each $w \in S$ is linear, note that there are no repetitions of black letters in each w (and each $w \in S$ contains at least one black letter).

Proof. With high probability, the properties of Lemmas 2.4.7 and 2.4.8 hold. Let g, x, y, a, b and U be as in the lemma. Fix $a' = yg$, $b' = x$, $c' = y$, $d' = x[a, b]g$ and note these are distinct and generic by assumption. Define $X = \{x, x[a, b]g, x[a, b], x[a, b]yg, e\}$, $Y = \{yg, y, g[a, b], yg[a, b], e\}$. By Lemma 2.4.8, we get R' and s with $R_0 \cup (XsY) \subseteq V(H_G) \setminus (U \cup N(G) \cup \{a', b', c', d'\})$ with R_0 1-absorbing $\{\{a', d'sc', b'\}, \{c', b'sa', d'\}\}$.

Define $w_1 = d'sc' = x[a, b]gsy = x[a, b]ygs[y, gs]$, $w_2 = x[a, b]ygs = x[a, b]syg[s, yg]$, $w_3 = x[a, b]syg = xsyg[a, b][syg, [a, b]]$, $w_4 = xsyg[a, b]$, and $w_4 = b'sa' = xsyg$, noting that all of these are in XsW and so are generic and disjoint from $R_0 \cup \{a', b', c', d'\}$. So we can use Lemma 2.4.7 to find disjoint sets $R_1, R_2, R_3, R_4 \subseteq R \setminus (R' \cup U)$ with R_i 1-absorbing $\{w_i, w_{i+1}\}$. We also have R_0 1-absorbing $\{\{a', w_1, b'\}, \{c', w_5, d'\}\}$. By Lemma 2.4.4, there is a subset $R' \subseteq \bigcup_{i=0}^{4} R_i \cup \{w_1, w_2, w_3, w_4, w_5\}$ which 1-absorbs $\{\{a', b'\}, \{c', d'\}\}$. □

The following lemma gives us a collection of distinct, generic elements.

Lemma 2.4.10. *Let $p \geq n^{-1/700}$ and $t \leq p^{800}n/10^{4020}$. Let G be a group and R a symmetric p-random subset. With high probability, the following holds:*

Let $g_1, \dots, g_t \neq e$ be distinct and $U \subseteq G$ with $|U| \leq p^{800}n/10^{4020}$. There are $y_1, \dots, y_{t-1} \in G$ such that the elements $y_1, y_2, \dots, y_{t-1}, y_1g_2, y_2g_3, \dots, y_{t-1}g_t$ are *distinct, generic elements in $R \setminus U$.*

Proof. Lemma 2.2.32 applies with $R_A = R$, and R_B, R_C arbitrary disjoint symmetric p-random subsets (which won't be used in the proof). Let $g_1, \dots, g_t \neq e$ be distinct and $U \subseteq G$ with $|U| \leq p^{800}n/10^{4020}$. Thinking of y as the free variable in $G * F_1$, let $S_i = \{y, yg_i\}$. Note all $w, w' \in S_i$ are linear and strongly separable (by part (b) since g_i is generic). For $i = 1, \dots, t$, use Lemma 2.2.32 to pick projections $\pi_1, \dots, \pi_i$ such that $\pi_i(S_i)$ is separated and disjoint from $U' := U \cup N(G) \cup \bigcup_{j<i} \pi_j(S_j)$ (noting that $|U'| \leq |U| + |N(G)| + 2t \leq 3p^{800}n/10^{4020} + 10^{9000} \leq n/10^{4010}$). Now $y_1 = \pi_1(y), \dots, y_{t-1} = \pi_{t-1}(y)$ satisfy the lemma. □

The following theorem of Gallagher is key to our approach and shows that elements of the commutator subgroup can be written as a product of a small number of commutators. Its proof uses character theory.

Theorem 2.4.11 (Gallagher, [27])**.** *Let G be a group. Any $g \in G'$ can be written as $g = \prod_{i=1}^{t} [a_i, b_i]$ for some $a_i, b_i \in G$ and $t \leq \log_4 |G'| \leq 10 \log n$.*

We now prove one of the main lemmas in this section. It strengthens several earlier lemmas and shows that we can 1-absorb any pair of elements as long as they are in the same coset of G'.

Lemma 2.4.12. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R_A^1 \cup R_B^2 \cup R_C^3$. With high probability, the following holds:*

For any generic $h, k \in G_A, G_B$, or G_C with $[h] = [k]$ and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4040}$, there is a set $R' \subseteq R \setminus U$ of size $\leq 400 \log n$ which 1-absorbs $\{h, k\}$.

Proof. With high probability, the properties of Lemmas 2.4.5, 2.4.7, 2.4.9, and 2.4.10 hold. We just prove the lemma for $h, k \in G_B$, the other two cases follow by symmetry. Let $h, k \in G_B$ be generic with $[h] = [k]$ and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4040}$. As in Lemma 2.4.6, the conclusion trivially follows from Lemma 2.4.5 if $h = k$, so assume $h \neq k$. Write $h = kg$ for some $g \in G'$. By Theorem 2.4.11, we have $g = \prod_{i=1}^{t}[a_i, b_i]$ for some $a_i, b_i \in G$ and $t \leq 10 \log n$. Assume this product is as short as possible i.e. that t is minimal. For each $s = 0, \dots, t-1$, define $g_s = \prod_{i=s}^{t}[a_i, b_i]$. By minimality of t, we have $[a_i, b_i] \neq e$ and $g_i \neq e$ for all i. Set $y_0 = k$ and note that $y_0, y_0 g_1$ are distinct, generic (since k, h are distinct, generic).

Note that $t \leq 10 \log n \leq n^{1/8}/10^{4010} \leq p^{800} n/10^{4010}$ (using that n is large which follows from "with high probability"). Use Lemma 2.4.10 with $U' = U \cup \{y_0, y_0 g_1\}$ to get elements $y_1, \dots, y_{t-1}$, noting that now

$$y_0, y_1, \dots, y_{t-1}, y_0 g_1, y_1 g_2, \dots, y_{t-1} g_t$$

are distinct, generic. $U' = U \cup \{y_i, y_i g_{i+1} : i = 1, \dots, t-1\}$. Using Lemma 2.4.9 with $y = y_i$, $x = y_{i-1}$, $g = g_{i+1}, a = a_i, b = b_i$ for all i (the conditions "$y_i g_{i+1}, y_{i-1}, y_i, y_{i-1}[a_i, b_i] g_{i+1}$ distinct, generic" coming from Lemma 2.4.10), gives a set R_i which 1-absorbs $\{\{y_i g_{i+1}, y_{i-1}\}, \{y_i, y_{i-1}[a_i, b_i] g_{i+1}\}\} = \{\{y_i g_{i+1}, y_{i-1}\}, \{y_i, y_{i-1} g_i\}\}$. Using Lemma 2.4.7 with $x = y_{t-1}$, $a = a_t, b = b_t$ (condition "$y_{t-1}, y_{t-1}[a_t, b_t]$ generic" coming from Lemma 2.4.10), gives a set R_t which 1-absorbs $\{y_{t-1}[a_t, b_t], y_{t-1}\} = \{y_{t-1} g_t, y_{t-1}\}$. By enlarging the set U during the application of these lemmas, we can assume that $R_1, \dots, R_{t-1}, R_t, \{y_i, y_i g_{i+1} : i = 1, \dots, t-1\}$ are all disjoint (there's space to do this because $|R_1 \cup \dots \cup R_{i-1}| \leq 400 \log n < 10^{800} n^{1/8}/10^{4020} \leq p^{800} n/10^{4030}$).

Set $R = R_1 \cup \dots \cup R_{t-1} \cup R_t \cup \{y_i, y_i g_{i+1} : i = 1, \dots, t-1\}$. Set $X_i = \{y_i g_{i+1}, y_{i-1}\}$, $Y_i = \{y_i, y_{i-1} g_i\}$ for $i = 1, \dots, t-1$, and $X_t = \{y_{t-1}\}$, $Y_t = \{y_{t-1} g_t\}$. Notice that the sets in the families $\{R_1, \dots, R_{t-1}\}$, $\{X_1, \dots, X_{t-1}\}$, $\{Y_1, \dots, Y_{t-1}\}$ are disjoint, that $\bigcup R_i$ is disjoint from $\bigcup X_i, \bigcup Y_i$ and that $(\bigcup X_i) \cap (\bigcup Y_i) = \{y_i, y_i g_{i+1} : i = 1, \dots, t-1\}$. By Lemma 2.4.3 R 1-absorbs $\{y_0 g_1, y_0\} = \{kg, k\} = \{h, k\}$. □

The following lemma is similar to the previous one, except it allows us to 1-absorb pairs of sets of size 2.

Lemma 2.4.13. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For any distinct, generic $a, b, c, d \in G_A, G_B$, or G_C with $[ab] = [cd]$ and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4050}$, there is a set $R' \subseteq R \setminus U$ of size $\leq 500 \log n$ which 1-absorbs $\{\{a, b\}, \{c, d\}\}$.

Proof. With high probability, the properties of Lemmas 2.4.8 and 2.4.12 hold. We just prove the lemma for $a, b, c, d \in G_B$. The other two cases follow by symmetry. Suppose that we have distinct, generic $a, b, c, d \in B(H_G)$ with $[ab] = [cd]$, and $U \subseteq G$ with $|U| \leq p^{800}n/10^{21}$. Let $X = \{d, b\}, Y = \{c, a\}$. Using Lemma 2.4.8, pick some R', s with $R' \cup XsY \subseteq R \setminus (U \cup N(G))$ and R' 1-absorbing $\{\{a, dsc, b\}, \{c, bsa, d\}\}$. When $dsc = bsa$, $R' \cup dsc$ 1-absorbs $\{\{a, b\}, \{c, d\}\}$ and so satisfies the lemma. So suppose $dsc \neq bsa$. Notice that $[ab] = [cd]$ implies $[bsa] = [dsc]$ and so using Lemma 2.4.12 we can choose a $Q \subseteq V(H_G) \setminus (U \cup R' \cup \{dsc, bsa\})$ which 1-absorbs $\{bsa, dsc\}$ (bsa, dsc are generic because they are contained in XsC). Set $R'' = R' \cup Q \cup \{bsa, dsc\}$ to get a set which 1-absorbs $\{\{a, b\}, \{c, d\}\}$ by Lemma 2.4.3. □

Distributive absorption for pairs

In this section we prove a variety of lemmas about h-absorbing sets of the form $\{\{a_1, b_1\}, \dots, \{a_k, b_k\}\}$ where $[a_1b_1] = \dots = [a_kb_k]$. The following does this for $h = k - 1$.

Lemma 2.4.14. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $k \leq 200$ and $x_1, y_1, \dots, x_k, y_k \in G_A, G_B$ or G_C be distinct, generic with $[x_1y_1] = \dots = [x_ky_k]$ and let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4060}$. Then there is a set $R' \subseteq R \setminus U$ of size $\leq 600k \log n$ which $(k-1)$-absorbs $\{\{x_1, y_1\}, \dots, \{x_k, y_k\}\}$.

Proof. With high probability, the property of Lemma 2.4.13 holds. Let $x_1, y_1, \dots, x_k, y_k$ be distinct, generic with $[x_1y_1] = \dots = [x_ky_k]$ and let $U \subseteq G$ with

$|U| \leq p^{800}n/10^{4060}$. Apply the property of Lemma 2.4.12 to get disjoint sets $R_1, \dots, R_{k-1} \subseteq R \setminus U$ which 1-absorb $\{\{x_i, y_i\}, \{x_{i+1}, y_{i+1}\}\}$ for each $i \in [k-1]$ (for disjointness use $U' = U \cup \bigcup_{j=1}^{i-1} R_j$ at each application. This set has size $\leq |U| + k500 \log n \leq p^{800}n/10^{4060}$). Set $R' = R_1 \cup \dots \cup R_{k-1}$. For each i, we have matchings R_i^- and R_i^+ with vertex sets $R_i \cup \{x_i, y_i\}$ and $R_i \cup \{x_{i+1}, y_{i+1}\}$. Now the matchings $M_i = (\bigcup_{j=1}^{i-1} R_i^-) \cup (\bigcup_{j=i}^{|Y|-1} R_i^+)$ have vertex sets exactly $V(M_i) = R' \cup Y \setminus \{x_i, y_i\}$, and so they satisfy the definition of R' $(k-1)$-absorbing Y. □

We'll need the following lemma which finds a common neighbour of a set of vertices in H_k.

Lemma 2.4.15. *Let $p \geq n^{-1/700}$ and let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R_A^1 \cup R_B^2 \cup R_C^3$. With high probability, the following holds:*

Let $k \leq 200$ and $a_1, \dots, a_k \in G_A$ be distinct and generic and let $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4001}$. Then there are distinct, generic $b, c_1, \dots, c_k \in G \setminus U$ such that for each i, $\{a_i, b, c_i\}$ is an edge of H_G.

Proof. With b the free variable in $G * F_1$, consider the set $S = \{b, b^{-1}a_1^{-1}, \dots, b^{-1}a_k^{-1}\}$, $S_A = \emptyset$, $S_B = \{b\}$, $S_C = \{b^{-1}a_1^{-1}, \dots, b^{-1}a_k^{-1}\}$. Notice that all $w \in S$ are linear, that S_A, S_B, S_C are pairwise separable (by part (b) of the definition since $a_1, \dots, a_k$ are generic), and all $w, w' \in S_C$ are weakly separable (by (b') since equalities between w, w' rearrange to $e = a_i a_j^{-1}$ and we know $a_i a_j^{-1} \neq e$ by distinctness). Thus the lemma follows from Lemma 2.2.32 applied with $U' = U \cup N(G)$. □

The following lemma 1-absorbs a set of k pairs.

Lemma 2.4.16. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R_A^1 \cup R_B^2 \cup R_C^3$. With high probability, the following holds:*

Let $k \leq 200$ and $x_1, y_1, \dots, x_k, y_k \in G_A$ be distinct, generic with $[x_1 y_1] = \dots = [x_k y_k]$ and let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4070}$. Then there is a set $R' \subseteq R$ of size

$\leq 600k \log n$ which 1-absorbs $\{\{x_1, y_1\}, \dots, \{x_k, y_k\}\}$.

Proof. With high probability, the conclusions of Lemmas 2.4.15 and 2.4.14 apply. Let $x_1, y_1, \dots, x_k, y_k \in G_A$ be distinct, generic with $[x_1y_1] = \dots = [x_ky_k]$ and let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4070}$. Use the conclusion of Lemma 2.4.15 twice to get elements b^x, b^y and $c_1^x, c_1^y, \dots, c_k^x, c_k^y$ outside U, such that $x_1, y_1, \dots, x_k, y_k$, b^x, b^y, $c_1^x, c_1^y, \dots, c_k^x, c_k^y$ are all distinct, generic and also $x_i b^x c_i^x, y_i b^y c_i^y$ are edges for all i (to get distinctness of all the vertices, enlarge U to include previously found vertices between the two applications). Notice that this implies that $[c_i^x c_i^y] = [x_i^{-1} b_x^{-1} y_i^{-1} b_y^{-1}] = [x_1^{-1} b_x^{-1} y_1^{-1} b_y^{-1}]$ for all i (using $[x_1y_1] = \dots = [x_ky_k]$). Thus using the conclusion of Lemma 2.4.14, we get a set R which $(k-1)$-absorbs $\{\{c_1^x, c_1^y\}, \dots, \{c_k^x, c_k^y\}\}$. In other words we have matchings $M_1, \dots, M_k$ with $V(M_i) = R \cup \{\{c_j^x, c_j^y\} : j \neq i\}$. Set $R'' = R' \cup \{b^x, b^y, c_1^x, c_1^y \dots, c_k^x, c_k^y\}$. Now for each i, $M_i \cup \{x_i b^x c_i^x, y_i b^y c_i^y\}$ is a matching with vertex set exactly $R'' \cup \{x_i, y_i\}$, verifying the definition of "1-absorbs". □

The below was shown by Montgomery in [45], and is the essence of the distributive absorption approach.

Lemma 2.4.17 (Montgomery, [45])**.** *There is a constant h_0 such that for every $h \geq h_0$ there exists a bipartite graph K with maximum degree at most 100 and vertex classes X and $Y \cup Y'$ with $|X| = 3h$, $|Y| = |Y'| = 2h$, so that the following holds. For any $Y_0 \subseteq Y'$ with $|Y_0| = h$, there is a perfect matching between X and $Y \cup Y_0$.*

Graphs produced by this lemma are called **robustly matchable bipartite graphs**. Combining this lemma with the previous one, we can h-absorb sets of pairs.

Lemma 2.4.18. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Consider sets Y, Y' of disjoint pairs of generic elements $(a_1, a_2) \in (A \setminus e) \times (A \setminus e)$ having $[a_1 a_2] = [a_1' a_2']$ for $(a_1, a_2), (a_1', a_2') \in Y, Y'$ and also $2h = |Y| = |Y|' \leq \frac{p^{800} n}{10^{4080} \log n}$. Let $U \subseteq G$ with $|U| \leq p^{800} n / 10^{4080}$. Then there is a subset $R' \subseteq R \setminus U$ of size $\leq 800|Y| \log n$ such that $R' \cup Y'$ h-absorbs Y.

Proof. With high probability the property of Lemma 2.4.16 holds.

Let Y, Y' be disjoint sets of pairs of elements $(a_1, a_2) \in A \times A$ having $[a_1 a_2] = [a_1' a_2']$ for $(a_1, a_2), (a_1', a_2') \in Y$ and also $2h := |Y| = |Y|' \leq \frac{p^{800} n}{10^{4080} \log n}$. Let $U \subseteq G$ with $|U| \leq p^{800} n / 10^{4080}$. Consider a robustly matchable bipartite graph D with $\Delta(D) \leq 100$ whose sets are Y, Y' and X as in Lemma 2.4.17 (here X is just an abstract set of size $3h$ unrelated to the hypergraph we have). For all $x \in X$ use Lemma 2.4.16 to pick a set $R_x \subseteq R \setminus (U \cup Y \cup Y')$ which 1-absorbs $N(x)$. We can choose all these sets to be disjoint by enlarging U to contain the union of previously picked sets at each application (whose total size is at most $100|X| \times 700 \log n \leq p^{800} n / 10^{4070}$).

Letting $R' \bigcup_{x \in X} R_x$ we claim that $R' \cup Y'$ will h-absorb Y, and so satisfy the lemma. Let $Y_0 \subseteq Y$ be a set of $|Y|/2 = h$ pairs. Since D is robustly matchable, there is a matching M with vertex set $Y_0 \cup Y' \cup X$. For each $x \in X$, let xy_x be the matching edge of M through x, and let N_x be a matching with vertex set $R_x \cup y_x$ (which exists because R_x 1-absorbs $N(x)$). Now $\bigcup_{x \in X} N_x$ is a matching with vertex set $Y_0 \cup R' \cup Y'$. □

The following is a version of the previous lemma with more versatility with choosing h.

Lemma 2.4.19. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Consider sets Y, Z of disjoint pairs of generic elements $(a_1, a_2) \in A \times A$ having $[a_1 a_2] = [a_1' a_2']$ for $(a_1, a_2), (a_1', a_2') \in Y, Z$ and $4|Y| \leq |Z| = \frac{p^{800} n}{10^{4090} \log n}$. Let $h \in \mathbb{N}$. Let $U \subseteq G$ with $|U| \leq p^{800} n / 10^{4090}$. Then there is a subset $R' \subseteq R \setminus U, Y' \subseteq Z$ of size $\leq 800|Y| \log n$ such that $R' \cup Y'$ h-absorbs Y.

Proof. With high probability Lemma 2.4.18 applies. If $h > |Y|$, then the definition of "h-absorbs Y" is vacuous, so we can suppose that $h \leq |Y|$. If $|Y|/2 \leq h \leq |Y|$, pick subsets $\hat{Y}, Y' \subseteq Z$ with $|\hat{Y}| = 2h - |Y|$ and $|Y'| = 2h$. Apply Lemma 2.4.18 to g, $Y \cup \hat{Y}$, Y', and h to get a set R' such that $R' \cup Y'$ h-absorbs $Y \cup \hat{Y}$. By Lemma 2.4.2 (1), $R' \cup Y'$ h-absorbs Y also.

If $h \leq |Y|/2$, pick subsets $\hat{Y}, Y' \subseteq Z$ with $|\hat{Y}| = |Y| - 2h$ and $|Y'| = 2|Y| - 2h$. Apply Lemma 2.4.18 to g, $Y \cup \hat{Y}$, Y', and $h' = |Y| - h$ to get a set R' such that $R' \cup Y'$ $(|Y| - h)$-absorbs $Y \cup \hat{Y}$. By Lemma 2.4.2 (2) with $t = |\hat{Y}| = |Y| - 2h$, $R' \cup Y' \cup \hat{Y}$ $(|Y| - h - |\hat{Y}|)$-absorbs Y. This implies the lemma since $|Y| - h - |\hat{Y}| = h$. □

Distributive absorption for singletons

Everything in this section is almost identical to the previous one (though often a bit easier). We give constructions of h-absorbers of sets vof vertices Y contained in a coset of G'. The following lemma does this with $h = |Y| - 1$.

Lemma 2.4.20. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For $g \in G$, let $Y \subseteq G_A, G_B$, or G_C with $Y \subseteq [g]$ be a set of generic elements with $|Y| \leq \frac{p^{800}n}{10^{4100}\log n}$ and let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4100}$. Then there is a set $R' \subseteq R \setminus U$ of size $\leq 700|Y|\log n$ which $(|Y| - 1)$-absorbs Y.

Proof. With high probability, the property of Lemma 2.4.12 holds.

Let $g \in G$ and $Y = \{y_1, \dots, y_{|Y|}\} \subseteq [g]$ with $|Y| \leq \frac{p^{800}n}{10^{4100}\log n}$ and U with $|U| \leq p^{800}n/10^{17}$. Apply the property of Lemma 2.4.12 to get disjoint sets $R_1, \dots, R_{|Y|-1} \subseteq R \setminus U$ which 1-absorb $\{y_i, y_{i+1}\}$ (for disjointness use $U' = U \cup \bigcup_{j=1}^{i-1} R_j$ at each application. This set has size $\leq |U| + |Y|500\log n \leq p^{800}n/10^{4090}$). Set $R' = R_1 \cup \dots \cup R_{|Y|-1}$. For each i, we have matchings R_i^- and R_i^+ with vertex sets $R_i \cup \{y_i\}$ and $R_i \cup \{y_{i+1}\}$. Now the matchings $M_i = (\bigcup_{j=1}^{i-1} R_i^-) \cup (\bigcup_{j=i}^{|Y|-1} R_i^+)$ have vertex sets exactly $V(M_i) = R' \cup Y \setminus \{y_i\}$, and so they satisfy the definition of R' $(|Y| - 1)$-absorbing Y. □

The next lemma builds 1-absorbers of small sets Y.

Lemma 2.4.21. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $k \leq 200$ and $s_1, \dots, s_k \in G_A$ be distinct, generic with $[s_1] = \cdots = [s_k]$ and let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4110}$. Then there is a set $R' \subseteq R \setminus U$ of size $\leq 700k \log n$ which 1-absorbs $\{s_1, \dots, s_k\}$.

Proof. With high probability, the conclusions of Lemmas 2.4.15 and 2.4.20 apply. Let $s_1, \dots, s_k \in G$ be distinct, generic with $[s_1] = \cdots = [s_k]$ and let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4110}$. Use the conclusion of Lemma 2.4.15 to get distinct, generic elements b and $c_1, \dots, c_k \in G \setminus U$, such that also $s_i b c_i$ are edges for all i. Thus using the conclusion of Lemma 2.4.20, we get a set R which $(k-1)$-absorbs $\{c_1, \dots, c_k\}$. From the definition of $(k-1)$-absorbs, we have matchings $M_1, \dots, M_k$ with $V(M_i) = R \cup \{c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_k\}$. Set $R'' = R' \cup \{b, c_1, \dots, c_k\}$. Now for each i, $M_i \cup s_i b c_i$ is a matching with vertex set exactly $R'' \cup \{s_i\}$, verifying the definition of "1-absorbs". □

The next lemma uses distributive absorption to build h-absorbers.

Lemma 2.4.22. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $g \in G$, and consider disjoint sets $Y, Y' \subseteq [g]$ of generic elements of G_A of size $2h := |Y| = |Y'| \leq \frac{p^{800}n}{10^{4120}\log n}$. Let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4120}$. Then there is are subset $R' \subseteq R \setminus U$ of size $\leq 10^5|Y|\log n$ such that $R' \cup Y'$ h-absorbs Y.

Proof. With high probability the property of Lemma 2.4.21 holds. Let $g \in G$, every disjoint sets $Y, Y' \subseteq [g]$ of size $2h = |Y| = |Y'| \leq \frac{p^{800}n}{10^{4120}\log n}$. Let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4120}$. Consider a robustly matchable bipartite graph D with $\Delta(D) \leq 100$ whose sets are Y, Y' and X as in Lemma 2.4.17 (here X is just an abstract set of size $3h$ unrelated to the hypergraph we have). For all $x \in X$ fix a set $R_x \subseteq R$

which 1-absorbs $N(x)$. We can choose all these sets to be disjoint by letting U be the union of previously picked sets at each application (whose total size is at most $100|X| \times 700 \log n \leq p^{800}n/10^{4110}$).

We claim that for $R' := \bigcup_{x \in X} R_x$, we have $R' \cup Y'$ h-absorbing Y, and so satisfying the lemma. Let $Y_0 \subseteq Y$ be a set of $|Y|/2 = h$ pairs. Since D is robustly matchable, there is a matching M with vertex set $Y_0 \cup Y' \cup X$. For each $x \in X$, let xy_x be the matching edge of M through x, and let N_x be a matching with vertex set $R_x \cup y_x$ (which exists because R_x 1-absorbs $N(x)$). Now $\bigcup_{x \in X} N_x$ is a matching with vertex set $Y_0 \cup R' \cup Y'$. □

The next lemma is a version of the previous one which allows for more flexibility in the value of h.

Lemma 2.4.23. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $g \in G$, and consider disjoint sets $Y, Z \subseteq [g]$ of generic elements of G_A with $4|Y| \leq |Z| = \frac{p^{800}n}{10^{4130}\log n}$. Let $h \in \mathbb{N}$. Let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4130}$. Then there is are subset $R' \subseteq R \setminus U, Y' \subseteq Z$ of size $\leq 10^5|Y|\log n$ such that $R' \cup Y'$ h-absorbs Y.

Proof. With high probability Lemma 2.4.22 applies. If $h > |Y|$, then the definition of "h-absorbs Y" is vacuous, so we can suppose that $h \leq |Y|$. If $|Y|/2 \leq h \leq |Y|$, pick subsets $\hat{Y}, Y' \subseteq Z$ with $|\hat{Y}| = 2h - |Y|$ and $|Y'| = 2h$. Apply Lemma 2.4.22 to g, $Y \cup \hat{Y}$, Y', and h to get a set R' such that $R' \cup Y'$ h-absorbs $Y \cup \hat{Y}$. By Lemma 2.4.2 (1), $R' \cup Y'$ h-absorbs Y also.

If $h \leq |Y|/2$, pick subsets $\hat{Y}, Y' \subseteq Z$ with $|\hat{Y}| = |Y| - 2h$ and $|Y'| = 2|Y| - 2h$. Apply Lemma 2.4.22 to g, $Y \cup \hat{Y}$, Y', and $h' = |Y| - h$ to get a set R' such that $R' \cup Y'$ $(|Y| - h)$-absorbs $Y \cup \hat{Y}$. By Lemma 2.4.2 (2) with $t = |\hat{Y}| = |Y| - 2h$, $R' \cup Y' \cup \hat{Y}$ $(|Y| - h - |\hat{Y}|)$-absorbs Y. This implies the lemma since $|Y| - h - |\hat{Y}| = h$. □

Main absorption lemma

The following lemma summarises everything from Section 2.4.1 that we use in other sections.

Lemma 2.4.24. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4140}$. Let $* = A/B/C$. Suppose that Y and h are one of the following:

(1) For some $g \in G$, $Y \subseteq [g] \cap * \setminus N(G)$ with $|Y| \leq \frac{p^{800}|G'|}{10^{4140}\log n}$, and $h \leq |Y|$.

(2) For some $g \in G$, $Y \subseteq [g] \cap * \setminus N(G)$ with $|Y| \leq \frac{p^{800}n}{10^{4140}\log n}$, and $h = |Y| - 1$.

(3) For some generic a_ϕ, we have $Y \subseteq G \backslash G$ a set of $|Y| \leq \frac{p^{800}n}{10^{4140}\log n}$ disjoint pairs of generic elements, with $[a_1 a_2] = [a_\phi]$ for $(a_1, a_2) \in Y$, and $h \leq |Y|$.

Then there is a subset $R' \subseteq R \setminus U$ of size $\leq 10^6|Y|\log n$ which h-absorbs Y.

Proof. With high probability, Lemmas 2.2.34, 2.4.20, 2.4.19, and 2.4.23 apply. Additionally when $|G'| \geq 10^{4140}p^{-800}\log n$, we can assume that $|R_i \cap [g]| \geq p|G'|/2 \geq 4 \times p^{800}|G'|/10^{4140}\log n$ for all cosets $[g]$ and $i = 1, 2, 3$ (using Chernoff's bound).

(1) We can suppose that $|G'| \geq 10^{4140}p^{-800}\log n$, since otherwise Y would have to be empty and the conclusion is vacuous. In this case, we have that for every g, there is a subset $Z_g \subseteq [g]$ of size $4 \times p^{800}|G'|/10^{4140}\log n$. Let $g \in G$, $Y \subseteq [g] \cap *$ of size $|Y| \leq \frac{p^{800}|G'|}{10^{4140}\log n}$, and $h \leq |Y|$. The result follows from Lemma 2.4.23 applied to Y, Z_g, U.

(2) This is strictly weaker than Lemma 2.4.20.

(3) We'll just prove the lemma when $Y \subseteq G_A \times G_A$, the other cases are symmetric. Use Lemma 2.2.34 in order to find a set $Z \subseteq (R_A \backslash U) \times (R_A \backslash U)$ of $4 \times \frac{p^{800}n}{10^{4140}\log n}$ pairs (a, a') of generic elements having $aa' = a_\phi$. Using Lemma 2.4.19, there is a $R' \subseteq R \setminus (U \cup Z)$ and $Z' \subseteq Z$ with $R' \cup Z'$ h-absorbing Y.

□

2.4.2 Absorbing coset-paired sets

In this section we prove results about absorbing coset-paired sets. The main lemmas in this section are all along the lines of "given a set Q, there exists a set R' with the property that for every coset-paired $S \subseteq Q$, there is a matching with vertex set $R' \cup (Q \setminus S)$".

For all $g \in G$, define $i_{[g]} = 2$ if $[g]$ is self-paired and 1 otherwise. The following lemma allows us to absorb the complement of a coset-paired set.

Lemma 2.4.25. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $s \in (0,1]$. Let $m \in \mathbb{N}$. Let $U \subseteq G$ and $Q \subseteq G_\diamond$ (for some $\diamond \in \{A, B, C\}$) be disjoint, and suppose $|U| \leq p^{800}n/10^{4150}$. Suppose also that Q is generic, $|Q| \leq \frac{p^{800}n}{10^{4150}\log n}$, and that Q satisfies:

(i) For all pairs g, h either $Q \cap [g] = Q \cap [h] = \emptyset$ or $|Q \cap [g]|, |Q \cap [h]| \geq i_{[g]}\lceil 12s|G'|\rceil$.

(ii) If $|G'| > s^{-1}/12$ then for all pairs g, h $|Q \cap [g]|, |Q \cap [h]| \leq \frac{p^{800}|G'|}{10^{4150}\log n}$

(iii) If $|G'| \leq s^{-1}/12$, then Q doesn't intersect any self-paired cosets.

Then, there exists an $R' \subseteq R \setminus U$ such that for all coset-paired, $\lceil s|G'|\rceil$-coset-bounded $S \subseteq Q$ of size m, there is a matching with vertex set $R' \cup (Q \setminus S)$.

Proof. With high probability Lemma 2.4.24 applies to R. Let s, m, U, Q be as in the lemma. Let K be the set of cosets $[g]$ with $Q \cap [g]$ nonempty. Partition $Q = Q_1 \cup Q_2$ where, for each coset $[g] \in K$, we have $|Q_1 \cap [g]| = i_{[g]}\lceil 12s|G'|\rceil$ (and so $|Q_2 \cap [g]| = i_{[g]}|Q \cap [g]| - \lceil 12s|G'|\rceil$). Note that Q_1 is coset-paired (since it has even intersection with self-paired cosets and intersection $\lceil 12s|G'|\rceil$ with other pairs of cosets) and so we can partition it into a set of $|Q_1|/2 \leq |Q| \leq \frac{p^{800}n}{10^{4150}\log n}$ pairs P_1. For each g, let $h_g = |Q \cap [g]| - i_{[g]}\lceil s|G'|\rceil$, noting that this is nonnegative by (i). Use Lemma 2.4.24 (1) or (2) to pick a subset $R_g \subseteq R$ which h_g-absorbs $Q \cap [g]$ (if $|G'| \geq s^{-1}/12$, then, using condition (ii), part (1) of that lemma gives this. Otherwise, we have that Q intersects no self-paired cosets and $\lceil s|G'|\rceil = 1$ and so part (2) gives it). Set $h_P = \sum_{[g]\in K} i_{[g]}\lceil s|G'|\rceil$, noting that this is even. Also note that we can assume that m is even and $\leq h_g$ (otherwise there could be no coset-paired, $\lceil s|G'|\rceil$-coset-bounded S of size m contained in Q). So, we can use Lemma 2.4.24 (3) to pick a subset R_P which $\frac{1}{2}(h_P - m)$-absorbs P_1. We can ensure that these are all disjoint by enlarging U as we choose them (the maximum total size of sets we need to avoid is $|U| + \sum_{[g]\in K} |R_g| \leq |U| + \sum_{[g]\in K} 10^6 \log n |Q \cap [g]| \leq |U| + 10^6 \log n|Q| \leq p^{800}n/10^{4150} + 10^6 \log n(\frac{p^{800}n}{10^{4150}\log n}) \leq p^{800}n/10^{4140}$). Let $R' = R_P \cup \bigcup P_1 \cup \bigcup_{g\in K} R_g$.

Consider some coset-paired, $\lceil s|G'|\rceil$-coset-bounded $S \subseteq Q$ of size m. Since S is coset-paired, $|S \cap [g]|/i_{[g]}$ is an integer for all $[g]$.

Claim 2.4.25.1. *There is a subset $P_1' \subseteq P_1$ of pairs disjoint from S which have $\lceil s|G'|\rceil - |S \cap [g]|/i_{[g]}$ pairs intersecting each $[g] \in K$.*

Proof. First suppose that $\lceil s|G'|\rceil = 1$. Note that for each $[g] \in K$ this means that we need to select at most one pair intersecting $[g]$, and we only need to do so when $S \cap [g], S \cap [\phi(g)] = \emptyset$. In this case we have $|Q_1 \cap [g]| = |Q_1 \cap [\phi(g)]| = i_{[g]}|\lceil 12s|G'|\rceil$ and so there is at least one pair in P_1 contained $[g] \cup [\phi(g)]$ (and choosing that pair works).

Next suppose that $\lceil s|G'|\rceil > 1$, which implies that $s|G'| \geq \lceil s|G'|\rceil - 1 \geq \lceil s|G'|\rceil/2$. Then $|Q_1 \cap [g]| = i_{[g]}\lceil 12s|G'|\rceil \geq 12s|G'| \geq 6\lceil s|G'|\rceil$. Thus there are at least $3\lceil s|G'|\rceil$ pairs contained in $[g]\cup[\phi(g)]$. At most $2\lceil s|G'|\rceil$ of them can intersect $S\cap([g]\cup[\phi(g)])$ and so we can choose $\lceil s|G'|\rceil - |S \cap [g]|/i_{[g]}$ of them disjointly from S. □

Note that $|P_1'| = \frac{1}{2}|\bigcup P_1'| = \frac{1}{2}\sum_{[g]\in K} i_{[g]}(\lceil s|G'|\rceil - |S \cap [g]|/i_{[g]}) = \frac{1}{2}(h_P - m)$. By the property of R_P, there is a matching M_P with vertex set $R_P \cup \bigcup P_1'$. For each $g \in K$, note that $|[g]\cap(S\cup\bigcup P_1')| = i_{[g]}\lceil s|G'|\rceil$ and so $|[g]\cap(Q\setminus(S\cup\bigcup P_1'))| = h_g$. By the property of R_g, there is a matching M_g with vertex set $R_g\cup([g]\cap(Q\setminus(S\cup\bigcup P_1')))$. The union of these matchings has vertex set $(R_P \cup \bigcup P_1') \cup \bigcup_{[g]\in K} R_g \cup ([g] \cap (Q \setminus (S \cup \bigcup P_1'))) = R' \cup \bigcup P_1' \cup (Q \setminus (S \cup \bigcup P_1')) = R' \cup (Q \setminus S)$ as required. □

The following simple lemma covers generic sets by matchings.

Lemma 2.4.26. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

For any $U \subseteq V(H_G)$ and any generic $X \subseteq V(H_G)$ with $|X|, |U| \leq p^{800}n/10^{4010}$, there is a matching M of size $|X|$ in H_G covering X and having all other vertices in $R \setminus U$.

Proof. With high probability, the property of Lemma 2.4.5 applies i.e. for any generic $v \in V(H_G)$ and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4000}$, there is an edge e of H_G passing through v and having the other two vertices in $R \setminus U$. Now let $U \subseteq V(H_G)$ and any generic $X \subseteq V(H_G)$ with $|X|, |U| \leq p^{800}n/10^{4010}$. Applying the property of Lemma 2.4.5 to each $v \in X$ we get edges e_v passing through v and having other vertices in $R \setminus U$. By enlarging U as we choose these to include previously found vertices, we can ensure that all e_v are disjoint i.e. they give us a matching like the lemma asks for. □

The following is a variant of Lemma 2.4.25 where Q is a random set.

Lemma 2.4.27. *Let $p \geq n^{-1/701}$ and $q \leq \frac{p^{800}}{10^{4160}\log n}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and Q a disjoint q-random subset. Set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $s \leq \frac{q^2p^{800}}{10^{4160}\log^3 n}$. For every $m \in \mathbb{N}$ and $U \subseteq G$ with $|U| \leq q^2p^{800}n/10^{4160}$, there is a U' with $U \subseteq U' \subseteq U \cup Q$, $|U'| \leq 5q^{-1}|U| + 50s^{-1}$, and $R' \subseteq R \setminus U$ such that for all coset-paired, $\lceil s|G'|\rceil$-coset-bounded $S \subseteq Q \setminus U'$ of size m, there is a matching with vertex set $R' \cup (Q \setminus (S \cup U'))$.

