

# PARTIAL ASSOCIATIVITY AND ROUGH APPROXIMATE GROUPS

W. T. GOWERS AND J. LONG



**Abstract.** Suppose that a binary operation  $\circ$  on a finite set  $X$  is injective in each variable separately and also associative. It is easy to prove that  $(X, \circ)$  must be a group. In this paper we examine what happens if one knows only that a positive proportion of the triples  $(x, y, z) \in X^3$  satisfy the equation  $x \circ (y \circ z) = (x \circ y) \circ z$ . Other results in additive combinatorics would lead one to expect that there must be an underlying ‘group-like’ structure that is responsible for the large number of associative triples. We prove that this is indeed the case: there must be a proportional-sized subset of the multiplication table that approximately agrees with part of the multiplication table of a metric group. A recent result of Green shows that this metric approximation is necessary: it is not always possible to obtain a proportional-sized subset that agrees with part of the multiplication table of a group.

## 1 Introduction

The following statement is a known result in additive combinatorics. Let  $n$  be a prime, let  $A \subset \mathbb{Z}/n\mathbb{Z}$  and let  $\phi : A \rightarrow \mathbb{Z}/n\mathbb{Z}$  be a map such that the number of quadruples  $(a, b, c, d) \in A^4$  with  $a + b = c + d$  and  $\phi(a) + \phi(b) = \phi(c) + \phi(d)$  is at least  $\alpha n^3$ . Then there is a subset  $A' \subset A$  of size at least  $\beta n$ , where  $\beta$  depends on  $\alpha$  only, such that  $\phi(a) + \phi(b) = \phi(c) + \phi(d)$  whenever  $a, b, c, d \in A'$  and  $a + b = c + d$ . A map with this last property is called a *Freiman homomorphism*, so this result is saying that a map that obeys the condition for a Freiman homomorphism a constant fraction of the time can be restricted to a dense set that obeys the condition all the time. One can then go further and show that  $\phi$  agrees on a further dense subset with the restriction of a ‘linear-like’ function, which gives a global structural characterization of functions that satisfy the initial local conditions. (For more details, see for example [Gow01], in particular Corollary 7.6).

There are several known statements of this general flavour, and the purpose of this paper is to prove another one. Here our starting point is a binary operation  $\circ$  defined on a finite set  $X$ . We assume that it is an injection (and therefore a bijection) in each variable separately, and that there exists a constant  $c > 0$ , independent of the size of  $X$ , such that the number of triples  $(x, y, z) \in X^3$  with  $x \circ (y \circ z) = (x \circ y) \circ z$  is at least  $c|X|^3$ . It is easy to see that if  $c = 1$  then these conditions are equivalent to the group axioms, so it is natural to ask whether if a binary operation has this

property for some smaller  $c$ , then there must be some underlying group structure that ‘explains’ the prevalence of associative triples. This question appears to have been known to various people—we heard about it from Emmanuel Breuillard, who attributed it to Ehud Hrushovski, and an essentially equivalent question arose out of work we ourselves had been doing—but it does not seem to have appeared in print.

The ‘99% case’ was dealt with by Elad Levi [Lev], who proved that if  $c$  is close to 1, then there must be a group  $G$  of size approximately equal to  $|X|$  and an injection  $\phi : X \rightarrow G$  such that  $\phi(x \circ y) = \phi(x)\phi(y)$  for almost all pairs  $x, y \in X^2$ . In other words, the multiplication table of  $\circ$  agrees almost everywhere with a group operation. In this paper we look at the ‘1% case’—that is, the case where  $c$  is a small fixed constant. We also weaken the hypothesis in a small way by considering binary operations that are only partially defined: this has no significant effect on our arguments, but it is convenient when discussing examples not to have to worry about whether they are defined everywhere. In the discussion that follows, we shall often use the word ‘operation’ to mean ‘partial binary operation’.

An easy way to create an operation with many associative triples is to take the operation  $\circ$  on a group  $G$  and turn it into a partial binary operation by restricting it to a dense subset  $X \subset G^2$ . This is not guaranteed to work, as there are not necessarily  $c|G|^3$  triples  $(x, y, z) \in G^3$  such that all of  $(x, y)$ ,  $(y, z)$ ,  $(x, y \circ z)$  and  $(x \circ y, z)$  belong to  $X$ . However, in many cases, such as when  $X$  is a random subset, it does. More generally, given any operation with many associative triples, one can find restrictions that still have many associative triples.

Another method is to take a subset  $A$  of a group  $G$  and restrict the group operation  $\circ$  to all pairs  $(a, b) \in A^2$  such that  $a \circ b \in A$ . Again, this is not guaranteed to work, but if  $A$  is an *approximate subgroup*, which roughly speaking means that it is closed ‘1% of the time’ (we shall discuss this condition in more detail in a moment), then this gives another source of examples.

A third method is based on structures that are approximately groups in a metric sense. For concreteness, we discuss a specific example.

**EXAMPLE 1.1.** Let  $\delta > 0$  and let  $X$  be a maximal  $\delta$ -separated subset of  $\mathrm{SO}(3)$  equipped with a suitable translation invariant metric. Now define a partial binary operation as follows. Let  $\theta > 0$  be a suitable absolute constant (as opposed to  $\delta$ , which is comparable to  $|X|^{-1/3}$ ) and then for  $x, y, z \in X$  let  $x \circ y = z$  if and only if  $d(xy, z) \leq \theta\delta$ . If  $d(xy, X) > \theta\delta$  then the product  $x \circ y$  is undefined. Importantly, translation invariance means that (assuming that  $\theta < 1/2$ ) the resulting operation is injective in each variable separately, in the sense that e.g.  $x \circ y = x' \circ y$  with both sides defined implies that  $x = x'$ .

We show in an appendix that no matter how the set  $X$  in Example 1.1 is chosen the resulting binary operation will be defined a positive proportion of the time, and there will necessarily be many associative triples. In order to prove this, we first prove a Bogolyubov-type lemma for  $\mathrm{SO}(3)$  (Lemma B.1 in the second appendix) that may be of independent interest. However, there is no obvious way of passing

to a subset of  $X^2$  where the operation is isomorphic to a restriction of a group operation. Indeed, in an earlier version of this paper, we conjectured that there was no such subset, and that conjecture has been proved by Ben Green [Gre20].

This example shows that a natural conjecture—that a partially associative binary operation agrees on a substantial set of pairs with a group operation—is false. However, the example has a suggestive structure that suggests an appropriate weakening of the conjecture.

By a metric group, we mean a group  $G$  equipped with a bi-invariant metric  $d$ , which we allow to take the value infinity. Our main result will be that if an operation has many associative triples (and is injective in each variable separately), then it agrees on a large set of pairs with a restriction of a small perturbation of the binary operation on a metric group.

The next theorem is not in fact our main theorem, but a consequence of it. However, to state the main theorem requires one more definition, so we shall state this result first. Loosely speaking, it says that the multiplication table of a partial binary operation with many associative triples must have a large subset that is approximately isomorphic to part of the multiplication table of a metric group  $G$ . The precise statement is as follows.

**Theorem 1.2.** *Let  $c > 0$ , let  $X$  be a finite set and let  $\circ$  be a partially defined binary operation on  $X$  that is injective in each variable separately. Let  $E \subset X^2$  be the set of pairs on which  $\circ$  is defined. Suppose that there are at least  $\epsilon|X|^3$  triples  $(x, y, z) \in X^3$  such that  $x \circ (y \circ z) = (x \circ y) \circ z$  (where this means in particular that all expressions and subexpressions are defined). Then for every positive integer  $b$  there exist a subset  $A \subset E$  of density at least  $\epsilon^{b^{26b}}$ , a metric group  $G$ , and maps  $\phi, \psi$  and  $\omega$  from  $X$  to  $G$ , such that the images  $\phi(X), \psi(X)$  and  $\omega(X)$  are 1-separated, and  $d(\phi(x)\psi(y), \omega(z)) \leq b^{-1}$  for every  $(x, y, z) \in X^3$  such that  $(x, y) \in A$  and  $x \circ y = z$ .*

**1.1 Quasigroups, the quadrangle condition, torsors, and our main theorem.** Our main result will have the same conclusion as that of Theorem 1.2 but a hypothesis that is both weaker and in some ways more natural. It arises out of the following simple question: suppose that an  $n \times n$  grid is filled with labels. Under what conditions is this labelled grid the multiplication table of some group?

We can ask the question more formally as follows. Suppose we are given three sets  $X, Y$  and  $Z$  with  $|X| = |Y| = n$ , and a function  $f : X \times Y \rightarrow Z$ . Under what conditions does there exist a group  $G$  of order  $n$  and bijections  $\phi : X \rightarrow G, \psi : Y \rightarrow G$  and  $\omega : Z \rightarrow G$  such that for every  $(x, y) \in X \times Y$  we have  $\phi(x)\psi(y) = \omega(f(x, y))$ ?

To discuss this, we use the following vocabulary. We call the elements of  $Z$  *labels*, sets of the form  $\{x\} \times Y$  *columns* and sets of the form  $X \times \{y\}$  *rows*. If  $f(x, y) = z$ , we say that  $z$  is the label in position  $(x, y)$ . A very obvious necessary condition is that  $Z$  should also have cardinality  $n$ . Another is that each label occurs exactly once in each row and each column.

A labelling of an  $n \times n$  grid that satisfies these two conditions is known as the *Latin square*  $(X, Y, Z, f)$ . If we think of the labelled grid as the multiplication table

of the binary operation  $f$ , then it has the property that for each  $x \in X$  the function  $y \mapsto f(x, y)$  is a bijection from  $Y$  to  $Z$ , and for each  $y \in Y$  the function  $x \mapsto f(x, y)$  is a bijection from  $X$  to  $Z$ . If we identify the sets  $X, Y$  and  $Z$  (using arbitrary bijections) and write  $x \circ y$  instead of  $f(x, y)$ , then we have a set  $X$  with a binary operation  $\circ$  with the property that for every  $a, b \in X$  the equations  $a \circ x = b$  and  $x \circ a = b$  have unique solutions. Such an algebraic structure is called a *quasigroup*. (Thus, quasigroups and Latin squares are essentially the same).

The question now becomes the following: when is a quasigroup a group? Equivalently, when is a Latin square the multiplication table of a group?<sup>1</sup>

The following definition allows us to give a satisfactory answer.

**DEFINITION 1.3.** *Let  $A = (X, Y, Z, f)$  be a Latin square. We say that  $A$  satisfies the quadrangle condition if for every configuration of the following form in  $A$ ,*

$$\begin{array}{cc} c & d \\ a & b \\ & c & d' \\ & a & b \end{array}$$

*we have  $d = d'$ . To put it a different way, we can define a ternary rectangle completion operation on the set  $Z$  of labels by mapping  $(a, b, c)$  to  $d$  whenever there exists a rectangle with labels  $a, b, c, d$  such that  $a$  is in the same row as  $b$  and the same column as  $c$ .*

Observe that if  $x_1, x_2, y_1, y_2$  are elements of a group  $G$ , and  $x_1 y_1 = a, x_2 y_1 = b$  and  $x_1 y_2 = c$ , then  $x_2 y_2 = b a^{-1} c$ . This simple observation shows that if a Latin square is a group multiplication table, then the rectangle completion operation is well-defined, and therefore group multiplication tables satisfy the quadrangle condition.

It turns out that the converse is true as well: a Latin square that satisfies the quadrangle condition is the multiplication table of a group. This is a well-known observation of Brandt [Bra27]. Since the proof is short, we give it here.