Proof. Randomly split each R^i into two disjoint, symmetric $p/2$-random sets R^i_-, R^i_+. Set $R^- = R^1_{-A} \cup R^2_{-B} \cup R^3_{-C}$ and $R^+ = R^1_{+A} \cup R^2_{+B} \cup R^3_{+C}$. Note that the following hold with high probability.

(A1) R^- satisfies Lemmas 2.4.25 while R^+ satisfies Lemma 2.4.26.

(A2) When $|G'| \geq q^{-2}\log^2 n$, then $|Q \cap [g]| \in [q|G'|/2, 2q|G'|]$ for all $g \in G$ (from Chernoff's bound).

(A3) $|Q| \leq qn + \sqrt{n}\log n \leq \frac{p^{800}n}{10^{4150}\log n}$ (from Chernoff's bound).

Now let $s \leq \frac{q^2 p^{800}}{10^{4160}\log^3 n} \leq q^2\log^{-2} n/12$, $m \in \mathbb{N}$, and $U \subseteq G$ with $|U| \leq q^2 p^{800} n/10^{4160}$. When $|G'| \leq s^{-1}/12$, let U_1 be the union of self-paired cosets, noting that $|U_1| \leq 30|G'| \leq 30s^{-1}$. When $|G'| > s^{-1}/12$, let U_1 be the union of cosets $[g]$ and their pairs for which $|(U \cup N(G)) \cap [g]| \geq q|G'|/10$, noting that $|U_1| \leq 2|G'|\frac{|U|+|N(G)|}{q|G'|/10}$. Set $U' = U \cup N(G) \cup U_1$, noting that $|U'| \leq 5q^{-1}|U| + 6q^{-1}|N(G)| + 30s^{-1} \leq 5q^{-1}|U| + 6q^{-1}10^{9000} + 30s^{-1} \leq 5q^{-1}|U| + 50s^{-1}$. Set $Q_1 = Q \setminus U'$. Let $Q_2 \subseteq Q_1$ be the subset formed by deleting all pairs of cosets for which $Q_1 \cap [g]$ or $Q_1 \cap [\phi(g)]$ is empty. Note that Q_2 satisfies all of the following properties.

1. Q_2 is generic (since it's disjoint from $N(G)$).

2. $|Q_2| \leq \frac{p^{800}n}{10^{4150}\log n}$ (by (A3) and $Q_2 \subseteq Q$).

3. When $|G'| \leq s^{-1}/12$, Q_2 doesn't intersect self-paired cosets (since it's disjoint from U_1).

4. When $|G'| \leq s^{-1}/12$, for all $g \in G$ we have that either $Q_2 \cap [g] = Q_2 \cap [h] = \emptyset$ or $i_{[g]}\lceil 12s|G'|\rceil = 1 \leq |Q_2 \cap [g]|, |Q_2 \cap [h]|$ (by construction of Q_2 from Q_1).

5. When $|G'| > s^{-1}/12$, for all $[g]$ either $Q_2 \cap [g] = Q_2 \cap [h] = \emptyset$ or we have $i_{[g]}12s|G'| \leq 24s|G'| \leq q|G'|/4 \leq q|G'|/2 - q|G'|/10 \leq |Q_2 \cap [g]|, |Q_2 \cap [h]| \leq 2q|G'| \leq \frac{p^{800}|G'|}{10^{4150}\log n}$ (by (A2), since Q_2 is disjoint from U_1, and since $s \leq \frac{q^2 p^{800}}{10^{4160}\log^3 n}$, $q \leq \frac{p^{800}}{10^{4160}\log n}$).

Thus Lemma 2.4.25 applies to Q_2. This gives us a set $R'^{-} \subseteq R^{-} \setminus U$. By Lemma 2.4.26, there is a matching M_1 covering $Q_1 \setminus Q_2$ with $V(M_1) \setminus (Q_1 \setminus Q_2) \subseteq R^{+} \setminus U$. Set $R' = R'^{-} \cup (M_1 \setminus Q) \subseteq R \setminus U$. This is possible as $Q_1 \setminus Q_2 \subseteq Q$ and (A3).

Now consider a coset-paired, $\lceil s|G'|\rceil$-coset-bounded $S \subseteq Q \setminus U' = Q_1$ of size m. Note that we must have $S \subseteq Q_2$ because S is coset paired (and for every $g \in Q_1 \setminus Q_2$ we have $[\phi(g)] \cap Q_1$ empty). Therefore, Lemma 2.4.25 gives us a matching M_2 with vertex set $R'^{-} \cup (Q_2 \setminus S)$. Combining this with M_1 gives a matching with vertex set $R'^{-} \cup (Q_2 \setminus S) \cup (M_1 \setminus Q) \cup (Q_1 \setminus Q_2) = R' \cup (Q_1 \setminus S) = R' \cup (Q \setminus (S \cup U'))$ as required. □

The following shows that large coset-paired matchings exist inside random sets.

Lemma 2.4.28. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4140}$. For any $k \in [1, |G'|]$, there is a coset-paired, k-coset-bounded matching of size $\frac{p^{800}nk}{10^{4140}|G'|}$ in $R \setminus U$.

Proof. With high probability, Lemma 2.2.32 applies. When $|G'| \leq n/\log^{10^{20}} n$, consider the set

$$S = \{x, x^{-1}a_\phi, y, b_\phi y^{-1}, y^{-1}x^{-1}, yc_\phi x\} \subseteq G * F_2.$$

When $|G'| > n/\log^{10^{20}} n$, consider the set

$$S = \{x_1, x_2a_\phi, y_1, y_2b_\phi, y_1^{-1}x_1^{-1}, y_2c_\phi x_2\} \subseteq G * F_4.$$

Note that in either case all words in S are linear and all pairs of words are separable (using part (a) or (c) of the definition of "separable". For checking (c), we use that for any $(w, w') = (x, x^{-1}a_\phi), (y, b_\phi y^{-1}), (y^{-1}x^{-1}, yc_\phi x)$ we have $\pi_0(ww'), \pi_0(w^{-1}w') \in \{a_\phi, b_\phi, c_\phi, a_\phi^{-1}, b_\phi^{-1}, c_\phi^{-1}\}$, which all satisfy (c) from Lemma 2.2.14).

Let $U \subseteq G$ with $|U| \leq p^{800}n/10^{4120}$. If $|G'| \leq p^{800}n/10^{4120}$ add all self-paired cosets to U in order to get a set U' with $|U'| \leq p^{800}n/10^{4110}$ (otherwise set $U' = U$). Use Lemma 2.2.32 to get a projection π which separates S and has $\pi(S) \subseteq R \setminus U'$. This gives us a coset-paired matching of size 2 as in the lemma (whose edges are $(\pi(x), \pi(y), \pi(y^{-1}x^{-1}))$ and $(\pi(x^{-1}a_\phi), \pi(b_\phi y^{-1}), \pi(yc_\phi x))$). To get one of size $\frac{p^{800}nk}{10^{4140}|G'|}$, keep selecting multiple matchings like this, enlarging U at every step in order to keep them disjoint (we can do this as long as $3|G'||M|k^{-1} \leq p^{800}n/10^{4130}$. Indeed for a matching M, letting U_M be the union of cosets $[g]$ having $|V(M) \cap [g]| \geq k$, note that $|U_M|/|G'| \leq |V(M)|/k$ which gives $|U_M| \leq 3|G'||M|/k$. For any pair of edges outside $U \cup U_M$, adding them to M gives a bigger matching like we want). □

Now we arrive at the main lemma of this section. The following again shows that we can absorb the complement of a coset-paired set. Unlike Lemmas 2.4.25 and 2.4.27, this one allows the coset-paired set to have variable size.

Lemma 2.4.29. *Let $q, p \geq n^{-1/701}$ with $q \leq \frac{p^{801}}{10^{4170}\log n}$. Let H_G be a multiplication hypergraph, let $R^1, R^2, R^3, Q_1, Q_2, Q_3$ be disjoint, symmetric subsets of G with R^1, R^2, R^3 p-random and Q_1, Q_2, Q_3 q-random subsets. Set $R = R^1_A \cup R^2_B \cup R^3_C$ and $Q = Q^1_A \cup Q^2_B \cup Q^3_C$. With high probability, the following holds:*

Let $s \leq \frac{q^{1600}p^{10^7}}{10^{10^7}\log^{2400} n}$, $U \subseteq G$ with $|U| \leq q^9 p^{800} n/10^{4170}$. There is a $U' \subseteq U$ with $|U'| \leq 500q^{-3}|U| + 10^{10}s^{-3}$, and $R' \subseteq R \setminus U$ such that for all coset-paired, $\lceil s|G'|\rceil$-coset-bounded, balanced $S \subseteq Q \setminus U'$, with $|S| \leq s^2 n$ there is a matching with vertex set $R' \cup (Q \setminus (S \cup U'))$.

Proof. Set $m := 8\lceil sn \rceil$ and $s' := s^{1/800}10^{4140}$. Split R into three disjoint, symmetric $p/3$-random subsets R_1, R_2, R_3 of $V(H_G)$ (by placing each $\hat{g}$ in $R_1/R_2/R_3$ with probability $1/3$, and making these choices independently of those for the random sets R^1, R^2, R^3 in the lemma. So now for all i $R_i \cap R^1_A$, $R_i \cap R^2_B$, $R_i \cap R^3_C$ are three disjoint, symmetric $p/3$-random subsets of G partitioning R_i). Lemma 2.4.27 applies to each pair $R' = R_i, Q' = Q_i$ for $i = 1, 2, 3$ (with $R'^1 = R_i \cap R^1_A$, $R'^2 = R_i \cap R^2_B$, $R'^3 = R_i \cap R^3_C$). Lemma 2.4.28 applies to Q. Let $U_0 := U$ be as in the lemma. For $i = 1, 2, 3$ build $U_1, U_2, U_3, R_1, R_2, R_3$ by applying Lemma 2.4.27 to R_i, Q_i with $U = U_{i-1}$, $m = m, s = 4s'$. The end result is sets $R'_1 \subseteq R_1, R'_2 \subseteq R_2, R'_3 \subseteq R_3$ and a set U' with $|U'| \leq 500q^{-3}|U| + 10^{10}s^{-3}$ such that for all coset-paired, $\lceil 4s'|G'|\rceil$-coset-bounded $S_i \subseteq Q_i \setminus U'$ of size m, there are matchings with vertex sets $R'_i \cup (Q_i \setminus (S_i \cup U'))$. Set $R' = R'_1 \cup R'_2 \cup R'_3$. Use Lemma 2.4.28 with $k = \lceil s'|G'|\rceil$ to find a coset-paired, $\lceil s'|G'|\rceil$-coset-bounded matching M_0 of size $\frac{p^{800}nk}{10^{4140}|G'|} \geq q^{800}s'n/10^{4140} \geq 8m$ in $Q_1 \cup Q_2 \cup Q_3 \setminus U'$.

Now, consider a coset-paired, $\lceil s|G'|\rceil$-coset-bounded, balanced $S \subseteq Q \setminus U'$ with $|S| \leq s^2 n$. Let $m_0 = |S \cap G_A| = |S \cap G_B| = |S \cap G_C| \leq s^2 n$, noting that these are even by coset-pairedness.

Claim 2.4.29.1. *There is a coset-paired submatching $M_0' \subseteq M_0$ of size $m - m_0$ disjoint from S. Additionally when $|G'| \leq s^{-1}$, then M_0' and S never intersect the same cosets.*

Proof. When $|G'| \leq s^{-1}$ let S' be the union of cosets intersecting S, otherwise let $S' = S$. Note that in either case $|S'| \leq |S|s^{-1} \leq sn \leq m$. Since $|M_0| \geq 8m$, there is a paired submatching $M_0' \subseteq M_0$ of size $m - m_0$ disjoint from S'. □

Now set $S_1' := (S \cup M_0') \cap G_A$, $S_2' := (S \cup M_0') \cap G_B$, $S_3' := (S \cup M_0') \cap G_C$ and note that these are coset-paired, $\lceil 4s'|G'|\rceil$-coset-bounded sets of size m. By the properties of R_1', R_2', R_3', we get matchings M_1, M_2, M_3 with vertex sets $V(M_i) = R_i' \cup (Q_i \setminus (S_i' \cup U'))$. Now $M = M_0' \cup M_1 \cup M_2 \cup M_3$ is a matching with vertex set $V(M_0') \cup \bigcup_{i=1}^{3} R_i' \cup (Q_i \setminus (S_i' \cup U')) = R' \cup V(M_0') \cup (Q \setminus (S \cup M_0' \cup U')) = R' \cup (Q \setminus (S \cup U'))$ as required. □

2.4.3 Absorbing zero-sum sets

The goal of this section is to prove a lemma which can absorb arbitrary balanced zero-sum sets (Lemma 2.4.32). To do this we first prove results about covering zero-sum sets using coset-paired matchings.

Covering zero-sum sets

Here we prove Lemma 2.4.31, which roughly states that given a random vertex subset R of H_G and a small zero-sum set S, there exists a coset-paired set $R' \subseteq R$ such that $S \cup R'$ contains a perfect matching. As explained in Section 2.1, this will allow us to reduce the task of absorbing arbitrary zero-sum sets to absorbing coset-paired sets.

The following lemma does the majority of the work for this section. It allows us to find the desired set R' iteratively by reducing the size of the set of vertices we need to cover by 3 at every step.

Lemma 2.4.30. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds.*

Let $F \subseteq V(H_G)$ be a balanced ϕ-generic set of size 6, $h \in [\prod F]$, and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4001}$. Then there is a matching M in H_G of size 15 and a disjoint set $\{a,b,c\}$ such that $F \subseteq V(M)$, $\{a,b,c\} \cup V(M) \setminus F \subseteq R \setminus U$, $\{a,b,c\} \cup V(M) \setminus F$ is coset-paired and $abc = h$. Furthermore, if $|G'| \leq \log^5 n/p^{2000}$, then for each $\diamond \in \{A,B,C\}$, and $g \in G$ we have that $|(\{a,b,c\} \cup V(M) \setminus F) \cap G_\diamond \cap [g]| \leq 1$.

Proof. With high probability, the property of Lemma 2.2.32 holds. Let $F \subseteq V(H_G)$ be a balanced ϕ-generic set of size 6, $h \in [\prod F]$, and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4001}$. Let $F = \{a_1, a_2, b_1, b_2, c_1, c_2\}$ with $a_i \in G_A, b_i \in G_B, c_i \in G_C$. Let $k = ha_2^{-1}c_1^{-1}c_2^{-1}b_2^{-1}b_1^{-1}a_1^{-1}$ and note that by the assumption that $h \in [\prod F]$, we know that $k \in G'$. Depending on whether $|G'| \geq \log^5 n/p^{2000}$ or not, consider the set of words S given in Figure 2.6 or 2.7, noting that blue/green/red vertices represent a partition as $S_A/S_B/S_C$. Note that in either case, all the words in S are linear and all pairs of words in S are weakly separable and words in S coming from different $S_A/S_B/S_C$ are strongly separable (see the figure captions for justification). In fact, it will be the case that every pair of words are strongly separable, meaning we never use condition (b'), so this distinction will not be essential in the justification.

For $* = A, B, C$, let $T_* = \{w^{-1}w' : w, w' \in S_*$ and w, w' do not have the same free variables$\}$, noting that $|T_*| \leq \binom{|S_*|}{2} = \binom{14}{2} = 91$ and that elements of T_* are all linear in at least one variable. If $|G'| > \log^5 n/p^{2000}$, let $U' = U \cup F$. If $|G'| \leq \log^5 n/p^{2000}$, then let U' be U together with all the self-paired cosets and all the cosets intersecting F, and G'. Note that in both cases, $|U'| \leq |U| + 36\log^5 n/p^{2000} \leq p^{800}n/10^{4000}$.

Use Lemma 2.2.32 to get a projection π with $\pi(S \cup T_A \cup T_B \cup T_C) \subseteq R \setminus (U \cup F)$

that separates $S \cup T_A \cup T_B \cup T_C$ and ensures that $S_A \subseteq R_1$, $S_B \subseteq R_2$, $S_C \subseteq R_3$ (so the matching is contained in R).

First, we prove the lemma without the "furthermore" part. Since all $w, w' \in S$ are separable, this means that all the vertices of $\pi(S)$ are distinct. Recall that $S_A/S_B/S_C$ are the blue/green/red vertices in the figures. Note that $F \cup \pi(S)$ has a partition into a matching M (given by the black triangles in Figures 2.8 and 2.9) and a set $\{a, b, c\}$ having $abc = h$ (given by the pink triangle in the same figures where a is the top left vertex, b is the bottom vertex, and c is the top right vertex). We claim that this matching M together with the set $\{a, b, c\}$ satisfy the lemma. The things that need to be verified by inspecting the figure are as follows.

- Each black triangle has that if its vertices are multiplied in the order blue/green/red ($G_A/G_B/G_C$), we obtain e. This shows that the black triangles are in fact edges of H_G.

- The product of the top left, bottom, and top right vertices of the yellow (central) triangle is h (here, use the definition of k). This gives that $abc = h$.

- The product of the blue/green/red edges belongs to $[a_\phi]/[b_\phi]/[c_\phi]$. In the case where $|G'| \geq 10^{-9}n$, note that we select the black (free) variables from G', hence they can be ignored while performing this check. This shows that $\{a, b, c\} \cup V(M) \setminus F$ is coset-paired.

For the "furthermore" part, we have that $|G'| \leq \log^5 n/p^{2000}$. For $* = A, B, C$, notice that for all $w, w' \in S_*$ we either have that $(\pi(w), \pi(w'))$ is a pair, or there is a free variable which appears in one of w/w', but not both (to check this, note that for any vertices w, w' of the same colour in Figure 2.8, either w, w' are joined by an edge or w and w' have different combinations of black letters). In the first case we have that $[\pi(w)] \neq [\pi(w')]$ since $\pi(S)$ is disjoint from all self-paired cosets. In the second case we have $w^{-1}w' \in T_*$, which implies $\pi(w^{-1}w') \notin G'$ (since $G' \subseteq U'$), or equivalently $[\pi(w)] \neq [\pi(w')]$. □

Now, we may prove the main result of this section.

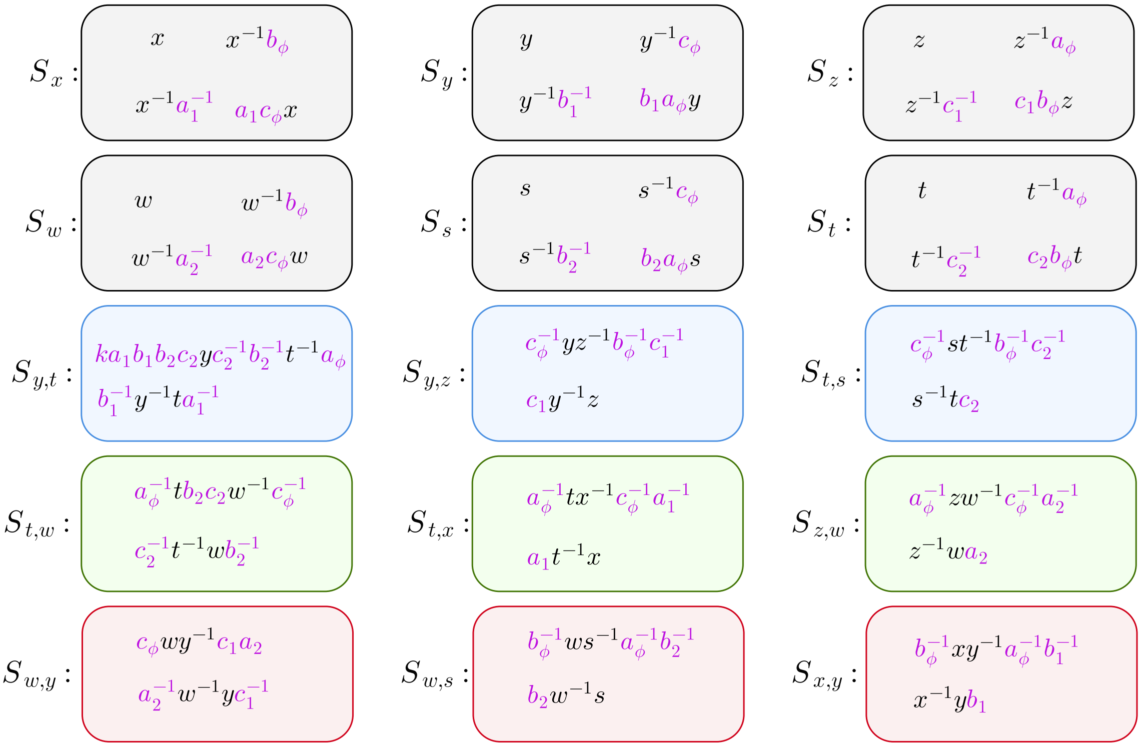

$S_x :$ x $x^{-1}b_\phi$ $x^{-1}a_1^{-1}$ $a_1c_\phi x$
$S_y :$ y $y^{-1}c_\phi$ $y^{-1}b_1^{-1}$ $b_1a_\phi y$
$S_z :$ z $z^{-1}a_\phi$ $z^{-1}c_1^{-1}$ $c_1b_\phi z$
$S_w :$ w $w^{-1}b_\phi$ $w^{-1}a_2^{-1}$ $a_2c_\phi w$
$S_s :$ s $s^{-1}c_\phi$ $s^{-1}b_2^{-1}$ $b_2a_\phi s$
$S_t :$ t $t^{-1}a_\phi$ $t^{-1}c_2^{-1}$ $c_2b_\phi t$
$S_{y,t} :$ $ka_1b_1b_2c_2yc_2^{-1}b_2^{-1}t^{-1}a_\phi$ $b_1^{-1}y^{-1}ta_1^{-1}$
$S_{y,z} :$ $c_\phi^{-1}yz^{-1}b_\phi^{-1}c_1^{-1}$ $c_1y^{-1}z$
$S_{t,s} :$ $c_\phi^{-1}st^{-1}b_\phi^{-1}c_2^{-1}$ $s^{-1}tc_2$
$S_{t,w} :$ $a_\phi^{-1}tb_2c_2w^{-1}c_\phi^{-1}$ $c_2^{-1}t^{-1}wb_2^{-1}$
$S_{t,x} :$ $a_\phi^{-1}tx^{-1}c_\phi^{-1}a_1^{-1}$ $a_1t^{-1}x$
$S_{z,w} :$ $a_\phi^{-1}zw^{-1}c_\phi^{-1}a_2^{-1}$ $z^{-1}wa_2$
$S_{w,y} :$ $c_\phi wy^{-1}c_1a_2$ $a_2^{-1}w^{-1}yc_1^{-1}$
$S_{w,s} :$ $b_\phi^{-1}ws^{-1}a_\phi^{-1}b_2^{-1}$ $b_2w^{-1}s$
$S_{x,y} :$ $b_\phi^{-1}xy^{-1}a_\phi^{-1}b_1^{-1}$ $x^{-1}yb_1$

Figure 2.6: The set S when $|G'| \leq 10^{-9}n$. Black letters represent free variables, while pink ones represent elements of G. The words are grouped into rectangles S_T based on which free variables occur where, as in Observation 2.2.24. To see that all words in S are linear, check that there are no repetitions of black letters in any word (and every word has at least one black letter). To see that any pair w, w' is strongly separable, note that by Observation 2.2.24, any w, w' coming from different rectangles fall under part (a) of the definition of separable. In the coloured rectangles, there are always two words w, w' which fall under part (c) of the definition of strongly separable. This leaves the grey rectangles. In each such rectangle the four elements are $v, v^{-1}d_i^{-1}, v^{-1}e_\phi, d_if_\phi v$ for a free variable v, $i \in \{1, 2\}$, and (d, e, f) some permutation of (a, b, c). The pair $w = v^{-1}d_i^{-1}, w' = d_if_\phi v$ falls under (c) because $\pi_0(ww') = \pi_0(v^{-1}d_i^{-1}d_if_\phi v) = f_\phi$. All other pairs fall under (b), as witnessed by the following equations $v = v^{-1}d_i^{-1}$, $v = d_if_\phi v$, $v = v^{-1}e_\phi$, $v^{-1}d_i^{-1} = (v^{-1}e_\phi)(e_\phi^{-1}d_i^{-1})$, $v^{-1}e_\phi = (d_if_\phi v)^{-1}(d_if_\phi e_\phi)$. The elements g in these equations are e_ϕ, d_i, d_if_ϕ, $e_\phi^{-1}d_i^{-1}$, $d_if_\phi e_\phi$, which are all generic (since d_i is ϕ-generic).

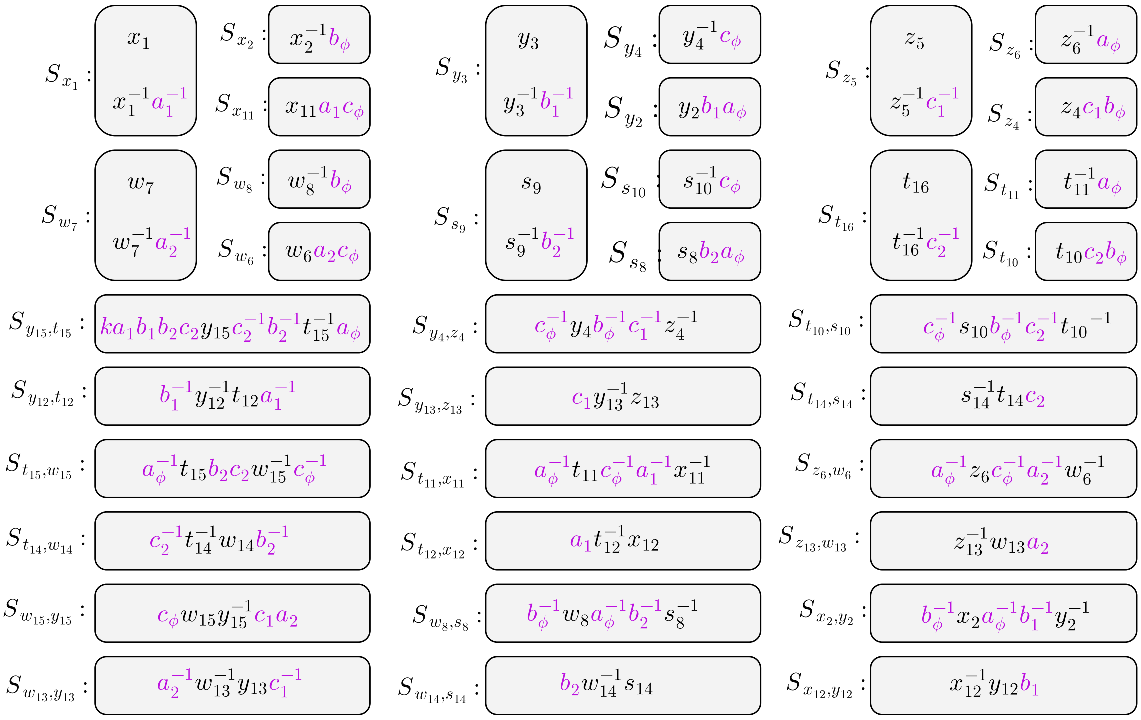

S_{x_1} : x_1 ; $x_1^{-1}a_1^{-1}$
S_{x_2} : $x_2^{-1}b_\phi$
$S_{x_{11}}$: $x_{11}a_1c_\phi$
S_{y_3} : y_3 ; $y_3^{-1}b_1^{-1}$
S_{y_4} : $y_4^{-1}c_\phi$
S_{y_2} : $y_2b_1a_\phi$
S_{z_5} : z_5 ; $z_5^{-1}c_1^{-1}$
S_{z_6} : $z_6^{-1}a_\phi$
S_{z_4} : $z_4c_1b_\phi$
S_{w_7} : w_7 ; $w_7^{-1}a_2^{-1}$
S_{w_8} : $w_8^{-1}b_\phi$
S_{w_6} : $w_6a_2c_\phi$
S_{s_9} : s_9 ; $s_9^{-1}b_2^{-1}$
$S_{s_{10}}$: $s_{10}^{-1}c_\phi$
S_{s_8} : $s_8b_2a_\phi$
$S_{t_{16}}$: t_{16} ; $t_{16}^{-1}c_2^{-1}$
$S_{t_{11}}$: $t_{11}^{-1}a_\phi$
$S_{t_{10}}$: $t_{10}c_2b_\phi$
$S_{y_{15},t_{15}}$: $ka_1b_1b_2c_2y_{15}c_2^{-1}b_2^{-1}t_{15}^{-1}a_\phi$
S_{y_4,z_4} : $c_\phi^{-1}y_4b_\phi^{-1}c_1^{-1}z_4^{-1}$
$S_{t_{10},s_{10}}$: $c_\phi^{-1}s_{10}b_\phi^{-1}c_2^{-1}t_{10}{}^{-1}$
$S_{y_{12},t_{12}}$: $b_1^{-1}y_{12}^{-1}t_{12}a_1^{-1}$
$S_{y_{13},z_{13}}$: $c_1y_{13}^{-1}z_{13}$
$S_{t_{14},s_{14}}$: $s_{14}^{-1}t_{14}c_2$
$S_{t_{15},w_{15}}$: $a_\phi^{-1}t_{15}b_2c_2w_{15}^{-1}c_\phi^{-1}$
$S_{t_{11},x_{11}}$: $a_\phi^{-1}t_{11}c_\phi^{-1}a_1^{-1}x_{11}^{-1}$
S_{z_6,w_6} : $a_\phi^{-1}z_6c_\phi^{-1}a_2^{-1}w_6^{-1}$
$S_{t_{14},w_{14}}$: $c_2^{-1}t_{14}^{-1}w_{14}b_2^{-1}$
$S_{t_{12},x_{12}}$: $a_1t_{12}^{-1}x_{12}$
$S_{z_{13},w_{13}}$: $z_{13}^{-1}w_{13}a_2$
$S_{w_{15},y_{15}}$: $c_\phi w_{15}y_{15}^{-1}c_1a_2$
S_{w_8,s_8} : $b_\phi^{-1}w_8a_\phi^{-1}b_2^{-1}s_8^{-1}$
S_{x_2,y_2} : $b_\phi^{-1}x_2a_\phi^{-1}b_1^{-1}y_2^{-1}$
$S_{w_{13},y_{13}}$: $a_2^{-1}w_{13}^{-1}y_{13}c_1^{-1}$
$S_{w_{14},s_{14}}$: $b_2w_{14}^{-1}s_{14}$
$S_{x_{12},y_{12}}$: $x_{12}^{-1}y_{12}b_1$

Figure 2.7: The set S when $|G'| \geq n/10^9$. Black letters represent free variables, while pink ones represent elements of G. The words here are exactly the same as in Figure 2.6, except that there are more free variables e.g. instead of the free variable x, we have free variables x_1, x_2, x_{11}, x_{12} with x replaced by one of x_1, x_2, x_{11}, x_{12} wherever it occurred. The words are grouped into rectangles S_T based on which free variables occur where, as in Observation 2.2.24. To see that all words in S are linear, check that there are no repetitions of black letters in every word (and every word has at least one black letter). To see that all $w, w' \in S$ are strongly separable, note that from Observation 2.2.24 this holds when w, w' come from different rectangles. This leaves only the pairs inside $S_{x_1}, S_{y_3}, S_{z_5}, S_{w_7}, S_{s_9}, S_{t_{16}}$, which are separable by (b) since $a_1, a_2, b_1, b_2, c_1, c_2$ are generic.

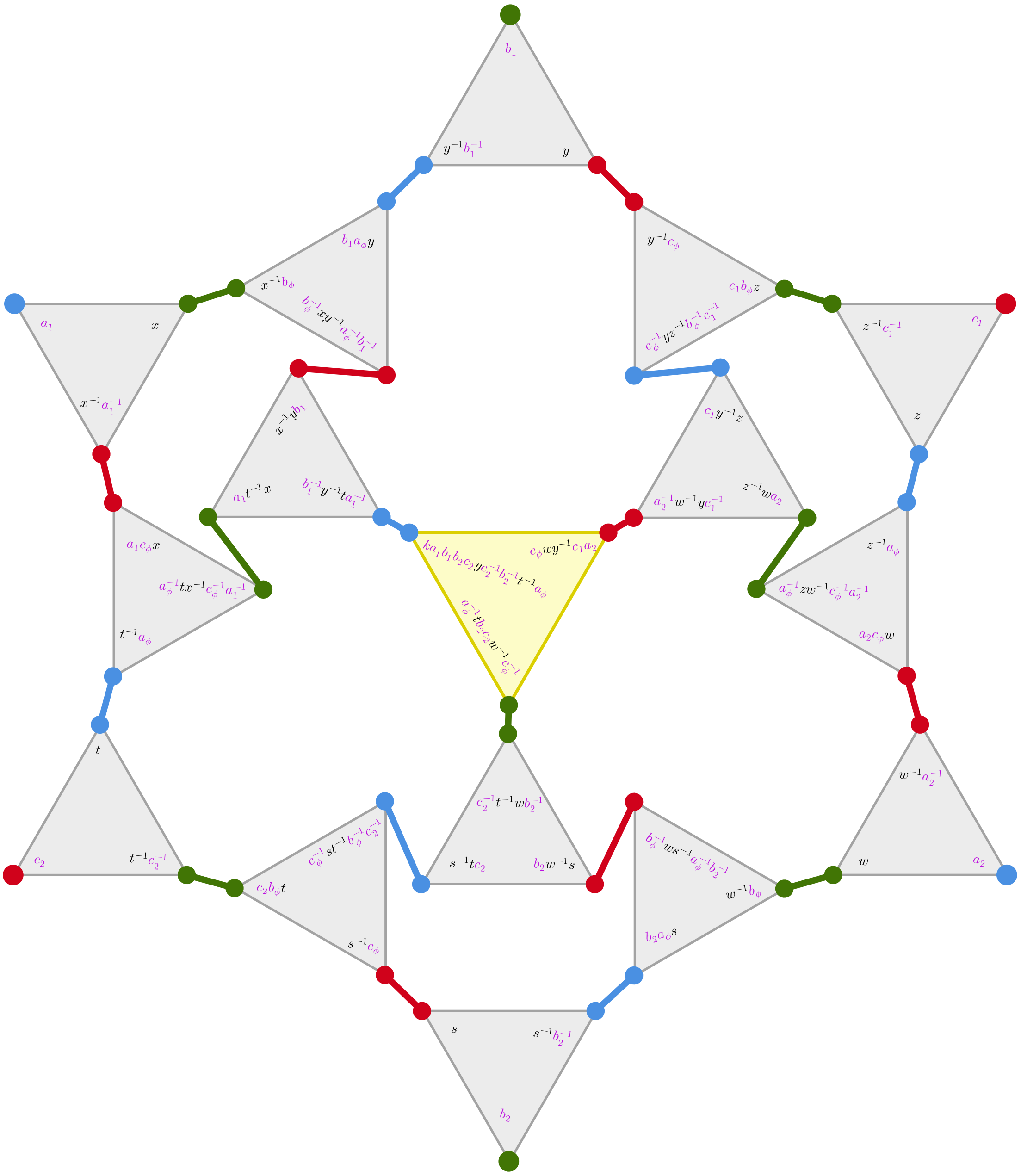


Figure 2.8: The set S when $|G'| \leq 10^{-9}n$. Blue, green, and red vertices give a partition of S into sets $S_A/S_B/S_C$ which are pairwise strongly separable. Black letters represent free variables, while pink ones represent elements of G. The elements $a_1, a_2, b_1, b_2, c_1, c_2$ are not part of S, they are depicted only to illustrate the matching. The black triangles are matching edges, while the yellow triangle is the set $\{a, b, c\}$. Note that the yellow triangle is an edge if and only if $k = e$.

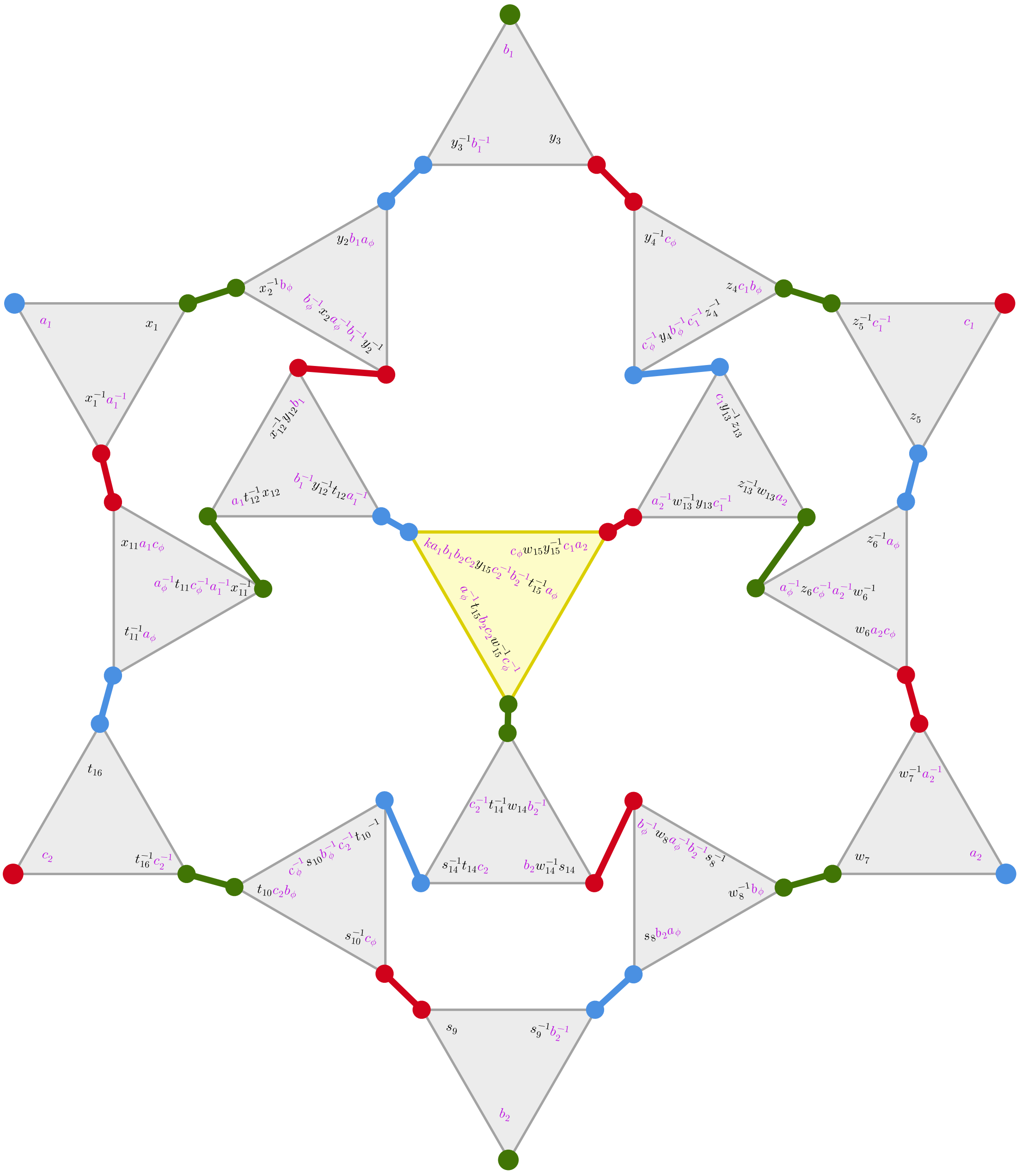

b_1
$y_3^{-1}b_1^{-1}$
y_3
$y_2 b_1 a_\phi$
$x_2^{-1}b_\phi$
$b_\phi^{-1}x_2 a_\phi^{-1}b_1^{-1}y_2^{-1}$
$y_4^{-1}c_\phi$
$z_4 c_1 b_\phi$
$c_\phi^{-1}y_4 b_\phi^{-1}c_1^{-1}z_4^{-1}$
a_1
x_1
$x_1^{-1}a_1^{-1}$
$z_5^{-1}c_1^{-1}$
c_1
z_5
$x_{12}^{-1}y_{12}b_1$
$a_1 t_{12}^{-1}x_{12}$
$b_1^{-1}y_{12}^{-1}t_{12}a_1^{-1}$
$c_1 y_{13}^{-1}z_{13}$
$a_2^{-1}w_{13}^{-1}y_{13}c_1^{-1}$
$z_{13}^{-1}w_{13}a_2$
$x_{11}a_1 c_\phi$
$a_\phi^{-1}t_{11}c_\phi^{-1}a_1^{-1}x_{11}^{-1}$
$t_{11}^{-1}a_\phi$
$ka_1 b_1 b_2 c_2 y_{15}c_2^{-1}b_2^{-1}t_{15}^{-1}a_\phi$
$c_\phi w_{15}y_{15}^{-1}c_1 a_2$
$a_\phi^{-1}t_{15}b_2 c_2 w_{15}^{-1}c_\phi^{-1}$
$z_6^{-1}a_\phi$
$a_\phi^{-1}z_6 c_\phi^{-1}a_2^{-1}w_6^{-1}$
$w_6 a_2 c_\phi$
t_{16}
c_2
$t_{16}^{-1}c_2^{-1}$
$w_7^{-1}a_2^{-1}$
w_7
a_2
$c_2^{-1}t_{14}^{-1}w_{14}b_2^{-1}$
$s_{14}^{-1}t_{14}c_2$
$b_2 w_{14}^{-1}s_{14}$
$c_\phi^{-1}s_{10}b_\phi^{-1}c_2^{-1}t_{10}^{-1}$
$t_{10}c_2 b_\phi$
$s_{10}^{-1}c_\phi$
$b_\phi^{-1}w_8 a_\phi^{-1}b_2^{-1}s_8^{-1}$
$w_8^{-1}b_\phi$
$s_8 b_2 a_\phi$
s_9
$s_9^{-1}b_2^{-1}$
b_2

Figure 2.9: The set S when $|G'| > 10^{-9}n$. Blue, green, and red vertices give a partition of S into sets $S_A/S_B/S_C$ which are pairwise strongly separable. Black letters represent free variables, while pink ones represent elements of G. The elements $a_1, a_2, b_1, b_2, c_1, c_2$ are not part of S. This set is exactly the same as Figure 2.8, except that there are more free variables. More precisely, for every black/pink triangle we introduce new variables and use them only in words occurring inside that triangle. The black triangles are matching edges, while the yellow triangle is the set $\{a, b, c\}$ where a is the top left vertex, b is the bottom vertex, and c is the top right vertex.

Lemma 2.4.31. *Let $p \geq n^{-1/10^{20}}$. Let H_G be a multiplication hypergraph, R^1, R^2, R^3 disjoint, symmetric p-random subsets of G and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds:*

Let $S \subseteq V(H_G)$ be a balanced and ϕ-generic subset with $\prod S \in G'$ and $|S| \leq \frac{p^{10^{13}}}{10^{10^6} \log(n)^{10^8}}$. Let $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4100}$. Then, there exists a matching M in H_G with the following properties.

Q1 $S \subseteq V(M)$

Q2 $V(M) \setminus S \subseteq R \setminus U$

Q3 $V(M) \setminus S$ *is coset-paired*

Q4 *If* $|G'| \leq \log(n)^{8000}/p^{10^{10}}$, *then for each* $\diamond \in \{A, B, C\}$, *and* $g \in G$ *we have that* $|(V(M) \setminus S) \cap G_\diamond \cap [g]| \leq 1$.

Proof. With high probability, the property of Lemma 2.4.30 holds and Lemma 2.2.34 applies to each R^1, R^2, R^3. Let $S \subseteq V(H_G)$ be a balanced and ϕ-generic subset with $\prod S \in G'$ and $|S| \leq \frac{p^{10^{13}}}{10^{10^6} \log(n)^{10^8}}$ and $U \subseteq V(H_G)$ with $|U| \leq p^{800}n/10^{4010}$. Without loss of generality, we can assume that $|S| \geq 6$ (if this doesn't hold, then pick ϕ-generic vertices $a, a', b, b', c, c' \in R \setminus (S \cup U)$ with $aa' = a_\phi, bb' = b_\phi, cc' = c_\phi$ using Lemma 2.2.34, and define $S' = S \cup \{a, a', b, b', c, c'\}$. Now S' is a set with $|S'| \geq 6$, so we can continue the proof with S' rather than S. (In the case when $|G'| \leq \log(n)^{8000}/p^{10^{10}}$ we can include in U the $30|G'| \ll \sqrt{n}$ many elements of G in self-paired cosets before applying Lemma 2.2.34 to guarantee that the pairs we are adding come from distinct cosets. This way, **Q4** is not violated.)

Set $t = |S|/3$. Since S is balanced we can write $S = \{a_1, \dots, a_t, b_1, \dots, b_t, c_1, \dots, c_t\}$ with $a_i \in G_A, b_i \in G_B, c_i \in G_C$.

We build matchings $M_2, \dots, M_t$ and ϕ-generic vertices $\{a'_2, \dots, a'_t, b'_2, \dots, b'_t, c'_2, \dots, c'_t\}$ as follows. Define $a'_2 = a_1, b'_2 = b_1, c'_2 = c_1$. Given a subset $S \subseteq G$, denote by $\psi(S)$ the set of all g in G such that there exists some $s \in S$ with $[s] = [g]$. Observe that $|\psi(S)| \leq |S| \cdot |G'|$. For $i = 2, \dots, t-1$ apply

Lemma 2.4.30 to $F_i = \{a_i, b_i, c_i, a'_i, b'_i, c'_i\}$ and

$$U = U \cup N(G) \cup \bigcup_{j<i} \psi(V(M_i))$$

in the case when $|G'| \leq \log(n)^{8000}/p^{10^{10}}$ and

$$U = U \cup N(G) \cup \bigcup_{j<i} V(M_i)$$

otherwise. This way, we obtain a matching M_i and a set $\{a'_{i+1}, b'_{i+1}, c'_{i+1}\}$ (use an arbitrary choice of h from the corresponding G'-coset for these applications). Further, **Q4** is maintained for $\bigcup_{j\leq i} M_i$ in the case when $|G'| \leq \log(n)^{8000}/p^{10^{10}}$ by our inclusion in U of all previously used G'-cosets. To see that the necessary upper bound on U holds for this application, note that when $|G'| \leq \log(n)^{8000}/p^{10^{10}}$, we have that $|\bigcup_{j<i} \psi(V(M_i))| \leq 100t|G'| \leq 100 \left(\frac{p^{10^{13}}}{10^{10^6} \log(n)^{10^8}}\right) (\log(n)^{8000}/p^{10^{10}}) \ll p^{800}n/10^{4001}$. Otherwise, when $|G'|$ is large, we have that $|\bigcup_{j<i} V(M_i)| \leq 100t \ll p^{800}n/10^{4001}$ as well.

Notice that $a'_t b'_t c'_t a_t b_t c_t \in G'$ (to see this, show that for all i we have $a'_i b'_i c'_i a_i b_i c_i a_{i+1} b_{i+1} c_{i+1} \dots a_t b_t c_t \in G'$ by induction. The initial case is just the assumption $\prod S \in G'$. For the induction step use that Lemma 2.4.30 gives $[a'_i b'_i c'_i a_i b_i c_i] = [a'_{i+1} b'_{i+1} c'_{i+1}]$). Now apply Lemma 2.4.30 to $F_t = \{a_t, b_t, c_t, a'_t, b'_t, c'_t\}$ with $h = e \in G'$ and $U = U \cup N(G) \cup \bigcup_{j<t} V(M_i)$ in order to obtain a matching M_t and a set $\{a, b, c\}$ with $abc = e$. Let $M = \bigcup_{i=2}^{t} M_i \cup \{abc\}$ in order to get a matching satisfying the lemma. Checking **Q1-Q2** is routine. To verify **Q3**, note that Lemma 2.4.30 tells us that $N_i := \{a'_{i+1}, b'_{i+1}, c'_{i+1}\} \cup V(M_i) \setminus F_i$ is coset-paired for $i = 2, \dots, t-1$ and also that $N_t := \{a, b, c\} \cup M_t \setminus F_t$ is coset-paired. Note that $\bigcup_{j<i} N_i \subseteq F_i \cup \bigcup_{j<i} V(M_i)$ for each i, which shows that $N_1, \dots, N_t$ are disjoint (N_i is trivially disjoint from F_i, and is disjoint from $\bigcup_{j<i} V(M_i)$ by choice of U when applying Lemma 2.4.30). Also, $\bigcup_{i=1}^{t} N_i = V(M) \setminus S$, which shows that this set is coset-paired as well.

□

The zero-sum absorption lemma

Now we arrive at the main lemma of this section.

Lemma 2.4.32. *Let $n^{-1/10^{100}} \leq p$. Let $R^1, R^2, R^3 \subseteq G$ be disjoint, symmetric p-random subsets and set $R = R^1_A \cup R^2_B \cup R^3_C$. With high probability, the following holds.*

Let $U \subseteq G$ with $|U| \leq p^{10^{14}} n / \log(n)^{10^{14}}$. Then, there exists a subset $R' \subseteq R \setminus U$ such that for all balanced and ϕ-generic subsets $S \subseteq V(H_G) \setminus R'$ with $|S| \leq \frac{p^{10^{13}}}{10^{10^8} \log(n)^{10^8}}$, $\sum S = 0$, there exists a matching with vertex set $R' \cup S$.

We remark that the set R' in this lemma is always balanced and zero-sum. To see this, notice that the conclusion of the lemma applies with $S = \emptyset$ (since the empty set is balanced and zero-sum). This gives a matching with vertex set $R' \cup \emptyset = R'$. Since matchings are balanced and zero-sum, we get that R' is balanced and zero-sum.

Proof. Let $q = \frac{p^{801}}{10^{4180} \log n}$ and $s = \frac{q^{1600} p^{10^7}}{10^{10^6} \log^{2400} n}$. For each $i \in \{1, 2, 3\}$, let Q^i, W^i be q-random and s^4-random, symmetric subsets of G respectively, satisfying $W^i \subseteq Q^i \subseteq R^i$. Note that Q^1, Q^2, Q^3 and W^1, W^2, W^3 are disjoint, symmetric q-random and disjoint, symmetric s-random subsets of G, respectively. The following all hold simultaneously with high probability.

1. Lemma 2.4.29 holds for R^1, R^2, R^3 and Q^1, Q^2, Q^3.

2. Lemma 2.4.31 holds for W^1, W^2, W^3 (observe that $s^4 \geq n^{-1/10^{20}}$).

3. If $|G'| \geq s^{-5}$, each W^i is $\lceil 2s^4|G'| \rceil$-coset-bounded. (This follows by Chernoff's bound, as $1/s \gg \log(n)^{10}$.)

4. $|\bigcup W_i| \leq 100 s^4 n$, by Chernoff's bound.

Fix some $U \subseteq G$ such that $|U| \leq p^{10^{14}} n / \log(n)^{10^{14}}$ as in the statement of the lemma. Note that it also follows that $|U| \leq q^9 p^{800} n / 10^{4170}$. Then, Lemma 2.4.29 gives us a set $U' \supseteq U$ with $|U'| \leq 500 q^{-3} |U| + 10^{10} s^{-3}$ and $R' \subseteq R \setminus U$ with the property that for all coset-paired, $\lceil s|G'|\rceil$-coset-bounded, balanced, $S \subseteq Q \setminus U'$, with $|S| \leq s^2 n$ there is a matching with vertex set $R' \cup (Q \setminus (S \cup U'))$.

We claim that $R' \cup \bigcup_{i\in[3]} Q^i \setminus U'$ satisfies the property required of R' in the statement of the lemma, and the rest of the proof will justify this. Fix some set S' with the properties of S as in the statement of the lemma, that is, S' is ϕ-generic, balanced, disjoint with $R' \cup \bigcup_{i\in[3]} W^i$, $|S'| \leq \frac{p^{10^{13}}}{10^{10^8}\log(n)^{10^8}}$, and S' is zero-sum. Note it easily follows that $|S| \leq p^{4000} n/\log(n)^{10}$, and we also have that $|U'| \leq p^{800}/10^{4100}$, so we may invoke the property of W^i coming from Lemma 2.4.31 with $S = S'$ and $U = U'$ to deduce the existence of a matching M_1 with properties **Q1-Q4**. Then, the set $S'' := \bigcup W^i \cap (V(M_1) \setminus S')$ is coset-paired by **Q3**, $\lceil s|G'|\rceil$-coset-bounded (if $|G'| \geq s^{-5}$, this follows by (3), and if $|G'| < s^{-5} \leq \log(n)^{8000}/p^{10^{10}}$ this follows by **Q4**, as we have 1-coset-boundedness in this case), balanced, and have size at most $s^2 n$ (as $S'' \subseteq \bigcup W^i$ and (4)). By property of the set R', we have that $R' \cup (Q \setminus (S \cup U'))$ has a perfect matching, M_2 say. Then, $M_1 \cup M_2$ is a perfect matching of $(R' \cup \bigcup_{i\in[3]} Q^i \setminus U') \cup S'$, as desired. □

2.4.4 Proof of Theorem 2.3.3

Our goal in this section is to prove the main result of the paper, Theorem 2.3.3, as stated in Section 2.3. We first prove the following slight variant which changes the 3rd bullet point from "$e \notin X, Y, Z$" to the more restrictive condition "X, Y, Z are ϕ-generic".

Theorem 2.4.33. *Let $\alpha \geq n^{-1/10^{101}}$. Let G be a group of order n. Let $R^1, R^2, R^3 \subseteq G$ be p-random, α-slightly-independent subsets. Then, with high probability, the following holds.*

Let X, Y, Z be equal-sized subsets of R^1_A, R^2_B, and R^3_C respectively, satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C) \setminus (X \cup Y \cup Z)| \leq \alpha^{10^{15}} n/\log(n)^{10^{15}}$
- $\sum X + \sum Y + \sum Z = 0$ *(in G^{ab})*

- *All elements of X, Y and Z are ϕ-generic.*

Then, $H_G[X, Y, Z]$ contains a perfect matching.

Proof. By the definition of R^1, R^2, R^3 being α-slightly-independent, we have $Q^1 \subseteq R^1, Q^2 \subseteq R^2, Q^3 \subseteq R^3$ such that the joint distribution of Q^1, Q^2, Q^3 is that of disjoint α-random subsets of G.

Partition G into a $(1/10^5)$-random subset G_- and disjoint $(1 - 1/10^5)$-random subset G_+, making the choices independently of $R^1, R^2, R^3, Q^1, Q^2, Q^3$. For each $i = 1, 2, 3$, let $R^i_+ = R^i \cap G_+$ to get three $(1/10^5)p$-random subsets. Use Lemma 2.2.17 to pick $Q^1_- \subseteq Q^1 \cap G_-, Q^2_- \subseteq Q^2 \cap G_-, Q^3_- \subseteq Q^3 \cap G_-$, so that their joint distribution is that of disjoint, symmetric $\alpha/10^{10}$-random subsets of G. Let $Q_- := (Q^1_-)_A \cup (Q^2_-)_B \cup (Q^3_-)_C$, $R := R^1_A \cup R^2_B \cup R^3_C$, and $R_+ := (R^1_+)_A \cup (R^2_+)_B \cup (R^3_+)_C$. With high probability, the property in Lemma 2.4.32 holds for Q_-. As the multiplication hypergraph H_G is $(0, 1, n)$-typical, with high probability, the property in Lemma 2.2.9 holds with the random sets $(A', B', C') = (R^1_+, R^2_+, R^3_+)$. Finally, as a consequence of Chernoff's bound, with high probability, $|Q_-| \leq |R_+|/100$, and for i we have $|R^i_+| = (p(1 - /10^5) \pm n^{-0.27})n$ and $|R^i|(p \pm n^{-0.27})n$

Now, given subsets X, Y, Z, define $U := (R^1_A \cup R^2_B \cup R^3_C) \setminus (X \cup Y \cup Z)$ and note $|U| \leq \alpha^{10^{15}} n / \log(n)^{10^{15}}$. From the property in Lemma 2.4.32, we find a set $Q'_- \subseteq Q_- \setminus U$ that can combine with balanced zero-sum sets to produce perfect matchings. By the remark after Lemma 2.4.32, Q'_- is balanced and zero-sum.

Let $L := R \setminus (Q'_- \cup R_+)$. Note that as R^1_+, R^2_+, R^3_+ each have size $(p(1 - /10^5) \pm n^{-0.27})n$, and Q'_- is a balanced set contained in R, there exists a $q \leq p/10^5$ such that $L \cap X$, $L \cap Y$, $L \cap Z$ each have size $(q \pm n^{-0.26})n$. Note that $q \leq (1 - 10^{-5})p/100$, so we may apply Lemma 2.2.9, to conclude that there exists a matching M'_1 with $V(M'_1) \subseteq L \cup R_+$ covering all but $n^{1-10^{-4}}$ vertices of $L \cup R_+$. Of the edges forming this matching, at most $\alpha^{10^{15}} n / \log(n)^{10^{15}}$ many of them can meet U, hence we find a matching M_1 of $H_G[X, Y, Z]$ covering all but at most $\alpha^{10^{15}} n / \log(n)^{10^{15}} + n^{1-10^{-4}} \leq (\alpha/10^{10})^{10^{14}} / \log(n)^{10^{14}}$ vertices of $(X \cup Y \cup Z) \setminus Q'_-$. Call this uncovered set of vertices S.

We claim that S is a balanced zero-sum set. This is as the sets Q'_{-}, $V(M_1)$, and $X \cup Y \cup Z$ are each balanced and zero-sum, and S is obtained by removing all of the former two sets from the latter set, and the former two sets are disjoint. Further, S is ϕ-generic, as by assumption, $X \cup Y \cup Z$ is ϕ-generic. Then, by property of the set Q'_{-}, $S \cup Q'_{-}$ spans a perfect matching, M_2 say. $M_1 \cup M_2$ then is the desired perfect matching of $X \cup Y \cup Z$. □

It remains to prove Theorem 2.3.3. We will need the following intermediary result.

Lemma 2.4.34. *Let $p \geq n^{-1/10^{100}}$. Let G be a group of order n. Let $Q^1, Q^2, Q^3 \subseteq G$ be disjoint, symmetric p-random subsets and set $Q = Q^1_A \cup Q^2_B \cup Q^3_C$. With high probability, the following holds.*

Let $S, U \subseteq V(H_G)$ with $|S|, |U| \leq p^{10^{10}} n / \log(n)^{10^{10}}$ and $e \notin S$. Then, there exists a matching M in $H_G[Q] \setminus U$ of size at most $2p^{10^{10}} n / \log(n)^{10^{10}}$ saturating S.

Proof. Let $g \in G$ be an arbitrary element of the group such that $g \neq e$. Then, by Proposition 2.2.12, it follows that g is contained in at least $n/5$ edges $\{g, x, y\}$ such that $x \notin \{y, y^{-1}\}$, regardless of whether $g \in G_A$, G_B, or G_C. For any i, j, and any such $\{x, y\}$, $x \in Q^i$ and $y \in Q^j$ with probability p^2. For each of the $n/5$ such edges, the corresponding pairs $\{x, y\}$ are disjoint. Hence, we may apply Chernoff's bound to deduce that with probability at least $1 - 1/n^2$, g has degree at least $p^2 n/1000$ in $H_G[Q^1_A, Q^2_B, Q^3_C]$, as long as $g \in V(H_G[Q^1_A, Q^2_B, Q^3_C])$. By a union bound, this property holds for all $g \in G$ such that $g \neq e$ in the case that $G = (\mathbb{Z}_2)^k$ for some k.

Now, let S, U be given. By the minimum degree property from the previous paragraph, all elements of S have degree at least $p^2 n/1000 - 2p^{10^{10}} n / \log(n)^{10^{10}} \geq p^2 n/2000$ in $H_G[Q] \setminus U$. As $p^2 n \gg |S|$, we may greedily pick a matching saturating S and disjoint from U of size $|S| \leq 2p^{10^{10}} n / \log(n)^{10^{10}}$ as claimed. □

We can now give the final proof of the section.

Proof of Theorem 2.3.3 from Theorem 2.4.33 and Lemma 2.4.34. By the definition of q-slightly-independent, we have disjoint, symmetric, q-random sets Q^1, Q^2, Q^3

with $Q^1 \subseteq R^1, Q^2 \subseteq R^2, Q^3 \subseteq R^3$. With high probability Theorem 2.4.33 holds R^1, R^2 and R^3 and Lemma 2.4.34 holds for Q^1, Q^2, Q^3. Given X, Y, Z, let $S = (X \cup Y \cup Z) \setminus (R^1_A \cup R^2_B \cup R^3_C)$ together with the non-ϕ-generic elements of $X \cup Y \cup Z$ (of which there are at most 10^{9010} many). Let $U = (R^1_A \cup R^2_B \cup R^3_C) \setminus (X \cup Y \cup Z)$. Note that $|S|, |U| \leq p^{10^{16}} n/\log(n)^{10^{16}}$ and $e \notin S$ in the case that $G = \mathbb{Z}_2^k$. Apply Lemma 2.4.34 to find a matching M, and set X', Y', Z' to be $X \setminus V(M), Y \setminus V(M), Z \setminus V(M)$ respectively. Observe that X', Y', Z' satisfy the hypothesis of X, Y, Z in Theorem 2.3.3. Indeed, $X'/Y'/Z'$ is contained in $R^1/R^2/R^3$ and does not contain non-ϕ-generic elements since all such elements are in $S \subseteq V(M)$. We have $\sum X' + \sum Y' + \sum Z' = 0$, as $\sum X + \sum Y + \sum Z = 0$, and $\sum V(M) = 0$ (as $V(M)$ spans a perfect matching). We have $|X'| = |Y'| = |Z'|$ for the same reason. Hence, $H_G[X', Y', Z']$ spans a perfect matching by Theorem 2.3.3, and this together with M gives a perfect matching of $H_G[X, Y, Z]$ as desired. □

2.5 Applications

2.5.1 Characterising subsquares with transversals

The goal of this section is to prove a far-reaching generalisation of Snevily's conjecture as stated in the introduction [55].

Preliminaries

The below lemma allows us to find zero-sum sets of prescribed size.

Lemma 2.5.1. *Let $k \in \{3, 4, 5\}$. Let $p \geq n^{-1/100}$. Let G be an abelian group, and let R be a p-random subset of G. Then, the following holds with high probability.*

Let $X \subseteq R$ with $|R \setminus X| \leq p^{8000} n/10^{6000}$. Then, there exists at least $p^{800} n/10^{4010}$ many disjoint zero-sum sets of size k in G.

Proof. With high probability, Lemma 2.2.32 apply to R (with $R = R_A$, set R_B and R_C arbitrarily for this application). Let X be given. If it is the case that $k = 3$, consider the set $S = \{xy, x^{-1}, y^{-1}\} \subseteq G * F_2$ and note that all pairs of words in S are linear and separable (by part (a) of the definition of separable). Thus, setting $U := R \setminus X$ there is a projection $\pi : G * F_2 \to G$ which has $\pi(S) \subseteq X$ and which separates S. Resetting U to include the vertices in the projection, and invoking Lemma 2.2.32 iteratively, we can conclude that there exists at least $(p^{800}n/10^{4000} - p^{8000}n/10^{6000})/10 \gg p^{800}n/10^{4010}$ disjoint zero-sum sets of size 3 contained in X. If $k = 4$ set $S = \{xyz, x^{-1}, y^{-1}, z^{-1}\}$, and if $k = 4$ set $S = \{xyzw, x^{-1}, y^{-1}, z^{-1}, w^{-1}\}$ to obtain sets of words which are linear and pairwise separable, and proceed in the same way as the case $k = 3$. □

For this section, we will only need the following easy corollary of Lemma 2.5.1.

Lemma 2.5.2. *Let G be a group with $|G| = n \geq 10^{50}$, let $3 \leq t \leq n/\log(n)$, and let $g \in G$. Then, there exists distinct $c_1, \ldots, c_t \in G$ such that $\sum c_i = g$ in G^{ab}.*

Proof. Apply Lemma 2.5.1 with $R = G$ $(p = 1)$. Write $t \geq 3$ as a sum of at most $n/\log(n)$ many copies of 3, 4, and 5. By Lemma 2.5.1, we can greedily choose the correct number of disjoint zero-sum sets of size 3, 4, or 5 as $p^{800}n/10^{4010} = n/10^{4010} \gg n/\log(n)$ to produce the desired zero-sum set of size t. □

The below result has essentially appeared in [47]. For our application here, we provide a more precise formulation as well as a full proof for the sake of completeness.

Lemma 2.5.3. *Let n be sufficiently large and set $\gamma := 1/\log(n)^{10^{100}}$. Let $S = A \times B$ be a subsquare of a multiplication table of a group G defined by two n-element sets $A, B \subseteq G$. Then, either S has at most $(1-\gamma)n$ symbols occurring more than $(1-\gamma)n$ times or there is a subgroup $H \subseteq G$ and elements $g, g' \in G$ such that $|A \Delta gH|, |B \Delta Hg'| \leq \gamma^{1/10}n$.*

Proof. We will use a theorem of Fournier as stated as Theorem 1.3.3 in the lecture notes of Green [34]. Towards that goal, we recall the definition of *multiplicative energy* of a subset A of a group. It is defined to be the quantity $E(A) = |\{(a_1, a_2, b_1, b_2) \in A^4 \colon a_1 a_2^{-1} = b_1 b_2^{-1}\}|$.

Suppose now that S has more than $(1-\gamma)n$ symbols occurring more that $(1-\gamma)n$ times. Call a subset S' of S a *monoversal* if all entries in S' are equal. By properties of groups, no two entry in a monoversal can share a row or a column. Our assumption is then equivalent to stating that there exists at least $(1-\gamma)n$ monoversals S' each of size $(1-\gamma)n$. Label these monoversals $S_1, \cdots, S_{(1-\gamma)n}$. Call an ordered pair of rows of S, (a_1, a_2) say, *good*, if there exists at least $(1-\sqrt{\gamma})n$ many monoversals S_i meeting both a_1 and a_2. Let α denote the number of good pairs of ordered rows. Let K denote the the number of tuples $(i, (s_1, s_2))$ where $i \in [(1-\gamma)n]$ and $\{s_1, s_2\} \subseteq S_i$, and observe that $K \geq (1-\gamma)n(1-\gamma)n((1-\gamma)n-1) \geq (1-5\gamma)n^3$. On the other hand, $\alpha n + (n^2-\alpha)(1-\sqrt{\gamma})n \geq K$. Combining the two inequalities, we find that $\alpha \geq (1-10\sqrt{\gamma})n^2$.

For each good ordered pair of rows, (a_1, a_2) say, we claim there exists $(1-5\sqrt{\gamma})n^2$ distinct elements of the set $\{(b_1, b_2, b_3, b_4) \in B^4 \colon b_1b_2^{-1} = b_3b_4^{-1}\}$. Indeed, for each monoversal meeting both a_1 and a_2, we have an ordered pair (b_1, b_2) and an equality of the form $a_1b_1 = a_2b_2$, equivalently $a_2^{-1}a_1 = b_2b_1^{-1}$. Hence, we have $(1-\sqrt{\gamma})n$ distinct ordered pairs $(b_1, b_2) \in B^2$ for which $b_2^{-1}b_1$ is constant. For each ordered pair of such ordered pairs, of which there are $(1-\sqrt{\gamma}n)^2 \geq (1-5\sqrt{\gamma})n^2$ many, we get a distinct element of $\{(b_1, b_2, b_3, b_4) \in B^4 \colon b_1b_2^{-1} = b_3b_4^{-1}\}$, as claimed.

Summing over all of the good ordered pair of rows, this gives $(1-10\sqrt{\gamma})n^2(1-5\sqrt{\gamma})n^2 \geq (1-20\sqrt{\gamma})n^4$ potential elements of $\{(b_1, b_2, b_3, b_4) \in B^4 \colon b_1b_2^{-1} = b_3b_4^{-1}\} = X$. We double count a 4-tuple (b_1, b_2, b_3, b_4) of X in two distinct ordered pair of rows (a_1, a_2) and (a_1', a_2') only if $a_2^{-1}a_1 = {a_2'}^{-1}a_1' = b_1b_2^{-1}$, and there are at most n such distinct ordered pair of rows. We deduce that $E(B) \geq (1-20\sqrt{\gamma})n^3$. By a symmetric argument, $E(A) \geq (1-20\sqrt{\gamma})n^3$ as well.

From Theorem 1.3.3 in [34], it follows immediately that there exists a subgroup H and $g \in G$ such that $|A \Delta gH| \leq \gamma^{1/9}n$. Consider an element t of S that repeats at least $(1-\gamma)n$ times. So we have $a_1b_1 = \cdots = a_{(1-\gamma)n}b_{(1-\gamma)n} = t$. Let $i \in [(1-\gamma)n]$ so that $a_i \in gH$, and note that there are at least $(1-2\gamma^{1/9})n$ indices i with this property. So we have that $t = (gh)b_i$ for some $h \in H$. Then, $b_i = h^{-1}g^{-1}t$, so $b_i \in H(g^{-1}t)$ for each such i. Then, the statement holds for $g' = g^{-1}t$.

□

Observation 2.5.4. *Let $A \times B$ be a subsquare of a multiplication table of a group G defined by two n-element sets $A, B \subseteq G$. Let $g, g' \in G$. Suppose that $A \times B$ contains a transversal. Then, $gA \times Bg'$ also contains a transversal.*

Proof. Let $(a_i, \phi(a_i))$, $i \in [n]$ be a transversal of $A \times B$, where ϕ is a bijection $A \to B$. Then, $(ga_i, \phi(a_i)g')$ is a transversal of $gA \times Bg'$, since $a_i\phi(a_i) \neq a_j\phi(a_j)$ implies that $ga_i\phi(a_i)g' \neq ga_j\phi(a_j)g'$. □

Lemma 2.5.5. *Let $\varepsilon > 0$ be sufficiently small and let n be sufficiently large. Let $\gamma \geq n^{-\varepsilon}$. Let S be a subsquare of a group G of order n such that S has at most $(1-\gamma)n$ symbols occuring more than $(1-\gamma)n$ times. Then, S contains a transversal.*

Proof. As subsquares of G correspond to edge-coloured balanced complete graphs, this lemma is a direct corollary of Lemma 8.1.3 from [47]. □

The characterisation

We first prove the following weakening of the main theorem which characterises when subsquares which are close to the whole multiplication table have transversals.

Lemma 2.5.6. *Let $t \leq n/\log(n)^{10^{30}}$. Let G be a sufficiently large group and $A, B \subseteq G$ with $|A| = |B| = n - t$. Then $A \times B$ has a transversal unless one of the following holds*

1. *H is a group that does not satisfy the Hall-Paige condition, and $A = B = H$.*
2. *$H \cong (\mathbb{Z}_2)^k$, $A = G \setminus \{a_1, a_2\}$, $B = G \setminus \{b_1, b_2\}$ for some distinct $a_1, a_2 \in H$ and distinct $b_1, b_2 \in H$ such that $a_1 + a_2 + b_1 + b_2 = 0$.*

Proof. If $t = 0$ then there is a transversal (by any proofs of the Hall-Paige Conjecture e.g. by Theorem 2.3.4 with $p = 1$). Thus, we can suppose that $1 \leq t \leq n/\log(n)^{10^{30}}$. Let $\alpha := \prod G(\prod A \prod B)^{-1}$ (where the terms in the products are multiplied in any fixed order).

Claim 2.5.6.1. *There are distinct elements $c_1, \dots, c_t \in G$ such that $c_1 + \dots + c_t = \alpha$ in H^{ab}.*

Proof. If $t = 1$, one can trivially choose $c_1 = \alpha$. If $t = 2$, and $\alpha \neq e$, then we could select $c_1 = e$ and $c_2 = \alpha$. The case $t = 2$, $\alpha = e$, $G = (\mathbb{Z}_2)^k$ is excluded by the lemma (indeed in this case we'd have $A = G \setminus \{a_1, a_2\}$, $B = G \setminus \{b_1, b_2\}$ for some distinct $a_1, a_2 \in H$ and distinct $b_1, b_2 \in H$. But then in G^{ab} we'd have $0 = \alpha = \sum H - \sum A - \sum B = a_1 + a_2 + b_1 + b_2 - \sum (\mathbb{Z}_2)^k = a_1 + a_2 + b_1 + b_2$). When $t = 2$, $\alpha = e$, $G \neq (\mathbb{Z}_2)^k$, then, there must exist distinct $c_1, c_2 \in G$ such that $c_1 c_2 = e$. Finally, when $t \geq 3$, then the claim follows from Lemma 2.5.2. □

Let $C = G \setminus \{c_1, \dots, c_t\}$ to get a set with $|C| = |A| = |B|$. Now Theorem 2.3.4 applies with $p = 1$ to give a perfect matching $\{(a_1, b_1 c_1'), \dots, (a_{n-1}), b_{n-t}, c_{n-t}'\}$ in $H_G[A, B, C^{-1}]$. The entries $(a_1, b_1), \dots, (a_{n-t}, b_{n-t})$ give a transversal in $A \times B$. □

We now prove the main result of this section.

Theorem 2.5.7. *There exists a $n_0 \in \mathbb{N}$ such that the following holds for all $n \geq n_0$. Let G be a group, and let $A, B \subseteq G$ with $|A| = |B| = n$. Then, $A \times B$ has a transversal, unless there exists some $k \geq 1$, $g_1, g_2 \in G$ and a subgroup $H \subseteq G$ such that one of the following holds.*

1. *H is a group that does not satisfy the Hall-Paige condition, and $A = g_1 H$ and $B = H g_2$.*

2. *$H \cong (\mathbb{Z}_2)^k$, $g_1 A \cong H \setminus \{a_1, a_2\}$, $g_2 B \cong H \setminus \{b_1, b_2\}$ for some distinct $a_1, a_2 \in H$ and distinct $b_1, b_2 \in H$ such that $a_1 + a_2 + b_1 + b_2 = 0$.*

Proof of Theorem 2.5.7. Set $\gamma = 1/\log(n)^{10^{100}}$. Suppose that $A \times B$ does not have a transversal. Then, by Lemma 2.5.5, we may assume that $A \times B$ contains more than $(1-\gamma)n$ symbols occuring more than $(1-\gamma)n$ times. By Lemma 2.5.3, it follows that there is a subgroup $H \subseteq G$ and elements $g, g' \in G$ such that $|A \Delta gH|, |B \Delta Hg'| \leq \gamma^{1/10} n$. As $A \times B$ does not contain a transversal, by Observation 2.5.4, $g^{-1} A \times B g'^{-1}$

does not contain a transversal either. Set $A' \times B' = g^{-1}A \times Bg'^{-1}$ and observe that $|A'\Delta H|, |B'\Delta H| \leq \gamma^{1/10}n$.

Set $A_1 = A' \cap H$, $A_2 = A' \setminus H$, $B_1 = B' \cap H$, $B_2 = B' \setminus H$, noting that $|A_2|, |B_2| \leq \gamma^{1/10}n$. If $A_2 = B_2 = \emptyset$, then the theorem follows from Lemma 2.5.6 applied with $G = H$. Thus we can suppose that A_2 and/or B_2 are nonempty. Note that all the elements in the multiplication table in $A_1 \times B_2$ and $A_2 \times B_1$ are outside H. Let $A_2 = \{a_1, \dots, a_{|A_2|}\}$, $B_2 = \{b_1, \dots, b_{|B_2|}\}$. We can greedily select a partial transversal $T_1 = \{(a_1, b'_1), \dots, (a_{|A_2|}, b'_{|A_2|})\ ,\ (a'_1, b_1), \dots, (a'_{|B_2|}, b_{|B_2|})\}$ by selecting elements $b'_1, \dots, b'_{|A_2|} \in B_1, a'_1, \dots, a'_{|B_2|} \in A_1$ in order (to see this note that there are at least $\min(|A_1|, |B_1|) \geq n/2$ choices for each element and so there's room to avoid the $|A_2| + |B_2| \leq \gamma^{1/10}n$ rows/columns/symbols previously used). Note that since there are at least $n/4$ choices for the last element $a'_{|B_2|}$, we can additionally ensure that $\sum A_1 \setminus \{a'_1, \dots, a'_{|B_2|}\} + \sum B_1 \setminus \{b'_1, \dots, b'_{|A_2|}\} \neq 0$ in G^{ab} in the case where G^{ab} has at least 100 elements. Thus Lemma 2.5.6 applies to give a transversal T_2 in $(A_1 \setminus \{a'_1, \dots, a'_{|B_2|}) \times (B_1 \setminus \{b'_1, \dots, b'_{|A_2|}\})$ (to apply Lemma 2.5.6, we need to know that we are not in cases 1 and 2. We're not in case 1 because we're assuming $A_2 \cup B_2 \neq \emptyset$. We're not in case 2 because in this case we have $|G^{ab}| \geq 100$, and also that $\sum A + \sum B = 0$, and we selected a'_i and b'_i to avoid this scenario). Now $T_1 \cup T_2$ is a transversal in $A' \times B'$ as required. □

2.5.2 Path-like structures in groups

Given a group G, by K_G^+ we denote the complete directed edge-coloured graph with vertex set G, edge set $\{\vec{ab} \colon a \neq b \text{ and } a, b \in G\}$, where the edge $\vec{ab}$ gets assigned the colour $ab \in G$. We call this the **multiplication digraph of** G. Similarly, by K_G^- we denote the **division digraph of** G. In the division digraph, the edge $\vec{ab}$ gets assigned the colour $a^{-1}b \in G$, and all other properties of K_G^- are same with those of K_G^+. We sometimes use the notation $K_G^\pm$ to make statements and definitions about K_G^+ and K_G^- simultaneously. For subsets $R, R' \subseteq G$, we will use $K_G^\pm[R; R']$ to denote the subgraph of $K_G^\pm$ induced on vertex set R consisting of all edges of colours in R'.