**PROPOSITION 1.4.** *Every Latin square that satisfies the quadrangle condition is the multiplication table of a group.*

*Proof.* Choose an arbitrary row  $R$  and column  $C$  and define a binary operation  $\circ$  on the set of labels as follows. Given labels  $a$  and  $b$ , find where  $a$  appears in row  $R$  and where  $b$  appears in column  $C$ , and then let  $a \circ b = c$ , where  $c$  is the label of the point in the same column as  $a$  and the same row as  $b$ . The label of the point where  $R$  and  $C$  intersect is then an identity for  $\circ$ , and the Latin square condition implies that every element has both a left and a right inverse. It remains to check

---

<sup>1</sup> It is important to clarify exactly what this question is asking. When we are presented with the Latin square, we are *not* given any correspondences between rows, columns and labels. Rather, we are given an arrangement of labels and asked to *find* correspondences in such a way that the resulting binary operation is a group operation.

associativity. To do this, consider the following picture, which is of a portion of the Latin square, chosen to demonstrate that  $(a \circ b) \circ c = a \circ (b \circ c)$ . We write  $d$  for  $a \circ b$ ,  $f$  for  $d \circ c$ ,  $g$  for  $b \circ c$ ,  $h$  for  $a \circ g$ , and  $e$  for the identity.

$$\begin{array}{cccc} g & h & & \\ c & g & f & \\ b & d & & \\ e & a & b & d \end{array}$$

For associativity we need  $f$  to equal  $h$ . But this follows from the quadrangle condition, since included in the above diagram are the points

$$\begin{array}{ccc} g & h & \\ & g & f \\ b & d & \\ & b & d \end{array}$$

Thus, the set of labels has an associative binary operation with an identity such that every element has a left and a right inverse, and we are done.  $\square$

A notable feature of the above argument is the arbitrary choice of the row  $R$  and the column  $C$ , and hence the arbitrary choice of which label would serve as the identity element. It shows that if we are presented just with the labelled grid and not with any correspondences between rows, columns and labels, then there is no way of telling which label corresponds to the identity. Another way of expressing this observation is to say that if  $G$  is a group and  $x$  is any element of  $G$ , then we can form a group  $G_x$  with identity element  $x$  by taking the binary operation  $a \circ b = ax^{-1}b$ , which is derived from the rectangle-completion operation discussed above.

If one wishes to avoid the artificiality of choosing an arbitrary element to be the identity, one can do so by working with an algebraic structure known as a *torsor*, which can be thought of as a group ‘except that we do not know which element is the identity’. The formal definition of a torsor is that it is a set  $X$  with a ternary operation  $\tau$  which has the following two properties.

- $\tau(x, x, y) = \tau(y, x, x) = y$  for every  $x, y \in X$ ;
- $\tau(x, y, \tau(z, u, v)) = \tau(\tau(x, y, z), u, v)$  for every  $x, y, z, u, v \in X$ .

The quantity  $\tau(x, y, z)$  should be thought of as  $xy^{-1}z$ . Indeed, the relationship between groups and torsors is closely analogous to the relationship between vector spaces and affine spaces, and the ternary map is also closely analogous to the (partially defined) map  $(a, b, c) \mapsto a - b + c$  that often appears in additive combinatorics when one has a set  $A$  with additive structure that is not ‘centred on zero’.

In order to draw out the relationship between the quadrangle condition and our problem, in which we are given a partially defined binary operation, we must turn our attention to *partial* Latin squares—that is, to grids that are partially labelled in such a way that no label occurs more than once in any row or column. The formal definition is given below.

DEFINITION 1.5. A partial Latin square is a quintuple  $(X, Y, Z, A, \phi)$ , where  $X, Y, Z$  are finite sets,  $A \subset X \times Y$ , and  $\phi : A \rightarrow Z$  is a function such that if  $\phi(a, b_1) = \phi(a, b_2)$  then  $b_1 = b_2$ , and if  $\phi(a_1, b) = \phi(a_2, b)$ , then  $a_1 = a_2$ .

We shall refer to the elements of  $X, Y$  and  $Z$  as *columns*, *rows* and *labels*, respectively, since we imagine drawing the object  $(X, Y, Z, A, \phi)$  as a labelled grid with columns indexed by  $X$  and rows indexed by  $Y$ , labelling the point  $(x, y) \in A$  with  $\phi(x, y)$ .

Given a partial Latin square  $(X, Y, Z, A, \phi)$  with  $|X| = |Y| = |Z| = n$ , we shall sometimes abuse notation and say that  $A$  is an  $n \times n$  partial Latin square (or simply that  $A$  is a partial Latin square). If  $(X, Y, Z, A, \phi)$  is a partial Latin square and  $B \subset A$ , we may also refer to the partial Latin square  $(X, Y, Z, B, \phi|_B)$  as  $B$ , calling it simply a subset of  $A$  (if it is clear from context that both objects are partial Latin squares).

The above definition does not treat the sets  $X, Y$  and  $Z$  in a symmetric way. For some of our arguments, which themselves give a special role to the label set, that is appropriate. However, later we shall need to make statements that are symmetric in the three sets, so we note here that a partial Latin square is nothing other than a linear tripartite 3-uniform hypergraph. (Recall that a 3-uniform hypergraph is a set of triples, and it is said to be linear if no pair is contained in more than one triple). We briefly prove that now.

LEMMA 1.6. *There is a one-to-one correspondence between partial Latin squares  $(X, Y, Z, A, \phi)$  and linear tripartite 3-uniform hypergraphs with vertex sets  $X, Y$  and  $Z$ .*

*Proof.* Given a partial Latin square  $P = (X, Y, Z, A, \phi)$ , let  $H$  consist of all triples  $(x, y, z) \in X \times Y \times Z$  such that  $(x, y) \in A$  and  $\phi(x, y) = z$ . Since  $\phi$  is a function, for any  $(x, y)$  there is at most one  $z$  such that  $(x, y, z) \in H$ , and since  $\phi$  is injective in each variable separately, for each  $(x, z)$  there is at most one  $y$  such that  $(x, y, z) \in H$  and for each  $(y, z)$  there is at most one  $x$  such that  $(x, y, z) \in H$ . Therefore,  $H$  is linear.

Conversely, given a linear tripartite 3-uniform hypergraph  $H$  with vertex sets  $X, Y, Z$ , let  $A$  be the set of all  $(x, y)$  such that there exists  $z$  with  $(x, y, z) \in H$ . Since  $H$  is linear, such a  $z$  is unique if it exists, so we can define a function  $\phi : A \rightarrow Z$  by setting  $\phi(x, y)$  to be the unique  $z$  such that  $(x, y, z) \in H$ . If  $\phi(x, y_1) = \phi(x, y_2)$ , then  $(x, y_1, z)$  and  $(x, y_2, z)$  both belong to  $H$ , so by the linearity property  $y_1 = y_2$ . Therefore,  $\phi$  is injective in the second variable. Similarly, it is injective in the first variable.  $\square$

With the above observations and definitions in mind, it is natural to formulate a torsor version of the question about binary operations with many associative triples. For reasons that we shall explain in the next subsection, we call a pair of identically labelled rectangles in a partial Latin square an octahedron. The following is a precise definition.

DEFINITION 1.7. *Given a partial Latin square  $(X, Y, Z, A, \phi)$ , an octahedron in  $A$  consists of a pair of rectangles*

$$R_1 = ((x_1, y_1), (x_2, y_1), (x_1, y_2), (x_2, y_2)) \in A^4$$

and

$$R_2 = ((x_3, y_3), (x_4, y_3), (x_3, y_4), (x_4, y_4)) \in A^4$$

such that we have the four identities

$$\begin{aligned} \phi(x_1, y_1) &= \phi(x_3, y_3), & \phi(x_2, y_1) &= \phi(x_4, y_3), \\ \phi(x_1, y_2) &= \phi(x_3, y_4), & \text{and } \phi(x_2, y_2) &= \phi(x_4, y_4). \end{aligned}$$

We allow degeneracies in this definition—for example, there might be equalities between values of  $\phi$  beyond those required by the definition itself.

As a subset of a partial Latin square  $(X, Y, Z, A, \phi)$  which has been drawn as a labelled grid, an octahedron is a configuration that looks like this (where we have chosen the example to emphasize that there is no ordering on  $X$  or  $Y$ , so all we care about are the relations ‘is in the same column as’, ‘is in the same row as’, and ‘has the same label as’).

$$\begin{array}{ccc} c & d & \\ & a & b \\ a & b & \end{array} \quad (a, b, c, d \in Z)$$

$$\begin{array}{ccc} & c & d \\ & & \end{array}$$

In a full Latin square, the relationship between the quadrangle condition (Definition 1.3) and octahedra is simple: an  $n \times n$  Latin square  $A$  satisfies the quadrangle condition if and only if the number of octahedra contained in  $A$  is equal to  $n^5$  (which is the maximal possible number in a Latin square). To see this, note that the number of rectangles in  $A$  is  $n^4$ , and if one wishes to find another rectangle with the same labelling, then there are at most  $n$  choices for the first corner (since its label is determined) and at most one choice for each remaining corner (since their labels are determined, as well as at least one of their row and column).

We now show that the multiplication table of a binary operation with many associative triples also contains many octahedra.

LEMMA 1.8. *Let  $X$  be a set of size  $n$  and let  $\circ$  be a partially defined binary operation on  $X$  that is injective in each variable separately and for which there are at least  $\epsilon n^3$  triples  $(x, y, z) \in X^3$  with  $x \circ (y \circ z) = (x \circ y) \circ z$ . Then the multiplication table of  $\circ$  contains at least  $\epsilon^4 n^5$  octahedra.*

*Proof.* For each  $b \in X$ , let  $W_b$  be the set of  $(a, c)$  such that  $a \circ (b \circ c) = (a \circ b) \circ c$ . Then the average size of  $|W_b|$  is at least  $\epsilon n^2$ . Writing  $\epsilon_b$  for the density of  $W_b$  in  $X^2$ , an easy Cauchy-Schwarz argument tells us that  $W_b$  contains at least  $\epsilon_b^4 n^4$  quadruples  $(a_0, a_1, c_0, c_1)$  such that all four points  $(a_i, c_j)$  belong to  $W_b$ . Therefore, by Jensen's inequality, the average number of such quadruples in  $W_b$  is at least  $\epsilon^4 n^4$ . Each such quadruple yields a diagram of the following form.

$$\begin{array}{ccccccc}
 g_1 & f_{01} & f_{11} & & & & \\
 g_0 & f_{00} & f_{10} & & & & \\
 c_1 & & & g_1 & f_{01} & f_{11} & \\
 c_0 & & & g_0 & f_{00} & f_{10} & \\
 b & d_0 & d_1 & & & & \\
 \\ 
 \circ & a_0 & a_1 & b & d_0 & d_1 & 
 \end{array}$$

where the left column and bottom row say which elements are being multiplied together. The associativity of the triples  $(a_i, b, c_j)$  is used to prove that  $a_i \circ (b \circ c_j) = (a_i \circ b) \circ c_j = f_{ij}$ , and the result is that each quadruple of triples gives us an octahedron. Note that from the octahedron we can reconstruct the pairs  $(a_0, d_0)$  and  $(a_1, d_1)$  from looking at which columns are used, and since the equation  $a_0 x = d_0$  has a unique solution, we can reconstruct  $b$ . Therefore, distinct  $b$  give rise to distinct octahedra, and putting all this together implies that there are at least  $\epsilon^4 n^5$  octahedra, as claimed. □

Observe that in an  $n \times n$  grid labelled completely at random from a set of size  $n$ , the expected number of octahedra is approximately  $n^4$ , since there are  $n^8$  pairs of rectangles, almost all of which are non-degenerate, and the probability that the labels agree is  $n^{-4}$  for each non-degenerate pair. It follows that the number of octahedra in a group multiplication table is far larger than the number of octahedra in a randomly labelled grid. The same is true of the number of octahedra in the multiplication table of a partially associative operation. These observations suggest that the octahedron count in a partial Latin square can serve as a measure of underlying group structure.

Thus, the hypothesis that we wish to consider is a weakening of the hypothesis of Theorem 1.2. Our main result is that we can obtain the same conclusion.