For disjoint subsets $V_1, V_2 \subseteq G$, by $K_G^{\pm}[V_1, V_2; R']$ we denote the bipartite subgraph of $K_G^{\pm}$ obtained by keeping only the edges between V_1 and V_2 with colour in R'.

The purpose of this section is to prove the following lemma, which has several further applications as discussed in [49]. We include the lemma here as it plays an important role in the proof of the Friedlander-Gordon-Tannenbaum conjecture.

Lemma 2.5.8. *Let $1/n \ll p \leq 1$, let t be a positive integer between $\log^7(n)$ and $\log^8(n)$, and let q satisfy $p = (t-1)q$. Let G be a group of order n. Let $V_{str}, V_{mid}, V_{end}$ be disjoint random subsets with V_{str}, V_{end} q-random and V_{mid} p-random. Let C be a $(q+p)$-random subset, sampled independently with the previous sets. Then, with high probability, the following holds.*

Let V'_{str}, V'_{end}, V'_{mid} be disjoint subsets of G, let C' be a subset of G, and let $\ell = |V'_{mid}|/(t-1)$. Suppose all of the following hold.

1. *For each random set $R \in \{V_{str}, V_{mid}, V_{end}, C\}$, we have that $|R \Delta R'| \leq n^{0.6}$.*

2. *Either **O.** $\sum V'_{end} - \sum V'_{str} = \sum C'$ or **C.** $\sum V'_{end} + \sum V'_{str} + \sum V'_{mid} + \sum V'_{mid} = \sum C'$ holds in the abelianization of G.*

3. *$e \notin C'$ if G is an elementary abelian 2-group.*

4. *$\ell := |V'_{str}| = |V'_{end}| = |V'_{mid}|/(t-1) = |C'|/t$*

*Then, given any bijection $f \colon V'_{str} \to V'_{end}$, we have that $K_G^{*}[V'_{str} \cup V'_{end} \cup V'_{mid}; C']$ has a rainbow $\vec{P}_t$-factor where each path starts on some $v \in V'_{str}$ and ends on $f(v) \in V'_{end}$ where $* = +$ if **C** holds and $* = -$ if **O** holds.*

The main trick in this section is to use our main theorem iteratively to build paths out of matchings. One key issue with this idea is that this does not give us to freedom to construct paths connecting specified end-points, which is critical for building a Hamilton path, instead of an arbitrary path/cycle-factor. To remedy this, in Section 2.5.2, based on ideas of Kühn, Lapinskas, Osthus, and Patel [42], we introduce a way of building path systems allowing us to construct path-factors with specified end-points.

Sorting networks

In [42] (see in particular Lemma 4.3), an ingenious method was introduced in order to construct path systems which can connect specified endpoints. The key idea is to consider use an appropriate *sorting network* as a template while building the path system. In this section, we adapt the arguments from [42] to our context. First, we introduce some terminology. For a more detailed treatment, we refer the reader to [14].

Definition 2.5.9. A *comparison network* is a union of four types of objects: input nodes $x_1, \ldots, x_m$, output nodes $y_1, \ldots, y_m$, comparators $C_1, \ldots, C_t$ and wires $w_1, \ldots, w_s$.

- Comparators are sets of 4 nodes $C_i = \{y_i^-, y_i^+, x_i^-, x_i^+\}$ (which are disjoint with the input and output nodes).
- Each wire joins an x-node to a y-node. Additionally, each node is in precisely one wire, and the directed graph formed by contracting comparators into single nodes is acyclic.

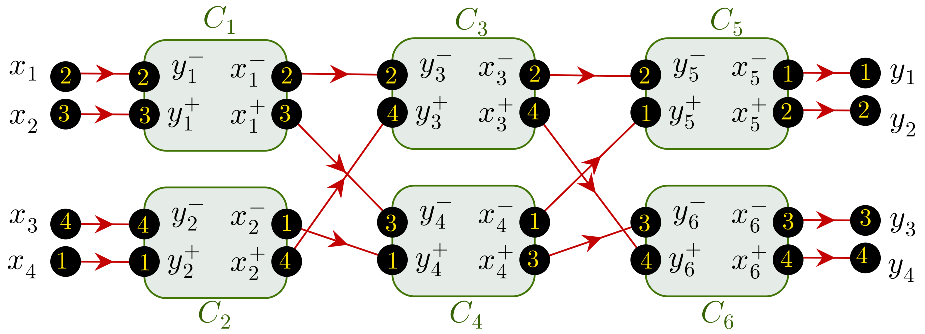


Figure 2.10: A comparison sorting network for sorting four numbers. The black circles represent nodes, the green rectangles are comparators, and the red arrows are wires. Here the network was given input values $v(x_1) = 1, v(x_2) = 3, v(x_3) = 4, v(x_4) = 1$ (represented by the yellow numbers inside those nodes). Then all other nodes get a value based on the rules in Definition 2.5.10 (represented by the yellow numbers in the other nodes). The network correctly sorted the numbers, which can be seen by the fact that each y_i contains yellow number i.

Definition 2.5.10. A *comparison sorting network* is a comparison network with the following additional property. Let σ be any permutation of $[m]$. Assign each x_i the value $v(x_i) = \sigma(i)$. Assign the values of the other nodes via the following rules.

1. If xy is a wire then $v(y) := v(x)$.

2. If $C_i = \{y_i^-, y_i^+, x_i^-, x_i^+\}$ is a comparator, then $v(x_i^-) := \min(v(y_i^-), v(y_i^+))$ and $v(x_i^+) := \max(v(y^-), v(y^+))$.

Then, all nodes get assigned a value and moreover, $v(y_i) = i$ for $i = 1, \dots, m$.

See Figure 2.10 for an example of these definitions. A classical result due to Batcher [9] states that for all $m \in \mathbb{N}$, there is a sorting network with m input/outputs and $100m \log^2 m$ comparators. In fact, there are sorting networks with $O(m \log m)$ comparators thanks to a celebrated result of Ajtai, Komlós, and Szemerédi [1] but we will not need this sharper bound here. However, it will be convenient for us to have sorting networks with symmetry in the following sense.

Lemma 2.5.11 ([1]). *For all $m \in \mathbb{N}$, there is a sorting network such that the length of every path from x_i to y_j is exactly $\lceil 100 \log^2 m \rceil$ (in the directed graph formed by contracting every comparator into a single node).*

The above can be proved by inspecting any common method of constructing a sorting network, for example the method of Batcher [9]. Indeed, the bound of Batcher is in terms of the *depth* of the network as opposed to the total number of comparators, so we can simply add redundant comparators to ensure the conclusion of Lemma 2.5.11.

We now show how to simulate the task of a comparator in a sorting network via a collection of paths.

Lemma 2.5.12. *Let $p \geq n^{-1/700}$. Let R^1, R^2 be disjoint p-random subsets of G.*

With high probability, for any $U \subseteq G$ with $|U| \leq p^{800} n / 10^{4010}$, there is a subgraph $C \subseteq K_G^{\pm}[R_1 \setminus U; R_2 \setminus U]$ consisting of 12 vertices and 10 colours containing vertices x^-, x^+, y^-, y^+ and directed paths $Q_{x^-,y^-}, Q_{x^+,y^+}, Q_{x^-,y^+}, Q_{x^+,y^-}$ with each $Q_{x,y}$ having *length 5 and going from x to y. Additionally the vertices and colours of the path pairs $(Q_{x^-,y^-}, Q_{x^+,y^+})$ and $(Q_{x^-,y^+}, Q_{x^+,y^-})$ both partition the 12 vertices and 10 colours of C.*

Proof. We prove the lemma when $K_G^{\pm} = K_G^{+}$. A slight change in variables in Figure 2.11 proves the lemma also when $K_G^{\pm} = K_G^{-}$. With high probability, Lemma 2.2.32 applies. Thinking of x, y, a, b, c, d, f, as free variables consider the set

$$S = \{x, y, a, b, xd, yd, d^{-1}b, d^{-1}, c, xf, yf, f^{-1}a, f^{-1}c, xa, yb, axd, byd, xb, ya, xc, yc, d^{-1}bxf, d^{-1}cyf\}$$

(see Figure 2.2). Note that all pairs w, w' are linear and separable (by (a), since they're all linear in different combinations of free variables). Lemma 2.2.32 gives a projection π which separates S and has $\pi(S) \subseteq R$. This means that $\pi(w)$ are distinct for all $w \in S$. Now the graph given in Figure 2.11 satisfies the lemma with $x^- = \pi(x), x^+ = \pi(y), y^- = \pi(f^{-1}a), y^+ = \pi(f^{-1}c)$, with paths as shown. □

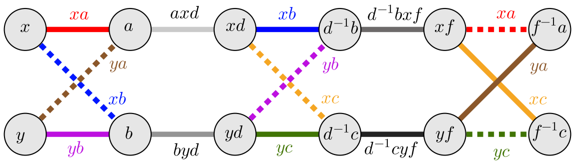


Figure 2.11: The coloured graph produced by Lemma 2.5.12 for K_G^+. Each edge is directed towards the right. Notice that all vertices/edges are labelled by elements of S. Since all $w, w' \in S$ are separable, this means that the π-image of this graph has all vertices/colours distinct (and so in particular has 12 vertices and 10 colours as required). To see that it satisfies the lemma, we need to exhibit paths $Q_{x^-,y^-}, Q_{x^+,y^+}, Q_{x^-,y^+}, Q_{x^+,y^-}$ between $x^- = x, x^+ = y, y^- = f^{-1}a, y^+ = f^{-1}c$. The solid lines in the picture give the two paths Q_{x^-,y^+}, Q_{x^+,y^-}. Replacing the coloured solid lines for the coloured dashed lines (and keeping all grey lines) gives the two paths Q_{x^-,y^-}, Q_{x^+,y^+}.

We now show how to simulate the task of a wire in a sorting network via short paths.

Lemma 2.5.13. *Let $p \geq n^{-1/700}$. Let H_G be a multiplication hypergraph, R^1, R^2 disjoint p-random subsets of G. With high probability, for any $x, y \in V(K_G^{\pm})$, $U \subseteq G$*

with $|U| \leq p^{800}n/10^{4010}$, there is a length 3 x to y path $xuvy$ in $K_G^{\pm}$ with $u, v \in R_1 \setminus U, c(xv), c(vy) \in R_2 \setminus U$.

Proof. We prove the lemma when $K_G^{\pm} = K_G^{+}$. A slight change in the upcoming definition of S proves the lemma also when $K_G^{\pm} = K_G^{-}$.

With high probability, Lemma 2.2.32 applies. Consider the set $S = \{u, v, uv, xu, uy\} \subseteq G * F_3$. Note that all pairs w, w' are linear also the pairs $(u, v), (xu, uy), (xu, uv), (uv, uy)$ are separable (since they're all linear in different combinations of free variables). Lemma 2.2.32 gives a projection π which separates S. Now the path $P = x\pi(u)\pi(v)y$ satisfies the lemma — the vertices are distinct since u, v were separable, while the colours are distinct since uv, xu, uy were pairwise separable. □

We now prove the analogue of Lemma 4.3 from [42] adapted to a setting where the host structure is $K_G^{\pm}$.

Lemma 2.5.14. *Let $p \geq n^{-1/800}$. Let $t = 8\lceil 100 \log^2 n \rceil$. Let R_V and R_C be p-random subsets of G, sampled independently. Then, the following holds with high probability.*

Let $A, B \subseteq G$ be subsets with $|A| = |B| \leq \frac{p^{1000}n}{10^{5000}\log(n)^3}$, and let $U \subseteq G$ with $|U| \leq \frac{p^{1000}n}{10^{5000}}$. Then, there exists $V \subseteq R_V \setminus U$ and $C \subseteq R_C \setminus U$ such that for any bijection $\phi\colon A \to B$, there exists a system of paths using exactly the vertices/colours of $K_G^{\pm}[A \cup B \cup V; C]$ from A to B, each of length t, and connecting a to $\phi(a)$ for each $a \in A$.

Proof. With high probability, Lemmas 2.5.12 and 2.5.13 apply. Let N be a sorting network given by Lemma 2.5.11, with $m := |A| = |B|$, noting this sorting network has $\leq 200m \log^2 m$ comparators. For each comparator $C_i = \{x_i^-, x_i^+, y_i^-, y_i^+\}$, use Lemma 2.5.12 to find a subgraph C_i' in $R \setminus (A \cup B)$. Identify the nodes $x_i^-, x_i^+, y_i^-, y_i^+$ of the comparator with the vertices $x_i^-, x_i^+, y_i^-, y_i^+$ of C_i'. Let $A = \{x_1, \ldots, x_m\}, B = \{y_1, \ldots, y_m\}$. For each wire xy of the sorting network use Lemma 2.5.13 to find a rainbow length 3 path P_{xy} joining corresponding vertices of $A \cup B \cup \bigcup C_i$. By enlarging the set U at all these applications, we can ensure that the subgraphs C_i' are all disjoint, and that the paths P_{xy} are all internally disjoint and colour-disjoint from the subgraphs C_i' and from each other. We claim that $V = A \cup B \cup \bigcup V(C_i) \cup \bigcup V(P_w)$ and $C = \bigcup C(C_i) \cup \bigcup C(P_w)$ satisfy the lemma.

Consider a bijection $\phi \colon A \to B$. This gives a permutation σ of $[m]$ so that $\phi(x_i) = y_{\sigma(i)}$. Assign value $v(a_i) := \sigma(i)$. This gives a value to each node and wire of the sorting network as in Definition 2.5.10. We now translate this into values for the corresponding paths/vertices in $K_G^{\pm}$. The values we assign come from the set $\{1, \dots, m\} \cup \{0\}$. For each wire xy of the sorting network, define $v(P_{xy}) = v(x) = v(y)$ and give all vertices/edges of P_{xy} this value. For each comparator $C = \{y_i^-, y_i^+, x_i^-, x_i^+\}$ we have either $v(y_i^-) = v(x_i^-), v(y_i^+) = v(x_i^+)$ or $v(y_i^-) = v(x_i^+), v(y_i^+) = v(x_i^-)$. In the former case, define $v(Q_{y_i^-, x_i^-}) = v(y_i^-) = v(x_i^-)$, $v(Q_{y_i^+, x_i^+}) = v(y_i^+) = v(x_i^+)$, and give the vertices and edges of the corresponding paths the same value. Give the paths $Q_{y_i^-, x_i^+}$ and $v(Q_{y_i^+, x_i^-})$ as well as the unused edges of C_i' (those edges of C_i' not on $Q_{y_i^-, x_i^-}$ or $Q_{y_i^+, x_i^+}$) value 0. In the latter case, define $v(Q_{y_i^-, x_i^+}) = v(y_i^-) = v(x_i^+)$, $v(Q_{y_i^+, x_i^-}) = v(y_i^+) = v(x_i^-)$, and give the vertices and edges of the corresponding paths the same value. As before, give the other two paths and the unused edges value 0. Note that this way every vertex/edge of U gets a value, and these values match those that corresponding nodes/wires have in N. For every $i = 1, \dots, m$, let

$$P_i = \bigcup_{v(P_{xy}) = i} P_{xy} \cup \bigcup_{v(Q_{y_i^{\diamond_1}, x_i^{\diamond_2}}) = i} Q_{y_i^{\diamond_1}, x_i^{\diamond_2}}.$$

We clarify that above xy is quantified over the set of wires, and $\diamond_j$ is quantified over $\{+, -\}$. We claim that $P_1, \dots, P_m$ are each paths and have all the required properties.

First note that every vertex in $V \setminus (A \cup B)$ has exactly one in-going edge of non-zero value and exactly one outgoing edge of non-zero value. The vertices a_i have no in-going edges and one out-going edge (whose value is $\sigma(i)$). The vertices b_i have no out-going edges and one in-going edge (whose value is i). Combined with the whole graph being acyclic (which holds due to the sorting network being acyclic), this shows that $\bigcup P_i$ is a union of paths.

Since $v(a_i) = \sigma(i)$ and $v(b_i) = i$, path $P_{\sigma(i)}$ goes from a_i to $b_{\sigma(i)}$. In particular, each P_i is a path. Also, this shows that the paths partition the vertices V and have the correct endpoints. We now show that their union is rainbow using exactly the colours C. The fact that every colour of $\bigcup C(P_{xy})$ is used exactly once comes from the fact that every wire has a value, and so every edge of each P_{xy} as a value. So such colours are used at least once (and hence exactly once because these colours occur once in the whole graph). The colours on the comparators C'_i are used once as a consequence of Lemma 2.5.12. To see this, note each such colour comes up exactly twice — once in the paths $Q_{x_i^-,y_i^-} \cup Q_{x_i^+,y_i^+}$ and once in the paths $Q_{x_i^-,y_i^+} \cup Q_{x_i^+,y_i^-}$. By the assignment of the values to the comparator one of these always has value 0.

Finally, to see that each P_i has length exactly t, observe that by Lemma 2.5.11 we know each P_i has length $\lceil 100\log^2 n\rceil$ when viewed as a path in the sorting network. As each wire gadget corresponds to a path of length 3 and each comparator corresponds to a path of length 5, it follows that each P_i has length $8\lceil 100\log^2 n\rceil$, as required. □

Deducing Lemma 2.5.8

We will need the following technical lemma, which allows us to use the nibble method to shrink sets of large deterministic vertices. This is necessary, as after applying Lemma 2.5.14 to a random subset, we will be left with a large set of deterministic vertices which do not immediately fit into the setting of our main theorem.

Lemma 2.5.15. *Let $a, b, c \geq n^{-1/10^{-20}}$ and let $\ell \in \mathbb{N}$ be such that $\ell \geq \max\{an, bn, cn\} - n^{0.7}$ and letting $(x, y, z) = (\ell - an, \ell - bn, \ell - cn)$ suppose that we have $x + y \leq cn/2$, $x + z \leq bn/2$ and $y + z \leq an/2$. Let $A, B, C \subseteq G$ be a, b, c-random subsets of G respectively, sampled with A and B disjoint, and C independent of A, B. Let $k := \lfloor n^{1-10^{-5}}\rfloor$. Then, with probability at least $1 - 1/n$ the following holds.*

Let $A', B', C' \subseteq G$ with $|B\backslash B'|, |A\backslash A'|, |C\backslash C'| \leq n^{0.78}$, $|C'| - k = |A'| = |B'| = \ell$.

Then, there is a perfect directed C'-matching in $K^{\pm}_G[A', B'; C']$.

To prove the above lemma, we will make use of the following result of Montgomery, Pokrovskiy, and Sudakov. We use the notation $x \overset{\text{POLY}}{\ll} y$ to mean that $x, y \in (0, 1]$ and there is some absolute constant $C \geq 1$ such that the proof works with $x \leq y^C/C$. Recall that an edge-coloured graph is **globally K-bounded** if each colour occurs at most K times in the colouring. When we say an edge-coloured bipartite graph is typical, we refer to the typicality of the underlying uncoloured bipartite graph.

Lemma 2.5.16 ([47], Corollary 8.12)**.** *Let n, δ, p, γ be such that $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} p, \delta \leq 1$. Every properly coloured, (γ, δ, n)-typical, globally $(1 - p)\delta n$-bounded, balanced bipartite graph G of order $2n$ has $(1 - p)\delta n$ edge-disjoint rainbow perfect matchings.*

Proof of Lemma 2.5.15. Let $\ell' = \max\{an, bn, cn\}$ noting that $\ell' \leq \ell + n^{0.7}$. Set $p = \ell'/n$ and $q = k/n$. Note that combining the given inequalities we obtain that $a + b + c \geq 12(\ell/n)/5 \geq 11p/5$. We define the following random sets.

- Let $A_1, A_2 \subseteq A$ be disjoint $p - b$ and $(p - c + q)$-random subsets of A. Let $A_3 := A \setminus A_1 \setminus A_2$ noting that this set is $(a - 2p + b + c - q) =: \alpha$-random.
- Let $B_1, B_2 \subseteq B$ be $(p - a)$ and $(p - c + q)$-random subsets of B. Let $B_3 := B \setminus B_1 \setminus B_2$ noting that this set is $(b - 2p + a + c - q) = \alpha$-random.
- Let $C_1, C_2 \subseteq B$ be $(p - a)$ and $(p - b)$-random subsets. Let $C_3 := C \setminus C_1 \setminus C_2$ noting that this set is $(c - 2p + a + b) = \beta$-random.

Note there is space to sample these sets disjointly due to the assumptions on the size of ℓ and a, b, c. In particular, $\alpha, \beta \geq p/5 - q \geq n^{-1/10^{-20}}/100$ and $p - a, p - c + q, p - b \geq 0$ so the random sets with these parameters are well-defined. For each pair of random sets (A_2, B_2), (A_1, C_2), and (B_1, C_1) such that the corresponding randomness parameter is at least $n^{-1/600}$, we have that with probability at least $1 - 5/n^{1.5}$, Lemma 2.2.8 holds (the linearity and typicality of the corresponding 3-uniform hypergraph of $K^{\pm}_G$ follows by a straightforward modification of Observation 2.2.3, which we omit). By Chernoff's bound, the sizes of all these random sets are at most $n^{0.6}$ elements away from their expectations with probability at least $1 - 1/n^5$.

Set $\gamma := n^{-0.25}$. For any $v \in G$, the expected number of (out-)neighbours of v in $K_G^{\pm}[\{v\} \cup B_3; C_3]$ is $\alpha\beta(n-1)$, as C is sampled independently with B and the analogous statement holds for expected number of neighbours of v in $K_G^{\pm}[A_3 \cup \{v\}; C_3]$. By Chernoff's bound and a union bound, both of these random variables are at most γn away from their expectation with probability at least $1 - 1/n^3$. Let $d_A(a, a')$ denote the pair degree of a and a' in $K_G^{\pm}[\{a, a'\} \cup B_3; C_3]$ (viewed as an uncoloured bipartite graph), and let $d_B(b, b')$ denote the pair degree of b and b' in $K_G^{\pm}[A_3 \cup \{b, b'\}; C_3]$. For each a, a' and b, b', we have that $\mathbb{E}(d_A(a, a')) = \mathbb{E}(d_B(b, b')) = \alpha\beta^2 n$. Further, observe that $d_A(a, a')$ and $d_B(b, b')$ are 2-Lipschitz random variables. Hence, by Azuma's inequality and a union bound, with probability at least $1 - 1/n^3$, for each a, a' and b, b', $d_B(b, b') = d_A(a, a') = (\alpha\beta^2 \pm \gamma)n$. This establishes in particular that $K_G^{\pm}[A_3, B_3; C_3]$ is $(\gamma, \beta, \alpha n)$-typical as an uncoloured bipartite graph. Let $d(c)$ denote the number of times c occurs in $K_G^{\pm}[A_3, B_3, \{c\}]$. Then $\mathbb{E}(d(c)) \leq \alpha^2 n$ for each c. By Chernoff's bound and a union bound, $d(c) \leq (1 \pm \gamma)\alpha^2 n$ for each $c \in G$. In particular, this implies that $K_G^{\pm}[A_3, B_3, C_3]$ is globally $(\alpha^2 + \gamma)n$-bounded, which implies that it is globally $(1-q)\beta(\alpha n)$-bounded.

With probability $\geq 1 - 1/n$, all of the previous properties hold. Let A', B', C' be given with the indicated properties. By properties coming from Lemma 2.2.8, $(A_2, B_2, C' \setminus C)$, $(A_1, B' \setminus B, C_2)$ and $(A' \setminus A, B_1, C_1)$ each contain matchings covering all but $n^{1-1/499}$ vertices. Here, if the corresponding random sets have parameters smaller than $n^{-1/600}$ (so we cannot apply Lemma 2.2.8) we simply take an empty matching. This is sufficient as in this case by Chernoff's bound, the random sets themselves cannot contain more than $n^{1-1/499}$ elements. Accounting for differences $A \setminus A'$, $B \setminus B'$, $C \setminus C'$ (each of size $\leq n^{0.78}$)., we have a matching covering all but $n^{1-1/498}$ vertices of $(A_2 \cap A', B_2 \cap B', C' \setminus C)$, $(A_1 \cap A', B' \setminus B, C_2 \cap C')$ and $(A' \setminus A, B_1 \cap B', C_1 \cap C')$.

The set of leftover vertices of A', B' and C' have a symmetric difference with A_3, B_3 and C_3 (respectively) of size at most $n^{1-1/495}$. By the pair-degree and vertex-degree bounds we obtained earlier, this implies that the associated properly coloured bipartite graph with the leftover vertices is $(\gamma', \beta, \alpha n)$-typical and globally $(1 - q/10)\beta(\alpha n)$-bounded where $\gamma' := n^{-1/498}$. Then, by Lemma 2.5.16 we can find a matching saturating the remaining vertices of A' and B' using the leftover colours from C' as desired. □

We can now give the proof of the key lemma of this section.

Proof of Lemma 2.5.8. Set $m = 8\lceil 100 \log^2 n \rceil$, $k = \lfloor n^{1-10^{-5}} \rfloor$, $r = 10^{-100}$.

Partition V_{mid} into random subsets $R_V, V_1, \dots, V_{t-m}$ where R_V is rp-random, V_1, V_2, V_3 and V_{t-m} are q-random and the rest are $((1-r)p - 4q)/(t-m-4)$-random. Note that $9q/10 \le ((1-r)p - 4q)/(t-m-4) \le q$. Independently with the previous sets, partition C into $R_C, C_1, \dots, C_{t-m}$ where R_C is rp-random, C_1, C_2 and C_{t-m} are q-random, and the rest are $((1-r)p-3q)/(t-m-3)$-random. Note similarly that $9q/10 \le ((1-r)p - 3q)/(t-m-3) \le q$.

With high probability, Theorem 2.3.9 applies with the q-random sets (V_2, V_3, C_2), Lemma 2.5.14 applies with R_V, R_C and Lemma 2.5.15 applies with all potential values of $\ell \in [qn - n^{0.61}, qn + n^{0.61}]$ (via a union bound) for

$$(V_{str}, V_1, C_1), (V_3, V_4, C_3), (V_4, V_5, C_4), \dots, (V_{t-m-1}, V_{t-m}, C_{t-m-1}), (V_{t-m}, V_{end}, C_{t-m})$$

We remark that here we use a union bound over all of the $< n^{0.8}$ potential values of ℓ as well as all of the $\le \log^5 n$ listed triples. This is possible as the failure probability of Lemma 2.5.15 is at most $1/n$. Also, an easy check shows that the hypothesis of Lemma 2.5.15 on the randomness parameters holds for each listed triple. Indeed, each random set listed above has parameter at most q, and we have $\ell \ge qn - n^{0.61}$ which checks the first inequality in Lemma 2.5.15. Also, each random set listed has parameter at least $9q/10$ and $\ell \le qn + n^{0.61}$ which implies the second inequality in Lemma 2.5.15. Finally, with high probability, all the random sets have sizes within $n^{0.6}$ of their expectations.

Fix all the random sets and the integer ℓ so they have all of the collected properties. Fix also the sets $V'_{str}, V'_{mid}, V'_{end}, C'$ so that they satisfy properties (1-4). Note that as the random sets have size close to their expectations, and property (1) holds for $V'_{str}/V'_{mid}/V'_{end}$, this implies in particular that $\ell = [qn - n^{0.6}, qn + n^{0.6}]$.

We can find disjoint subsets $R'_V, V'_1, V'_2, \ldots, V'_{t-m} \subseteq G$ and a subset $R'_C \subseteq R_C \cap C'$ with the following properties.

1. $R'_V, V'_1, V'_2, \ldots, V'_{t-m}$ partition V'_{mid}, $R'_V \subseteq R_V$, $|V'_i| = \ell$, $|V_i \setminus V'_i| \leq n^{0.65}$ and $|V_i \Delta V'_i| \leq n^{0.7}$ for each $i = 1, 2, 3$ and $t - m$.

2. R'_V has the property from Lemma 2.5.14 with respect to V'_1 and V'_2 with the colour set $R'_C \subseteq R_C$. That is, for any bijection $\phi\colon V'_1 \to V'_2$, there exists a system of rainbow paths using exactly the vertices/colours of $K^{\pm}_G[V'_1 \cup V'_2 \cup R'_V; C']$ from V'_1 to V'_2, each of length $m = 8\lceil 100 \log^2 n \rceil$, and connecting v to $\phi(v)$ for each $v \in V'_1$.

Note that the relevant inequalities hold while applying Lemma 2.5.14 with $(A, B) = (V'_1, V'_2)$ and R_V and R_C, since $|V'_1| = |V'_2| = \ell \leq n/\log^7 n \ll (rp)^{1000} n/\log^4 n$ since $1/n \ll rp$.

Similarly, we can find disjoint subsets $C'_1, C'_2, \cdots, C'_{t-m} \subseteq G$ partitioning $C' \setminus R'_C$ with the following property. $|C'_i| = \ell + k$ for each $i \neq 2$, $|C_i \setminus C'_i| \leq n^{0.72}$ and $|C_2 \Delta C'_2| \leq n^{1-10^{-10}}$.

Invoke Lemma 2.5.15 with sets (V'_{str}, V'_1, C'_1), (V'_3, V'_4, C'_3), (V'_4, V'_5, C'_4), $\ldots$, $(V'_{t-m-1}, V'_{t-m}, C'_{t-m-1}), (V'_{t-m}, V'_{end}, C'_{t-m})$ to find perfect matchings (saturating the vertex sets, and using all but k of the colours from each of the colour sets). For each $i \neq 2$, denote the colour subset of C'_i used in the corresponding perfect matching by C''_i. Let C''_2 be the union of C'_2 and all of the unused colours from each C'_i, that is, $C'_i \setminus C''_i$ $(i \neq 2)$, noting that there are $k(t - m - 1) \ll n^{1-10^{-4}}$ such unused colours.

Claim 2.5.16.1. *$|C''_0| = \ell$ and if **O** holds, we have $\sum V'_3 - \sum V'_2 = \sum C''_2$, and if **C** holds, we have $\sum V'_3 + \sum V'_2 = \sum C''_2$.*

Proof. To see this, it is convenient to apply the property of R'_V and R'_C coming from Lemma 2.5.14 with an arbitrary choice of ϕ. This gives a packing of rainbow paths in $K_G^{\pm}$ with the set of endpoints of the paths being V_1', V_2' and R_C' being the set of colours used on the paths. This, together with assumption (4), and that the colours in $C' \setminus R_C' \setminus C_2''$ have been used to find perfect matchings in the specified sets implies the first part of the claim. In the case that we work with the division digraph, considering all the directed paths we have found so far, this implies that $\sum V_2' - \sum V_{str} = \sum C_1'' + \sum R_C'$ and $\sum V_{end} - \sum V_3' = \sum C_3'' \cup C_4'' \cup \cdots \cup C_{t-m}''$. Combining these equalities with $\mathbf{O}$, we obtain the second part of the claim. In the case that we work with the multiplication digraph instead, a similar argument shows the second part of the claim. □

The previous claim allows us to apply Theorem 2.3.9 with (V_2', V_3', C_2'') to find a perfect matching directed from V_2' towards V_3' both in the case of division digraphs and multiplication digraphs (note C_2'' cannot contain e if G is Boolean by assumption). Now, we invoke the property of the sets R_V' and R_C' with a choice of ϕ so that we produce a $\vec{P}_t$-factor connecting each $v \in V_{str}$ to $f(v) \in V_{end}$. To define such a ϕ, for each $v \in V_1$, let v' be the matched neighbour of v in V_{str}. Consider the vertex v'' which is obtained by starting with $f(v') \in V_{end}$, and following each matched edge until we reach a vertex of V_2. Set $\phi(v) = v''$. It is easy to see that this function ϕ has the desired behaviour, concluding the proof.

□

Chapter 3

The Friedlander-Gordon-Tannenbaum conjecture

3.1 Main theorem and overview of the proof

3.1.1 Definitions of key auxiliary graphs and hypergraphs

It is customary in combinatorics to rephrase statements such as Conjecture 1.3.5 in terms of finding perfect matchings in hypergraphs, or finding rainbow structures in edge-coloured graphs, and we follow this tradition in the current paper.

Given a group of order n, we denote by $\vec{K}_G$ the edge-coloured directed graph defined as follows. $V(\vec{K}_G) := G$, and $E(\vec{K}_G) := \{(a, b) \in G \times G \colon a \neq b\}$, and the *colour* of an edge (a, b) is the group element ab^{-1}. Given subsets $V, C \subseteq G$, by $\vec{K}_G[V; C]$ we denote the subgraph of $\vec{K}_G$ obtained by keeping only the vertices in V, and the directed edges with colours in C. Occasionally, the following related definition will also be useful. Given multiple subsets $V_1, V_2, \dots, V_k \subseteq G$, we denote by $\vec{K}_G[V_1, V_2, \dots, V_k]$ the edge-coloured directed graph with vertex set $V_1 \sqcup V_2 \sqcup \dots \sqcup V_k$ ($\sqcup$ indicates that we are taking a disjoint union) and edge set consisting of edges of the form $e = (v, w) \in V_i \times V_{i+1}$ (with colour vw^{-1}) for some $i \in \{1, 2, \dots, k\}$ (where

$k+1=1$). By $\vec{K}_G[V_1, V_2, \ldots, V_k; C]$, we denote the same graph obtained by keeping only edges whose colour is in C.

Recall that a subgraph of an edge-coloured graph is called **rainbow** if all edges have distinct colours. Given $V, C \subseteq G$, let $\mathcal{H}_k[V; C]$ be the $2k$-uniform hypergraph on the vertex set $V \sqcup C$ where $v \sqcup c$ is an edge whenever $v \subseteq V$ induces a rainbow directed cycle of length k in $\vec{K}_G$ with the colour set of the cycle being precisely c. $\mathcal{H}_k$ denotes $\mathcal{H}_k[G; G]$. Sometimes we overload the terms vertex and colour by referring to elements of $V(\mathcal{H}_k[V; C])$ which come from V as **vertices** and those which come from C as **colours**. The following observation is quite critical.

Observation 3.1.1. *If c is the colour set of an edge in $\mathcal{H}_k$, or the colour set of some directed rainbow cycle in $\vec{K}_G$ (of any length), then the sum of all the elements of c must equal 0, i.e. c is a zero-sum set.*

Proof. As in a directed cycle each vertex has one in-edge and one out-edge, when we take a sum of all the colours of a cycle in $\vec{K}_G$, each vertex appears twice, once positive, and once negative. The statement follows. □

Given graphs H and G, we say that G contains an H-**factor** if there exists a collection of copies of H in G that partition the vertex set of G. For example, a K_2-factor in a graph is a perfect matching. $\vec{P}_k$ denotes a directed path of length k (meaning with k edges). $\vec{C}_k$ denotes a directed cycle of length k (meaning with k vertices and k edges). The following proposition follows from all the definitions presented thus far.

Proposition 3.1.2. *Let G be a finite abelian group and let k be an integer with $k \geq 2$. The following are equivalent.*

- *G admits an orthomorphism fixing the identity and permuting the remaining elements as products of disjoint k-cycles.*
- *$\vec{K}_G[G \setminus \{0\}; G \setminus \{0\}]$ contains a rainbow $\vec{C}_k$-factor.*
- *$\mathcal{H}_k[G \setminus \{0\}; G \setminus \{0\}]$ has a perfect matching.*

We invite the reader to verify the above proposition. Thanks to Proposition 3.1.2, we can phrase our main result in the language of hypergraph matchings in the next subsection.

3.1.2 Main theorem and its proof modulo key lemmas

The $k = 2$ case of Conjecture 1.3.5 is proven implicitly in [25], where the authors give orthomorphisms of odd-order cyclic groups which are products of disjoint transpositions (see also [23] for a proof of the $k = 2$ case). The case of $k > \log^{10} n$ (the "high-girth case"), on the other hand, can be resolved by using some tools from Chapter 2.5.2. We give the details for this in Section 3.5. We remark that for the methods of Chapter 2.5.2, this polylogarithmic lower bound on the cycle length is a hard barrier, essentially because any sorting network (see [9, 1]) must have depth at least $\log n$ (we discuss this in more detail later on). Hence, our main theorem is concerned with the $3 \le k \le \log^{10} n$ case of the FGT conjecture.

We recall the following conventions before stating our main theorem. Recall that a p-**random** subset of set S is one obtained by sampling each element of S independently with probability p. Similarly, we say a collection of random sets $R_1, \dots, R_k \subseteq S$ is **disjoint** p-**random** if each element of S belongs to each R_i with probability p, and to none of the R_i with probability $1 - pk$, and these decisions are made independently for each element of S. We reserve the letter n for the size of the ambient group throughout the paper. When we say that an event holds "with high probability", we mean that the probability of the event approaches 1 as n tends to infinity.