**Theorem 1.9.** *Let  $X, Y, Z$  be sets of size  $n$ , let  $E \subset X \times Y$ , and let  $\lambda : E \rightarrow Z$  be a partial Latin square with at least  $\epsilon n^5$  octahedra. Then for every positive integer  $b$  there exist a subset  $A \subset E$  of density at least  $\epsilon^{b^{25b}}$ , a metric group  $G$ , and maps  $\phi : X \rightarrow G$ ,  $\psi : Y \rightarrow G$  and  $\omega : Z \rightarrow G$ , such that the images  $\phi(X), \psi(Y)$  and  $\omega(Z)$  are 1-separated, and  $d(\phi(x)\psi(y), \omega(z)) \leq b^{-1}$  for every  $(x, y, z) \in X \times Y \times Z$  such that  $(x, y) \in A$  and  $\lambda(x, y) = z$ .*

Theorem 1.2 follows immediately by applying Lemma 1.8 followed by Theorem 1.9.



While the quadrangle condition as defined for Latin squares in Definition 1.3 makes sense also for partial Latin squares, it is not symmetric in  $X, Y$  and  $Z$ , and that turns out to be inconvenient. Instead, we make the following definitions.

DEFINITION 1.10. *Let  $P = (X, Y, Z, A, \phi)$  be a partial Latin square. Then  $P$  satisfies the column, row, or label quadrangle condition if it contains no configuration of the form*

$$\begin{array}{ccc} \begin{array}{cc} c & d \\ a & b \end{array} & , & \begin{array}{cc} c & d \\ a & b \end{array} & d & , & \text{or} & \begin{array}{cc} c & d \\ a & b \end{array} \\ & & & c & & & c & d' & , \\ & & & a & b & & a & b \end{array}$$

respectively. It satisfies the quadrangle condition if it satisfies all three of the column, row, and label quadrangle conditions.

The column/row/label quadrangle condition is saying that there is no configuration obtained from an octahedron by taking one of its elements and changing just its column/row/label. For a full Latin square, the three quadrangle conditions are equivalent, but for a partial Latin square they are not, and it is convenient to distinguish between them.

The following combinatorial statement, which is of independent interest, is in fact a special case of Theorem 1.9. We shall treat it separately in Sect. 5.4 as an introduction to the more general argument.

**Theorem 1.11.** *Let  $X, Y$  and  $Z$  be sets of size  $n$ , let  $A \subset X \times Y$ , and let  $\phi : A \rightarrow Z$  be a partial Latin square with at least  $\epsilon n^5$  octahedra. Then there is a subset  $B \subset A$  of size at least  $\alpha n^2$ , where  $\alpha = \alpha(\epsilon) > 0$ , that satisfies the quadrangle condition.*

In order to prove it, we first obtain a dense subset that satisfies the label quadrangle condition, and then use symmetry to pass to further dense subsets that satisfy the column and row quadrangle conditions as well.

One might at first think that Theorem 1.11 (with a suitable bound) would imply not just Theorem 1.9, but even a stronger result where  $H$  is a  $k$ -approximate subgroup rather than an  $(\epsilon, k)$ -approximate subgroup. However, while a Latin square that satisfies the quadrangle condition must be the multiplication table of a group, a partial Latin square that satisfies the quadrangle condition is not necessarily part of the multiplication table of a group: indeed, Example 1.1 of approximate multiplication on a  $\delta$ -net of  $SO(3)$  gives a counterexample. (This is significantly easier to prove than Green’s result [Gre20], which says that one cannot even restrict it to a dense set that is isomorphic to part of a group multiplication table). More elementary counterexamples can be obtained by observing that if a group multiplication table ever contains the following configuration,

$$\begin{array}{cc} e & d \\ f & c \\ a & b \end{array}$$

then  $ab^{-1}cd^{-1}ef^{-1}$  is equal to the identity, so any five of the labels determine the sixth. Thus, in a group multiplication table we have not only the (label) quadrangle condition but also a natural ‘pair of 6-cycles’ generalization, which states that in a configuration such as the following,  $f$  must equal  $f'$ .

$$\begin{array}{cc}
 e & d \\
 f & c \\
 a & b \\
 & e & d \\
 & f' & c \\
 & a & b
 \end{array}$$

Note that that configuration itself satisfies the quadrangle condition (for trivial reasons) even if  $f \neq f'$ .

What we therefore need to do in order to prove Theorem 1.9 is find a subset of the partial Latin square that satisfies a generalized quadrangle condition that applies to all configurations up to a certain size. What those configurations are will be explained in the next subsection.

**1.2 Where the metric group comes from.** It turns out that the metric group we obtain in Theorem 1.9 (and therefore also in Theorem 1.2) is given by a simple universal construction. That does not mean that the proof of our main result is simple, however, because it is not obvious how to pass to a dense subset of the partial Latin square for which the universal construction has the desired properties.

Suppose we are given a partial Latin square  $(X, Y, Z, A, \lambda)$  and would like it to satisfy the conclusion of Theorem 1.9. Then we want a metric group  $G$  and maps  $\phi : X \rightarrow G, \psi : Y \rightarrow G$  and  $\omega : Z \rightarrow G$  such that the images of  $\phi, \psi$  and  $\omega$  are 1-separated and  $d(\phi(x)\psi(y), \omega(z)) \leq b^{-1}$  for every  $x \in X, y \in Y$  and  $z \in Z$  such that  $\lambda(x, y) = z$ . An obvious approach is to let  $G$  be the free group on  $X \cup Y \cup Z$  (we assume that  $X, Y$  and  $Z$  are disjoint—if not, we make disjoint copies), to take the largest metric such that the second condition holds, and hope that it is large enough for the first condition to hold as well.

This metric can be described explicitly in a standard way using van Kampen diagrams. We take the group presentation with generators  $X \cup Y \cup Z$  and all the relations given by the partial Latin square when it is thought of as a multiplication table. That is, the relations are all words of the form  $xyz^{-1}$  such that  $\lambda(x, y) = z$ . Given two words  $w_1, w_2$  in the free group on  $X \cup Y \cup Z$ , their distance is  $b^{-1}$  times the area of the smallest van Kampen diagram with these relations and with boundary word  $w_1w_2^{-1}$ . Of course, if the relations do not imply that  $w_1 = w_2$ , then there is no such van Kampen diagram, in which case we set the distance to be infinite. (It is not hard to see that a necessary and sufficient condition for the partial Latin square to arise as part of a group multiplication table is that *all* distances between distinct generators are infinite).

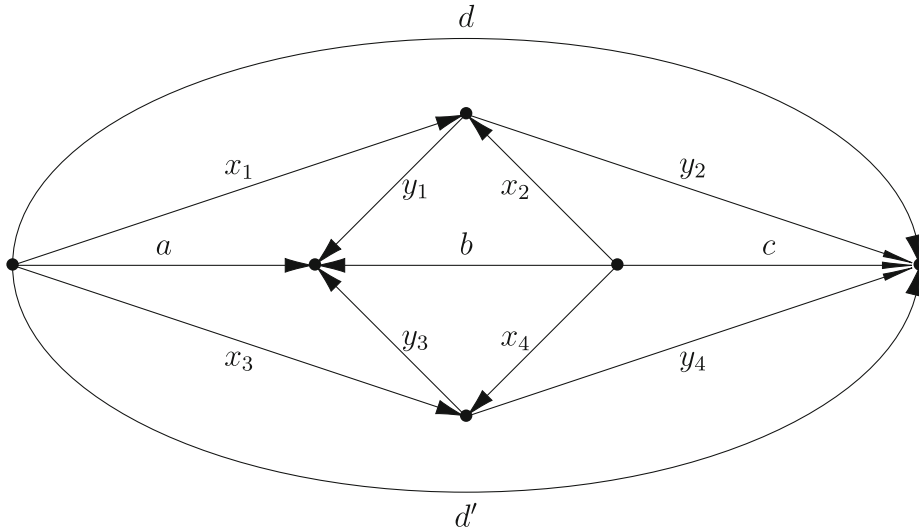


Figure 1: A van Kampen diagram of area 8 that corresponds to a proof that  $d$  and  $d'$  are equal in any group that satisfies the given relations

In order to prove Theorem 1.9, we therefore need to pass to a dense subset of the partial Latin square with the property that given any two distinct elements  $x_1, x_2$  of  $X$ , there is no van Kampen diagram of area less than  $b$  with boundary word  $x_1x_2^{-1}$ , and similarly for  $Y$  and  $Z$ .

We shall discuss van Kampen diagrams more fully later in the paper, but for the benefit of the reader unfamiliar with them, we illustrate here how a partial Latin square that fails the label quadrangle condition gives rise to a van Kampen diagram of area 8 with boundary word of the form  $z_1z_2^{-1}$ . Suppose, then, that the partial Latin square contains the configuration

$$\begin{array}{cc} c & d \\ a & b \\ & c & d' \\ & a & b \end{array}$$

and that  $d \neq d'$ . If we give appropriate names to the rows and columns, then the relations we obtain from this configuration give us the van Kampen diagram in Fig. 1 below. For example, the triangle towards the left with edges labelled  $x_1, y_1$  and  $a$  is telling us that the first  $a$  in the configuration belongs to column  $x_1$  and row  $y_1$ . (We direct edges so that if  $xy = z$  then there is a directed path with edges  $x$  then  $y$ , and a directed edge  $z$  with the same start and end points). Since there is also a triangle with edges labelled  $x_3, y_3$  and  $a$ , we can read off that  $x_1y_1 = x_3y_3$ . More generally, any cycle corresponds to a word that can be proved to be the identity if we use the given relations: in this case, the word is  $x_1y_1y_2^{-1}x_2^{-1}$ .

Since this van Kampen diagram has area 8 (which simply means that it has eight faces), it shows that the distance between  $d$  and  $d'$  is at most  $8b^{-1}$ . Therefore, for

$b > 8$  we need to avoid it. The generalized quadrangle condition we mentioned earlier is then that a partial Latin square should contain no configuration that corresponds to a van Kampen diagram with boundary word of the form  $uv^{-1}$  for distinct generators  $u, v$ , and area less than  $b$ .

Note that if  $d = d'$ , then the edges labelled  $d$  and  $d'$  become identified, and the resulting diagram is a triangulation of the 2-sphere. It is the triangulation given by the faces of an octahedron, which is why we use the word ‘octahedron’ in Definition 1.7 to describe a pair of identically labelled rectangles in the grid.

## 2 Obtaining $C$ -well-defined completion operations

In this section, we take an important first step by proving a weaker form of Theorem 1.11, or more precisely a generalization of a weaker form. Recall, as discussed in Definition 1.3, that we can think about the quadrangle conditions in terms of certain functions being well-defined. For instance, a partial Latin square satisfies the label quadrangle condition if and only if there is a well-defined partial ternary operation on the set of labels that takes a triple  $(a, b, c)$  to  $d$  if and only if there exists a rectangle with its corners labelled  $\begin{smallmatrix} c & d \\ a & b \end{smallmatrix}$ . We called this the *rectangle completion operation*. We now make the following definitions. It will be convenient to think of a partial Latin square as a hypergraph.

**DEFINITION 2.1.** *Let  $H$  be a linear tripartite 3-uniform hypergraph. A label  $2r$ -cycle in  $H$  is a sequence of triples*

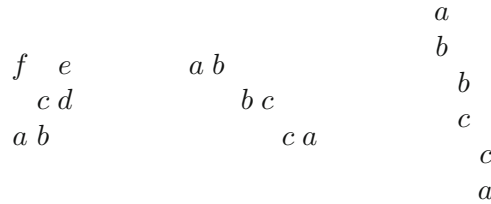
$$(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_{2r}, y_{2r}, z_{2r})$$

such that  $x_i = x_{i+1}$  if  $i$  is even, and  $y_i = y_{i+1}$  if  $i$  is odd (and  $x_{2r} = x_1$ ). The sequence  $(z_1, \dots, z_{2r})$  is the label sequence of the  $2r$ -cycle. A row  $2r$ -cycle and a column  $2r$ -cycle are defined in the same way after appropriate cyclic permutations of the coordinates. (For instance, for a column  $2r$ -cycle we require  $y_i = y_{i+1}$  if  $i$  is odd and  $z_i = z_{i+1}$  if  $i$  is even, and  $(x_1, \dots, x_{2r})$  is the column sequence).

The label  $2r$ -cycle completion operation is the not necessarily well-defined partial function that takes a  $(2r - 1)$ -tuple  $(z_1, \dots, z_{2r-1})$  to  $z_{2r}$  if  $(z_1, \dots, z_{2r})$  is the label sequence of some  $2r$ -cycle. We say that this operation is well-defined if there is always at most one possibility for  $z_{2r}$  given  $z_1, \dots, z_{2r-1}$ , and we say that it is  $C$ -well-defined if the number of possibilities is always at most  $C$ .

Definition 2.1 generalises the notion of a rectangle in a partial Latin square (which is simply a label 4-cycle). Here are pictures of a label 6-cycle, a column 6-cycle, and

a row 6-cycle.



In this section, we shall be using the language of partial Latin squares, so we shall omit the word ‘label’ when referring to label  $2r$ -cycles and the label  $2r$ -cycle completion operation. (Of course, a label  $2r$ -cycle in a partial Latin square is just a label  $2r$ -cycle in the corresponding hypergraph).

When referring to a  $2r$ -cycle  $C$ , we will sometimes abbreviate the list of triples to a string, e.g.  $C = x_1y_1 \dots x_ry_r$  where the  $x_i$  and  $y_j$  are triples and the row of  $x_1$  is shared with the row of  $y_1$ , etc.

The main result of this section is the following.

**Theorem 2.2.** *Let  $0 < \epsilon < 10^{-3}$  and let  $k \geq 2$  be a positive integer. Let  $A$  be a partial Latin square containing at least  $\epsilon n^5$  octahedra (see Definition 1.7). Then we can find a subset  $B \subset A$  of density  $\beta \geq \epsilon^{10}$  with the property that for each  $2 \leq r \leq k$  the  $2r$ -cycle completion operation in  $B$  is  $\epsilon^{-33k^3}$ -well-defined.*

We begin with a well-known bound for the number of  $2r$ -cycles in a bipartite graph, which will underlie many of the calculations throughout this section.

**LEMMA 2.3.** *Let  $A$  be a subset of the  $n \times n$  grid of density  $\alpha$ . Then  $A$  contains at least  $\alpha^{2r} n^{2r}$  and at most  $\alpha^r n^{2r}$  distinct labelled  $2r$ -cycles.*

*Proof.* We may view  $A$  as a bipartite graph with vertex sets  $X$  and  $Y$  of size  $n$  and  $\alpha n^2$  edges. Let  $\lambda_1, \dots, \lambda_n$  be the singular values of the adjacency matrix of this graph. Then the number of  $2r$ -cycles is equal to  $\sum_i \lambda_i^{2r}$ . But the largest singular value is at least  $\alpha n$ , so this sum is at least  $\alpha^{2r} n^{2r}$ .

For the upper bound we observe that the number of  $2r$ -cycles can be counted by summing, over all (ordered)  $r$ -tuples  $(x_1, \dots, x_r) \in A^r$ , the indicator that there is a  $2r$ -cycle  $x_1y_1 \dots x_ry_r$ . This sum is clearly at most  $|A|^r = \alpha^r n^{2r}$ , since that is the number of ways of choosing  $(x_1, \dots, x_r)$  without the additional condition.  $\square$

The lower bound on the octahedron count in  $A$  requires that the labelling of a random rectangle is repeated, on average, many times. This motivates the following definition.

**DEFINITION 2.4.** *Given a partial Latin square  $A$ , a  $2r$ -cycle is  $\theta$ -popular in  $A$  if the labelling of the cycle occurs at least  $\theta n$  times in  $A$ .*

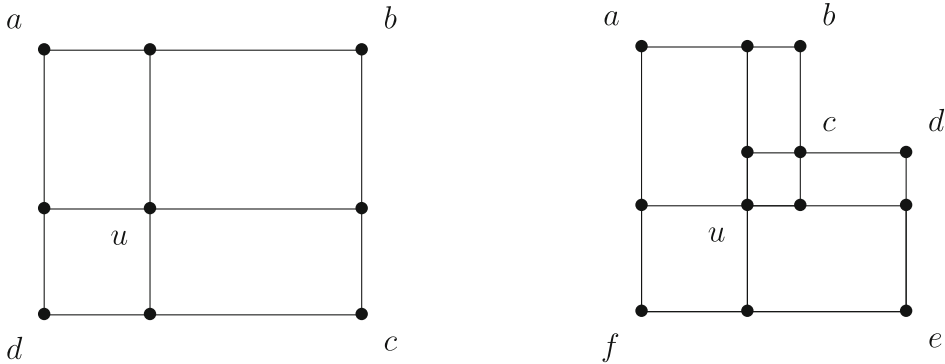


Figure 2: A point decomposition of a rectangle  $(a, b, c, d)$  and a 6-cycle  $(a, b, c, d, e, f)$ . If all rectangles with  $u$  and a vertex from the cycle as opposite corners are  $\epsilon$ -popular, then the decomposition is  $\epsilon$ -popular

Note that the trivial maximum for the number of occurrences of a  $2r$ -cycle with a given labelling is  $n$ , since once one has chosen which of at most  $n$  points to choose with the first label, the condition that no label is repeated in any row or column implies that rest of the  $2r$ -cycle, if it exists, is determined by the labelling.

The first step towards obtaining the decompositions we need is a dependent random selection that ensures that almost all  $2r$ -cycles can be decomposed into popular rectangles in many ways. The next definition explains what we mean by ‘decomposed’ here.

**DEFINITION 2.5.** *Given a  $2r$ -cycle  $C = x_1y_1 \dots x_ry_r$  in a partial Latin square  $A$ , a point decomposition of  $C$  in  $A$  is a collection of  $2r$  rectangles, all belonging to  $A$  and all sharing a point  $u$ , with the corners opposite to  $u$  being the  $x_i$  and  $y_i$ . We call the point decomposition  $\epsilon$ -popular if each of the  $2r$  rectangles is  $\epsilon$ -popular in  $A$  (see Definition 2.4).*

Point decompositions for a rectangle and a 6-cycle are shown in Fig. 2.

**LEMMA 2.6.** *Let  $0 < \epsilon, \delta < \frac{1}{100}$  and let  $k > 1$  be a fixed integer. Given a partial  $n \times n$  Latin square  $A$  containing at least  $\epsilon n^5$  octahedra, we can find a subset  $B_1 \subset A$  of density  $\beta_1 \geq \epsilon/2$  such that for each  $2 \leq r \leq k$ , a proportion at least  $1 - \delta$  of  $2r$ -cycles in  $B_1$  have at least  $\delta \epsilon^{4k} n^2$  different  $\epsilon/2$ -popular point decompositions in  $A$ .*

*Proof.* We define a graph  $G$  with vertex set given by  $[n]^2$  corresponding to the cells of the  $n \times n$  grid, and edges given by joining  $x$  to  $y$  if the rectangle with opposite corners  $x$  and  $y$  has all its vertices in  $A$  and is  $\epsilon/2$ -popular.

Let  $X$  be the number of edges in  $G$  and  $Y$  be the number of non-edges. An edge in  $G$  can be associated to a set of at least  $\epsilon n/2$  (and at most  $n$ ) octahedra, by combining the rectangle corresponding to the edge with one of the other rectangles

with the same labelling. Similarly, a non-edge in  $G$  can be associated to a set of less than  $\epsilon n/2$  octahedra. In such a way, all octahedra of  $A$  are accounted for. Therefore

$$\begin{aligned} Xn + Y\epsilon n/2 &\geq \epsilon n^5 \\ \Rightarrow Xn + \epsilon n^5/2 &\geq \epsilon n^5 \end{aligned}$$

so  $G$  has average degree at least  $\epsilon n^2$ .

Let  $\eta = \eta(\delta, k) = \delta\epsilon^{4k}$ . A  $2r$ -cycle has at least  $\eta n^2$  different  $\epsilon/2$ -popular point decompositions in  $A$  if the common neighbourhood (in  $G$ ) of the  $2r$  corner vertices has size at least  $\eta n^2$ .

We choose a vertex  $v$  in  $G$  uniformly at random, and let  $N(v)$  be the neighbourhood of  $v$  in  $G$ . This is our dependent random selection. It remains to prove that it works with positive probability.

Let  $C = x_1y_1 \dots x_r y_r$  be a given  $2r$ -cycle in  $A$ . Let  $N(C)$  be the set of vertices in  $G$  that are joined to all of  $x_1, \dots, y_r$ . We shall say that  $C$  is *bad* if  $|N(C)| < \eta n^2$ . If  $C$  is bad, we have that

$$\mathbb{P}(C \subset N(v)) = \frac{|N(C)|}{n^2} < \eta.$$

Let  $Z_r$  count the number of bad  $2r$ -cycles in  $N(v)$ . We have  $\mathbb{E}Z_r \leq \eta n^{2r}$ . Let  $Z = \sum_{r=2}^k n^{-2r} Z_r$ . Then

$$\mathbb{E}Z \leq \sum_{r=2}^k \eta \leq k\eta.$$

Our lower bound on the average degree of  $G$  also gives us that

$$\mathbb{E}(|N(v)|) \geq \epsilon n^2.$$

In particular, we have

$$\mathbb{E}\left(|N(v)|n^{-2} - \epsilon/2 - \epsilon Z(2k\eta)^{-1}\right) \geq 0$$

so there is a choice of vertex  $v$  such that  $|N(v)|n^{-2} \geq \epsilon/2$  and  $|N(v)|n^{-2} \geq \epsilon\eta^{-1}Z/2k$ . The first inequality gives us that the total count,  $X_r$ , of  $2r$ -cycles in  $N(v)$  is at least  $(\epsilon/2)^{2r}n^{2r}$ , while the second inequality implies that  $Z \leq 2k\eta\epsilon^{-1}$ . So for each  $2 \leq r \leq k$ ,

$$Z_r n^{-2r} \leq 2k\eta\epsilon^{-1} \leq 2k\eta\epsilon^{-1}(\epsilon/2)^{-2r}n^{-2r}X_r,$$

which implies that

$$Z_r \leq k\eta(\epsilon/2)^{-(2r+1)}X_r.$$

Therefore, letting  $B_1 = N(v)$  for this choice of  $v$ , we have  $\beta_1 n^2 = |N(v)| \geq \epsilon n^2/2$  and the proportion of  $2r$ -cycles in  $N(v)$  which are bad is at most  $k\eta(\epsilon/2)^{-(2r+1)} \leq \delta$ .  $\square$

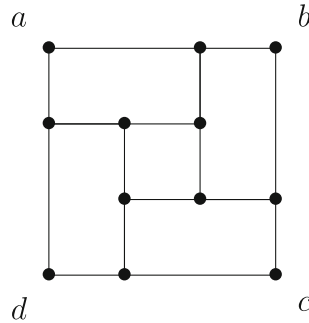


Figure 3: A ring decomposition of a 4-cycle  $(a, b, c, d)$

Using Lemma 2.6 we may pass to a dense subset  $B_1$  of  $A$  such that almost all  $2r$ -cycles have many (within a constant factor of the trivial maximum) popular point decompositions in  $A$ . However, for our purposes the ‘almost all’ is not sufficient, and we need to use a more complicated decomposition to boost Lemma 2.6 into an ‘all’ statement.

The following definitions introduce these more complex decompositions.