Theorem 3.1.3 (Main theorem)**.** *There exists an absolute constant $\varepsilon_{3.1.3} > 0$ such that the following holds. Let G be an abelian group of order n, let $p \ge n^{-\varepsilon_{3.1.3}}$, and suppose k is some integer such that $3 \le k \le \log^{10} n$. Let $R_1, R_2 \subseteq G$ be p-random subsets, sampled independently. Then, the following holds with high probability.*

Let $V, C \subseteq G$ be equal-sized subsets with $|V \Delta R_1|, |C \Delta R_2| \le n^{3/4}$. Suppose k divides $|V|$ (and thus, $|C|$), and suppose $\sum C = 0$ and $0 \notin C$. Then, $\mathcal{H}_k[V; C]$ has *a perfect matching.*

Theorem 3.1.3 turns into a deterministic statement when applied with $p = 1$. This statement, when n is sufficiently large, implies Conjecture 1.3.5 (when $3 \leq k \leq \log^{10} n$, the "low-girth case") by setting $V = C = G \setminus \{0\}$. Theorem 3.1.3 can thus be interpreted as a randomised version of Conjecture 1.3.5. As far as our proof method is concerned, it does not take extra work to prove Theorem 3.1.3 compared to Conjecture 1.3.5. Theorem 3.1.3 also has further applications. Using its full strength, one can find orthomorphisms with other cycle types, see Section 3.6 for more details.

Comparison with the random Hall-Paige conjecture

It is worth discussing how Theorem 3.1.3 is different with the main result of Chapter 2.5.2 and the applications given therein. The main result of Chapter 2.5.2, or the random Hall-Paige conjecture, is concerned with finding perfect matchings in hypergraph whose vertices are group elements, and edges are given by triples (a, b, c) where $a + b + c = 0$. In comparison, the hypergraph in Theorem 3.1.3, for example when $k = 3$, is defined by 6-tuples (a, b, c, d, e, f) where $a = d - e$, $b = e - f$, $c = f - d$ (note this implies in particular that $a + b + c = 0$). With extra ideas, it is sometimes possible to glue together simpler equations to deduce information about hypergraphs with more complicated structure. For example, the random Hall-Paige conjecture guarantees existence of perfect rainbow matchings in random subsets of $\vec{K}_G$. Collecting multiple disjoint random sets, and finding rainbow matchings between them (using disjoint colour sets), we can build long rainbow path forests in $\vec{K}_G$. However, when the lengths of the rainbow paths approach k, we need to "stitch" the endpoints of these rainbow paths together in order to obtain k-cycles. It turns out that one can set aside some structure in the beginning which allows us to stitch arbitrary endpoints together, using a specified set of vertices and colours. The catch is that for the stitching, we need to use paths of length at least $\log n$. This barrier comes from, roughly speaking, the stitching process being equivalent to running a sorting algorithm. Due to limitations on how many comparisons any sorting algorithm has to make, there unfortunately is no room for improvement for this part of the argument, and therefore we need novel ideas.

In fact, it turns out that when k is small (say $k = 3$), the stitching statement we require (corresponding to Lemma 2.5.8 when $t = 2$) simply becomes false, in the sense that conditions coming from simply summing elements together, as in the Hall-Paige conjecture, are not sufficient to characterise the existence of the desired spanning structure. This is closely related to obstructions which arise in the toroidal n-queens problem for which Bowtell and Keevash recently made a break-through [12]. We won't give much context for this problem here, mentioning only that the problem is about finding perfect matchings in hypergraphs where edges are of the form (a, b, c, d) where $c = a+b$ and $d = a-b$, similar in spirit to the FGT conjecture. In order to characterise when such perfect matchings arise, even in cyclic groups, one needs a condition about the sum of *squares* in the group. This is in stark contrast with both the Hall-Paige and FGT conjectures, where the obstructions can be expressed using only the additive structure of the group in question.

Proof ideas

As outlined before, the low-girth case of the FGT conjecture is highly sensitive, and several novel ideas are required compared to the random Hall-Paige conjecture from Chapter 2.5.2. To illustrate just one specific challenge we need to over-come, consider the following question: For which k-subsets $C \subseteq G$ does $\mathcal{H}_k$ have an edge with C as its colour set? Equivalently, which colour sets induce directed rainbow k-cycles in $\vec{K}_G$? This is highly relevant as a necessary condition for solving the FGT conjecture is coming up with a *partition* of $G \setminus \{0\}$ into such colour sets. It is not hard to see that we need that $\sum C = 0$ (see Observation 3.1.1) and that $0 \notin C$. Further thought reveals that to find a directed rainbow k-cycle (and not just a closed k-walk) in $\vec{K}_G$ with colour set C, we need to be able to order C as $(c_1, \dots, c_k)$ such that c_1, $c_1 + c_2$, $c_1 + \cdots c_k$ are all distinct. We call $(c_1, \dots, c_k)$ a **cycle-candidate** if this property holds. Can a zero-sum subset not containing 0 be ordered to be a cycle-candidate? It turns out that even for cyclic groups of prime order, this is an open problem, posed initially by Ronald Graham in 1971. Surprisingly little is known about this problem (see Problem 10 from [32], see also [7, 17, 16, 15]). For example, the problem is already open for $k = 13$ and cyclic groups of prime order.

Due to difficulties surrounding Graham's problem, our understanding of the structure of sequences yielding cycle-candidates is rather limited. A central aspect of our approach (which we believe translates well to adjacent problems, see Section 3.6.1) capitalises on the fact that there is a rich and well-behaved subset of cycle-candidates, namely, those which can be constructed by gluing together short, carefully curated families of subsequences (see for example the definition of *dissociable* in Section 3.2.2, as well as Lemma 3.2.24). A key tension in the proof is that the definitions of these building blocks need to be strong enough so that when they are combined, they produce cycle-candidates, but also weak enough that they exist in abundance throughout G. The latter need arises because we will have to construct *absorbers* for these building blocks from the beginning (see Lemma 3.4.18), in order to have some flexibility in how we use them towards the end of the proof.

Absorption is a general framework in probabilistic combinatorics designed to turn almost spanning structures into fully spanning structures (see [48, 12, 43] for some recent breakthroughs using this framework). The method has been developed immensely in the past decade, making it difficult to pinpoint exactly what the absorption method is. However, the common denominator in all implementations of the method is relying on the existence of small scale structures (gadgets) with various properties in order to build a large scale, flexible structure. A relevant question is which small scale structures can even be found in the host structure. Lemma 3.2.16 is a key tool we use in the current paper which gives a general method to produce *substructure patterns* which exist in abundance throughout host structures which are algebraically defined. We refer the reader to Figure 3.1 for examples of such patterns we work with in the current paper.

We emphasise that this result is quite flexible (in particular, it is possible extend Definition 3.2.7 with further sufficient conditions if a future problem requires a richer class of substructures). We refer the reader to Section 3.6.1 for a discussion about a potential connection with the study of graceful and harmonious labellings.

Proof of Theorem 3.1.3

For the rest of this section, we focus on Theorem 3.1.3, which is concerned with the "low-girth" case of the FGT conjecture. The key lemma used to prove Theorem 3.1.3 is the following, which states the existence of an "absorber for zero-sum subsets". Roughly speaking, this lemma states that random subsets contain "absorbers" which have the ability to combine with any small enough set to produce matchings (we say that the small set is "absorbed"), provided that this small set satisfies some straightforward necessary conditions.

Lemma 3.1.4 (Zero-sum absorption)**.** *There exist absolute constants $\varepsilon = \varepsilon_{3.1.4} > 0$ and $K = K_{3.1.4} \geq 1$ with $\varepsilon K \leq 10^{-10}$ such that the following holds. Let $3 \leq k \leq \log^{10} n$, $p \geq n^{-\varepsilon}$. Let $R_1, R_2 \subseteq G$ be p-random subsets, sampled independently. Let $m \in k \cdot \mathbb{N}$ with $m \leq (p/k \log n)^K n$. Then, the following holds with high probability.*

Let $U \subseteq G$ with $|U| \leq n^{4/5}$. Then, there exist $V \subseteq R_1 \setminus U$ and $C \subseteq R_2 \setminus U$ with the following property. For any $V' \subseteq G \setminus V$ and $C' \subseteq G \setminus C$ with $|V'| = |C'| = m$, $\sum C' = 0$, $0 \notin C'$, we have that $\vec{K}_G[V \cup V'; C \cup C']$ has a rainbow $\vec{C}_k$-factor, or equivalently, $\mathcal{H}_k[V \cup V'; C \cup C']$ has a perfect matching.

To finish, we need a version of Rödl nibble that works with regular hypergraphs plus a few "junk vertices", which correspond to the leftover in the smaller random sets after the absorber is removed. The following lemma encapsulates this idea.

Lemma 3.1.5. *There exists an absolute constant $\varepsilon_{3.1.5} > 0$ such that the following holds. Let G be an abelian group of order n. Let $3 \leq k \leq \log^{10} n$, and let $p \geq n^{-\varepsilon_{3.1.5}}$. Let R_1 and R_2 be p-random subsets of G, sampled independently. The following holds with high probability.*

For any $|V_D| = |C_D| \subseteq G$ with $|V_D|, |V_C| \leq \varepsilon_{3.1.5} p^3 n/k^{100}$, $\mathcal{H}_k[R_1 \cup V_D; R_2 \cup C_D]$ contains a matching covering all but at most $n^{1-1/10^8}$ vertices.

The proof of Lemma 3.1.5 comes down to establishing certain pseudorandomness properties of $\mathcal{H}_k$. Checking pseudorandomness in hypergraphs is notoriously tricky, for example see [36] for a useful criterion for dense hypergraphs. Unfortunately, $\mathcal{H}_k$ is quite sparse, and potentially has large uniformity, so [36] is not immediately useful in our set-up. For this reason, we have to put a fair bit of care into the proof of Lemma 3.1.5.

We can now give the proof of our main theorem, assuming these two lemmas. We remark that often in our proofs, we have random subsets $R' \subseteq R$ where R itself is a random subset of the group G. When we say that R' is a q-random subset, we always mean that R' is a q-random subset of the group G, and not of R.

Proof of Theorem 3.1.3. Pick a value of $\varepsilon_{3.1.3}$ such that $0 < \varepsilon_{3.1.3} \ll \varepsilon_{3.1.4}, \varepsilon_{3.1.5}, 1/K_{3.1.4}$. For each $i \in \{1, 2\}$, partition R_i into $R_i^{(1)}$ and $R_i^{(2)}$ which are disjoint p_1-random and p_2-random sets respectively, where $p_1 = (1/10)\varepsilon_{3.1.5} p^4/k^{100}$ (and $p_2 = p - p_1$). We have that $p_1 \geq n^{-\varepsilon_{3.1.4}}$ and $p_2 \geq n^{\varepsilon_{3.1.5}}$ if $\varepsilon_{3.1.3}$ is small enough. Select some $m \in k \cdot \mathbb{N}$ such that $10n^{1-1/10^8} \leq m \leq (p_1/k \log n)^K n$ (there exists such values of m as $\varepsilon_{3.1.4} K_{3.1.4} \leq 10^{-10}$). With high probability, Lemma 3.1.4 holds with $(R_1^{(1)}, R_2^{(1)})$ with this value of m and Lemma 3.1.5 holds with $(R_1^{(2)}, R_2^{(2)})$. Also with high probability, the size of each random set is at most $\sqrt{n} \log n$ away from its expectation (by Chernoff's bound). With high probability, all of these properties hold simultaneously.

Now, fix random sets having all these properties and let V and C be given as in the statement of the theorem. Set $U := (R_1 \setminus V) \cup (R_2 \setminus C)$ noting $|U| \leq 2n^{3/4}$. Apply Lemma 3.1.4 to find absorbing subsets $V_A \subseteq R_1^{(1)} \setminus U \subseteq V$ and $C_A \subseteq R_2^{(1)} \setminus U \subseteq C$ which can combine with m-sized vertex-sets and m-sized zero-sum colour-sets to produce perfect matchings. Note this implies in particular that $|V_A| = |C_A|$. Set $V_D := (R_1^{(1)} \cap V) \setminus V_A$ and $C_D := (R_2^{(1)} \cap C) \setminus C_A$, noting $|V_D|, |C_D| \leq 2p_1 n \leq$

$\varepsilon_{3.1.5}p_2^3 n/k^{100}$. Note also that $||V_D| - |C_D|| \leq 10n^{3/4}$. By Lemma 3.1.5, $\mathcal{H}_k[V_D \cup (R_1^{(2)} \cap V); C_D \cup (R_2^{(2)} \cap C)]$ has a matching M_1 covering all but at most $10n^{1-1/10^8}$ vertices (formally, we delete $\leq 10n^{3/4}$ elements from C_D or V_D so that $|V_D| = |C_D|$, also we initially find the matching inside $\mathcal{H}_k[V_D \cup R_1^{(2)}; C_D \cup R_2^{(2)}]$ and delete the $\leq |U|$ matched edges that use a vertex/colour from U). If necessary, unmatch some edges of M_1 so that the number of leftover vertices $V' := V \setminus V(M_1)$ and colours $C' := C \setminus V(M_1)$ are both equal to m (possible as k divides $|V|, |C|$ and m and $|V_A| = |C_A|$). Note that the matching M_1 guarantees that all colours in $C \setminus C'$ admit a partition $C_1, C_2, \ldots$ where each C_i is the colour set of a rainbow cycle in $\vec{K}_G$, meaning that $\sum C_i = 0$ for each i (see Observation 3.1.1). So we must have $\sum C \setminus C' = 0$ also. As $\sum C = 0$ by assumption, this implies that $\sum C' = 0$, so we can invoke the property coming from Lemma 3.1.4. This means that V' and C' combine with V_A and C_A to produce a matching, M_2. $M_1 \cup M_2$ is then the desired perfect matching. □

3.1.3 Organisation of the rest of the paper

We collect some preliminary tools in Section 3.2. We have already broken up the task of proving Theorem 3.1.3 into proving Lemma 3.1.5 and Lemma 3.1.4. The former is proved in Section 3.3 and the latter is proved in Section 3.4. In Section 3.5, we show how the high-girth case of the FGT conjecture can be derived from results from [49], as promised earlier on in this section. In Section 3.6, we discuss some directions for future research.

3.2 Preliminaries

3.2.1 Probabilistic tools

Concentration inequalities

Nibble-type lemmas

We say that a r-partite r-uniform hypergraph H is (γ, p, n, k)-**regular** if every part has $(1\pm\gamma)n$ vertices and every vertex has degree $(1\pm\gamma)pn^k$. For a 3-uniform 3-partite hypergraph H, vertices u, v and a subset $U \subseteq V(H)$, we define the **pair degree** of (u, v) into U as the number of vertices in U which are in the neighbourhood of both u and v, i.e. the number of vertices z in U such that there exists $y, w \in V(H)$ such that $\{u, z, y\}$ and $\{v, z, w\}$ are both edges of H. A 3-uniform 3-partite hypergraph H is (γ, p, n)-**typical** if it is $(\gamma, p, n, 1)$-**regular** and every pair of vertices u, v coming from the same part has pair degree $(1 \pm \gamma)p^2n$ into every other part of H. A hypergraph is **linear** if it maximum co-degree at most 1.

The following nibble-type result due to Ehard, Glock, and Joos is convenient to use for our application here.

Theorem 3.2.1 ([20]). *Suppose $\delta \in (0, 1)$ and $r \in \mathbb{N}$ with $r \geq 2$, and let $\varepsilon := \delta/50r^2$. Then there exists Δ_0 such that for all $\Delta \geq \Delta_0$, the following holds. Let $\mathcal{H}$ be an r-uniform hypergraph with $\Delta(\mathcal{H}) \leq \Delta$ and $\Delta^c(\mathcal{H}) \leq \Delta^{1-\delta}$ as well as $e(\mathcal{H}) \leq \exp(\Delta^{\varepsilon^2})$. Suppose that $\mathcal{W}$ is a set of at most $\exp(\Delta^{\varepsilon^2})$ weight functions on $E(\mathcal{H})$. Then, there exists a matching $\mathcal{M}$ in $\mathcal{H}$ such that $\omega(\mathcal{M}) = (1 \pm \Delta^{-\varepsilon})\omega(E(\mathcal{H}))/\Delta$ for all $\omega \in \mathcal{W}$ with $\omega(E(\mathcal{H})) \geq \max_{e\in E(\mathcal{H})} \omega(e)\Delta^{1+\delta}$.*

Applying Theorem 3.2.1 with a single uniform weight function, we obtain the following.

Corollary 3.2.2. *Let n be sufficiently large and let $p \geq n^{-1/10000}$.*

1. *Let $\mathcal{H}$ be a 6-uniform 6-partite hypergraph on n vertices which is $(n^{-0.01}, p, n, 2)$-regular with maximum co-degree at most $10n$. Then, $\mathcal{H}$ has a*

matching covering all but $n^{1-10^{-5}}$ vertices.

2. *For any $\gamma \geq 0$, every (γ, δ, n)-regular linear tripartite hypergraph has a matching covering all but at most $n^{1-1/500} + 3\gamma n$ vertices.*

3.2.2 Group theoretic tools

Given a sequence $\vec{c} = (c_1, c_2, \ldots, c_k)$ of group elements, and another group element v, we define the following sequences:

- $P_{out}(v, \vec{c}) := (v, v - c_1, v - c_1 - c_2, \ldots, v - c_1 - c_2 - \cdots - c_k)$
- $P_{in}(v, \vec{c}) := (v, v + c_1, v + c_1 - c_2, \ldots, v + c_1 + c_2 + \cdots + c_k)$

Observe that in $\vec{K}_G$, $P_{out}(v, \vec{c})$ denotes the vertex sequence obtained by starting a walk from v, and following the out-edges given by the sequence $\vec{c}$. $P_{in}(v, \vec{c})$ is analogous, except it follows the in-edges.

We call a sequence of group elements $\vec{c} = (c_1, c_2, \ldots, c_k)$ a **path-candidate** if the partial sums $\sum_{i \in [j]} c_i$ for each $j \in [k]$ (including $j = 0$) are all distinct. Equivalently, all non-empty partial sums (of consecutive elements) are non-zero. Observe that $\vec{c}$ being a path-candidate simply means that for any vertex $v \in G$, the walks $P_{out}(v, \vec{c})$ and $P_{in}(v, \vec{c})$ both give paths in $\vec{K}_G$.

We call a sequence of group elements $\vec{c} = (c_1, c_2, \ldots, c_k)$ a **cycle-candidate** if $(c_1, c_2, \ldots, c_{k-1})$ is a path-candidate and $\sum_{i \in [k]} c_i = 0$. This means that $P_{out}(v, \vec{c})$ and $P_{in}(v, \vec{c})$ both give cycles (of length k) in $\vec{K}_G$.

We call a sequence of group elements **rainbow** if all coordinates are distinct. Notice that a necessary condition for solving the FGT conjecture is a partition of $G \setminus \{0\}$ into rainbow cycle-candidates, each of length k.

The following two definitions only come up in the cover-down strategy for $k \geq 10$.

We call a collection of length k sequences **dissociable** if for any two distinct sequences $\vec{c} = (c_1, c_2, \ldots, c_k)$ and $\vec{b} = (b_1, b_2, \ldots, b_k)$ and $j, j' \in [k]$, $\sum_{i \in [j]} c_i \neq \sum_{i \in [j']} b_i$. This means that $P_{out}(v, \vec{c})$ and $P_{out}(v, \vec{b})$ are disjoint except on v. We call such a collection **near-dissociable** if the previous property holds for each $j, j' \leq$

$k-1$ (or equivalently, the sequences obtained by removing the last element from each tuple gives a dissociable family). This means that the corresponding directed walks are disjoint except on the endpoints.

We call two length k sequences $(c_1, c_2, \ldots, c_k)$ and $\vec{b} = (b_1, b_2, \ldots, b_k)$ **separable at distance** d if for all $j, j' \in [k]$, $\sum_{i \in [j]} c_i + \sum_{i \in [j']} b_i \notin \{-d, d\}$. This means that for v and w where $v - w = d$, $P_{out}(v, \vec{c})$ and $P_{in}(w, \vec{b})$ are disjoint (except potentially on v or w).

The following simple lemma is key to the gadget finding strategy presented in Section 3.2.2.

Lemma 3.2.3. *Let G be abelian of order n. Then either the map $x \to 2x$ or the map $x \to 3x$ has an image of size at least $n^{1/5}$.*

Proof. For an integer k and an abelian group G let G_k denote the number of distinct images of the map $G \to G$ via $x \to k \cdot x$. Observe that $(G \times H)_k = G_k \cdot H_k$. By the fundamental theorem of finite abelian groups, $G \cong (\mathbb{Z}_2)^\ell \times (\mathbb{Z}_3)^j \times H$ where H is a product of cyclic groups none of which are of order 2 or 3. Note that $(\mathbb{Z}_t)_2 \geq t/2$ if $t \neq 2$ and $(\mathbb{Z}_t)_3 \geq t/3$ if $t \neq 3$. This, together with the observation, implies that $H_2 \geq |H|/2^{\log_4 |H|} = |H|^{0.5}$ and $H_3 \geq |H|/3^{\log_4 |H|} \geq |H|^{0.2}$ (using that H has at most $\log_4 |H|$ many cyclic factors in its decomposition as each factor needs to have size at least 4). Also, $(\mathbb{Z}_2)_3^\ell = 2^\ell$ and $(\mathbb{Z}_3)_2^j = 3^j$, as multiplying by 3 in $\mathbb{Z}_2$ and multiplying by 2 in $\mathbb{Z}_3$ are both bijections.

Note that $|G| = n = 2^\ell \cdot 3^j \cdot |H|$. If $j \geq \ell$, we have that $n \leq 6^j |H|$ and $|G_2| \geq 3^j |H|^{0.5}$, so $|G_2| \geq n^{1/2}$. Otherwise, if $j < \ell$, we have that $n \leq 6^\ell |H|$ and $|G_3| \geq 2^\ell |H|^{0.2}$, so $|G_3| \geq n^{1/5}$. □

Finding gadgets

In this section, we refine some tools from Section 2.2.6 in order to prove Lemma 3.2.16, a versatile tool to find small substructures in $\vec{K}_G$.

By F_k, we denote the free abelian group on k generators, the free variables are denotes as $v_1, \dots, v_k$ (recall that $F_k \cong \mathbb{Z}^k$). $G * F_k$ denotes the free product, and $(G * F_k)^{\mathrm{ab}}$ denotes the abelianization of the free product (recall that the abelianization G^{ab} of a group G is defined by the property that any homomorphism $G \to H$ where H is abelian factors uniquely through G^{ab}). A **word** is simply an element of $(G * F_k)^{\mathrm{ab}}$. As all groups G are abelian in this paper, $(G * F_k)^{\mathrm{ab}} \cong G \times F_k$, where the latter denotes a direct product. However, the former perspective makes it clear that each word w can be represented as

$$w = z_1 \cdot v_1 + \cdots + z_t \cdot v_t + g$$

where each v_i is a free variable, each z_i is a (non-zero) integer, and $g \in G$, and this representation is unique up to reordering the summands.

A word w is **constant** if $w \in G$, i.e. w does not include any free variables. We say that z_i is the **coefficient** of v_i. We say that w is **linear in** v_i if $z_i \in \{1, -1\}$. We say that w is **linear** if each $z_i \in \{1, -1\}$, and w is not constant. That is, w is linear in each free variable, and there exists at least one free variable in w.

A homomorphism $\pi : (G * F_k)^{\mathrm{ab}} \to G$ is a **projection** if $\pi(g) = g$ for all $g \in G$. We show two basic properties of projections. We remind the reader that throughout, G is a finite abelian group of order n.

Lemma 3.2.4. *For each function $f : \{v_1, \dots, v_k\} \to G$, there is precisely one projection $\pi_f : (G * F_k)^{\mathrm{ab}} \to G$ which agrees with f on $\{v_1, \dots, v_k\}$. In particular, there are precisely n^k projections $(G * F_k)^{\mathrm{ab}} \to G$.*

Proof. By the universal property of free abelian groups, there is a unique homomorphism $g \colon F_k \to G$ which agrees with f on $\{v_1, \dots, v_k\}$. By the universal property of free products, there is a unique homomorphism $h \colon G * F_k \to G$ that agrees with g on F_k and with the identity homomorphism $G \to G$. As G is abelian, h can be written uniquely as $h = h' \circ p$ where $p\colon (G * F_k) \to (G * F_k)^{\mathrm{ab}}$ is the quotient map, and $h'\colon (G * F_k)^{\mathrm{ab}} \to G$ is a projection that agrees with f on $\{v_1, \ldots, v_k\}$. This gives the desired one to one correspondence. □

Lemma 3.2.5. *Let $w \in \{v_1, \ldots, v_k\}$ be linear in some free variable v_i and let $g \in G$. Then there are exactly n^{k-1} projections $\pi : (G * F_k)^{\mathrm{ab}} \to G$ having $\pi(w) = g$.*

Proof. Suppose that $i = k$, without loss of generality. By Lemma 3.2.4 there are exactly n^{k-1} projections $\pi : (G * F_{k-1})^{\mathrm{ab}} \to G$. For each such π, we show that there is a unique projection π' that agrees with π and additionally has $\pi'(w) = g$. By linearity of w in v_k, the equation $w = g$ rearranges into $v_k = h$ for some $h \in (G*F_k)^{\mathrm{ab}}$ and v_k does not appear in h. So, $\pi'(w) = g$ is equivalent to $\pi'(w) = \pi'(g)$ (as π' is a projection) which is equivalent to $\pi'(v_k) = \pi'(h)$ (as π is a homomorphism). As π' agrees with π and $h \in (G * F_{k-1})^{\mathrm{ab}}$, we have that $\pi'(v_k) = \pi(h) \in G$. Therefore, π' has that $\pi'(v_k) = \pi(h)$, and $\pi'(v_i) = \pi(v_i)$ for $1 \leq i < k$. By Lemma 3.2.4, there is a unique projection with this property. □

The following is a simple consequence of the previous lemma.

Lemma 3.2.6. *Let $S \subseteq (G * F_k)^{\mathrm{ab}}$ be a set of elements which are each linear in at least one variable, and let $U \subseteq G$. Then the number of projections $\pi : (G*F_k)^{\mathrm{ab}} \to G$ for which $\pi(S)$ intersects U is $\leq |S||U|n^{k-1}$.*

Definition 3.2.7. Let $w, w' \in (G * F_k)^{\mathrm{ab}}$. We say that w and w' are **separable** if any of the following hold.

(a) $w' - w$ is linear in some free variable v_i. Note that this is equivalent to asking that there exists a free variable v with coefficient z in w and z' in w' and we have $|z - z'| = 1$.

(b) The equation $w = w'$ rearranges into $g = 0$ for some non-zero group element $g \in G$.

(c) The equation $w = w'$ rearranges into $3v_i - 2v_j = g$ for some group element $g \in G$ and distinct free variables v_i and v_j.

Definition 3.2.8. Let $S \subseteq G * F_k$. We say that a homomorphism $\phi : (G * F_k)^{\mathrm{ab}} \to G$, **separates** S if for every separable $w, w' \in S$ we have $\phi(w) \neq \phi(w')$.

Lemma 3.2.9. *Let $n \geq 10^{100}$. Let $S \subseteq (G * F_k)^{\mathrm{ab}}$ be a set of size ≤ 1000. Then there are at most $|S|^2 n^{k-1/5}$ projections $\pi : (G * F_k)^{\mathrm{ab}} \to G$ which do not separate S.*

Proof. Let w and w' be two separable words in S. We case on which of the conditions (a)/(b)/(c) makes w and w' separable, and count the projections which do not separate them in each case.

(a) In this case, $\pi(w) = \pi(w')$ is equivalent to $\pi(w - w') = e$ (using that π is a homomorphism). By Lemma 3.2.5, there are n^{k-1} projections π satisfying this latter identity.

(b) We can rearrange $\pi(w) = \pi(w')$ into $\pi(g) = \pi(e)$ using that π is a homomorphism. The latter implies that $g = e$ using that π is a projection, which is a contradiction. Hence there can be no projections π with $\pi(w) = \pi(w')$ in this case.

(c) Similarly to the previous cases, we can rearrange $\pi(w) = \pi(w')$ into $3\pi(v_i) - 2\pi(v_j) = g$. Using Lemma 3.2.3, suppose first that $x \to 3x$ has at least $n^{1/5}$ images, and suppose π also satisfies $\pi(v_j) = g'$ for some $g' \in G$, so we have $3\pi(v_i) = g'' \in G$ (where $g'' = g + 2g'$). As $x \to 3x$ is a homomorphism (G is abelian) and has at least $n^{1/5}$ images, the preimage of g'' under the map $x \to 3x$ has size at most $n^{4/5}$ (each non-empty preimage must have the same size in a group homomorphism). This means $\pi(v_i)$ must live in a set $T_{g'}$ of size at most $n^{4/5}$ assuming that $\pi(w) = \pi(w')$ and $\pi(v_j) = g'$. Thus, if $\pi(w) = \pi(w')$, π must agree with one of $n^{k-1}n^{4/5}$ functions $f\colon \{v_1, \ldots, v_k\} \to G$, meaning that there are at most $n^{k-1/5}$ such projections, using Lemma 3.2.4. A symmetric argument works when $x \to 2x$ has at least $n^{1/5}$ images.

As there are at most $\binom{|S|}{2}$ pairs of separable words in S, the desired bound follows. □

Lemma 3.2.10. *Let $n \geq 10^{10}$, and let $S \subseteq (G * F_k)^{\mathrm{ab}}$ be a set of at most 100 elements which are all linear in at least one variable. Then, there are projections $\pi_1, \dots, \pi_{n/200}$ which separate S and have $\pi_1(S), \dots, \pi_{n/200}(S)$ disjoint.*

Proof. Call a projection *good* if it separates S. Let $\pi_1, \dots, \pi_t$ be a maximal collection of good projections with the sets $\pi_i(S)$ being pairwise disjoint. Set $T = \pi_1(S) \cup \cdots \cup \pi_t(S)$ noting $|T| = |S|t$. For any good projection π, we must have $\pi(S) \cap T \neq \emptyset$ by maximality, so by Lemma 3.2.6, we have that there are at most $100tn^{k-1}$ good projections. On the other hand, there are at most $|S|^2 n^{k-1/5} \leq n^k/2$ projections which are not good by Lemma 3.2.9, so there are at least $n^k/2$ good projections (there are n^k projections total). Combining, we have $n^k/2 \leq 100tn^{k-1}$, meaning $t \geq n/200$, as desired. □

Combining the previous lemma with a standard application of Chernoff's bound, we obtain the following.

Lemma 3.2.11. *Let $p \geq n^{-1/700}$. Let R be p-random subset of G. With high probability, the following holds.*

Let $S \subseteq (G * F_k)^{\mathrm{ab}}$ a set of ≤ 100 elements which are each linear in at least one variable, and let $U \subseteq G$ with $|U| \leq p^{100}n/1000$. Then there is a projection $\pi : (G * F_k)^{\mathrm{ab}} \to G$ which separates S, has $\pi(S) \cap U = \emptyset$ and $\pi(S) \subseteq R$.

We now package everything we have so far into a lemma (Lemma 3.2.16) that fits nicely with our application in the setting of $\vec{K}_G$.

Definition 3.2.12. Given a group G, a **pattern** P is a directed (simple) graph equipped with a vertex and edge labelling ϕ with the following properties.

1. ϕ maps vertices and edges to $(G * F_k)^{\mathrm{ab}}$ for some positive integer k.

2. Each vertex gets a distinct label via ϕ (i.e. $\phi|_{V(P)}$ is injective, but distinct edges can potentially receive the same label)

3. If $\vec{e} \in E(P)$ is a directed edge from v to w for $v, w \in V(P)$, we have that $\phi(v) - \phi(w) = \phi(\vec{e})$.

In Figure 3.1 we have several examples of patterns. We can naturally view the edge-labels as colours, hence each pattern can also be viewed as an edge-coloured graph.

A **pairwise separable** subset S is a subset where any two distinct words w and w' are separable.

Definition 3.2.13. We call a pattern (P, ϕ) **well-distributed** if the following two conditions hold.

1. The subsets (viewed as sets, not multisets) $\{\phi(v) \colon v \in V(P)\}$ and $\{\phi(\vec{e}) \colon \vec{e} \in E(P)\}$ are both pairwise separable subsets of $G * F_k$.
2. Each label is either a constant, or linear in at least one free variable.

Notice that we are not insisting that any $\phi(v)$ and $\phi(\vec{e})$ are separable for a vertex v and edge $\vec{e}$. This is because in our applications vertex sets and colour sets are sampled independently, hence we don't need any separability properties.

Definition 3.2.14. A **copy** of a well-distributed pattern (P, ϕ) is a subgraph S of $\vec{K}_G[V; C]$ such that there exists a projection $\pi \colon (G * F_k)^{\mathrm{ab}} \to G$ (where k is the number of free variables used in ϕ) with the following properties.

1. π maps $\phi[V(P)]$ (the vertex labels) to $V(S) \subseteq V$ and $\phi[E(P)]$ (the edge labels) to C.
2. π separates $\phi[V(P)]$ and π separates $\phi[E(P)]$. In particular, π is injective when restricted to $\phi[V(P)]$.
3. The (v, w) is a directed edge of S if and only if there is a directed edge from the vertex with the label $\pi^{-1}(v)$ to the vertex with the label $\pi^{-1}(w)$ in P.

An **edge-coloured directed graph isomorphism** ψ between two edge-coloured simple directed graphs G_1, G_2 is a graph isomorphism mapping vertices of G_1 to vertices of G_2 and mapping edges (v, w) of G_1 to $(\psi(v), \psi(w))$ that preserves the direction of each edge and respects colours. This means that e_1 and e_2 of G_1 have the same colour if and only if $\psi(e_1)$ and $\psi(e_2)$ have the same colour.

Observation 3.2.15. *Let S be a copy of P. Then, there is a edge-coloured directed graph isomorphism ψ between P and S. Furthermore, if x is the label of a vertex or colour of P, and x is a constant, $\psi(x) = x$.*

Proof. The projection π witnessing that S is a copy naturally corresponds to a ψ with the desired properties, as π fixes elements of $(G * F_k)^{\mathrm{ab}}$ which are constants by definition of a projection. □

The following is a consequence of the definition of well-distributed, copy, and applying Lemma 3.2.11 to R_1 and R_2.

Lemma 3.2.16. *Let $p \geq n^{-1/700}$. Let R_1 and R_2 be p-random subsets of G, sampled independently. With high probability, the following holds.*

Let P be a well-distributed pattern with $V(P) + E(P) \leq 100$. Let $U \subsetneq G$ with $|U| \leq p^{150}n/10^7$. Let V' and C' be the set of labels of vertices and colours in P which are constants. Then, there is a copy of P in $\vec{K}_G[(R_1 \setminus U) \cup V'; (R_2 \setminus U) \cup C']$.

As an example application of the above result, we recommend the reader to inspect the proof of Lemma 3.4.2.

Partitioning into sets with fixed sum

In this subsection, we prove some lemmas designed to "cover-down" part of the absorption strategy. See the proof overview for more context. For the reader interested in the $k = 3$ case, the $k = 3$ case of Lemma 3.2.22 is all that is required, and this case follows directly from Lemma 3.2.18 (without having to use Lemma 3.2.21). We cite the following three results from [49] which are consequences of variants of Theorem 1.2.1 as proved in Chapter 2.5.2 of the thesis, but we omit the proofs due to space considerations. Recall that $H_G[X, Y, Z]$

denotes the 3-uniform hypergraph whose vertex set is a disjoint union of X, Y and Z, and $(x, y, z) \in X \times Y \times Z$ is an edge whenever $x + y + z = 0$.

Theorem 3.2.17 ([49]). *Let $p \geq n^{-1/10^{100}}$. Let G be an abelian group of order n. Let $R^1, R^2 \subseteq G$ be disjoint p-random subsets, and let $R^3 \subseteq G$ be a p-random subset, sampled independently with R^1 and R^2. Then, with high probability, the following holds.*

Let X, Y, Z be equal-sized subsets of G_A, G_B, and G_C respectively, satisfying the following properties.

- $|(R^1_A \cup R^2_B \cup R^3_C)\Delta(X \cup Y \cup Z)| \leq p^{10^{10}} n/\log(n)^{10^{10}}$
- $\sum X + \sum Y + \sum Z = 0$
- *Suppose that* $0 \notin X \cup Y \cup Z$

Then, $H_G[X, Y, Z]$ contains a perfect matching.

Lemma 3.2.18 ([49]). *Let $p \geq n^{-1/10^{100}}$ and $3 \leq k \leq 100$. Let R be a p-random subset of an abelian group G. With high probability the following holds.*

Let $X \subseteq G$ with $|X \triangle R| \leq p^{10^{10}} n/\log(n)^{10^{18}}$, $0 \notin X$, $\sum X = 0$, and $|X| \equiv 0 \pmod{k}$. Then, X can be partitioned into zero-sum sets of size k.

Lemma 3.2.19 ([49]). *Let $p \geq n^{-1/700}$ and let R be a p-random subset of an abelian group G. With high probability the following holds.*

Let $\epsilon \in [2 \log n/\sqrt{n}, p^{800}/10^{4010}]$. For any m with $|m - pn| \leq \epsilon n$, $g \in G$ and Z with $|Z| \geq m + 3$, $|R \setminus Z| \leq \epsilon n$, there is a set $R' \subseteq Z$ with $|R'| = m$, $|R' \triangle R| \leq 6\epsilon n$, and $\sum R' = g$.

Corollary 3.2.20. *Let $p \geq n^{-1/10^{100}}$. Let R be a p-random subset of an abelian group G. With high probability the following holds.*

Let $X \subseteq G$ with $|X \triangle R| \leq p^{10^{10}} n/\log(n)^{10^{22}}$. Let $\alpha \in G$. $0 \notin X$, $|X| \equiv 0 \pmod{4}$, $\sum X = (|X|/4) \cdot \alpha$. Then, X can be partitioned into sets of size 4 with sum α.

Proof. Let R_1, R_2, R_3, R_4 be disjoint $(p/4)$-random subsets of G which partition R. Let S be a $(p/4)$-random subset of G, sampled independently with the previous sets. Note that the set $-S-\alpha$ is also a $(p/4)$-random subset of G, which is independent with the previous random sets (not including S). With high probability, Theorem 3.2.17 holds with the sets (R_1, R_2, S) and $(R_3, R_4, -S-\alpha)$, Lemma 3.2.19 holds for each random set, and by Chernoff's bound, each random set is within a $n^{0.6}$ term of its expectation.