**DEFINITION 2.7.** Let  $X$  be a partial Latin square. Given a  $2r$ -cycle  $C = x_1y_1 \dots x_ry_r$ , let  $C' = y'_1x'_2 \dots x'_ry'_1$  be a second  $2r$ -cycle in  $X$  such that all the rectangles  $R_1, \dots, R_{2r}$  with opposite corner pairs either  $(x_i, x'_i)$  or  $(y_i, y'_i)$  belong to  $X$ . We call the collection of  $C'$  and the rectangles  $R_1, \dots, R_{2r}$  a ring decomposition of  $C$  in  $X$ . An example (for  $r = 2$ ) is shown in Fig. 3. As in Definition 2.5, we say that a ring decomposition  $(C', R_1, \dots, R_{2r})$  of  $C$  is  $\epsilon$ -popular if  $C'$  and all the rectangles  $R_1, \dots, R_{2r}$  are  $\epsilon$ -popular in  $X$  (see Definition 2.4).

**REMARK 2.8.** The reason we listed the points of the cycle  $C'$  in the order  $y'_1x'_2 \dots x'_ry'_1$  is that according to Definition 2.1 of a  $2r$ -cycle, the first two points of a cycle must share a row rather than a column, so had we listed the vertices in the more obvious order, then it would not technically have been a cycle. This is slightly more than hair splitting: it is important that if two points in  $C$  share a row, then the corresponding points in  $C'$  share a column, and vice versa.

**DEFINITION 2.9.** Let  $X$  be a partial Latin square. Let  $C$  be a  $2r$ -cycle in  $X$ , and let  $(C', R_1, \dots, R_{2r})$  be a ring decomposition of  $C$ . A full decomposition of  $C$  in  $X$  consists of a point decomposition of  $C'$ , together with point decompositions of each rectangle  $R_i$ . Since a point decomposition of a  $2r$ -cycle itself consists of  $2r$  rectangles, a full decomposition of a  $2r$ -cycle involves  $10r$  rectangles as shown (for  $r = 2$ ) in Fig. 4. As usual, we say that a full decomposition is  $\epsilon$ -popular if all of these  $10r$ -rectangles are  $\epsilon$ -popular in  $X$  (see Definition 2.4).

**REMARK 2.10.** It will be important to keep track of the order (in  $n$ ) of the trivial maxima for the number of ring decompositions and full decompositions of a  $2r$ -cycle



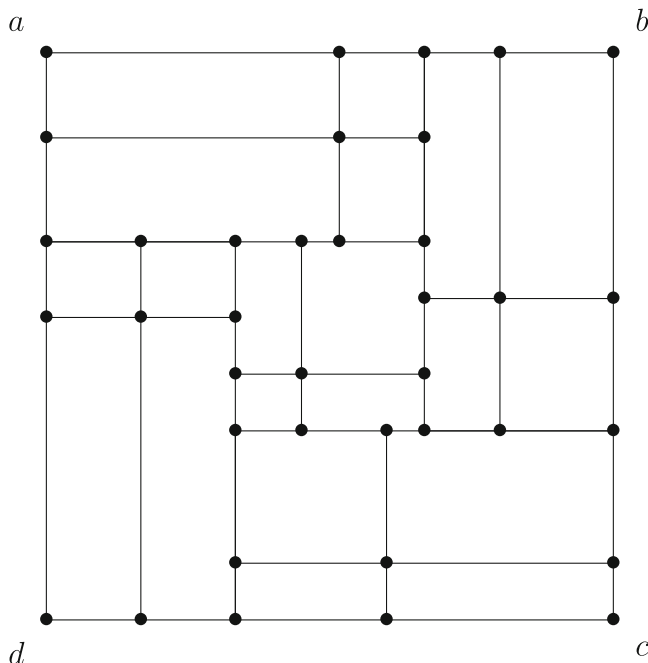


Figure 4: A full decomposition of the rectangle with labels  $(a, b, c, d)$ . If each of the 20 small rectangles in the figure is  $\epsilon$ -popular, then we say that the decomposition is  $\epsilon$ -popular

in a dense subset of an  $n \times n$  Latin square. The number of ring decompositions is at most  $n^{2r}$ , since a ring decomposition of a  $2r$ -cycle  $C$  is uniquely defined by a  $2r$ -cycle  $C'$ . In a full decomposition,  $C'$  and all the rectangles in the ring decomposition are given point decompositions, each of which can be chosen in at most  $n^2$  ways. So the number of full decompositions is at most  $n^{2r+2(2r+1)} = n^{6r+2}$ .

Our next step is to pass to a subset  $B_2$  of  $B_1$  such that all  $2r$ -cycles in  $B_2$  have within a constant factor of the trivial maximum number of ring decompositions. Since almost all  $2r$ -cycles in  $B_1$  have many popular point decompositions, we will then be able to pass to a further subset  $B_3$  of  $B_2$  so that all  $2r$ -cycles in  $B_3$  have within a constant factor of the trivial maximum number of  $\epsilon$ -popular full decompositions.

In order to achieve the first step of this process, we will again apply a dependent random selection argument. The following lemma is in fact far more general than we need, but the full statement is more natural to prove than the special case that we will use.

LEMMA 2.11. *Let  $k$  be a positive integer. Let  $G$  be a bipartite graph with vertex classes  $X, Y$  of size  $n$  and edge density  $\delta$ , with  $0 < \delta < \frac{1}{100}$ . Then we can pass to subsets  $X' \subset X$  and  $Y' \subset Y$ , each of size at least  $\delta^2 n / 16$ , such that the edge density in  $G' = G|_{X' \times Y'}$  is at least  $\delta / 4$  and for any  $2 \leq r \leq k$  and any choice of  $r$  vertices  $x_1, \dots, x_r \in X'$  and  $y_1, \dots, y_r \in Y'$  we have at least  $\delta^{5k^2+4k} n^{2r}$  choices of vertices*

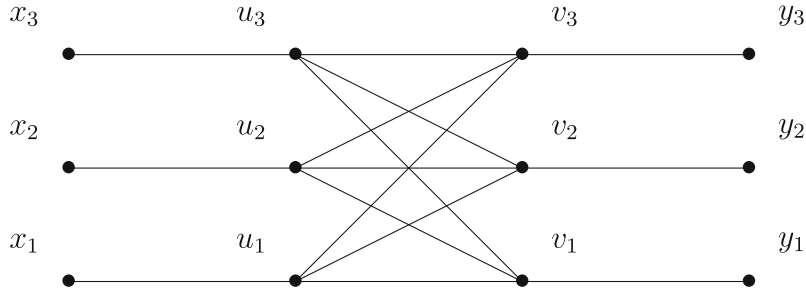


Figure 5: The subgraph found in Lemma 2.11 when  $k = r = 3$

$u_1, \dots, u_r, v_1, \dots, v_r$  in  $G$  with  $u_i v_j \in E(G)$ ,  $x_i u_i \in E(G)$  and  $y_i v_i \in E(G)$  for each  $i, j \in \{1, \dots, r\}$ .

Figure 5 shows the vertices  $x_i, y_j, u_k$  and  $v_l$  with the corresponding edges when  $k = r = 3$ .

*Proof.* Let us begin by discarding all vertices from  $X$  of degree smaller than  $\delta n/2$ . This leaves a set  $X_1 \subset X$  of size at least  $\delta n/2$ .

Let  $c_1 = \delta^{2k}$  and  $c_2 = \delta^{5k}$ . We will use a dependent random selection argument that allows us to pass to a subset  $X_2 \subset X_1$  of size at least  $(\delta^2/8)n$  with the property that for a  $(1 - c_1)$  proportion of choices  $(x_1, \dots, x_{k+1})$  from  $X_2$  we have at least  $c_2 n$  vertices in the shared neighbourhood  $\Gamma(x_1, \dots, x_{k+1}) \subset Y$ .

We do this by picking a vertex  $y \in Y$  at random and considering  $\Gamma(y)$ . Observe that

$$\mathbb{E}(|\Gamma(y)|) \geq \delta |X_1|/2 \geq \delta^2 n/4.$$

Let us call a  $(k + 1)$ -tuple  $(x_1, \dots, x_{k+1})$  *bad* if  $|\Gamma(x_1, \dots, x_{k+1})| < c_2 n$ . Let  $B$  be the number of bad tuples in  $\Gamma(y)$ . The probability that a given bad  $(k + 1)$ -tuple belongs to  $\Gamma(y)$  is less than  $c_2$ , since for this to happen  $y$  must be picked from  $\Gamma(x_1, \dots, x_{k+1})$ . Therefore

$$\mathbb{E}(B) < c_2 n^{k+1}$$

and so

$$\mathbb{E}(c_1 |\Gamma(y)|^{k+1} - c_1 (\delta^2/8)^{k+1} n^{k+1} - B) > (c_1 (\delta^2/4)^{k+1} - c_1 (\delta^2/8)^{k+1} - c_2) n^{k+1}.$$

Since  $c_2 = c_1 \delta^{3k}$ , this expectation is positive and so there is some choice of  $y$  for which both  $c_1 |\Gamma(y)|^{k+1} \geq c_1 (\delta^2/8)^{k+1} n^{k+1}$  and  $c_1 |\Gamma(y)|^{k+1} \geq B$ . These inequalities imply that  $|\Gamma(y)| \geq (\delta^2/8)n$  and that at most a proportion  $c_1$  of the  $(k + 1)$ -tuples from  $\Gamma(y)$  are bad. So we may take  $X_2 = \Gamma(y)$ .

Now we let  $X_3$  be the subset of  $X_2$  consisting of all vertices  $x_1 \in X_2$  with the property that for a proportion  $(1 - 2c_1)$  of the choices of  $x_2, \dots, x_{k+1} \in X_2$ , the shared neighbourhood  $\Gamma(x_1, \dots, x_{k+1}) \subset Y$  contains at least  $c_2 n$  vertices. Since

$|\Gamma(x_1, \dots, x_{k+1})| \geq c_2n$  for at least a proportion  $(1 - c_1)$  of all  $(k + 1)$ -tuples,  $|X_3| \geq |X_2|/2 \geq \delta^2n/16$ .

Since each vertex in  $X_3$  has at least  $\delta n/2$  neighbours in  $Y$ , the number of edges from  $Y$  to  $X_3$  is at least  $\delta n|X_3|/2$ . We now pass to the subset  $Y_1 \subset Y$  that consists of all vertices with at least  $\delta|X_3|/4$  edges into  $X_3$ . We note that  $|Y_1| \geq \delta n/4$ .

Now let  $x_1, \dots, x_k$  be chosen from  $X_3$  and  $y_1, \dots, y_k$  from  $Y_1$ . Let  $A_1, \dots, A_k$  be the neighbourhoods of the  $y_i$  in  $X_3$ —note that  $|A_i| \geq \delta|X_3|/4$ . Let  $T = A_1 \times \dots \times A_k$  and note that it has cardinality at least  $(\delta|X_3|/4)^k \geq (\delta|X_2|/8)^k$ .

By the choice of  $X_3$ , we know that the number of choices of  $u_1, \dots, u_k \in X_2$  such that  $|\Gamma(x_i, u_1, \dots, u_k)| < c_2n$  is at most  $2c_1|X_2|^k$  for each  $i = 1, \dots, k$ . Letting  $c_1 = \delta^{2k}$  so that  $2c_1k < (\delta/8)^k/2$  and noting that  $|T| = (\delta|X_2|/8)^k$ , we see that there must be at least  $(\delta|X_2|/8)^k/2$  choices of  $(a_1, \dots, a_k) \in T$  such that  $|\Gamma(x_i, a_1, \dots, a_k)| \geq c_2n$  for each  $i = 1, \dots, k$ . Observe that for any such choice of  $(a_1, \dots, a_k)$  and for any choice of  $b_i \in \Gamma(x_i, a_1, \dots, a_k)$  we get a complete bipartite graph between the  $a_i$  and the  $b_i$  as well as the edges  $x_i b_i$  and  $y_i a_i$  for each  $i$ .

The number of choices of the  $a_i$  and  $b_i$  from the above paragraph is at least

$$\begin{aligned} & \left( (\delta|X_2|/8)^k/2 \right) \times \left( (c_2n)^k \right) \geq (\delta^3/64)^k (\delta^{5k})^k n^{2k} \\ & \geq \delta^{5k^2+4k} n^{2k}. \end{aligned}$$

Observe that the subgraph induced by the  $x_i, y_j, a_k$  and  $b_l$  contains a  $2r$ -cycle  $a_1 b_1 \dots a_r b_r$  as well as the edges  $x_i a_i$  and  $y_i b_i$  for each  $i$ . Moreover, the edge density in  $X_3 \times Y_1$  is at least  $\delta/4$ , so taking  $X' = X_3$  and  $Y' = Y_1$ , the result follows.  $\square$

REMARK 2.12. It is well known that given a dense bipartite graph  $G$ , we may pass to a dense subgraph  $H$  such that any two vertices of  $H$  are joined by many  $P_3$ s in  $G$  [FS11]. Lemma 2.11 shows that a considerable generalization of this statement is available for relatively little extra effort: given any fixed bipartite graph  $H'$  with  $t$  special vertices  $v_1, \dots, v_t$  such that the shortest path from any  $v_i$  to any  $v_j$  has length at least 3, we may pass to a dense subgraph  $H$  of  $G$  such that for any  $u_1, \dots, u_t$  the number of isomorphic copies  $\phi(H')$  of  $H'$  in  $H$  with  $\phi(v_i) = u_i$  for all  $i$  is within a constant of the trivial maximum. The  $P_3$  statement is the special case where  $H'$  is a path of length 3 and  $v_1$  and  $v_2$  are its endpoints. (A similar observation was made in a blog post of Tao [Taob], but he was content to discuss just the special case he needed, and he left the proof as an exercise for the reader).

As an immediate corollary we obtain the following result, which will soon be applied in order to help guarantee the presence of many ring decompositions.

LEMMA 2.13. *Let  $k > 1$  be a positive integer. Let  $G$  be a bipartite graph with vertex classes  $X, Y$  of size  $n$  and edge density  $\delta$ , with  $0 < \delta < \frac{1}{100}$ . Then we can pass to subsets  $X' \subset X$  and  $Y' \subset Y$ , each of size at least  $\delta^2n/16$ , such that the edge density in  $G' = G|_{X' \times Y'}$  is at least  $\delta/4$  and for any  $2 \leq r \leq k$  and any choice of  $r$  vertices  $x_1, \dots, x_r \in X'$  and  $y_1, \dots, y_r \in Y'$  we have at least  $\delta^{7k^2} n^{2r}$  choices of  $2r$ -cycle  $u_1 v_1 \dots u_r v_r$  in  $G$  with  $x_i u_i \in E(G)$  and  $y_i v_i \in E(G)$  for each  $i = 1, \dots, r$ .*

*Proof.* The result follows by applying Lemma 2.11, and noting that the complete bipartite graph on  $r + r$  vertices contains a  $2r$ -cycle.  $\square$

When viewed as a statement about subsets of the grid, Lemma 2.13 states that we may pass to a dense subset  $B_2 \subset B_1$  such that all  $2r$ -cycles in  $B_2$  have many ring decompositions in  $B_1$  (see Definition 2.7). We must now pass to a further subset in which all  $2r$ -cycles have many popular full decompositions.

In the statement of the following lemma, we introduce a set  $\mathcal{C}$  of cycles. Each  $C \in \mathcal{C}$  is a  $2r$ -cycle for some  $2 \leq r \leq k$ . At this point in the argument, it may help to think about  $\mathcal{C}$  as the set of  $\theta$ -popular cycles (for some suitable  $\theta$ ), as this will be how  $\mathcal{C}$  is defined in our first application of the lemma. However, we will apply the lemma again later in the paper with a different collection  $\mathcal{C}$  of cycles, and hence we state the result in this more general way.

LEMMA 2.14. *Let  $0 < \beta, \delta, \gamma < \frac{1}{100}$  and  $k > 1$ . Let  $B$  be an  $n \times n$  partial Latin square of density at least  $\beta$ , and let  $\mathcal{C}$  be a collection of cycles in  $B$  with the property that for each  $2 \leq r \leq k$  at least a proportion  $1 - \delta$  of  $2r$ -cycles in  $B$  belong to  $\mathcal{C}$ . If  $\delta \leq \beta^{9k^2}$  then we can find a subset  $B'$  of  $B$  with density  $\beta' \geq \beta^8$  with the property that any  $2r$ -cycle in  $B'$  has at least  $\beta^{8k^2} n^{2r}$  different ring decompositions (see Definition 2.7) into cycles belonging to  $\mathcal{C}$ .*

*Proof.* Recall from Definition 2.7 that a ring decomposition of a  $2r$ -cycle  $C$  involves a  $2r$ -cycle  $C'$  and  $2r$  rectangles  $R_1, \dots, R_{2r}$  between these cycles, as shown in Fig. 3. The ring decomposition is fully determined by the choice of cycle  $C'$ . We shall call a ring decomposition of a cycle  $C$  *good* if all of  $C', R_1, \dots, R_{2r}$  belong to  $\mathcal{C}$ . We call a  $2r$ -cycle *indecomposable* if it has fewer than  $\beta^{8k^2} n^{2r}$  good ring decompositions in  $B$ , so that our goal is to pass to a dense subset of  $B$  in which no  $2r$ -cycle is indecomposable for  $2 \leq r \leq k$ .

In parallel with the partial Latin square  $B$ , we shall also consider the corresponding bipartite graph  $G$  in which the rows and columns form the vertex sets and the points of  $B$  form the edges (the labels here are ignored). The  $2r$ -cycles in  $B$  correspond precisely to the  $2r$ -cycles in  $G$ .

We begin by applying Lemma 2.13 to  $G$ . This allows us to pass to a subset  $B'$  of  $B$  of density at least  $(\beta^2/16)^2(\beta/4) \geq \beta^7$  with the property that each  $2r$ -cycle in  $B'$  has at least  $\beta^{7k^2} n^{2r}$  ring decompositions in  $B$ .

Let  $C$  be a  $2r$ -cycle in  $B'$ . If  $(C', R_1, \dots, R_{2r})$  is a ring decomposition of  $C$  which is bad (i.e. not good), then either  $C'$  does not belong to  $\mathcal{C}$  or some  $R_i$  does not belong to  $\mathcal{C}$ , or both. Since only a proportion  $\delta \leq \beta^{9k^2}$  of all  $2r$ -cycles in  $B$  are bad, and the maximum possible number of  $2r$ -cycles in  $B'$  is trivially bounded by  $n^{2r}$ , the number of ring decompositions of  $C$  for which  $C'$  is bad is at most  $\beta^{9k^2} n^{2r}$ . Therefore  $C$  has at least  $(\beta^{7k^2} - \beta^{9k^2}) n^{2r}$  ring decompositions  $(C', R_1, \dots, R_{2r})$  in  $B$  in which the  $2r$ -cycle  $C'$  is good.

If for each  $2 \leq r \leq k$  there are no more than  $\beta^7 n^2 / 4k^2$  disjoint indecomposable  $2r$ -cycles in  $B'$ , then discarding all points from a maximal disjoint set of indecomposable

cycles we discard at most  $\beta^7 n^2/2$  points, leaving a set of density at least  $\beta^7/2 \geq \beta^8$  with no indecomposable cycles and so we are done.

Thus, for some  $r$  it must be possible to find at least  $\beta^7 n^2/4k^2$  disjoint indecomposable  $2r$ -cycles  $C_1, \dots, C_t$  in  $B'$ . Each of these cycles  $C_i$  has at least

$$(\beta^{7k^2} - \beta^{9k^2})n^{2r} > \beta^{7k^2} n^{2r}/2$$

ring decompositions in  $B$  in which the  $2r$ -cycle  $C'$  is good but also has fewer than  $\beta^{8k^2} n^{2r}$  good ring decompositions. It follows that each  $C_i$  has at least

$$(\beta^{7k^2} - 2\beta^{8k^2})n^{2r}/2 > \beta^{7k^2} n^{2r}/3$$

ring decompositions  $(C'_i, R_{i,1}, \dots, R_{i,2r})$  in which some rectangle  $R_{i,j}$  does not belong to  $\mathcal{C}$ .

Since the cycles  $C_i$  are disjoint, the collections  $R_{i_1,j}$  and  $R_{i_2,j}$  of rectangles are disjoint when  $i_1 \neq i_2$  (since a rectangle  $R_{i_1,j}$  contains points from  $C_{i_1}$  which is disjoint from  $C_{i_2}$ ). Furthermore, a rectangle  $R_{i,j}$  can belong to at most  $n^{2r-2}$  different ring decompositions of  $C_i$  (since  $R$  determines two vertices of  $C'_i$ ). This means that  $B$  contains at least

$$\left(\beta^{7k^2} n^2/3\right) \left(\beta^7 n^2/4k^2\right) > \beta^{9k^2} n^4$$

bad rectangles in  $B$ .

But the number of bad rectangles is at most  $\delta\beta^2 n^4$ , so if  $\delta \leq \beta^{9k^2}$  then we have a contradiction.  $\square$

By applying Lemmas 2.6, 2.13 and 2.14 we will be able to pass to a dense subset  $B$  of  $A$  in which all  $2r$ -cycles have many popular full decompositions (see Definition 2.9 and Fig. 4). Before we give the details, we give one more technical lemma which draws a connection between popular full decompositions and the  $2r$ -cycle completion operation described in Definition 2.1.

**LEMMA 2.15.** *Let  $A$  be a partial Latin square and let  $B$  be a subset of  $A$ . Suppose that every  $2r$ -cycle in  $B$  has at least  $\gamma n^{6r+2}$  different  $\epsilon$ -popular full decompositions in  $A$  (see Definition 2.9). Then the  $2r$ -cycle completion operation in  $B$  is  $\epsilon^{-10r}\gamma^{-1}$ -well-defined (see Definition 2.1).*

*Proof.* Suppose that we have a tuple  $(a_1, \dots, a_{2r-1})$  such that the set  $\{x_i\}$  of possible labelling completions has size at least  $K$ . For each completion we can find  $\gamma n^{6r+2}$   $\epsilon$ -popular full decompositions.

Let us think about a typical one of these decompositions as follows. (For the discussion that follows, it may help to look at Fig. 6). We begin with a  $2r$ -cycle  $C$  with points  $x_1, y_1, \dots, x_r, y_r$ , where  $x_i$  has label  $a_{2i-1}$  when  $1 \leq i \leq k$ ,  $y_i$  has label  $a_{2i}$  when  $1 \leq i \leq k-1$ , and we do not know about the label attached to the point  $y_r$ . (It is important to be clear that the  $x_i$  and  $y_i$  are elements of  $[n]^2$  and not of  $[n]$  in this discussion).

Next, we have another  $2r$ -cycle  $C' = y'_1 x'_2 \dots x'_r y'_r x'_1$  (see Remark 2.8 for an explanation of why we list its points in this order).

Now we complete the cycles  $C$  and  $C'$  to a ring decomposition by adding in  $2r$  points  $u_1, v_1, \dots, u_r, v_r$ , where  $u_i$  is in the row that contains  $x_i$  and  $y_i$  and the column that contains  $x'_i$  and  $y'_i$ , and  $v_i$  is in the column that contains  $y_i$  and  $x_{i+1}$  and the row that contains  $y'_i$  and  $x'_{i+1}$ . (The points  $u_i$  and  $v_i$  do not form a  $2r$ -cycle).