Let $X \subseteq G$ be given. By Lemma 3.2.19, we can partition X into equal sized sets X_1, X_2, X_3, X_4 such that $|X_i \Delta R_i| \leq p^{10^{10}} n/\log(n)^{10^{21}}$ and $\sum X_1 + \sum X_2 = 0$. This readily implies that $\sum X_3 + \sum X_4 = (|X|/4)\cdot\alpha$ by the sum condition on X. Similarly, via Lemma 3.2.19, we can fix a set S' with $\sum S' = 0$, $|S'| = |X|/4$, and such that S' has small symmetric difference with S. This implies that $\sum(-S'-\alpha) = -(|X|/4)\cdot\alpha$, and also we have that $-S'-\alpha$ has small symmetric difference with $-S-\alpha$. Thus we have that $\sum X_1 + \sum X_2 + \sum S' = 0$ and $\sum X_3 + \sum X_4 + \sum(-S'-\alpha) = 0$. Also, we remark that S' can be chosen so that both S' and $-S'-\alpha$ do not contain 0. So we can apply Theorem 3.2.17 twice to deduce that both $H_G[X_1, X_2, S']$ and $H_G[X_3, X_4, -S'-\alpha]$ has a perfect matching. For each $s' \in S'$, consider the edges (x_1, x_2, s') and $(x_4, x_4, -s'-\alpha)$ guaranteed by the two perfect matchings, and observe that $x_1 + x_2 + x_3 + x_4 = \alpha$. Combining 4-tuples of this form, we obtain the desired partition of X. □

For technical reasons, our absorption strategy for large k requires the assumption that $k \geq 10$. This leaves the case of $3 \leq k \leq 9$ open. The previous lemmas already give us a way to partition sets into k-sets which are zero sum in this regime. Once we have access to such a partition, a natural strategy is to look for an ordering of the k-set yielding a cycle-candidate, in order to be able to perform the cover-down step (see proof overview). We rely on the following result of Alspach and Liversidge to find suitable orderings. Similar results for cyclic groups were obtained in [17, 37].

Lemma 3.2.21 (Alspach-Liversidge, [7], Corollary 5.2)**.** *Let G be any abelian group (not necessarily finite). Let $S \subseteq G$ be of size at most 9. with $\sum S = 0$. Then, S*

admits an ordering yielding a rainbow cycle-candidate if $\sum S = 0$, and otherwise S admits an ordering yielding a rainbow path-candidate.

We can now prove the main lemma of this section.

Lemma 3.2.22. *There exists an absolute constant $\varepsilon_{3.2.22}$ such that the following holds. Let $3 \leq k \leq 9$. Let G be an abelian group of order n, let $p \geq n^{-\varepsilon_{3.2.22}}$. Let R be a p-random subset of G. With high probability, the following holds. Let $R' \subseteq G$ such that $|R' \Delta R| \leq p^{10^{10}} n/\log(n)^{10^{23}}$. Suppose k divides $|R'|$ and that $0 \notin R'$.*

1. *Suppose that $\sum R' = 0$. Then, R' can be partitioned into k-tuples which are rainbow cycle-candidates.*

2. *Suppose that $k = 4$ and for some $\alpha \in G \setminus \{0\}$, $\sum R' = (|R'|/k) \cdot \alpha$. Then, R' can be partitioned into k-tuples which are rainbow path-candidates with sum α.*

Proof. Choose some $\varepsilon_{3.2.22} \leq 10^{-1000}$. With high probability, Lemma 3.2.18 and Corollary 3.2.20 both hold for R. Let R' be given. For part (1), we apply Lemma 3.2.18 to partition R' into k-sets which are zero-sum. Then, Lemma 3.2.21 implies that each of these k sets can be ordered to obtain a rainbow cycle candidate, as $k \leq 9$. For part (2), we apply Corollary 3.2.20 to partition into 4-tuples each with sum α. As $\alpha \neq 0$, we can order each tuple to be path-candidates by Lemma 3.2.21. This concludes the proof. □

Good families of colours

In this section we have some lemmas designed to deal with the $k \geq 10$ case of the cover-down step.

Lemma 3.2.23. *Let G be an abelian group of order n, let $s \in G \setminus \{0\}$ let $\bar{T}$ be the collection of k-tuples $(g_1, \ldots, g_k)$ with $\sum g_i = s$ (note $|\bar{T}| = n^{k-1}$). Suppose $n^{0.01} \geq k \geq 2$ and $n \geq 10^{10}$. Let S be a subset of G of size at most $n/(20k)$. Then, all but at most $n^{k-1}/4$ tuples in $\bar{T}$ are all rainbow path-candidates disjoint with S.*

Proof. As $s \neq 0$, we can count that there are at least $(n-1)(n-2)\cdots(n-k+1) \geq (n-k)^{k-1} \geq n^{k-1}/1.01$ (using that k is small for the final inequality) path-candidates in $\bar{T}$.

If $k \geq 3$, by a direct counting we can see that there are at most $k^2 n n^{k-3} \leq k^2 n^{k-2}$ tuples in $\bar{T}$ with two coordinates being equal. If $k = 2$, using that G is an abelian group and $s \neq 0$, we see that there are at most $n/2$ tuples (generously) in $\bar{T}$ with two coordinates being equal. In either case, all but $n^{k-1}/2$ tuples in $\bar{T}$ are rainbow (using that $n \gg k$).

For each $g \in G$, there are at most kn^{k-2} elements of $\bar{T}$ having g in some coordinate, here we used that $k \geq 2$. So, there are at most $|S|kn^{k-2} \leq n^{k-1}/20$ many tuples $\bar{t}$ not disjoint with $\mathcal{S}_G$.

We derive that there are at least $n^{k-1}/1.01 - n^{k-1}/2 - n^{k-1}/20 \geq n^{k-1}/4$ tuples in $\bar{T}$ satisfying all the desired properties. □

Lemma 3.2.24. *There exists some absolute constant $C_{3.2.24}$ such that the following holds. Let G be an abelian group of order n, let $n \geq 10^{100}$, and let k be a integer such that $10 \leq k \leq n^{0.001}$. Then, G contains two families $\mathcal{F}_G = \mathcal{F}_G(k)$ and $\mathcal{S}_G = \mathcal{S}_G(k)$ of disjoint tuples $\mathcal{F}_1, \dots, \mathcal{F}_{\lfloor n/kC_{3.2.24}\rfloor}$ and $\mathcal{S}_1, \dots, \mathcal{S}_{\lfloor n/kC_{3.2.24}\rfloor}$ with the following properties.*

(1) Each $\mathcal{F}_i$ is of size 4 and has the same sum $f = f(G,k)$.

(2) Each $\mathcal{S}_i$ has the same size and sum $s = s(G,k)$. In fact, $|\mathcal{S}_i| =: z_{\mathcal{S}} \in \{2,3,4,5\}$.

(3) Each $\mathcal{S}_i$ is a rainbow path candidate, and $\mathcal{S}_G$ is near-dissociable.

(4) $k-4-z_{\mathcal{S}}$ is divisible by 4. Furthermore, set $q := q_{G,k} = -((k-4-z_{\mathcal{S}})/4)f - s$. We have that $q \neq 0$.

(5) Each $\mathcal{F}_i$ can be partitioned into two tuples, $\mathcal{F}_i^+ = (f_i^{+,1}, f_i^{+,2})$ and $\mathcal{F}_i^- = (f_i^{-,1}, f_i^{-,2})$, both of which are rainbow path candidates. The resulting collection of $\mathcal{F}_i^+$ and $\mathcal{F}_i^-$ are both dissociable. Also, each $\mathcal{F}_i$ is a rainbow path candidate, and $\mathcal{F}_G$ is near-dissociable.

(6) *For each $m \in \{0, 1, 2, \ldots, k\}$ and $i \in \lfloor n/kC_{3.2.24} \rfloor$, we have that $\mathcal{F}_i^+$ and $\mathcal{F}_i^-$ are separable at a distance $q + mf$.*

Proof. Pick some $z_\mathcal{S}$ between 2 and 5 so that $k - 4 - z_\mathcal{S}$ is positive and divisible by 4, note that this is possible as $k \geq 10$. Pick any $f \neq 0$. Pick some s so that $q + m \cdot f \neq 0$ for any $m \in \{0, 1, 2, \ldots, k, k+1\}$ for $q := -((k - 4 - z_\mathcal{S})/4)f - s$. Indeed, there are $0.9n$ such choices of s, as $k \leq n^{0.001}$.

Claim 3.2.24.1. *We can find $\mathcal{S}_G$ satisfying (2) and (3).*

Proof. Set $k' = z_\mathcal{S}$, recalling that $k' \geq 2$. Suppose that we have found a maximal family $\mathcal{S}_G$ satisfying (2) and (3) and suppose that $|\mathcal{S}_G| < n/(kC)$ for some C. We will derive a contradiction for C sufficiently large.

Let $\bar{G}$ be the collection of rainbow path candidate k'-tuples $(g_1, \ldots, g_{k'})$ with $g_1 + \cdots + g_{k'} = s$ which are disjoint with $\bigcup \mathcal{S}_G$ (observing this set has size at most $5n/(kC)$ by assumption). $|\bar{G}| \geq n^{k'-1}/4$ by Lemma 3.2.23 (supposing $C \geq 100$). If for some $\bar{t} \in \bar{G}$ we have that $\mathcal{S}_G^*$ becomes non-dissociable upon the addition of $\bar{t}$, there must exist some $\bar{t}' \in \mathcal{S}_G$ $j, j' \in [k'-1]$ such that $\sum_{i \in [j]} \bar{t}'_i = \sum_{i \in [j']} \bar{t}_i$. There are at most $(n/kC)k'n^{k'-2} \leq n^{k'-1}/C$ such $\bar{t}$, meaning that there is a $\bar{t} \in \bar{G}$ that we can add to $\mathcal{S}_G$ without breaking (2) and (3), a contradiction. □

It remains to construct $\mathcal{F}_G$. Suppose $\mathcal{F}_G$ is a family of 4-tuples satisfying the properties with size at most $n/(kC) - 1$ (where C is a sufficiently large constant). We will show that $\mathcal{F}_G$ can be extended.

Fix some $f^+, f^- \in G \setminus \{0\}$ such that $f^+ + f^- = f$, and the following two properties hold.

1. For any $\mathcal{F}_i \in \mathcal{F}_G$ we have that $f^+, f^- \neq \sum \mathcal{T}_i$ for any $\mathcal{T}_i \subseteq \mathcal{F}_i$.

2. $f^+ + f^- \neq \pm q + mf$ for any $m \in \{0, 1, \ldots, k, k+1\}$.

Such f^+, f^- with the first property exist as long as $C > 50$, as $\sum \mathcal{T}_i^\pm$ can take at most $20n/C$ distinct values due to the assumption on the size of $\mathcal{F}_G$. Such f^+, f^- automatically satisfy the second property as $q + m \cdot f \neq 0$ for any $m \in \{0, 1, \ldots k\}$.

Let F^+ denote the set of ordered triples with sum f^+ and F^- denote the set of ordered triples with sum f^-, noting $|F^+| = |F^-| = n$.

Claim 3.2.24.2. *Suppose we delete all triples from $F \in F^+$ such that the collection $\mathcal{F}_+ \cup \{F\}$ fails to be dissociable. This deletes at most $10n/C$ triples.*

Proof. If for some $F \in F^+$ we have that $\{\mathcal{F}_i^+\} \cup F$ is not dissociable, there must exist some $F' \in \mathcal{F}^+$ and $j' \in \{1, 2\}$ and $j \in \{1, 2\}$ such that $\sum_{i \in [j']} F'(i) = \sum_{i \in [j]} F(i)$. It cannot be that $j = 2$ by the first property coming from our choice of f^+. By the bound on $|\mathcal{F}_G|$, there are at most $10n/C$ distinct values the quantity $\sum_{i \in [j']} F'(i) =: w$ can take. For each such w, there is at most one $F \in F^+$ with $F'(1) = w$. This implies that in total there are at most the claimed number of triples which make the corresponding collection not dissociable. □

Claim 3.2.24.3. *Suppose we delete all tuples from $F \in F^+$ such that F and the 1-tuple (f^-) are not separable at a distance $q + m \cdot f$ for some $m \in \{0, \cdots, k\}$. This deletes at most $2k$ tuples.*

Proof. If for some $F \in F^+$ we have that F and (f^-) are not separable at a distance $q + m \cdot f$ for some $j \in \{1, 2\}$, then we must have $\sum_{i \in [j]} F(i) + f^- = \pm(q + m \cdot f)$ for some $m \in \{0, \ldots, k\}$. Here, $j = 2$ is precluded by the second property of f^+ and f^-. For each of the $2k$ possible values of $\pm(q + m \cdot f) - f^-$, there exists at most one $F \in F^+$ such that $F(1) = \pm(q + m \cdot f) - f^-$, which implies the claim. □

Claim 3.2.24.4. *There are at least $n/4$ tuples in F^+ which are rainbow path-candidates and which contain no coordinate $F(i)$ also present in an element of $\mathcal{S}_G$ or $\mathcal{F}_G$.*

Proof. This is immediate by Lemma 3.2.23 and bounding $|\bigcup \mathcal{S}_G \cup \bigcup \mathcal{F}_G|$. □

Claim 3.2.24.5. *Deleting all tuples $F \in F^+$ with $F(1) = \pm \sum \mathcal{T}_i$ or $F(2) = \pm \sum \mathcal{T}_i$ for some $\mathcal{T}_i \subseteq \mathcal{F}_i \in \mathcal{F}_G$, we delete at most $80n/C$ elements.*

Proof. There are at most $20n/C$ possible values for the quantity $\pm\sum \mathcal{T}_i$ by the upper bound on the size of $\mathcal{F}_G$. This implies the claim, as $F(1)$ (and $F(2)$) is a distinct value for each $F \in F^+$. □

By the bounds coming from the claims, we can fix $\mathcal{F}^+_{new}$ to be a 2-tuple from F^+ which is a rainbow path candidate disjoint with the earlier sets, keeps $\mathcal{F}^+$ dissociable, and is separable with (f^-) at a distance $q + m \cdot f$ for each $m \leq k$.

Now, we perform the analogous steps for F^-. Claim 3.2.24.2 and 3.2.24.4 (thinking of F^+_{new} as an element of $\mathcal{F}_G$ to ensure disjointness) also hold when $+$ is replaced by $-$, giving us at least $n/5$ potential elements of F^- we can select while maintaining dissociability of $\mathcal{F}^-$, disjointness with previous tuples, and rainbow path candidacy.

In addition, we delete the elements of F^- which are not separable with $\mathcal{F}^+_{new}$ at a distance $q+mf$ for some $m \in \{0, 1, \ldots, k\}$. For any $F \in F^-$, it is already impossible for $\sum_{i\in[j]} \mathcal{F}^+_{new}(i) + \sum_{i\in[j']} F(i) = \pm(q + m \cdot f)$ when $j' = 2$ (by the property from Claim 3.2.24.3). When $j' = 1$, note that $\pm(q+m\cdot f) - \sum_{i\in[j]} \mathcal{F}^+_{new}(i)$ can take at most $4k$ distinct values v, and we only need to delete at the at most $4k$ many $F \in F^-$ with $F(i) = v$.

For each $F \in F^-$, consider the 4-tuple $\mathcal{F}_{new} = (\mathcal{F}^+_{new}(1), \mathcal{F}^+_{new}(2), F(1), F(2))$, and note that this is always a rainbow sequence. If $F \in F^-$ makes $\mathcal{F}_G \cup \{\mathcal{F}_{new}\}$ not near-dissociable, we delete F from F^-. To count how many such F there are, suppose that for some $\mathcal{F} \in \mathcal{F}_G$, we have that $\sum_{i\in[j]} \mathcal{F}(i) = \sum_{i\in[j']} \mathcal{F}_{new}(i)$ where $j, j' \in [3]$. It is impossible that $j' \in \{1, 2\}$ due to Claim 3.2.24.5 and the first property of f^+. Note there are at most $20n/C$ potential values of $\sum_{i\in[j]} \mathcal{F}(i)$ due to the bound on the size of $\mathcal{F}_G$. This implies that for the relevant equality to hold, $F(1)$ needs to belong to a set of size $20n/C$, so in this step we delete at most $20n/C$ elements from F^-. Similarly, if $(\mathcal{F}^+_{new}(1), \mathcal{F}^+_{new}(2), F(1), F(2))$ is not a path candidate, it must be that a partial sum of the sequence is 0. This partial sum cannot contain both of $F(1)$ and $F(2)$, as the whole sum is $f \neq 0$, and $\mathcal{F}^+_{new}(2) \neq -F(1) - F(2)$ by Claim 3.2.24.5, and $(F(1), F(2))$ is a path candidate. But the partial sum has to contain $F(1)$, as

$(\mathcal{F}^+_{new}(1), \mathcal{F}^+_{new}(2))$ alone gives a path candidate. This means that at most 2 extra values of $F(1)$ are forbidden if $(\mathcal{F}^+_{new}(1), \mathcal{F}^+_{new}(2), F(1), F(2))$ is to be a rainbow path candidate.

Selecting C large, we can fix a value of $F \in F^-$ so that setting $\mathcal{F}^-_{new} = F$, we successfully extend $\mathcal{F}_G$, as desired. □

3.3 Nibble with some determinism

In this section we give a proof of Lemma 3.1.5.

Observation 3.3.1. *Let $\mathcal{E}$ be an equation of the form $\pm a \pm b \pm c = 0$. Let H be a tripartite hypergraph obtained by taking three copies of some group G of order n, and letting $(a, b, c) \in G^3$ be an edge whenever it is a solution to $\mathcal{E}$. Then, H is $(0, 1, n)$-typical.*

Proof. For a proof for when $\mathcal{E}$ is $a + b + c = 0$, see Observation 3.3 in [49]. For other equations of this form, the proof is essentially identical. □

Lemma 3.3.2. *Let $p \geq n^{-1/600}$ and let $X \subseteq G$ be p-random. Then, with probability at least $1 - 8/n^3$, the following holds. For any $Y \subseteq G$, for all but at most $8n^{9/10}$ vertices $g \in G$, and for each equation of the form $\mathcal{E} := \pm g \pm x \pm y = 0$ (where g is a constant and x and y are free variables), we have that there are $p|Y| \pm n^{9/10}$ many $(x, y) \in X \times Y$ such that x, y and g satisfy $\mathcal{E}$.*

Proof. Thanks to Observation 3.3.1, we can apply Lemma 2.2.7 to the corresponding hypergraph defined by each of the $2^3 = 8$ possible equations $\mathcal{E}$, and with probability at least $1 - 8/n^3$, we ensure that the conclusion of Lemma 2.2.7 holds for each of these hypergraphs. The desired statement follows immediately. □

We say that $g \in G$ is **generic** if $g \neq e$ and there are at most $n^{1/2}$ solutions to $x^2 = g$ in G. Let $N(G)$ denote the set of non-generic elements and note that $|N(G)| \leq n^{1/2}$.

Observation 3.3.3. *Let G be an abelian group of order n and let $A \subseteq G$ be a multiset of order k. Consider the sets $A+g$ for each $g \in G$. Then, at most $kn^{-1/10}$ many such sets have more than $n^{3/5}$ many non-generic elements. Also, there are at most $kn^{-1/10}$ sets $A-g$ with more than $n^{3/5}$ many non-generic elements.*

Proof. There are $\leq kn^{1/2}$ tuples $(a,g) \in A \times G$ where $a+g$ is non-generic. Let $\#$ be the number of $g \in G$ such that there are $\geq n^{3/5}$ many $a \in A$ such that $a+g$ is non-generic. Then, $\# \cdot n^{3/5} \leq kn^{1/2}$, so $\# \leq kn^{-1/10}$. The same argument applies when $+$ is replaced by $-$. □

Recall that given a fixed graph F, a **packing** of F in some other graph G is just a collection of vertex-disjoint copies of F in G. When we talk about rainbow packings, we always mean that there is no colour repetition in edges across all copies of F in the packing.

Lemma 3.3.4. *There exists an absolute constant $\varepsilon_{3.3.4} > 0$ such that the following holds. Let $p \geq n^{1-\varepsilon_{3.3.4}}$. Let G be a group of order n. Let $V_2, V_3 \subseteq G$ be disjoint p-random, let $C_1, C_2, C_3 \subseteq G$ be disjoint p-random, sampled independently with V_2, V_3. The following holds with probability at least $1 - 1/n^{2.9}$.*

Let $V_1, V_4 \subseteq G \setminus (V_2 \cup V_3)$ with $|V_1| = |V_2| = (p \pm n^{-0.1})n$. Let $f\colon V_1 \to V_4$ be a bijection. Then, $\vec{K}_G[V_1, V_2, V_3, V_4; C]$ contains a rainbow packing of at least $n^{1-1/10^5}$ paths of length 3, directed $V_1 \to V_2 \to V_3 \to V_4$, such that for all paths $\vec{P}$ in the packing and $v_1 \in V_1 \cap V(\vec{P})$, we have that $f(v_1) \in V_4 \cap V(\vec{P})$.

Proof. Each of the following holds with probability at least $1 - O(1/n^3)$, thus they all simultaneously hold with probability at least $1 - 1/n^{2.9}$.

(1) Lemma 3.3.2 holds for X set to be each of V_2, V_3, C_1, C_2 and C_3.

(2) For each i, $|V_i|, |C_i| = (p \pm n^{-0.1})n$, by Chernoff's bound.

(3) For every colour $c \in G \setminus \{0\}$ and vertex pair $v, w \in G$ such that $v - w - c$ is generic, we have that there exists $(p^4 \pm n^{-0.1})n$ many rainbow paths of length 3 directed $v \to V_2 \to V_3 \to w$ with edge colours (c_1, c, c_3) for some

$(c_1, c_3) \in C_1 \times C_3$. If $v - w - c$ is not generic, we have that there exists at most $(p^4 + n^{-0.1})n$ such paths.

Proof. Consider all tuples $(v, v_2, v_3, w, c_1, c, c_3)$ where $v_2, v_3, c_1, c_3 \in G$, and $v - v_2 = c_1$, $v_2 - v_3 = c$ and $v_3 - w = c_3$. There are n such tuples. Note $c_1 + c_3 = v - v_2 + v_3 - w = v - w - c$ which is generic. This means for all but at most $n^{1/2}$ tuples, $c_1 \neq c_3$. As $c \neq 0$, for all tuples $v_2 \neq v_3$. For each of the $n - n^{1/2}$ tuples where $c_1 \neq c_3$, the probability of $(v_2, v_3, c_1, c_3) \in V_2 \times V_3 \times C_1 \times C_3$ is p^4. Letting X denote the expected number of paths of the desired form, we obtain that $\mathbb{E}[X] = (p^4 \pm n^{-1/2})n$. Further, X is 2-Lipschitz, so the desired concentration follows from Azuma's inequality. When $v - w - c$ is not generic, the same argument applies except we only have an upper bound on $\mathbb{E}[X]$. □

(4) For every pair of vertices $v, w \in G$ such that $v - w$ is generic, we have that there exists $(p^3 \pm n^{-0.05})n$ many rainbow paths directed $v \to V_3 \to w$ with edge colours from $C_2 \times C_3$. If $v - w$ is not generic, we have that there exists at most $(p^3 + n^{-0.1})n$ such paths.

(5) For every pair of vertices $v, w \in G$ such that $v - w$ is generic, we have that there exists $(p^3 \pm n^{-0.05})n$ many rainbow paths directed $v \to V_2 \to w$ with edge colours from $C_1 \times C_2$. If $v - w$ is not generic, we have that there exists at most $(p^3 + n^{-0.1})n$ such paths.

The proofs for (4) and (5) are essentially identical to the proof for (3), hence we omit them.

Now, suppose V_i and C_i all of the properties, and fix a bijection $f \colon V_1 \to V_4$. Let $\mathcal{H}$ be the hypergraph consisting of edges $(v_1, v_2, v_3, v_4, c_1, c_2, c_3) \in V_1 \times V_2 \times V_3 \times V_4 \times C_1 \times C_2 \times C_3$ where $v_1 - v_2 = c_1$, $v_2 - v_3 = c_2$, $v_3 - v_4 = c_3$ and $f(v_1) = v_4$. Our goal is to find a matching covering all but $n^{1-1/1000}$ vertices in this hypergraph. We sometimes refer to the edges of this hypergraph as paths. We will show that there is a set S of $\leq n^{1-1/100}$ vertices we can delete from $\mathcal{H}$ so that the resulting hypergraph $\mathcal{H}'$ is almost regular, that is, for all $v \in V(\mathcal{H})$, $d(v) = (p^5 \pm n^{-0.05})n^2$.

Towards that goal, set S to include

- the $\leq 90n^{9/10}$ vertices of G coming from Lemma 3.3.2 applied with each of V_2, V_3, C_1, C_2, C_3
- the $\leq 2|V_1|n^{-1/10} \leq 2n^{9/10}$ elements of G coming from Observation 3.3.3 applied with the multiset $\{v - f(v) \colon v \in V_1\}$ and both $+$ and $-$

so we have $|S| \leq 100n^{9/10}$. We will show all vertices of $\mathcal{H}$ not in S have degree $(p^5 \pm n^{-0.05})n^2$. Since S is small, this shows that S has the desired property. We consider several cases.

Let $v_1 \in V_1 \setminus S$, and set $v_4 = f(v_1) \in V_4 \setminus S$. For all but $n^{1/2}$ many $c_2 \in C_2$ we have that $v - w - c_2$ is generic. For such c_2, we have by 3. that there are $(p^4 + n^{-0.1})n$ paths passing through both v_1 and c_2. Combined with the bound on the size of C_2 coming from (2), this shows the desired upper and lower bound on $d(v_1)$ because through the few c_2 such that $v - w - c_2$ is non-generic, there exists at most $10n$ paths passing through both v_1 and c_2, giving in total $O(n^{3/2})$ such paths.

If $v_4 \in V_4 \setminus S$, set $v_1 = f^{-1}(v_4)$ and apply the result from the previous paragraph.

Let $c_1 \in C_1 \setminus S$. From Lemma 3.3.2, we have that there exists $p^2n \pm n^{9/10}$ directed c_1 coloured edges from V_1 to V_2. As $c_1 \notin S$, for all but $n^{3/5}$ many $v \in V_1$, $v - w - c_1$ is generic, and so for such v, we have that $(v - c_1) - w$ is generic, so we can apply (4) to obtain that there exists $(p^4 + n^{-0.1})n$ paths to $f(v)$ passing through c_1. Combined with the bound on $|V_1|$, this gives the desired bound on $d(c_1)$, as the number of paths going through the c_1 such that $v - w - c_1$ is non-generic is too small to influence the count, as before.

Let $c_3 \in C_3 \setminus S$. This case follows by a symmetric argument with the $c_1 \in C_1 \setminus S$ case, using (5) in place of (4).

Let $c_2 \in C_2 \setminus S$. Let $(v, w = f(v)) \in V_1 \times V_2$ and suppose that $v - w - c_2$ is generic. Then by (3) there are $(p^4 + n^{-0.1})n$ paths passing through v, w, and c_2. As $c_2 \notin S$, we have that all but $n^{3/5}$ values of $v \in V_1$, $v - w - c_2$ is generic. This, with the bound on $|V_1|$ implies the desired bound on $d(c_2)$, again because there are few paths passing through c_2 with $v - w - c_2$ non-generic.

So S has the desired properties, making $\mathcal{H}'$ almost-regular. Let $\mathcal{H}''$ be the hypergraph obtained by contracting v and $f(v)$ to a single vertex for each $v_1 \in V_1$. Note that as f is a bijection, and any edge through v has to pass through $f(v)$ as well, this does not change the regularity parameters of any of the other vertices in $\mathcal{H}'$. To see that this satisfies the hypotheses of Corollary 3.2.2(1), the only thing left to check is the co-degree condition. This is equivalent to obtaining an upper bound on the number of tuples $(v_1, v_2, v_3, v_4, c_1, c_2, c_3) \in \mathcal{H}$ where the values of 2 coordinates are fixed, and it is not the case that these two coordinates are the first and the fourth (since the corresponding vertices have been contracted). This means that we are counting solutions to a system of equations with 7 free variables and 6 independent constraints, hence there are at most n such solutions. This gives the desired co-degree bound. Corollary 3.2.2(1) then gives the desired result. □

Proof of Lemma 3.1.5. Let $q \leq \varepsilon_{3.1.5} p^3/k^{100}$ be a rational number with denominator at most n, observing that there are at most n values of such q. First, we will show that with probability at least $1/n^2$, the statement holds for any $V_D, C_D \subseteq G$ with $|V_D| = |C_D| = qn$.

Suppose first that $k = 3$. Let $r = (p - 2q)/3$ and let $R_1^{(1)}, R_1^{(2)}, R_1^{(3)}$ be disjoint r, r and $r + 2q$ random (respectively) sets partitioning R_1. Let $R_2^{(1)}, R_2^{(2)}, R_2^{(3)}, R_2^{(4)}, R_2^{(5)}$ be disjoint q, q, r, r and r-random (respectively) sets partitioning R_2. With probability at least $1 - O(1/n^{2.9})$ (we assume here that $q \geq n^{-1/100}$, otherwise the argument up to finding M_1 can be discarded, and the M_2 found at the end of the argument satisfies the requirements), Lemma 3.3.2 holds for $R_2^{(1)}$ and $R_2^{(2)}$ and Lemma 3.3.4 holds with $(V_2, V_3) = (R_1^{(1)}, R_1^{(2)})$ and $(C_1, C_2, C_3) = (R_2^{(3)}, R_2^{(4)}, R_2^{(5)})$. Also, by Chernoff's bound the following holds for all cycle-candidate triples (a, b, c) simultaneously with probability at least $1 - 1/n^{10}$: there exists at least $p^3 n/1000$ vertex-disjoint 3-cycles in $\vec{K}_G[R_1^{(3)}]$ with colour sequence (a, b, c) (this holds with high probability by Lemma 3.2.16 as well, indeed see Lemma 3.4.2, but here we cite Chernoff's bound directly to obtain an explicit bound on the probability). Finally, with probability at least $1 - 1/n^{10}$, all random sets are at most $n^{0.6}$ elements away from their expectations. With probability at least $1 - 1/n^2$ all of these properties hold simultaneously.

Now let V_D, C_D be given. Let $H_G[R_2^{(1)}, R_2^{(2)}, C_D]$ denote the 3-partite 3-uniform hypergraph on the indicated parts where triples are edges if and only if they are zero-sum. From Lemma 3.3.2 applied with the equation $g + x + y = 0$, we have that all but $n^{99/100}$ vertices of $H_G[R_2^{(1)}, R_2^{(2)}, C_D]$ do not satisfy the regularity hypothesis from Corollary 3.2.2(2). Deleting such vertices, we obtain a $(n^{-0.01}, q^2, qn)$-regular linear tripartite hypergraph, so Corollary 3.2.2(2) implies that all but $n^{1-1/700}$ elements of $R_2^{(1)} \cup R_2^{(2)} \cup C_D$ can be covered by disjoint zero-sum triples, denote these triples by $\mathcal{T}$. If necessary, delete at most one edge from $\mathcal{T}$ so that 0 is not used on any triple, meaning that the remaining triples can be ordered to be cycle-candidates (see, for example, Lemma 3.2.21). Using the property of $R_1^{(3)}$ repeatedly for each triple in $\mathcal{T}$, we can find a matching M_1 saturating all triples in $\mathcal{T}$ (and nothing else) in $\mathcal{H}_k[R_1^{(3)}; \bigcup \mathcal{T}]$. Now, invoke Lemma 3.3.4 with $V_1 = V_4 = (R_1^{(3)} \setminus V(M_1)) \cup V_D$ (noting $|V_1| = |R_1^{(3)}|$) and f set to be the identity function. This gives that $\mathcal{H}_k[(R_1^{(1)} \cup R_1^{(2)} \cup R_1^{(3)} \setminus V(M_1)) \cup V_D; R_2^{(3)}, R_2^{(4)}, R_2^{(5)}]$ has a matching covering all but $10n^{1-1/10^5}$ vertices, say M_2. Then, $M_1 \cup M_2$ is the desired matching.

Union bounding over all potential values of q, we obtain that with high probability, for each q the statement holds. This is sufficient to deduce the assertion, as qn is always an integer.

Now, suppose that $k \geq 4$. For each $i \in [k]$, and $j \in \{1, 2\}$ let $R_j^{(i)}$ be a $((p+q)/k)$-random set for $i \geq 2$ and $((p+q)/k - q)$-random set for $i = 1$, partitioning R_j. Similarly to the $k = 3$ case, we will fix some rational $q \leq pn/k^{100}$ with denominator at most n, and prove that the desired statement holds with probability at least $1 - 1/n^{1.1}$.

With probability at least $1 - 1/n^{1.3}$, Lemma 2.5.15 holds for $(A, B, C) = (R_1^{(i)}, R_1^{(i+1)}, R_2^{(i)})$ with $\ell = ((p+q)/k)n$ for each $i \in [k-3]$ (using that k is polylogarithmic in n for the union bound). With probability at least $1 - 1/n^{10}$, Lemma 3.3.4 holds with $(V_1, V_2) = (R_1^{(k-1)}, R_1^{(k)})$, $(C_1, C_2, C_3) = (R_2^{(k-2)}, R_2^{(k-1)}, R_2^{(k)})$. With probability at least $1 - 1/n^2$, each random set is within $n^{0.6}$ elements of its expected size. With probability at least $1 - 1/n^{1.1}$, all these properties hold simultaneously.

Let V_D, C_D be given. Apply Lemma 2.5.15 with random sets $(R_1^{(1)}, R_1^{(2)}, R_2^{(1)})$ and $(A', B', C') = (R_1^{(1)} \cup V_D, R_1^{(2)}, R_2^{(1)} \cup C_D)$ to find a matching that saturates all but $n^{1-1/10^7}$ vertices. Continue invoking Lemma 2.5.15 with corresponding random sets and $(A', B', C') = (R_1^{(i)}, R_1^{(i+1)}, R_2^{(i)})$ for each $i \in [k-3] \setminus \{1\}$. In both of these applications, we may delete/add $O(n^{0.78})$ elements from the corresponding sets so that they have size precisely ℓ or $\ell + \lfloor n^{1-10^{-5}} \rfloor$ (depending on whether they are vertex or colours sets), so that the hypotheses of Lemma 2.5.15 are satisfied, and then if necessary we can delete all edges passing through a dummy vertices/colours. Deleting all vertices that fail to be covered by one of the $k-3$ matchings found via the previous applications of Lemma 2.5.15, we delete at most $kn^{1-1/10^7} \le n^{1-1/(2\cdot 10^7)}$ vertices. The remaining vertices form directed paths following sets $R_1^{(1)} \to R_1^{(2)} \to \cdots \to R_1^{(k-2)}$. Let $V_1 \subseteq R_1^{(1)}$ and $V_4 \subseteq R_1^{(k-2)}$ be the vertices used by these directed paths, noting $|V_1| = |V_4|$, and let $f\colon V_1 \to V_4$ be the bijection induced by the two endpoints of each directed path. Now, Lemma 3.3.4 allows us to complete all but $n^{1-1/10^5}$ of these paths of length $k-3$ into a k-cycle using the remaining random sets, which gives the desired matching in $\mathcal{H}_k$. Union bounding over the potential values of q, we obtain the desired result, as in the $k = 3$ case. □

3.4 Zero-sum absorption

In this section we prove Lemma 3.1.4. Throughout this section, whenever a constant C appears inside the statement of a lemma, this should be read as "there is a sufficiently large absolute constant C so that the statement holds with this value of C".

3.4.1 Cover-down step: saturating vertices and colours

Covering vertices

The next lemma gives us a way to find edges of $\mathcal{H}_k$ that pass through a specific set of vertices.

Lemma 3.4.1. *Let $p \geq n^{-1/700}$. Let $3 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

1. *Let $U \subseteq G$ be a set with $|U| \leq p^{300}n/C_{3.4.1}$ (recall the convention set in the beginning of Section 3.4). Let $u, v \in G$ be two vertices, not necessarily distinct. Let k' be such that $2 \leq k' \leq k$. Then, if $u \neq v$, there exists a rainbow path of length k' directed from u to v in $\vec{K}_G[(R_1 \setminus U) \cup \{u, v\}; R_2 \setminus U]$. If $u = v$ and $k' \geq 3$, there exists a directed rainbow cycle of length k' using the vertex v.*

2. *Let $U \subseteq G$ be a set with $|U| \leq p^{300}n/4C_{3.4.1}$. Let $V \subseteq G$ be a set of vertices with $|V| \leq p^{300}n/(4kC_{3.4.1})$. Then, $\mathcal{H}_k[(R_1 \setminus U) \cup V; R_2 \setminus U]$ has a matching saturating V where each matched edge uses exactly one vertex from V.*

Proof. Fix a large constant $C \geq 10^7$ and fix $C_{3.4.1} \gg C$. Fix some distinct vertices of $\vec{K}_G$, u and v. Fix a set of $n/10$ triples (c_1, x, c_2) where $u \to x \to v$ is a rainbow path of length 2 with edge sequence (c_1, c_2), and the resulting collection of x and $\{c_1, c_2\}$ are both pairwise disjoint. Such a collection exists because there are at least $n/4$ disjoint (c_1, c_2) with $c_1 \neq c_2$ and $c_1 + c_2 = u - v \neq 0$ and G is an abelian group. By Chernoff's bound, with exponentially high probability, $\vec{K}_G[R_1 \cup \{u, v\}; R_2]$ contains at least $p^3n/10$ such paths. By a union bound over all distinct u and v, we have that $\vec{K}_G[R_1 \cup \{u, v\}; R_2]$ contains at least $p^3n/10$ such paths for any choice of u and v with high probability. Call this property $(*)$. Also, Lemma 3.2.16 holds with high probabilty.

We claim that a stronger version of part (1) holds when $k' \in \{2, 3\}$, with $C_{3.4.1}$ replaced with C. For $k' = 2$, this already follows from the property $(*)$ as each element of U (other than u or v) can eliminate at most 1 path. We claim the $k' = 3$

case follows from an application of Lemma 3.2.16. In the case that $u \neq v$, we can see this by defining a pattern that is a directed path of length 3, first vertex labelled u, last vertex labelled v, and colour sequence $(c_1,\ c_2,\ u - v - c_1 - c_2)$ (labelled in order of proximity to u) where c_1, c_2 are free variables (and u and v are constants). Note this implies that the vertex sequence is $(u, u - c_1, u - c_1 - c_2, v)$ (so that the third property in the definition of a pattern holds). This is a well-defined pattern as each vertex gets a distinct label. Furthermore, the pattern is well-distributed. The colours are separable by (a), u and v are separable by (b) (as $u \neq v$) and the rest of the vertex pairs are separable by (a). Also, each label is either a constant, or linear in c_1 or c_2. Then, Lemma 3.2.16 gives us a copy of this pattern, which corresponds to the desired rainbow path thanks to Observation 3.2.15. In the case that $u = v$, we can proceed similarly, this time using a pattern that is a directed cycle of length 3, with colour sequence $(c_1, c_2, -c_1 - c_2)$ and vertex sequence $(u, u - c_1, u - c_1 - c_2, u)$, where c_1, c_2 are free variables.