The rectangles  $R_1, \dots, R_{2r}$  of this ring decomposition are given by  $S_i = (x_i, u_i, x'_i, v_{i-1})$  and  $T_i = (y_i, v_i, y'_i, u_i)$ . To form a point decomposition, we now add points  $p_i$  and  $q_i$ , and form the four rectangles that have a vertex in  $S_i$  and the opposite vertex at  $p_i$ , and the four rectangles that have a vertex in  $T_i$  and the opposite vertex at  $q_i$ . As well as the point  $p_i$ , we have to add four more points to  $S_i$  in order to complete the decomposition into four rectangles. Of these four points, let  $r_i$  and  $s_i$  be the ones in the same row and the same column as  $x_i$ ; we shall not bother giving names to the other two. Similarly, let  $w_i$  and  $z_i$  be the points in the same column and row as  $y_i$  that are part of the decomposition of  $T_i$  into four rectangles.

Now let us consider a certain subset of the (variable set of) points of the full decomposition. We shall take the points  $u_i$  and  $v_i$ , the points  $r_i$  and  $s_i$ , and the points  $w_i$  and  $z_i$  with  $1 \leq i \leq r - 1$ . We shall also take the two points from the point decomposition of  $C'$  that are in the same row and column as  $x'_1$ , and the two points from the decomposition of the rectangle  $T_r$  that are in the same row and column as  $y'_r$ . This makes a total of  $6r + 2$  points, so by the pigeonhole principle we can find some choice of labellings of these  $6r + 2$  points that occurs at least  $K\gamma$  times amongst the set of  $\epsilon$ -popular full decompositions of  $2r$ -cycles  $C$  for which the points  $x_1, y_1, \dots, x_r$  are labelled  $a_1, \dots, a_{2r-1}$ .

Observe that a full decomposition of a given cycle is uniquely determined by the way it is labelled, since once a point has been specified, any other point in the same row or column is then determined by its label. Observe also that since each rectangle in an  $\epsilon$ -popular full decomposition is  $\epsilon$ -popular, given three labels of any rectangle there are at most  $1/\epsilon$  different choices of label for the fourth, since otherwise there would be more than  $n$  rectangles that shared three labels, which is impossible.

Our aim now is use this observation to show that once the labellings of the  $6r + 2$  points specified earlier are given, the number of possible labellings of the remaining points is at most  $\epsilon^{-10r}$ . Since we know that it is also at least  $K\gamma$ , this will give us our desired upper bound on  $K$ .

To do this, we consider the 4-cycle completion operation as described in Definition 2.1. The observation above implies that if we know the labels at some subset of the points of the  $\epsilon$ -popular full decomposition from which we can repeatedly apply the 4-cycle completion operation to generate the entire decomposition, and if there are  $t$  other points, then the number of possible ways of completing the labelling is at most  $\epsilon^{-t}$ . We apply this to the set of  $6r + 2$  points we have chosen.

Note first that the rectangles  $R_i$ , apart from the rectangle containing the unfixed point of  $C$ , each contain five points from the set in their point decompositions, and

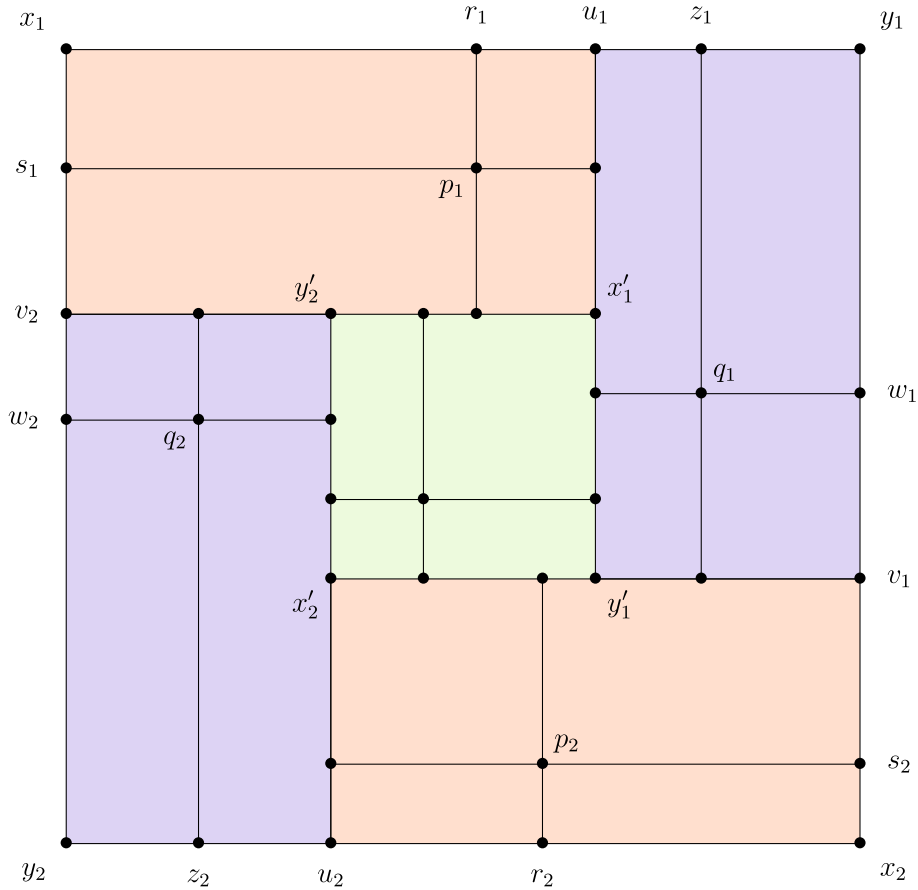


Figure 6: A full decomposition of the 4-cycle  $(a, b, c, d)$ , with labelling as described in the proof of Lemma 2.15. The rectangles  $S_1$  and  $S_2$  are shaded in red,  $T_1$  and  $T_2$  in blue, and the cycle  $C'$  in green

furthermore these five generate (using the 4-cycle completion operation) the other four. Therefore the closure of the set contains all the points in all the point decompositions of the rectangles  $S_1, \dots, S_r$  and  $T_1, \dots, T_{r-1}$ . These include the points  $x'_1, \dots, x'_r$  and  $y'_1, \dots, y'_{r-1}$ . Since we also have the points in the same row and column as  $x'_1$ , we obtain the central point of the point decomposition of  $C'$ , and using this we can work round  $C'$  and obtain all the points in its point decomposition. And now we have five points of the rectangle  $T_r$  that generate the others (since they lie along two edges), which shows that the  $6r + 2$  points we choose generate all the points of the full decomposition. It is not hard to check that a full decomposition contains  $18r + 1$  points, so we find, as promised, that the number of labellings given the labels at the  $6r + 2$  points and  $2r - 1$  of the points of  $C$  is at most  $\epsilon^{-10r}$ , as claimed, and this proves that  $K \leq \epsilon^{-10r}\gamma$ .  $\square$

We are now ready to prove the result we stated at the beginning of the section.

*Proof of Theorem 2.2.* Apply Lemma 2.6 with  $\delta = (\epsilon/2)^{9k^2}$ . This allows us to pass to a subset  $B_1 \subset A$  of density  $\beta_1 \geq \epsilon/2$  such that for each  $2 \leq r \leq k$  a proportion at least  $1 - \delta$  of  $2r$ -cycles in  $B_1$  have at least

$$(\epsilon/2)^{9k^2} \epsilon^{4k} n^2 \geq \epsilon^{11k^2} n^2$$

different  $\epsilon/2$ -popular point decompositions.

From here we apply Lemma 2.14, where we take the collection  $\mathcal{C}$  of cycles to be those  $2r$ -cycles  $C$  with  $2 \leq r \leq k$  and such that  $C$  has at least  $\epsilon^{11k^2} n^2$  different  $\epsilon/2$ -popular point decompositions. We can do this since  $B_1$  has density  $\beta_1 \geq \epsilon/2$ , so  $\delta \leq \beta_1^{9k^2}$ . The lemma gives us a subset  $B_2$  of  $B_1$  of density (in the original  $n \times n$  grid)  $\beta_2 \geq \beta_1^8 \geq \epsilon^{10}$  in which every  $2r$ -cycle in  $B_2$  has at least

$$\begin{aligned} \left( (\epsilon/2)^{8k^2} n^{2r} \right) \left( \epsilon^{11k^2} n^2 \right)^{2r+1} &\geq \epsilon^{20k^2+20k^3} n^{6r+2} \\ &\geq \epsilon^{30k^3} n^{6r+2} \end{aligned}$$

different  $\epsilon/2$ -popular full decompositions. (The first bracket on the left is a lower bound for the number of ring decompositions in which every cycle belongs to  $\mathcal{C}$ , and the second is a lower bound for the number of ways of converting each one into an  $\epsilon/2$ -popular full decomposition).

This allows us to apply Lemma 2.15 with  $\gamma = \epsilon^{30k^3}$ , which implies the result with  $B = B_2$  (since  $(\epsilon/2)^{-10r} \gamma^{-1} \leq (\epsilon/2)^{-10k-30k^3} \leq \epsilon^{-33k^3}$ ). □

We draw attention here to an analogy with the additive combinatorics result mentioned at the beginning of the paper, which states that if  $\phi : \mathbb{Z}_N \rightarrow \mathbb{Z}_N$  is a map such that  $\phi(x) + \phi(y) = \phi(z) + \phi(w)$  for a positive proportion of the quadruples  $x + y = z + w$ , then we can pass to a dense subset  $A \subset \mathbb{Z}_N$  such that the restriction of  $\phi$  to  $A$  is a Freiman homomorphism. One way of proving this result begins by showing that it is possible to pass to a set  $A'$  such that for each  $w$ , the number of values that  $\phi(x) + \phi(y) - \phi(z)$  can take when  $x + y - z = w$  is bounded by some constant  $C$  that is independent of  $N$ . This first step mirrors what we have achieved thus far. It is then necessary to find a separate argument to pass to a further subset where  $C$  is reduced to 1.

We have to do the same here, though at this point the analogy breaks down somewhat, since in the additive problem, Plünnecke’s inequality is used, but our setting does not involve an ambient group so we do not appear to have an analogous tool. Thus, while Theorem 2.2 constitutes significant progress towards our positive result, it turns out that we are still quite a long way from reducing  $C$  to 1.

### 3 Simplifying the decompositions

Perhaps surprisingly, the first step towards reducing  $C$  to 1 will involve abandoning full decompositions. The reason is that when we start to consider van Kampen



diagrams, we want the configurations that we consider in a partial Latin square to correspond to triangulated surfaces. Each labelled point in a partial Latin square corresponds to a triangle, and the triangles corresponding to two points share an edge if and only if the points share a column, a row, or a label. The problem with full decompositions is that they give rise to edges that are contained in more than two faces, and hence not to surfaces.

In this section, we shall use Theorem 2.2 as a tool in a ‘second pass’ through the arguments in Sect. 2. Our first lemma for this section shows that the property of  $C$ -well-definedness (see Definition 2.1) is sufficient to ensure that almost all of the cycles in  $B$  are popular, for a lower threshold of popularity (see Definition 2.4). This is significant because it enables us to repeat the process of the previous section, but eliminates the need for Lemma 2.6 and point decompositions, which are the source of the edges that are contained in too many faces. We will simply be able to reapply Lemma 2.14 to the subset  $B$  with a different set  $\mathcal{C}$  of cycles:  $\mathcal{C}$  will now simply be the set of  $\theta$ -popular cycles, for some appropriate  $\theta$ , rather than the set of those cycles with many popular point decompositions.

**LEMMA 3.1.** *Let  $B$  be an  $n \times n$  partial Latin square of density  $\beta$ . Suppose that the  $2r$ -cycle completion operation in  $B$  is  $C$ -well-defined (see Definition 2.1). Let  $\delta, \theta$  be such that  $\beta^{2r} \delta \theta^{-1} > C$ . Then the proportion of  $2r$ -cycles in  $B$  that are not  $\theta$ -popular is at most  $\delta$ .*

*Proof.* By Lemma 2.3, the number of  $2r$ -cycles in  $B$  is at least  $\beta^{2r} n^{2r}$ . Therefore, given a tuple  $(a_1, \dots, a_{2r-1})$  of labels, the number of  $2r$ -cycles with first  $2r - 1$  labels  $(a_1, \dots, a_{2r-1})$  is on average at least  $\beta^{2r} n$ . However, since the  $2r$ -cycle completion operation is  $C$ -well-defined we have further that the number of different  $a_{2r}$  completing a  $2r$ -cycle labelling  $(a_1, \dots, a_{2r})$  in  $B$  is at most  $C$ .