For larger k', (1) follows from repeated applications of the cases of $k' \in \{2, 3\}$. Indeed, any directed path/cycle of length ≥ 4 can be broken up into directed paths of length 2 and 3. While iteratively invoking (1) with $k' \in \{2, 3\}$, we extend U at each step, in total adding at most k new elements to U. As we know (1) holds with $k' \in \{2, 3\}$ with the smaller constant C, this reduction is valid, as $p^{900}n/C_{3.4.1} + k < 2p^{900}n/C_{3.4.1} < p^{900}n/C$ (using that k is small and that $C \ll C_{3.4.1}$, respectively).

Now, let a U be given as in the part (2) of the statement, and include in U all vertices of V (without relabelling). Given a vertex $z \in V$, we can invoke part (1) with $z = u = v$ to find a cycle of length k using z. For the next iterations, we apply (1) adding to U the vertices we've used so far, which adds to U at most $k|V| \leq p^{300}n/(4C_{3.4.1})$ elements. This means that U never exceeds a size of $p^{300}n/C_{3.4.1}$ in any of the iterations, making the applications of (1) valid. □

Covering colours: Small k

Lemma 3.4.2. *Let $p \geq n^{-1/700}$. Let $k \leq 50$. Let R_1 be a p-random subset of G. With high probability, the following holds.*

Let S be a k-tuple which is a rainbow path candidate. Let $U \subseteq G$ with $|U| \leq p^{350}n/C_{3.4.2}$. Then, $\vec{K}_G[R_1 \setminus U]$ contains a path with colour sequence S. Similarly, if S is a cycle candidate, $\vec{K}_G[R_1 \setminus U]$ contains a cycle with colour sequence S.

Proof. With high probability, Lemma 3.2.16 holds with R_1 (and R_2 set to be a p-random set independent with R_1 – R_2 will not be relevant in the proof). Consider a pattern (as in Definition 3.2.12) consisting of a directed path on $k+1$ vertices, and label ith edge with c_i and label the first vertex of the path with v where v is a free variable. This is enough information to determine the label of the remaining k vertices: the label of the ith vertex on the path for $2 \leq i \leq k+1$ has to be $v - \sum_{1 \leq j < i} c_i$ for the pattern to be well-defined (note also that each vertex gets a distinct label).

This pattern is well-distributed. For pairs of colours, separability follows by (b) as $c_i \neq c_j$ when $i \neq j$. For pairs of vertices, separability follows from (b) once again, this time using that S is a path-candidate (recall being a path-candidate implies that all partial sums of S are non-zero). Also, all vertices of this pattern are linear in at least one variable, namely, v. Also, all colours are constants. Therefore, by Lemma 3.2.16, there exists a copy of this pattern in $\vec{K}_G[R_1 \setminus U, (R_2 \setminus U) \cup S]$, which corresponds to a directed path in $\vec{K}_G$ with S as its colour sequence. To justify this correspondence, recall Observation 3.2.15.

If S was a cycle-candidate instead, an analogous argument works, this time starting with a pattern that is a directed cycle of length k, with colour labels coming from S, and one of the vertices labelled v which is a free variable. Separability of colours and vertices are by (b), using the definition of a cycle-candidate. □

Our cover-down statement for small groups of colours is the following. The proof is omitted, it follows easily by iteratively invoking Lemma 3.4.2.

Lemma 3.4.3. *Let $p \geq n^{-1/700}$. Let $3 \leq k \leq 9$. Let R_1 be a p-random subset of G. With high probability, the following holds.*

Let $C \subseteq G$ be a set of $k\ell$ colours, admitting a partition into tuples $C_1, \dots, C_\ell$ where each C_i is a rainbow cycle-candidate. Suppose $\ell \leq p^{400}n/kC_{3.4.3}$. Then, $\mathcal{H}_k[R_1 \setminus U; C]$ contains a matching of size ℓ.

Saturating colours: Large k

When k is large, the strategy in the previous section fails for two reasons. Firstly, assuming that the C_i are cycle-candidates as in Lemma 3.4.2 would be too much of an ask, as in general we do not have a good way of finding orderings of zero-sum sets in this way (for $k \leq 9$, we get to assume this without loss of generality, relying on Lemma 3.2.21). Even if we were able to find such orderings, there is a second issue which is probabilistic which comes up only when $k \geq \log n$. The issue is that the expected number of k-cycles using a particular colour sequence contained in a p-random set is $\leq p^k n$ – this is too small to have any useful analogue of Lemma 3.2.16. Therefore, we pursue a more complicated strategy as follows.

Recall that the families $\mathcal{F}_G$ and $\mathcal{S}_G$ were defined in Lemma 3.2.24.

Lemma 3.4.4. *Let $10 \leq k \leq \log^{10} n$. Let G be a group of order n, where n is sufficiently large. The following statements both hold.*

1. *Let $v, w \in G$ be distinct vertices with $v - w = q_{G,k} + m \cdot f_{G,k}$ for some natural $m \leq k$. Then, there exists a family $\mathcal{P}$ of size $\geq n/(kC_{3.4.4})$ of pairs of rainbow paths (P, P') in $\vec{K}_G$ with the following properties.*

   (a) For each $(P, P') \in \mathcal{P}$, P starts on v, and P' ends on w.

   (b) Each path in $\mathcal{P}$ is rainbow, pairwise colour disjoint, and pairwise vertex disjoint except on $\{v, w\}$.

   (c) For each $(P, P') \in \mathcal{P}$, $C(P) \cup C(P') \in \mathcal{F}_G$

2. *Let $v, w \in G$ be distinct vertices with $v - w = s_{G,k}$. Then, there exists a family of rainbow paths $\mathcal{P}$ of size $\geq n/(kC_{3.4.4})$ in $\vec{K}_G$ with the following properties.*

(a) For each $P \in \mathcal{P}$, P starts on v, and P ends on w.

(b) Each path in $\mathcal{P}$ is rainbow, pairwise colour disjoint, and pairwise vertex disjoint except on $\{v, w\}$.

(c) For each $P \in \mathcal{P}$, $C(P) \in \mathcal{S}_G$

Proof. We suppose $C_{3.4.4}$ is sufficiently large for the following calculations to go through.

Let $\mathcal{P}$ be a maximal family with properties $1(a)$, $1(b)$, and $1(c)$. Suppose $|\mathcal{P}| < n/(kC_{3.4.4})$. Supposing that $C_{3.4.4} \ll C_{3.2.24}$, there exists at least $n/(2kC_{3.2.24})$ many $\mathcal{F}_i \in \mathcal{F}(G)$ which are unused by $\mathcal{P}$. For each such unused $\mathcal{F}_i \in \mathcal{F}(G)$, consider the path P_i with vertex sequence $P_{out}(v, \mathcal{F}_i^+)$ and path P_i' with vertex sequence $P_{in}(w, \mathcal{F}_i^-)$ (these are defined in Section 3.2.2). Note that P_i and P_i' are in fact paths as the corresponding colour sequences are path-candidates by Lemma 3.2.24.

Also by Lemma 3.2.24, the paths P_i and P_i' are vertex disjoint except possibly on v and w, as $\mathcal{F}_i^+$ and $\mathcal{F}_i^-$ are separable at a distance $q + mf$. By the dissociability property coming from Lemma 3.2.24, for each P_i, P_j with $i \neq j$, P_i and P_j are vertex disjoint except on v, and similarly P_i', P_j' are vertex-disjoint except on w. Using these two properties and assuming that $C_{3.4.4} \ll C_{3.2.24}$ we can find some i for which $\mathcal{F}_i$ is unused in $\mathcal{P}$, (P_i, P_i') is vertex-disjoint with all vertices included in $\mathcal{P}$ (note there are at most $10n/kC_{3.4.4}$ such vertices), and $V(P_i) \cap V(P_i') = \emptyset$, contradicting maximality of $\mathcal{P}$.

An analogous argument shows the second part of the statement. □

Lemma 3.4.4 combined with Chernoff's bound implies the following easily.

Lemma 3.4.5. *Let $p \geq n^{-1/700}$, let $10 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

Let $U \subseteq G$ be a set with $|U| \leq p^{50} n/(kC_{3.4.5})$. Let $v, w \in G$ be distinct vertices.

1. *Suppose $v - w = q_{G,k} + m \cdot f_{G,k}$ for some $0 \leq m \leq k$. There exists some $\mathcal{F} \in \mathcal{F}_G$ such that $\vec{K}_G[(R_1 \backslash U) \cup \{v, w\}; \mathcal{F}]$ has two vertex disjoint rainbow directed paths*

of length 2, one directed away from v, one directed into w. Furthermore, $\mathcal{F}$ is disjoint with U.

2. *Suppose $v - w = s_{G,k}$. There exists some $\mathcal{S} \in \mathcal{S}_G$ such that $\vec{K}_G[(R_1 \setminus U) \cup \{v, w\}; \mathcal{S}]$ contains a rainbow directed path (of length $z_{\mathcal{S}}$) from v to w. Furthermore, $\mathcal{S}$ is disjoint with U.*

The next lemma summarises our cover-down strategy for large k. The proof simply iterates parts (1) and (2) of Lemma 3.4.5, and this works due to properties acquired in Lemma 3.2.24.

Lemma 3.4.6. *Let $p \geq n^{-1/700}$, let $10 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

Let $U \subseteq G$ be a set with $|U| \leq p^{600}n/(kC_{3.4.6})$. Suppose ℓ is some positive integer with $\ell \leq p^{600}n/(k^2C_{3.4.6})$. Let $C \subseteq G$ be a set of 4ℓ colours admitting a partition into 4-tuples $C_1, \dots, C_\ell$ where each C_i is a rainbow path candidate with sum $q_{G,k}$. Then, $\mathcal{H}_k[R_1 \setminus U; (R_2 \setminus U) \cup C]$ has a matching saturating C, and the set of colours C^* used on the matching other than C is closed under $\mathcal{F}_G$ and $\mathcal{S}_G$. Moreover, C^* uses exactly $\ell(k - 4 - z_{\mathcal{S}})/4$ tuples from $\mathcal{F}_G$ and exactly ℓ tuples from $\mathcal{S}_G$. As a consequence, the matching consists of exactly ℓ edges of $\mathcal{H}_k$.

Proof. Fix some large K and some $C_{3.4.6} \gg K$. With high probability, Lemma 3.4.5 and Lemma 3.4.2 both hold.

We will first prove the statement when $\ell = 1$ with $C_{3.4.6}$ replaced with K (note this strengthens the statement). Initialise $U_1 = U$, and for each $i \in \{1, \dots, \ell\}$, do the following.

In $\vec{K}_G[R_1 \setminus U_i; (R_2 \setminus U_i) \cup C]$, find a rainbow path P_0 with colours C_1 via Lemma 3.4.2. Note that $\mathrm{start}(P_0) - \mathrm{end}(P_0) = q_{G,k} + m \cdot f_{G,k}$ for some natural $m \leq k$ (in fact $m = 0$). For a natural i, while $k - |P_i| > z_{\mathcal{S}}$, apply Lemma 3.4.5(1) to extend P_i into another rainbow path P_{i+1} using a set of 4 extra colours coming from an element of $\mathcal{F}_G$ (disjoint with U_i). Note that this preserves that $\mathrm{start}(P_{i+1}) - \mathrm{end}(P_{i+1}) = q_{G,k} + m \cdot f_{G,k}$ for some $0 \leq m \leq k$. Update U_{i+1} to include the new colours and vertices used in P_{i+1}.

In at most k steps, this procedure yields a path P with $k - |E(P)| = z_{\mathcal{S}}$ due to the divisibility conditions coming from Lemma 3.2.24, note also $|U_i| \leq |U| + 10i$, so we add at most $10k$ new elements to U throughout the process. At this point, we can apply Lemma 3.4.5(2) to complete P into a rainbow cycle, say $\mathcal{C}$, and therefore an edge of $\mathcal{H}_k$. Also, observe that while building $\mathcal{C}$, we used $(k - 4 - z_{\mathcal{S}})/4$ tuples from $\mathcal{F}_G$ and one tuple from $\mathcal{S}_G$.

When $2 \leq \ell \leq p^{600}n/(k^2 C_{3.4.6})$, we can repeat the above procedure for each of $C_2, \dots, C_\ell$, at each iteration including in U the set of at most $10k$ vertices and colours used in the previous step. This would add at most $10k\ell \leq 10p^{600}n/kC_{3.4.6}$ new elements to U, which means U will never exceed a size of $11p^{600}n/kC_{3.4.6} \ll p^{600}n/kK$ throughout the process. Hence at each step, we can invoke (the stronger version of) the $\ell = 1$ case. □

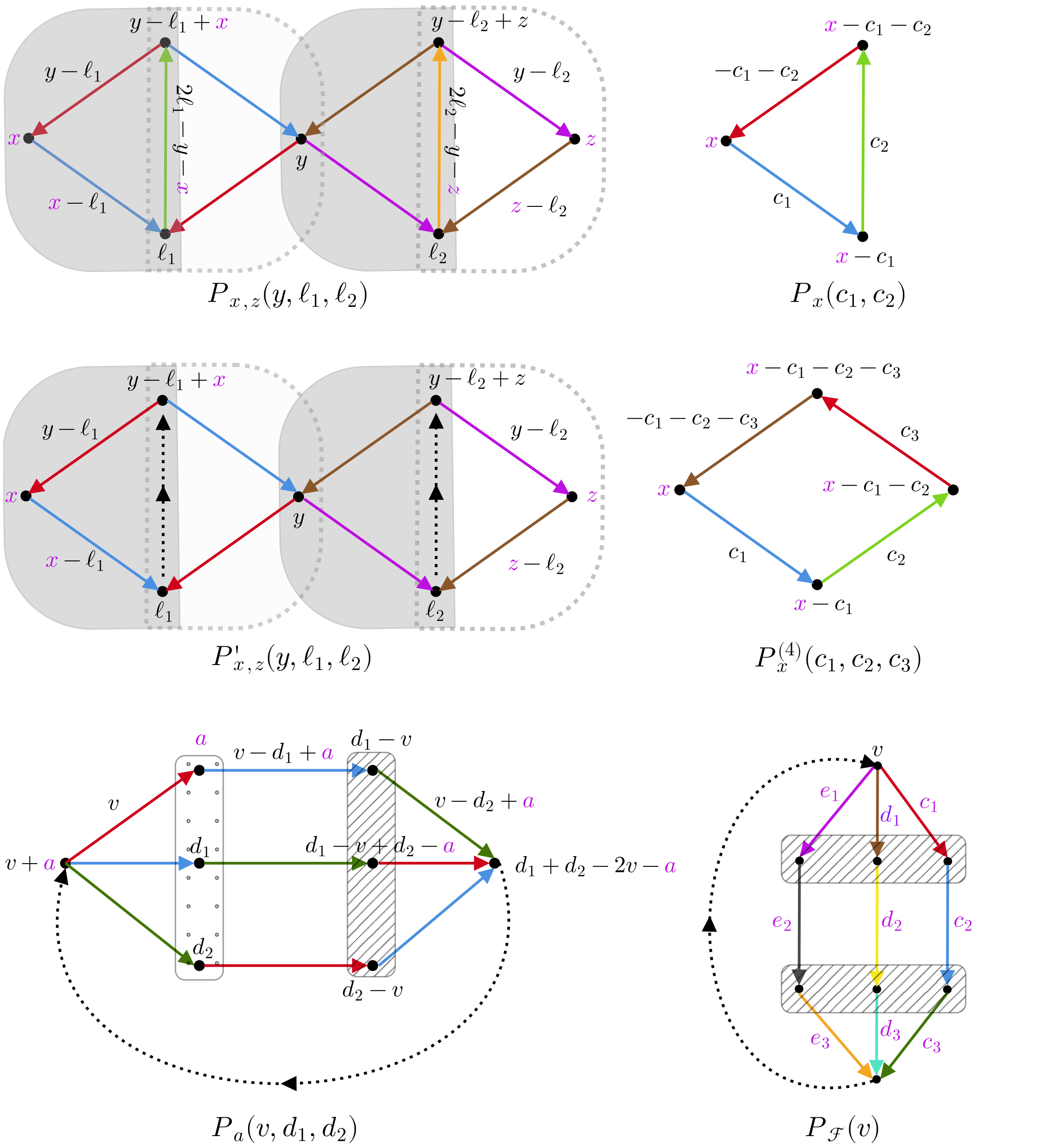
3.4.2 Distributive absorption in $\mathcal{H}_k$

Definition 3.4.7. Let H be a hypergraph and let $\mathcal{F} = \{S_1, \dots, S_t\}$ be a family of subsets of $V(H)$. We say that a set of vertices R **m-absorbs** $\mathcal{F}$ if for every subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size m, there is a hypergraph matching whose vertex set is exactly $R \cup \bigcup_{S_i \in \mathcal{F}'} S_i$.

We will build the desired absorbing structures by finding collections of small subgraphs with certain properties. Each structure found in this section is formed by combining patterns coming from Figure 3.1. Therefore, the following result is crucial, as it will allow us to use Lemma 3.2.16 numerous times throughout the section (we remark that $P_{\mathcal{F}}(V)$ is excluded in the statement of the following lemma as we treat this pattern separately).

Lemma 3.4.8. *In Figure 3.1, all patterns depicted except for $P_{\mathcal{F}}(V)$ are well-defined patterns which are well-distributed.*

Proof. The definition of a pattern and the second part of the definition of well-distributed is easy to verify by inspecting the figure, so we focus on checking



$P_{x,z}(y,\ell_1,\ell_2)$

$P_x(c_1,c_2)$

$P'_{x,z}(y,\ell_1,\ell_2)$

$P_x^{(4)}(c_1,c_2,c_3)$

$P_a(v,d_1,d_2)$

$P_{\mathcal{F}}(v)$

Figure 3.1: Several patterns (see Definition 3.2.12) used in Section 3.4.2. Edge colours correspond to edge labels, so for a collection of edges with the same colour, the label is written for only one of the edges. Free variables are denoted in black letters, and elements of G are denoted in pink letters. The dashed arrows indicate that after a copy of the pattern is found, a rainbow directed path (of the appropriate length) between the indicated vertices will be found. For triples tiled with the diagonal lines, in the proof we find a 2-absorber for the indicated vertices.

pairwise separability for vertices and colours. Hence, we focus on the first part of the statement for each individual pattern.

- $P_{x,z}(y, \ell_1, \ell_2)$. For this pattern and the next, we implicitly assume that $x \neq z$, making x and z separable by (b). All other pairs of vertices are separable by (a). Except for the two pairs of colours $(2\ell_1 - y - x, y - \ell_1)$ and $(2\ell_2 - y - z, z - \ell_2)$ which are separable by (c), all pairs of colours are separable by (a).
- $P'_{x,z}(y, \ell_1, \ell_2)$. This pattern is well-distributed as it is a strict subset of $P_{x,z}(y, \ell_1, \ell_2)$.
- $P_x(c_1, c_2)$. All pairs of vertices and colours are separable by (a).
- $P_x^{(4)}(c_1, c_2, c_3)$. All pairs of vertices and colours are separable by (a).
- $P_a(v, d_1, d_2)$. All pairs of vertices and colours are separable by (a).

Hence, each pattern is well-distributed as desired. □

The following remark is crucial to keep in mind as we will make many references to Figure 3.1 in the rest of the section.

Remark 4. On numerous occasions, we will consider multiple *instances* of the same *pattern type* in Figure 3.1. For example, $P_{x,z'}(y', \ell_1, \ell_2)$ refers to the pattern depicted in Figure 3.1 denoted $P_{x,z}(y, \ell_1, \ell_2)$ with y replaced with y', z replaced with z', but ℓ_1 and ℓ_2 unchanged. So $P_{x,z}(y, \ell_1, \ell_2)$ and $P_{x,z}(y', \ell_1, \ell_2)$ are different patterns with some overlap in the vertex and edge labels they receive.

Vertex-switchers

We first show how to 1-absorb a set of 2 vertices. We emphasise that when we say a vertex of $\mathcal{H}_k$ in this section, we specifically mean a vertex which corresponds to a vertex of $\vec{K}_G$, as opposed to a colour of $\vec{K}_G$. This is crucial, as in fact it is impossible to build a 1-absorber for a set of 2 vertices of $\mathcal{H}_k$ which correspond to colours of $\vec{K}_G$. This follows from Observation 3.1.1.

Lemma 3.4.9. *Let $p \geq n^{-1/700}$. Let $3 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

For any distinct vertices $x, z \in \mathcal{H}_k$ and $U \subseteq G$ with $|U| \leq p^{300}n/C_{3.4.9}$, $\mathcal{H}_k[(R_1 \setminus U) \cup \{x, z\}, (R_2 \setminus U)]$ contains a subgraph of size at most $10k$ that 1-absorbs $\{x, z\}$.

Proof. With high probability, Lemma 3.2.16 and Lemma 3.4.1 both hold.

Suppose first that $k = 3$. Consider the pattern $P = P_{x,z}(y, \ell_1, \ell_2)$ from Figure 3.1. A copy of P can be found in $\vec{K}_G[(R_1 \setminus U) \cup \{x, z\}, (R_2 \setminus U)]$ by Lemma 3.2.16. Inspecting the two sets of matchings (the solid matching and the dashed matching) in Figure 3.1 (top-left), we see that such a copy of P corresponds to a desired absorbing subgraph of $\mathcal{H}_k[(R_1 \setminus U) \cup \{x, z\}, (R_2 \setminus U)]$ of size ≤ 30. Recall Observation 3.2.15 to justify this correspondence.

Suppose now that $4 \leq k \leq \log^{10} n$. In this case consider a copy of the pattern $P = P'_{x,z}(y, \ell_1, \ell_2)$ from Figure 3.1 given by Lemma 3.2.16. To complete the absorber, we need to find rainbow paths of size $k - 2 \geq 2$ from ℓ_1 to $y - \ell_1 + x$ and from ℓ_2 to $y - \ell_2 + x$ (when we say ℓ_1, we mean the copy of the vertex with the label ℓ_1 in $P'_{x,z}(y, \ell_1, \ell_2)$, similarly for the other variables. We use this convention in the rest of this section). Two applications of Lemma 3.4.1(1) allows us to find these paths. As in the previous case, the solid matching and the dashed matching demonstrates that the desired absorption property holds. □

Chaining together gadgets which 1-absorb pairs as in the previous lemma, we can construct gadgets which $(s-1)$-absorb sets of size s.

Lemma 3.4.10. *Let $p \geq n^{-1/700}$. Let $3 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

Let S be a vertex-subset of size at most 100 and let $U \subseteq G$ with $|U| \leq p^{310}n/C_{3.4.10}$. Then, there are sets $V' \subseteq R_1 \setminus U$ and $C' \subseteq R_2 \setminus U$ of size $\leq 10^4 k$ such that $V' \cup C'$ $(|S| - 1)$-absorbs S in $\mathcal{H}_k$.

Proof. With high probability, Lemma 3.4.9 holds.

Let $S = \{a_1, \ldots, a_\ell\}$ be given for some ℓ with $2 \leq \ell \leq 100$. For every $i \in \{1, \ldots, \ell - 1\}$, apply Lemma 3.4.9 with $(a, b) = (a_i, a_{i+1})$, each time finding a

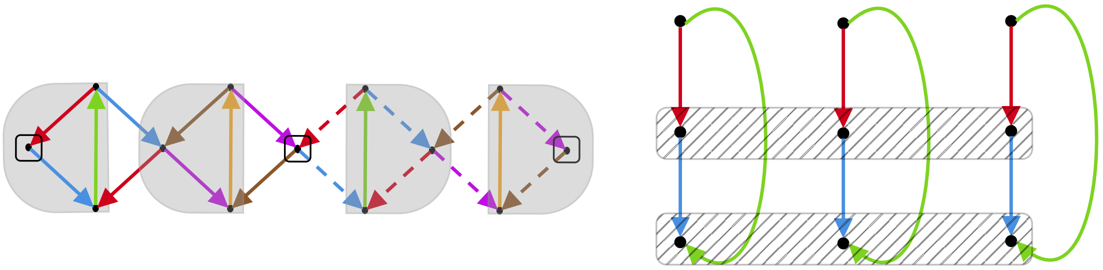


Figure 3.2: *On the left:* A 2-absorber for the for the 3 vertices contained in boxes (from the proof of Lemma 3.4.9). Dashed versions of a coloured edge are to be interpreted as having a distinct colour. The matching which absorbs the outer two vertices is indicated. *On the right:* The pattern P_S (from the proof of Lemma 3.4.11 Case 1) consisting of 3 copies of the pattern $P_x(c_1, c_2)$ ($\ell = 3$), and the sets S_2 and S_3 shaded with the diagonal lines. For the vertices covered with the diagonal lines we have a 2-absorber for the indicated vertices. The union of these two 2-absorbers and the illustrated directed graph 1-absorbs the 3-vertices on the top row of the diagram.

subgraph F_i which 1-absorbs $\{a_i, a_{i+1}\}$ disjointly with U. By extending U in each application to include $F_1, \ldots, F_{i-1}$, we can also ensure the collection of F_i are vertex and colour disjoint (except for elements of S). The union of the subgraphs F_i ($i \in [\ell]$) now has the desired absorption property. For an illustration of the case when $k = \ell = 3$, see Figure 3.2 (left). □

Now we show how to 1-absorb sets of size $s \leq 100$. For small k, essentially all the necessary ideas are included in the previous two lemmas, however for large k we introduce some new ideas.

Lemma 3.4.11. *Let $p \geq n^{-1/700}$ and let $3 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

Let S be a vertex-subset of size at most 100 and let $U \subseteq G$ with $|U| \leq p^{340}n/C_{3.4.11}$, there are sets $V' \subseteq R_1 \setminus U$ and $C' \subseteq R_2 \setminus U$ of size $\leq 10^8k^2$ such that $V' \sqcup C'$ 1-absorbs S in $\mathcal{H}_k$.

Proof. With high probability, Lemma 3.2.16, Lemma 3.4.1 and Lemma 3.4.10 all hold. Denote $S = \{a_1, a_2, \ldots, a_\ell\}$ where $2 \leq \ell \leq 100$.

Case 1: $k \in \{3,4\}$. We write the details of the argument for $k=3$, for $k=4$ a proof can be obtained by replacing $P_{a_i}(c_1,c_2)$ with $P_{a_i}^{(4)}(c_1,c_2,c_3)$ in the below argument[1].

Consider a pattern P_S formed by the union of patterns $P_{a_i}(c_1,c_2)$ for each $i \in [\ell]$ (recall Remark 4). This pattern is well-defined as the vertices of $P_{a_i}(c_1,c_2)$ and $P_{a_j}(c_1,c_2)$ get different labels when $i \neq j$ (as $a_i \neq a_j$). To see that this pattern is well-distributed we only need to check vertices and edges belonging to different copies, as each $P_{a_i}(c_1,c_2)$ is well-distributed by Lemma 3.4.8 already. There is nothing to check with edge labels as they are all identical. Vertices in the same position in the triangle are separable by (b) (this is as $a_i \neq a_j$ when $i \neq j$), and vertices in different positions are separable by (a). By Lemma 3.2.16, P_S admits a copy in $\vec{K}_G[(R_1 \setminus U) \cup S; R_2 \setminus U]$, say P'_S. For $j \in \{2,3\}$, denote by S_2 the vertices of F'_S coming from copies of top vertices of $P_{a_i}(c_1,c_2)$ and denote by S_3 the vertices coming from bottom vertices of $P_{a_i}(c_1,c_2)$ for each $i \in [\ell]$. Apply Lemma 3.4.10 first with $S=S_2$, and then $S=S_3$, to find sets A_2 and A_3 $(|S|-1)$-absorbing S_2 and S_3, respectively. We can also ensure that A_2, A_3 and F'_S are vertex and colour disjoint by extending U in each successive application of Lemma 3.4.10. Now $F'_S \cup A_2 \cup A_3$ is the desired absorber. See Figure 3.2 for a demonstration of this when $\ell=3$, the boxes shaded with diagonal lines represent the sets S_2 and S_3 which we can 2-absorb.

Case 2: $k \geq 5$. We begin by the following observation which will help to motivate the choice of parameters for the rest of the argument. We remind the reader that when we say d_1+d_2-2v-a below, what we mean is the copy of the vertex of $P_a(v,d_1,d_2)$ with the label d_1+d_2-2v-a, and similarly for the other expressions.

Observation 3.4.12. *Consider a copy P of $P_a(v,d_1,d_2)$ from Figure 3.1. Let Q be a rainbow path from d_1+d_2-2v-a to $v+a$, colour-disjoint with P, and vertex-disjoint with P except on the endpoints. Let A be a set that 2-absorbs $Y := \{d_1-v, d_1-v+d_2-a, d_2-v\}$ (the column shaded by the diagonal lines), vertex and colour disjoint with $P \cup Q$, except on Y. Then, $P \cup A \cup Q \setminus \{a, d_1, d_2\}$ 1-absorbs*

[1] The same argument works for each $k = O(1)$, the reason why the second case exists is the range when $k = \Omega(\log n)$.

the vertex-triple $\{a, d_1, d_2\}$.

The previous observation essentially shows that we can find sets 1-absorbing triples, provided that 2 elements of the triples are *free variables* (it is important that d_1 and d_2 are free variables while ensuring that $P_a(v, d_1, d_2)$ is well distributed, and that d_1+d_2-2v-a is linear in at least one variable). The remainder of the proof is focused on using this property many times with carefully chosen sets of triples to find subgraphs that 1-absorb sets of size at most 100 constants (not free variables). The following observation motivates the choice of triples.

Observation 3.4.13. *Consider a collection of sets* $\{a_1, d_\ell, d_1\}$, $\{a_2, d_1, d_2\}$, $\{a_3, d_2, d_3\}$, $\ldots$, $\{a_{\ell-1}, d_{\ell-2}, d_{\ell-1}\}$, $\{a_\ell, d_{\ell-1}\}$. *Suppose for some* a_i, *the set containing* a_i *is deleted from the collection. Then, there exists a choice of an element from each of the remaining sets in the collection so that overall the chosen elements are precisely* $\{d_1, \ldots, d_{\ell-1}\}$.

Proof. When $i = 3$, the correct choices are displayed in the below table.

$$\begin{array}{ccccccc}
a_1 & a_2 & \mathbf{a_3} & a_4 & \cdots & a_{\ell-1} & a_\ell \\
d_\ell & d_1 & d_2 & \mathbf{d_3} & \cdots & \mathbf{d_{\ell-2}} & \mathbf{d_{\ell-1}} \\
\mathbf{d_1} & \mathbf{d_2} & d_3 & d_4 & \cdots & d_{\ell-1} &
\end{array}$$

For other values of i, one can similarly choose the value on the bottom row for the jth column where $j < i$, and choose the value on the middle row for the jth column where $j > i$. $\square$

Consider a pattern P formed by the union of patterns $P_{a_1} := P_{a_1}(v_1, d_\ell, d_1)$, $P_{a_2} := P_{a_2}(v_2, d_1, d_2)$, $P_{a_3} := P_{a_3}(v_3, d_2, d_3)$, $\ldots$, $P_{a_{\ell-1}} := P_{a_{\ell-1}}(v_{\ell-1}, d_{\ell-2}, d_{\ell-1})$ (recall Remark 4). Formally, P is obtained by taking the (disjoint) union of each of the graphs P_{a_i} and identifying together vertices which share the same label (this identification step ensures that the second property in the definition of pattern is satisfied, so P is indeed a well-defined pattern). We remark that the vertex/edge labels live in the set $(G * F_{2\ell-1})^{\text{ab}}$.

Claim 3.4.13.1. *P is well-distributed.*

Proof. Each $P_{a_i}(\cdot,\cdot,\cdot)$ is well-distributed by Lemma 3.4.8, so we only need to check separability for pairs of vertices and pairs of colours coming from different $P_{a_i}(\cdot,\cdot,\cdot)$. For pairs of such colours, the word corresponding to a label includes v_i as a free variable for some i, and the other label does not include v_i. This makes the pair separable by (a). Similarly, for pairs of vertices, if one of the words has a v_i as a free variable for some i, we have separability by (a). If this is not the case the pair of words could be of the form (d_i, d_j) for some i and j. If $i \neq j$, we have separability by (a), and otherwise, the words are identical and in this case we do not need to check separability, as in the definition of the union we identify such vertices together. The pair of words could also be of the form $(d_i, d_j + d_{j+1} - 2v_j - a_j)$, in which case either d_j or d_{j+1} is different with d_i, giving separability by (a). Finally, pair of words could also be of the form $(d_i + d_{i+1} - 2v_i - a_i, d_j + d_{j+1} - 2v_j - a_j)$ where i and j are distinct, so we again have separability by (a). □

By Lemma 3.2.16, we can find a copy of P, say P', in $\vec{K}_G[R_1 \setminus U; R_2 \setminus U]$. In addition, by Lemma 3.4.9 applied with $\{a, b\} = \{a_\ell, d_{\ell-1}\}$, we obtain a subgraph A which 1-absorbs $\{a_\ell, d_{\ell-1}\}$. By extending U in this application, we can ensure A is disjoint with P'.

Recall that $k - 3 \geq 2$ by assumption. For each $i \in [\ell - 1]$, apply Lemma 3.4.1 with u set to be the copy of the rightmost vertex in P_{a_i} and v set to be the copy of the leftmost vertex in P_{a_i} to find a rainbow path of length $k-3$ directed from u to v that does not clash with any of the forbidden colours or vertices (see the dashed line in Figure 3.1). We can achieve this by iteratively invoking Lemma 3.4.1, extending U at each step. Note that $k \leq \log^{10} n$, so we never add more than $200 \log^{10} n$ elements to U during this process. Again, for each $i \in [\ell - 1]$, apply Lemma 3.4.10 with S set the be the subset corresponding to the copies of the 3 vertices of P_{a_i} which correspond to the column indicated with the dots to obtain a subgraph A_{a_i} which 2-absorbs this subset. We can ensure that the collection of A_{a_i} are pairwise disjoint (except for the vertices corresponding to highlighted vertices plugged into S), and also disjoint with U, again by extending U in each application of Lemma 3.4.10. By the bound coming from Lemma 3.4.10, we never have to extend U by more than $10^6k^2\ell$ elements during this process.

For each $i \in [\ell-1]$, let P'_{a_i} be the copy of P_{a_i} combined with the path found by applying Lemma 3.4.1 and A_{a_i}, and let Z_{a_i} be the set of vertices corresponding to the copies of the vertices of P_{a_i} (the column indicated by dots). The following is a rephrasing of Observation 3.4.12.

Observation 3.4.14. *We have that $P'_{a_i} \setminus Z_{a_i}$ 1-absorbs Z_{a_i}.*

Now, we claim that $(\bigcup P'_{a_i} \setminus S) \cup A$ is the desired absorber. To see this, let $a_i \in S$. We wish to show that the vertices and colours of $(\bigcup P'_{a_i} \setminus S) \cup A \cup \{a_i\}$ induce a perfect matching in $\mathcal{H}$. To see this, first take a perfect matching in $(P'_{a_i} \setminus Z_{a_i}) \cup \{a_i\}$ (which exists by definition of 1-absorbing and Observation 3.4.14).

Now, from each set Z_{a_j} (for $i \neq j$) and $\{a_\ell, d_{\ell-1}\}$ (coming from the 1-absorbing A), we can select exactly one $d_{j'}$ so that each $d_{j'}$ (for each $j' \in [\ell-1]$) is selected precisely once, using Observation 3.4.13. Using either Observation 3.4.14 or the 1-absorbing property of A, we can find a perfect matching of $(\bigcup P'_{a_i} \setminus S) \cup A \cup \{a_i\}$ as required. □

Colour-switchers

Lemma 3.4.15. *Let $p \geq n^{-1/700}$. Let $5 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

Let $\alpha \in G$, and $s \in \mathbb{N}$ with $2 \leq s \leq \min\{k-2, 100\}$. Let S be a disjoint and near-dissociable family of rainbow s-tuples of colours, each tuple sums to α, and each tuple is a path-candidate, and suppose $|S| \leq 100$. Let $U \subseteq G$ with $|U| \leq p^{400}n/C_{3.4.15}$. Then, there are sets $V' \subseteq R_1 \setminus U$ and $C' \subseteq R_2 \setminus U$ of size $\leq 10^{11}k^2$ such that $V' \cup C'$ 1-absorbs S in $\mathcal{H}_k$.

Proof. With high probability, Lemma 3.2.16, Lemma 3.4.10 and Lemma 3.4.1 hold. Let S and U be given as in the statement. Consider a pattern P constructed as follows. Take $|S|$ directed paths, each of length s, with the same start and end vertices, but internally vertex-disjoint. Label the start vertex with the free variable

v. Label the edges of the ith directed path with the ith s-tuple of S (counting in order of proximity to v). Note this induces a labelling on each of the remaining vertices of P. This labelling is well-defined on the end-vertex, because each tuple in S has the same sum, namely, α. In particular, the end-vertex receives the label $v - \alpha$. For an illustration of the pattern when $s = 3$ and $S = \{(e_1, e_2, e_3), (d_1, d_2, d_3), (c_1, c_2, c_3)\}$, inspect the bottom-right pattern in Figure 3.1.

Observation 3.4.16. *P is well-distributed.*

Proof. Each pair of colours is separable by (b) as distinct coordinates of elements of S are distinct, and the elements of S are pairwise disjoint. We claim each pair of vertices is separable by (b). For vertices belonging to the same directed path, this follows as elements of S are path-candidates. For vertices belonging to different directed paths, this follows as S is near-dissociable. □

Thus, we may apply Lemma 3.2.16 to find a copy of P, say P', in $\vec{K}_G[R_1 \setminus U; (R_2 \setminus U) \cup \bigcup S]$. For each $i \in [s-1]$, let P_i denote the vertices of distance i from v, noting $|P_i| = |S| \leq 100$. Apply Lemma 3.4.10 for each P_i to find (disjointly) sets P_i' which $|S| - 1$ absorb P_i. Finally, noting that $k - s \geq 2$, apply Lemma 3.4.1 with u as the end-vertex of the path of length s from v, and $v = v$, and $k' = k - s$, to find a rainbow path of length $k - s$. It is easy so see that the resulting structure has the desired absorption property. □

Putting the gadgets together

So far we have lemmas allowing us to find sets 1-absorbing arbitrary sets of size ≤ 100. How can we go from to sets which can h absorb sets of size $(1+\beta)h$, where h and βh are potentially linear in n? The bipartite graph given in the next lemma gives us a nice collection of subsets of size at most 100 to 1-absorb, which together have the desired property. In this sense the utility of this bipartite graph is similar in spirit to that of the sequence constructed in Observation 3.4.13 (which could be viewed as a bipartite graph with maximum degree 3 with much weaker properties).

Lemma 3.4.17 (Montgomery, [45])**.** *Let $0 < \beta \leq 1$. There is a positive integer h_0 such that for every $h \geq h_0$ there exists a bipartite graph K with maximum degree at most 100 and vertex classes X and $Y \cup Y'$ with $|X| = 3h$, $|Y| = 2h$, $|Y'| = h + \beta h$ so that the following holds. For any $Y_0 \subseteq Y'$ with $|Y_0| = h$, there is a perfect matching between X and $Y \cup Y'$.*

Graphs produced by this lemma are called *robustly matchable bipartite graphs.*

Lemma 3.4.18. *Let $p \geq n^{-1/700}$, $3 \leq k \leq \log^{10} n$. Let R_1, R_2 be p-random subsets of G sampled independently. With high probability, the following holds.*

Let $0 < \beta \leq 1$. Let $h \in \mathbb{N}$ with $\sqrt{n} \leq h \leq p^{400}n/C_{3.4.18}k^2$. Let $U \subseteq G$ with $|U| \leq n^{999/1000}$. Let Y' be a subset of size $(1+\beta)h$ with one of the following forms.

1. *$Y' \subseteq \mathcal{H}_k$ is a vertex-subset of $\mathcal{H}_k$.*

2. *Let $\alpha \in G$, and $s \in \mathbb{N}$ with $2 \leq s \leq \min\{k-2, 100\}$. Y' is a disjoint and near-dissociable family of s-tuples of colours of $\mathcal{H}_k$ where each tuple sums to α, and each tuple is a path-candidate (hence $\alpha \neq 0$).*

Then, there exists a set $A \subseteq \mathcal{H}_k \setminus U$ of size $\leq 10^{15}k^2 3h$ where A h-absorbs Y'.

Proof. We show how to prove part (1) of the statement using Lemma 3.4.11. The proof for part (2) is essentially the same, using Lemma 3.4.15 instead.

With high probability, Lemma 3.4.11 holds and R_2 has size at least $10h$.

Let β, h, U, and Y' be given. As h is sufficiently large, we can apply Lemma 3.4.17 to construct a bipartite graph G with parameters β and h with vertex classes X and $Y \cup Y'$. Here, we associate the given set of vertices Y' with the Y' that denotes a set of vertices of G. We arbitrarily associate, disjointly with Y' and U, a subset of vertices of $\mathcal{H}_k$ from R_2 with Y ($|R_2| \geq 10h$, so there is space to do this). Now, each element of $x \in X$ is linked, via the graph G, to a subset of vertices of $\mathcal{H}_k$ of size at most 100, i.e. the neighbourhood which we denote $N_G(x)$. For each element of $x \in X$, we will find a $A_x \subseteq \mathcal{H}_k$ that 1-absorbs $N_G(x)$, and the collection of A_x we find will be disjoint except on elements of $Y \cup Y'$. The property from Lemma 3.4.11

allows us to do this greedily, extending U with 10^8k^2 elements at each step, adding to U at most

$$10^8k^2|X| \leq 10^8k^23h \leq 3 \cdot 10^8k^2p^{400}n/(C_{3.4.18}k^2) \leq 3 \cdot 10^8p^{400}n/C_{3.4.18}$$

elements. Combined with the initial elements of U, this means that U never exceeds a size of $p^{340}n/C_{3.4.11}$ if $C_{3.4.18}$ is sufficiently large. This means that the applications of Lemma 3.4.11 are valid.

We claim that the union of the A_x with Y have the desired absorption property. To see this, take some subset $Y_0 \subseteq Y'$ of size h. In G, we have perfect matching f from X to $Y \cup Y_0$. For each A_x, use the matching of $A_x \cup \{f(x)\}$ which exists by the absorption property of A_x. These matchings together give a matching of $\bigcup A_x \cup Y \cup Y_0$, as required.

□

3.4.3 Proof of Lemma 3.1.4

Now we combine the distributive absorption strategy with the cover-down strategy to give a proof of Lemma 3.1.4. Recall the convention about random subsets of random sets given before the proof of Theorem 3.1.3.

Proof. Let $K = K_{3.1.4} \geq 1$ be sufficiently large, and fix $\varepsilon = \varepsilon_{3.1.4} \ll 1/K$, so that in particular, $\varepsilon K \leq 10^{-10}$ holds.

Case 1: $3 \leq k \leq 9$. Write $p = p_1 + p_2$ where $p_2 = p^{500}/(10C_{3.4.18}k^3)$. Partition R_i into disjoint p_1 and p_2-random sets $R_i^{(1)}$ and $R_i^{(2)}$ for each $i \in [2]$. Let $R_i^* \subseteq R_i^{(2)}$ be a r-random subset of G where $r = p^{500}/(1000C_{3.4.18}k^3)$. With high probability, $R_1^{(2)}$ satisfies Lemma 3.4.3, R_2^* satisfies Lemma 3.2.22 as well as Lemma 3.2.19, (R_1^*, R_2^*) satisfies Lemma 3.4.1, and $(R_1^{(1)}, R_2^{(1)})$ satisfies Lemma 3.4.18 (the necessary lower bounds for the corresponding randomness parameters in each of these applications is satisfied for a small enough value of $\varepsilon_{3.1.4}$). With high probability, the size of each random set is at most $n^{0.6}\log n$ away from its expectation. All these properties hold simultaneously with high probability.

Now, let $U \subseteq G$ with $|U| \le n^{4/5}$, without relabelling, include 0 in U. By Lemma 3.2.19, we can find a subset $R_2^{**} \subseteq R_2^* \backslash U$ using all but at most k elements of $R_2^* \backslash U$ such that $\sum R_2^{**} = 0$ and k divides $|R_2^{**}|$. Set β and h so that they satisfy the two identities $(1+\beta)h = |R_1^{(2)} \setminus U|$ and $\beta h = |R_2^{**}|$ (so $h = |R_1^{(2)} \setminus U| - |R_2^{**}| \ge \sqrt{n}$, and $0 < \beta \le 1$ by choice of r). Apply Lemma 3.4.18(1) with these values of β and h and $Y' := R_1^{(2)} \setminus U$ to obtain an absorbing set A contained in $R_1^{(1)} \cup R_2^{(1)} \setminus U$ (the necessary upper bound on h holds by definition of p_1, p_2 using that each random set has size close to its expectation).

We claim now that $A \cup R_2^{**} \cup (R_1^{(2)} \setminus U)$ has the desired property (of $V \cup C$ in the statement). To see this, take $V', C' \subseteq G$ with $|V'| = |C'| = m$ as in the statement of the lemma. As $m \ll r^{300}n/kC_{3.4.1}$ (supposing $K_{3.1.4}$ is sufficiently large), by Lemma 3.4.1(2), there exists a matching M_1 of size exactly m in $\mathcal{H}_k$ saturating V' and using exactly $(k-1)m$ vertices from $R_1^* \backslash U$ and km vertices from R_2^{**}. $C'' := C' \cup (R_2^{**} \backslash V(M_1))$ is a zero-sum set whose order is divisible by k with small symmetric difference with R_2^* (note that $|C''| = m + |R_2^{**}| - km$, so $|C'' \Delta R_2^*| \le 10km + n^{4/5} \le 10k(p/k \log n)^K n + n^{4/5} \le r^{10^{10}} n/\log(n)^{10^{23}}$ supposing K is sufficiently large). Hence C'' can be partitioned into k-sets which are cycle-candidates by Lemma 3.2.22(1). This partition allows us to apply Lemma 3.4.3 to deduce that there exists a matching M_2 saturating the colours C'' using exactly $|C''|$ vertices from the set $R_1^{(2)} \backslash U \backslash V(M_1)$. Observe that in total we used exactly $|R_2^{**}| = \beta h$ vertices from $R_1^{(2)} \backslash U$, and therefore the remaining vertices in $R_1^{(2)} \backslash U$ combined with A admits a perfect matching M_3 by the absorption property of A. Then, $M_1 \cup M_2 \cup M_3$ is the desired perfect matching of $A \cup R_2^{**} \cup V' \cup C' \cup (R_1^{(2)} \setminus U)$.

Case 2: $k \ge 10$. Set $q_1 = p^{500}/(10^{10}C_{3.4.18}k^2)$, $q_2 = (p - q_1)/3$, $r_* = q_1/1000k^{10}$. Let $R_1^{(1)}$, $R_1^{(2)}$, $R_1^{(3)}$, $R_1^{(4)}$ be disjoint subsets of R_1, and q_1, q_2, q_2, q_2-random, respectively. Let $R_1^{(1,1)}, R_1^{(1,2)} \subseteq R_1^{(1)}$ be r_*-random and $(q_1 - r_*)$-random and disjoint. Let R_2^*, $R_2^{(1)}$, $R_2^{(2)}$, $R_2^{(3)}$, $R_2^{(4)}$ be disjoint subsets of R_2 and r_*, $(q_1 - r_*)$, q_2, q_2 and q_2-random, respectively.

With high probability, Lemma 3.4.1 holds for $(R_1^{(1,1)}, R_2^*)$, Lemma 3.2.22 holds for

R_2^*, Lemma 3.4.6 holds for $(R_1^{(1,2)}, R_2^{(1)})$, Lemma 3.4.18 holds for each of $(R_1^{(2)}, R_2^{(2)})$, $(R_1^{(3)}, R_2^{(3)})$, $(R_1^{(4)}, R_2^{(4)})$, Lemma 3.2.19 holds for R_2^*, and the size of each random set is at most $n^{0.6}\log n$ away from its expectation. These applications are valid supposing $\varepsilon_{3.1.4}$ is small enough, i.e. p is large enough.

Let U be given, as before, include 0 in U. Fix f to be the largest integer bounded above by $|R_2^* \setminus U|$ with the property that $f-(k-1)m$ is divisible by 4. By Lemma 3.2.19, we can fix a f-subset $R_2^{**} \subseteq R_2^* \setminus U$ using all but at most 4 vertices from the latter set such that $\sum R_2^{**} = ((f-(k-1)m)/4) \cdot q_{G,k}$.

Set β_1 and h_1 be so that $(1+\beta_1)h_1 = |R_1^{(1)} \setminus U|$ and $\beta_1 h_1 = (k-1)m + k(f-(k-1)m)/4$. Denote by $\mathcal{F}'_G$ the family of sets from $\mathcal{F}_G$ which are entirely contained in $R_2^{(1)} \setminus U$. Similarly, denote by $\mathcal{S}'_G$ the family of sets from $\mathcal{S}_G$ which are entirely contained in $R_2^{(1)} \setminus U$. Set β_2 and h_2 so that $(1+\beta_2)h_2 = |\mathcal{F}'_G|$ and $\beta_2 h_2 = ((f-(k-1)m)/4)(k-4-z_{\mathcal{S}})/4$ (recall this is an integer by Lemma 3.2.24). Set β_3 and h_3 so that $(1+\beta_3)h_3 = |\mathcal{S}'_G|$ and $\beta_3 h_3 = (f-(k-1)m)/4$.

Apply Lemma 3.4.18(1) with $(R_1^{(2)}, R_2^{(2)})$ and $Y' = R_1^{(1)} \setminus U$ with parameters (β_1, h_1) to obtain a set A_1 (disjoint with U) with a vertex-absorption property. Apply Lemma 3.4.18(2) with $(R_1^{(3)}, R_2^{(3)})$ and $Y' := \mathcal{F}'_G$ with parameters (β_2, h_2) to obtain a set A_2 (disjoint with U and A_1) with a colour-absorption property. Similarly, apply Lemma 3.4.18(2) with $(R_1^{(4)}, R_2^{(4)})$ and $Y' := \mathcal{S}'_G$ with parameters (β_3, h_3) to obtain a set A_3 (disjoint with U, A_1, and A_2) with a colour-absorption property. For the last two applications, we use that $\mathcal{F}_G$ and $\mathcal{S}_G$ are near-dissociable, contain only path-candidates, and that $k-2 \geq z_{\mathcal{S}}, 4$ as $k \geq 10$. These properties come from Lemma 3.2.24. For all three applications, the necessary upper bound on h holds by definition of q_1, q_2 using that each random set has size close to its expectation. The lower bounds on h and that $0 < \beta \leq 1$ for the latter two applications follow from lower bounds on the sizes of $\mathcal{F}'_G$ and $\mathcal{S}'_G$ which can be derived from Lemma 3.4.6 (this is done implicitly in the rest of the argument).

We claim that $A_1 \cup A_2 \cup A_3 \cup R_2^{**} \cup (R_1^{(1)} \setminus U) \cup \bigcup \mathcal{F}'_G \cup \bigcup \mathcal{S}'_G$ has the desired absorption property. To see this, let V' and C' be given as in the lemma. By Lemma 3.4.1(2), there exists a matching M_1 in $\mathcal{H}_k$ saturating V' and using exactly

$(k-1)m$ vertices from $R_1^{(1,1)} \setminus U$ and km vertices from R_2^{**}. $C' \cup (R_2^{**} \setminus V(M_1)) := C''$ then has size $m+f-km = f-(k-1)m$ which is divisible by 4 by choice of the integer f. Furthermore, $\sum C'' = (|C''|/4) \cdot q_{G,k}$ by the sum property on the set R_2^{**}. Hence, C'' can be partitioned into 4-tuples with sum $q_{G,k}$ (recall this is not 0) which are path-candidates by Lemma 3.2.22(2) (as in the previous case, to check that C'' has small symmetric difference with R_2^{**}, recall that K is sufficiently large). This partition of C'' allows us to apply Lemma 3.4.6 (with $\ell = (f-(k-1)m)/4$) to deduce that there exists a matching M_2 saturating C'' using (exactly $k\ell = k(f-(k-1)m)/4$ many) vertices from $R_1^{(1,2)} \setminus U$ and colours from $R_2^{(1)} \setminus U$ which are closed under the families $\mathcal{F}_G$ and $\mathcal{S}_G$, and hence also closed under the families $\mathcal{F}'_G$ and $\mathcal{S}'_G$ (as the colours come from the set $R_2^{(1)}$). Lemma 3.4.6 also guarantees that M_2 uses $\ell(k-4-z_{\mathcal{S}})/4$ elements of $\mathcal{F}'_G$ and ℓ elements of $\mathcal{S}'_G$. Thus, there are exactly h_1 elements of $R_1^{(1)} \setminus U$, h_2 elements of $\mathcal{F}'_G$, and h_3 elements of $\mathcal{S}'_G$ that are unused by $M_1 \cup M_2$, so the leftovers of these sets combine with A_1, A_2 and A_3 (respectively) to produce perfect matchings, say M_3, M_4 and M_5. Then, $\bigcup_{i\in[5]} M_i$ is the desired matching. □

3.5 The high-girth case

In this section, we show how the high girth case of the FGT conjecture follows by results from Chapter 2.5.2, namely, Lemma 2.5.8.

Theorem 3.5.1. *Let G be an abelian group of order n, where n is sufficiently large. Suppose k is some integer such that $k \geq \log^9 n$, and k divides $n-1$. Suppose $\sum G = 0$. Then, $\mathcal{H}_k[G \setminus \{0\}; C \setminus \{0\}]$ has a perfect matching.*

Proof. If $k \leq n^{1/10^{10^{10}}}$, set $s = k$, otherwise set $s = \lceil \log^{10} n \rceil$.

Partition the group G into disjoint sets twice, independently, as $V_1, \dots, V_s$ and $C_0, \dots, C_{s-1}$ where each set is $(1/s)$-random, noting $1/s \geq n^{-1/10^{10^{10}}}$ in either case for n large. Set $t := \lceil \log^7 n \rceil$.

Lemma 2.5.8 holds with high probability with t, $V_{mid} = \bigcup_{1 \leq i \leq t-2} V_i$ and $C = \bigcup_{0 \leq i \leq t-2} V_i$. Lemma 2.5.15 holds with random sets (V_i, V_{i+1}, C_i) for each i (where indices are viewed in a cyclic order) and each integer value of $\ell = n/s \pm n^{1-1/10^{10}}$ (we achieve this via a union bound over many applications of Lemma 2.5.15). Also with high probability, all random sets are within $n^{0.6}$ elements of their expectations via Chernoff's bound. By the probabilistic method, fix the random sets so they have all the aforementioned properties.

Suppose first that $k \leq n^{1/10^{10^{10}}}$, so $s = k$. By the divisibility assumption and the property coming from Chernoff's bound, we can move $O(n^{0.7})$ elements between the sets V_i without relabelling so that each set V_i has size exactly $(n-1)/k$. Similarly, moving around at most $O(n^{1-10^6})$ elements, we can make sure each C_i where $i \geq t-1$ has $(n-1)/k + \lfloor n^{1-10^5} \rfloor$ elements. Now, apply Lemma 2.5.15 (with $\ell = (n-1)/k$) with the triples

$$(V_{t-1}, V_t, C_{t-1}), (V_t, V_{t+1}, C_t), \ldots, (V_{k-1}, V_k, C_{k-1})$$

to find rainbow matchings saturating the corresponding vertex sets (and missing $\lfloor n^{1-10^5} \rfloor$ colours from each C_i, $i \geq t-1$). Note that the union of the matchings found give a rainbow $\vec{P}_{k-t}$ factor where each path is directed from V_{t-1} to V_k. Now, we apply Lemma 2.5.8 with $V_{str} = V_k$ and $V_{end} = V_{t-1}$, and C' set to be the union of C and the $(k-t+1)\lfloor n^{1-10^5} \rfloor$ unused colours in each C_i, $i \geq t-1$. V_{mid} remains unchanged. All but the second hypothesis of Lemma 2.5.8 follow easily from our choice of sets. To see that $\sum V_k - \sum V_{t-1} = \sum C'$, first note that $\sum V_{t-1} - \sum V_k = \sum C''$ where C'' is all of the colours used via applications of Lemma 2.5.15 (this comes from the fact that we have a rainbow directed path factor in $\vec{K}_G$ where each path starts in V_{t-1} and ends in V_k). As $\sum G \setminus \{0\} = 0$ by assumption, and $C' = G \setminus \{0\} \setminus C''$, the desired equality follows. Thus we can indeed apply Lemma 2.5.8. In our application, we set f to be the bijection that maps the last endpoint of each $\vec{P}_{k-t}$ to the first endpoint of the same directed path. This allows us to complete each $\vec{P}_{k-t}$ into a cycle of length k, giving us a cycle-factor that corresponds to the desired matching in $\mathcal{H}_k$.

Suppose now that $k > n^{1/10^{10^{10}}}$, so $s = \lceil \log^{10} n \rceil$. If it was the case that s divides k, then we can proceed exactly like the previous case, with the only difference being in the choice of f in the previous paragraph (we would choose f so that when the connecting paths are found we end up with a C_k-factor as opposed to a C_s-factor). So suppose that r, the remainder when k is divided by s, is positive, noting that $r < s$. We start by finding $(n-1)/k$ vertex/colour disjoint rainbow $\vec{P}_r$ in $\vec{K}_G$, calling this collection of paths $\mathcal{P}$. Note this can be done greedily, and the resulting collection of paths occupies $2n^{1-1/10^{10^{10}}}$ vertices, due to our assumption on k and s. Let P_1 and P_2 denote collection of first endpoints of each of the paths in $\mathcal{P}$, respectively. We remove the vertices in $\mathcal{P} \setminus P_2$ from the graph, and proceed exactly as in the previous case to redistribute the sets so that they are of the right size, with the additional condition that $P_2 \subseteq V_{t-1}$. In the end, while applying Lemma 2.5.8, we set V_{end} to be $(V_{t-1} \setminus P_2) \cup P_1$. We can then select an appropriate bijection f so that after an application of Lemma 2.5.8, the resulting structure is a $\vec{C}_k$-factor. □

3.6 Concluding remarks

In this section we outline some directions for further research.

3.6.1 Further applications in graph labelling

As alluded to previously, our methods are quite flexible and can potentially be used to make advances on other embedding problems with an algebraic flavour. The language of *graph labellings* makes this connection explicit. Suppose we have a graph T with m edges, and a set L of $\geq m$ labels with some algebraic structure. For example, when L is the cyclic group on m elements, a labelling of the vertices of T with L is called a **harmonious labelling** if for each edge $e = \{x, y\}$ of T, we have that $x + y \pmod{m}$ is distinct. Similarly, if L is the first m positive integers, a vertex-labelling where for each edge we have that $|x - y|$ is distinct is called a

graceful labelling. The famous harmonious tree and the graceful tree conjectures respectively assert that all trees have harmonious and graceful labellings [33, 28]. Harmonious and graceful labellings are also heavily investigated for other classes of graphs due to intimate connections with the theory of error correcting codes (here we refer the reader to [28]).

Section 3.1.1 establishes that the FGT conjecture can be framed as a problem about labelling the vertices of a collection of vertex-disjoint directed cycles, where the edge labelling rule is $a - b$ (if ab is an edge of a directed cycle) within an abelian group G. We develop general methods for labellings of short paths and cycles in the current paper, and it is likely that our methods could give well-behaved labellings of other sparse structures build up of short paths and cycles.

For example, we would not expect that working with the labelling rule $a + b$ (as opposed to $a - b$) creates dramatic complications (see Remark 1), therefore our methods seem applicable to the study of harmonious labellings. Harmonious labellings of various sparse graph classes (such as powers of paths, which in particular contain collections of vertex-disjoint cycles) have been separately investigated, using ad-hoc methods, and “few general results are known” as Gallian notes in his survey [28]. We believe the methods we develop here could be used as a unifying method to produce classes of harmonious graphs (those graphs which admit harmonious labellings).

The connection with the study of graceful labellings is less direct, as the function $(x, y) \to |x - y|$ over the integers behaves rather differently compared to the function $x + y$ over the cyclic group. That said, the function $(x, y) \to |x - y|$ still exhibits a lot of algebraic symmetry, hence there is potential for our methods to be applicable.

3.6.2 Non-abelian groups, Latin squares, and Ryser's conjecture

Now that the Friedlander-Gordon-Tannenbaum conjecture is verified, at least for sufficiently large groups, we propose the following extension for general groups.

Conjecture 3.6.1. *Let G be a sufficiently large group satisfying the Hall-Paige condition, and suppose $k \geq 3$ and k divides $n-1$. Then, there exists an orthomorphism of G that fixes the identity element, and permutes the remaining elements of disjoint cycles of length k.*

It is also sensible to replace orthomorphisms with complete mappings in the above conjecture, due to the assumption that $k \geq 3$ (recall Remark 1 from the Introduction). One way of attacking the above conjecture would be to try to combine the methods from this paper with the methods developed for non-abelian groups in [49]. Also, we remark that in Section 3.5, we did not actually use that the group G is abelian. Therefore, the above conjecture is true in the high-girth case.

More generally, we can turn our attention to Latin squares, which are also known as quasi-groups. These objects can be described as n by n arrays filled with n symbols so that no symbol repeats in a row or a column. For us, it will be more natural to view Latin squares in the following way (see the survey by Pokrovskiy from [50] for a more detailed discussion). We first take a complete directed graph $\vec{K}_n$ with edges in both directions between all vertices and a loop at every vertex. We then equip this graph with a proper edge-colouring using n colours. The most famous conjecture in the area is the following.

Conjecture 3.6.2 (Ryser's conjecture)**.** *Suppose n is odd. Then, $\vec{K}_n$ contains a rainbow spanning subgraph where every vertex has in-degree and out-degree equal to one. Equivalently, $\vec{K}_n$ can be packed with directed cycles in a rainbow fashion.*

In analogy with the Friedlander-Gordon-Tannenbaum conjecture, it makes sense to strengthen Ryser's conjecture to ask for cycles of specific lengths. There are numerous conjectures in this area which focus on finding a single cycle which covers the entirety of the vertex set, which is analogous to the $k = n - 1$ case of the Friedlander-Gordon-Tannenbaum conjecture. For more information about these conjectures, we refer the reader to Pokrovskiy's survey about rainbow subgraphs in [50] and a recent paper by Gould and Kelly which includes a nice unifying conjecture [30]. We pose a conjecture in the other extreme, where the cycle lengths are as small as possible. This is analogous to the $k = 3$ case of the Friedlander-Gordon-Tannenbaum conjecture.

Conjecture 3.6.3. *Let K_n be a complete graph properly coloured with n colours. Then, K_n contains a rainbow subgraph which is a disjoint union of triangles covering all but at most C vertices, for some absolute constant C.*

We are not aware of any examples that would rule out the possibility that one can take $C = 2$ above. In the other direction, one can prove a relaxed version of the above conjecture with C replaced with $n^{1-\varepsilon}$ for some $\varepsilon > 0$ by using the Rödl nibble (see for example Corollary 3.2.2(1)). Improving this bound, for example by replacing C with a polylogarithmic term, could be an interesting challenge, see [41] for an analogous result in the setting of Ryser's conjecture.

3.6.3 Other cycle types

To prove the FGT conjecture, we only used the $p = 1$ case of Theorem 3.1.3. Applying Theorem 3.1.3 with different values of p, we can derive that many other cycle types for orthomorphisms are possible. Suppose for example that G is an abelian group of order n with the Hall-Paige property, $n-1 = 3k+4\ell$, and we want to find an orthomorphism fixing the identity and permuting the remaining elements as k many disjoint 3-cycles and ℓ many 4-cycles. Let's also suppose for simplicity that $k, \ell = \Omega(n)$. Then, we can partition the vertices of $\vec{K}_G$ into a $3k/n$-random set V_1 and $4\ell/n$-random set V_2, and we can partition the colours of $\vec{K}_G$ into $3k/n$-random set C_1 and $4\ell/n$-random set C_2. With positive probability, Theorem 3.1.3 holds with (V_1, C_1), $k = 3$ and with (V_2, C_2), $k = 4$. We can then do a few exchanges between the sets of vertices and colours so that they satisfy the divisibility condition as well as the sum condition $\sum C_1 = \sum C_2 = 0$. Then, by Theorem 3.1.3 we obtain the desired cycle partition.

We can go further and ask the following question. Suppose that $s_1, s_2, \dots, s_j$ is a sequence of integers where $s_i \geq 2$ and $\sum s_i = n - 1$, and suppose that G is an abelian group with the Hall-Paige property. When is it true that G has an orthomorphism fixing the identity and permuting the remaining elements as cycles of lengths $s_1, s_2, \dots, s_j$? Note that a necessary condition for the existence of such an orthomorphism is a partition of $G \setminus \{0\}$ into zero-sum sets of size $s_1, s_2, \dots, s_j$ (recall Observation 3.1.1). Characterising pairs of sequences $s_1, s_2, \dots, s_j$ and abelian groups that admit such a partition is known as Tannenbaum's problem. This problem was solved for large groups in [49]. Perhaps the methods from the current paper could be sufficient to solve the more general problem of characterising which cycle types are feasible for orthomorphisms.

3.6.4 Other equations

As discussed in Section 3.1, there is a connection between the Hall-Paige conjecture, the FGT conjecture, and toroidal version of the n-queens problem [12]. We can make this connection more formal as follows. Suppose A is a $\ell \times m$ matrix with integer entries, and G is an abelian group of order n. Can we find a collection of n-many vectors $\vec{v}$ in G^m with $A \cdot \vec{v} = \vec{0}_\ell$ (meaning the ℓ-dimensional 0-vector) such that for each $i \in \{1, 2, \dots, m\}$, the collection of ith coordinates of the vectors $\vec{v}$ is equal to G (i.e. contains no repetitions). If this is possible, let us call the pair (A, G) **matchable**. This term is motivated by the fact that we can equivalently phrase this as a hypergraph matching problem in m-partite m-uniform hypergraphs where the edge set is governed by a collection of ℓ linear equations given by the matrix A.

For example, in the Hall-Paige conjecture, the corresponding matrix A is $[[1, -1, -1]]$, in the $k = 3$ case of the FGT conjecture, the matrix is $[[1, -1, 0, -1, 0, 0], [0, 1, -1, 0, -1, 0], [-1, 0, 1, 0, 0, -1]]$, and in the n-queens problem, the matrix is $[[1,1,-1,0],[1,-1,0,-1]]$. Characterising integer matrices A and abelian groups G such that (A,G) is matchable is a natural unifying problem. This would be interesting already when A consists only of $\{-1,0,1\}$-entries.

3.6.5 Controlling the cycle type of both bijections

It is also natural to investigate the existence of orthomorphisms/complete mappings ϕ where one makes a restriction on the cycle type of ϕ as well as the cycle type of the permutation $g \to g^{-1}\phi(g)$. Several partial results as well as open problems in this direction are given in [10, 11] by Bors and Wang. It would be interesting to see if our methods can be adapted to address this more restrictive variant of the problem.

Bibliography

[1] M. Ajtai, J. Komlós, and E. Szemerédi. Sorting in $c \log n$ parallel steps. *Combinatorica*, 3(1):1–19, Jan. 1983.

[2] Saieed Akbari and Alireza Alipour. Transversals and multicolored matchings. *Journal of Combinatorial Designs*, 12(5):325–332, 2004.

[3] Noga Alon. Combinatorial nullstellensatz. *Combinatorics, Probability and Computing*, 8(1-2):7–29, 1999.

[4] Noga Alon. Additive Latin transversals. *Israel Journal of Mathematics*, 117(1):125–130, 2000.

[5] Noga Alon and Joel Spencer. *The Probabilistic Method.* John Wiley & Sons, 2004.

[6] Brian Alspach, Donald L Kreher, and Adrián Pastine. The Friedlander-Gordon-Miller conjecture is true. *Australas. J Comb.*, 67:11–24, 2017.

[7] Brian Alspach and Georgina Liversidge. On strongly sequenceable abelian groups. *Art Discrete Appl. Math.*, 2020.

[8] Bodan Arsovski. A proof of Snevily's conjecture. *Israel Journal of Mathematics*, 182(1):505–508, 2011.

[9] K. E. Batcher. Sorting networks and their applications. In *Proceedings of the April 30–May 2, 1968, Spring Joint Computer Conference*, AFIPS '68

(Spring), page 307–314, New York, NY, USA, 1968. Association for Computing Machinery.

[10] Alexander Bors and Qiang Wang. Coset-wise affine functions and cycle types of complete mappings. *Finite Fields and Their Applications*, 83:102088, 2022.

[11] Alexander Bors and Qiang Wang. Cycle types of complete mappings of finite fields. *Journal of Algebra*, 591:577–610, 2022.

[12] Candida Bowtell and Peter Keevash. The n-queens problem. *arXiv preprint arXiv:2109.08083*, 2021.

[13] John N. Bray, Qi Cai, Peter J. Cameron, Pablo Spiga, and Hua Zhang. The Hall–Paige conjecture, and synchronization for affine and diagonal groups. *J. Algebra*, 545:27–42, 2020.

[14] Thomas H. Cormen. *Introduction to Algorithms.* The MIT Press, 2009.

[15] Simone Costa, Stefano Della Fiore, and MA Ollis. Sequencings in semidirect products via the polynomial method. *arXiv preprint arXiv:2301.09367*, 2023.

[16] Simone Costa, Stefano Della Fiore, MA Ollis, and Sarah Z Rovner-Frydman. On sequences in cyclic groups with distinct partial sums. *arXiv preprint arXiv:2203.16658*, 2022.

[17] Simone Costa and Marco Antonio Pellegrini. Some new results about a conjecture by Brian Alspach. *Archiv der Mathematik*, 115(5):479–488, 2020.

[18] Samit Dasgupta, Gyula Károlyi, Oriol Serra, and Balázs Szegedy. Transversals of additive Latin squares. *Israel Journal of Mathematics*, 126(1):17–28, 2001.

[19] Sean Eberhard, Freddie Manners, and Rudi Mrazović. An asymptotic for the Hall–Paige conjecture. *Advances in Mathematics*, 404:108423, 2022.

[20] Stefan Ehard, Stefan Glock, and Felix Joos. Pseudorandom hypergraph matchings. *Combinatorics, Probability and Computing*, 29(6):868–885, 2020.

[21] Paul Erdős, András Gyárfás, and László Pyber. Vertex coverings by monochromatic cycles and trees. *Journal of Combinatorial Theory, Series B*, 51(1):90–95, 1991.

[22] Anthony Evans. The admissibility of sporadic simple groups. *J. Algebra*, 321(1):105–116, 2009.

[23] Anthony B Evans. *Orthogonal Latin squares based on groups*, volume 57. Springer, 2018.

[24] Péter Frankl and Vojtech Rödl. Near perfect coverings in graphs and hypergraphs. *European Journal of Combinatorics*, 6(4):317–326, 1985.

[25] Richard Friedlander, Basil Gordon, and Peter Tannenbaum. Partitions of groups and complete mappings. *Pacific Journal of Mathematics*, 92(2):283–293, 1981.

[26] Richard J Friedlander, Basil Gordon, and Michael D Miller. On a group sequencing problem of Ringel. *Congr. Numer*, 21:307–321, 1978.

[27] PX Gallagher. Group characters and commutators. *Mathematische Zeitschrift*, 79(1):122–126, 1962.

[28] Joseph Gallian. Graph labelling. *The Electronic Journal of Combinatorics*, 2021.

[29] Luis Goddyn and Kevin Halasz. All group-based latin squares possess near transversals. *Journal of Combinatorial Designs*, 28(5):358–365, 2020.

[30] Stephen Gould and Tom Kelly. Hamilton transversals in random Latin squares. *arXiv preprint arXiv:2104.12718*, 2021.

[31] W Gowers. Probabilistic combinatorics and the recent work of peter keevash. *Bulletin of the American Mathematical Society*, 54(1):107–116, 2017.

[32] RL Graham. On sums of integers taken from a fixed sequence. In *Proceedings, Washington State University Conference on Number Theory*, pages 22–40, 1971.

[33] Ronald L Graham and Neil James Alexander Sloane. On additive bases and harmonious graphs. *SIAM Journal on Algebraic Discrete Methods*, 1(4):382–404, 1980.

[34] Ben Green and A Wigderson. *Lecture notes for the 22nd McGill invitational workshop on computational complexity*. 2010.

[35] M. Hall and L.J. Paige. Complete mappings of finite groups. *Pacific J. of Math.*, 5:541–549, 1955.

[36] Julie Haviland and Andrew Thomason. On testing the 'pseudo-randomness' of a hypergraph. *Discrete mathematics*, 103(3):321–327, 1992.

[37] Jacob Hicks, MA Ollis, and John R Schmitt. Distinct partial sums in cyclic groups: polynomial method and constructive approaches. *Journal of Combinatorial Designs*, 27(6):369–385, 2019.

[38] Eugene C Johnsen and Thomas Storer. Combinatorial structures in loops I. elements of the decomposition theory. *Journal of Combinatorial Theory, Series A*, 14(2):149–166, 1973.

[39] A Donald Keedwell and József Dénes. *Latin Squares and Their Applications: Latin Squares and Their Applications*. Elsevier, 2015.

[40] Peter Keevash. The existence of designs. *arXiv preprint arXiv:1401.3665*, 2014.

[41] Peter Keevash, Alexey Pokrovskiy, Benny Sudakov, and Liana Yepremyan. New bounds for Ryser's conjecture and related problems. *Transactions of the American Mathematical Society, Series B*, 9(08):288–321, 2022.

[42] Daniela Kühn, John Lapinskas, Deryk Osthus, and Viresh Patel. Proof of a conjecture of Thomassen on Hamilton cycles in highly connected tournaments. *Proceedings of the London Mathematical Society*, 109(3), 2014.

[43] Matthew Kwan, Ashwin Sah, Mehtaab Sawhney, and Michael Simkin. High-girth steiner triple systems. *arXiv preprint arXiv:2201.04554*, 2022.

[44] Michael Molloy and Bruce Reed. Near-optimal list colorings. *Random Structures & Algorithms*, 17(3-4):376–402, 2000.

[45] Richard Montgomery. Spanning trees in random graphs. *Advances in Mathematics*, 356, 2019.

[46] Richard Montgomery. A proof of the ryser-brualdi-stein conjecture for large even n. *arXiv preprint arXiv:2310.19779*, 2023.

[47] Richard Montgomery, Alexey Pokrovskiy, and Benjamin Sudakov. Decompositions into spanning rainbow structures. *Proceedings of the London Mathematical Society*, 119, 04 2019.

[48] Richard Montgomery, Alexey Pokrovskiy, and Benny Sudakov. A proof of Ringel's conjecture. *Geometric and Functional Analysis*, 31, 2021.

[49] Alp Müyesser and Alexey Pokrovskiy. A random Hall-Paige conjecture. *arXiv preprint arXiv:2204.09666*, 2022.

[50] Anthony Nixon and Sean Prendiville. *Surveys in Combinatorics 2022*, volume 481. Cambridge University Press, 2022.

[51] MA Ollis. Sequenceable groups and related topics. *The Electronic Journal of Combinatorics*, 1000:DS10–Aug, 2002.

[52] Gerhard Ringel. Cyclic arrangements of the elements of a group. *Notices of the American Mathematical Society*, 21(1):A–95, 1974.

[53] Gerhard Ringel. *Map color theorem*, volume 209. Springer Science & Business Media, 2012.

[54] Vojtěch Rödl, Endre Szemerédi, and Andrzej Ruciński. An approximate Dirac-type theorem for k-uniform hypergraphs. *Combinatorica*, 28(2):229–260, 2008.

[55] Hunter S. Snevily. The Cayley addition table of $\mathbb{Z}_n$. *The American Mathematical Monthly*, 106(6):584–585, 1999.

[56] GH from MO (https://mathoverflow.net/users/11919/gh-from mo). How many square roots can a non-identity element in a group have? MathOverflow. URL:https://mathoverflow.net/q/410789 (version: 2021-12-15).

[57] Andrew Thomason. Dense expanders and pseudo-random bipartite graphs. In *Annals of Discrete Mathematics*, volume 43, pages 381–386. Elsevier, 1989.

[58] CD Wang. On harmoniousness and complete mappings decomposable into disjoint cycles of the same length. *Combinatorics, Graph Theory, Algorithms and Applications (Beijing, 1993)*, pages 347–353, 1994.

[59] Ian Wanless. *Transversals in Latin squares: a survey*, page 403–437. London Mathematical Society Lecture Note Series. Cambridge University Press, 2011.

[60] S. Wilcox. Reduction of the Hall-Paige conjecture to sporadic simple groups. *J. Algebra*, 321(5):1407–1428, 2009.