If a proportion greater than  $\delta$  of  $2r$ -cycles are not  $\theta$ -popular, then by averaging there must be some  $(a_1, \dots, a_{2r-1})$  such that a proportion greater than  $\delta$  of  $2r$ -cycles starting with these labels are not  $\theta$ -popular. But that means that there must be more than  $\beta^{2r} \delta \theta^{-1} > C$  completions which is a contradiction to the assumption that the  $2r$ -cycle completion operation in  $B$  is  $C$ -well-defined.  $\square$

We are now ready to put together our technical lemmas to prove the following proposition, which will be the main tool for passing to a dense subset of a partial Latin square that avoids the configurations we wish to avoid.

**PROPOSITION 3.2.** *Let  $0 < \epsilon < 10^{-3}$  and let  $A$  be a partial Latin square containing at least  $\epsilon n^5$  octohedra (see Definition 1.7). Let  $k \geq 100$  be an integer. Then we can find  $B \subset A$  of density  $\beta \geq \epsilon^{80}$  such that for each  $r = 2, \dots, k$  we have that every label  $2r$ -cycle in  $B$  has at least  $\epsilon^{80k^2} n^{2r}$  different  $\theta$ -popular ring decompositions (see Definition 2.7) in  $A$ , where  $\theta \geq \epsilon^{35k^3}$ . Moreover, the number of octahedra in  $B$  is at least  $\epsilon^{70k^3} n^5$ .*

*Proof.* We begin by applying Theorem 2.2. This allows us to pass to a subset  $B_1 \subset A$  of density  $\beta_1 \geq \epsilon^{10}$  with the property that for each  $2 \leq r \leq k$  the  $2r$ -cycle completion operation in  $B$  is  $C$ -well-defined, where  $C = \epsilon^{-33k^3}$ .

By Lemma 3.1 we see that a proportion greater than  $1 - \delta$  of  $2r$ -cycles (for each  $2 \leq r \leq k$ ) in  $B$  are  $\theta$ -popular for any choice of  $\theta < \beta_1^{2k} \delta / C$ .

We now apply Lemma 2.14 again, but taking the collection  $\mathcal{C}$  to consist of those  $2r$ -cycles for  $2 \leq r \leq k$  which are  $\theta$ -popular. To do this, we take  $\delta = (\beta_1)^{9k^2}$ . With this value of  $\delta$ , we may take some  $\theta \geq \epsilon^{20k+90k^2+33k^3} \geq \epsilon^{35k^3}$ .

The lemma then gives us a subset  $B_2$  of density  $\beta_2 \geq \beta_1^8 \geq \epsilon^{80}$  in which every  $2r$ -cycle in  $B_2$  has at least  $\beta_1^{8k^2} n^{2r} \geq \epsilon^{80k^2} n^{2r}$  many  $\theta$ -popular ring decompositions in  $A$ .

Since  $B_2$  is a subset of  $B_1$ , the rectangle completion operation in  $B_2$  is still  $C$ -well-defined. By Lemma 2.3 the number of rectangles in  $B_2$  is at least  $\beta_2^4 n^4$ , and since octahedra are counted by pairs of rectangles with the same labelling, the octahedron count is minimized when the the number of rectangles with each labelling is as balanced as possible (by convexity). For each triple of labels  $(a, b, c)$  the number of possible completions  $d$  is at most  $C$ , so the number of octahedra is at least

$$(\beta_2^4 n / C)^2 n^3 = (\beta_2^8 / C^2) n^5 \geq \epsilon^{70k^3} n^5$$

as required. □

We now observe that an octahedron, which consists of two identically labelled rectangles, still corresponds to an octahedron if we permute the coordinates of the points. Indeed, using the hypergraph point of view we can define it more symmetrically as a sequence of eight triples

$$(x_{000}, y_{000}, z_{000}), (x_{001}, y_{001}, z_{001}), \dots, (x_{111}, y_{111}, z_{111})$$

such that  $x_\epsilon = x_\eta$  if  $\epsilon_2 = \eta_2$  and  $\epsilon_3 = \eta_3$ ,  $y_\epsilon = y_\eta$  if  $\epsilon_1 = \eta_1$  and  $\epsilon_3 = \eta_3$ , and  $z_\epsilon = z_\eta$  if  $\epsilon_1 = \eta_1$  and  $\epsilon_2 = \eta_2$ . (Note that this is *not* an octahedron in the usual hypergraph sense. Indeed, it cannot be, since adjacent faces of an octahedron intersect in an edge, so that hypergraph is not linear. Rather, it is the hypergraph formed by the triangular faces of a cuboctahedron. However, as mentioned in Sect. 1.2, this object does correspond to an octahedron when viewed in the framework of van Kampen diagrams, so to keep terminology to a minimum we use the word ‘octahedron’ even when referring to the corresponding hypergraph just described.)

This observation implies that Proposition 3.2 remains true if the word ‘label’ is replaced by either ‘column’ or ‘row’ and we interpret ‘ring decompositions’ in the appropriate way—that is, after permuting the coordinates. (To put it slightly differently, to perform a column ring decomposition, one can interchange the column and label coordinates, perform a label ring decomposition, and interchange the column and label coordinates again). Thus, we will in fact be able to find decompositions of all three kinds of  $2r$ -cycles. This will be crucial for our argument in Sect. 5.

**Theorem 3.3.** Fix  $\epsilon \leq 10^{-3}$  and  $k \geq 100$ . Let  $A$  be a 3-uniform, linear hypergraph that contains at least  $\epsilon n^5$  octahedra. Then there exists a sequence  $A = A_0 \supset A_1 \supset \dots$  such that each  $A_i$  has density at least  $\alpha_i(\epsilon, k)$  and  $A_i$  contains at least  $\epsilon_i(\epsilon, k)n^5$  octahedra, and for each  $r = 2, \dots, k$ , every  $2r$ -cycle in  $A_i$  has at least  $\gamma_i(\epsilon, k)n^{2r}$  different  $\theta_i(\epsilon, k)$ -popular ring decompositions in  $A_{i-1}$ . Each of the parameters  $\alpha_i, \epsilon_i, \theta_i, \gamma_i$  may be chosen to be at least  $\epsilon^{k^{15i}}$ .

*Proof.* Let  $A_0 = A$ . We now repeatedly use Proposition 3.2. Once we have chosen  $A_i$ , we apply Proposition 3.2 to pass to a dense subset  $B_i^{(1)}$  in which all label  $2r$ -cycles have at least  $\gamma_i n^{2r}$  different  $\theta_i$ -popular ring decompositions in  $A_i^{(1)}$  for  $2 \leq r \leq k$ . We then apply the proposition to pass to a dense subset  $B_i^{(2)}$  of  $B_i^{(1)}$  in which all column  $2r$ -cycles have at least  $\gamma'_i n^{2r}$  different  $\theta'_i$ -popular ring decompositions in  $B_i^{(1)}$  for  $2 \leq r \leq k$ . Finally, we apply the proposition to pass to a dense subset  $B_i^{(3)}$  of  $B_i^{(2)}$  such that all row  $2r$ -cycles have at least  $\gamma''_i n^{2r}$  different  $\theta''_i$ -popular ring decompositions in  $B_i^{(2)}$  for  $2 \leq r \leq k$ . The dependence between these parameters, which we shall discuss in more detail in a moment, is given by Proposition 3.2. We now set  $A_{i+1}$  to be  $B_i^{(3)}$ .

If the density of  $A_i$  is  $\alpha_i$  and the number of octahedra in  $A_i$  is  $\epsilon_i n^5$ , then the density of  $B_i^{(1)}$  is at least  $\epsilon_i^{80}$ . Moreover, the octahedron count of  $B_i^{(1)}$  is at least  $\epsilon_i^{70k^3} n^5$ . Therefore, the density of  $B_i^{(2)}$  is at least

$$(\epsilon_i^{70k^3})^{80} \geq \epsilon_i^{2^{13}k^3}$$

and the octahedron count of  $B_i^{(2)}$  is at least

$$(\epsilon_i^{70k^3})^{70k^3} \geq \epsilon_i^{2^{13}k^6}.$$

This implies that the density of  $B_i^{(3)}$  is at least

$$(\epsilon_i^{2^{13}k^6})^{80} \geq \epsilon_i^{2^{20}k^6} \geq \epsilon_i^{k^{15}}$$

and the octahedron count is at least

$$(\epsilon_i^{2^{13}k^6})^{70k^3} n^5 \geq \epsilon_i^{2^{20}k^9} n^5 \geq \epsilon_i^{k^{15}} n^5.$$

Lastly, we also have

$$\theta''_i \geq (\epsilon_i^{2^{13}k^6})^{35k^3} \geq \epsilon_i^{2^{19}k^9} \geq \epsilon_i^{k^{15}}$$

and

$$\gamma''_i \geq (\epsilon_i^{2^{13}k^6})^{80k^2} \geq \epsilon_i^{2^{20}k^8} \geq \epsilon_i^{k^{15}}.$$

Note also that  $\gamma_i, \gamma'_i \geq \gamma''_i$  and  $\theta_i, \theta'_i \geq \theta''_i$  since the octahedron counts of  $B_i^{(1)}$  and  $B_i^{(2)}$  are larger than that of  $B_i^{(3)}$ .

Therefore,  $A_{i+1}$  is still dense, and has the property that any  $2r$ -cycle (for  $2 \leq r \leq k$ ) in  $A_{i+1}$  is popularly decomposable in  $A_i$ .

After each step of the inductive construction, the density  $\alpha_{i+1}$  is at least  $\epsilon_i^{k^{15}}$  and the octahedron count  $\epsilon_{i+1}n^5$  is at least  $\epsilon_i^{k^{15}}n^5$ . The threshold for popularity  $\theta_{i+1}$  is at least  $\epsilon_i^{k^{15}}$ , and  $\gamma_{i+1} \geq \epsilon_i^{k^{15}}$  also.

Therefore, starting at  $A_0 = A$  with  $\epsilon n^5$  octahedra, we find that for  $i \geq 1$  we have

$$\epsilon_i \geq \epsilon^{k^{15i}}.$$

This gives us

$$\alpha_i \geq (\epsilon^{k^{15(i-1)}})^{k^{15}} = \epsilon^{k^{15i}}$$

and similarly  $\theta_i \geq \epsilon^{k^{15i}}$  and  $\gamma_i \geq \epsilon^{k^{15i}}$ .

Therefore every  $2r$ -cycle in  $A_i$  is  $\epsilon^{k^{15i}}$ -popularly decomposable in  $A_{i-1}$  in at least  $\epsilon^{k^{15i}}n^{2r}$  different ways. □

To close this section, we observe that saying that a  $2r$ -cycle is popularly decomposable in many different ways is equivalent to saying that the  $2r$ -cycle has many decompositions of another kind, which we now define.

**DEFINITION 3.4.** *Let  $C$  be a  $2r$ -cycle  $x_1y_1 \dots x_r y_r$  with  $x_1$  and  $y_1$  sharing a row. A dispersed ring decomposition of  $C$  consists of a  $2r$ -cycle  $x'_1y'_1 \dots x'_r y'_r$  with  $x'_1$  and  $y'_1$  sharing a column, together with rectangles  $R_i = x''_i u_i x'''_i v_i$  and  $S_i = y''_i w_i y'''_i z_i$  (where  $u_i$  shares a row with  $x''_i$  and  $w_i$  shares a column with  $y''_i$ ) such that for each  $i$ ,  $x_i$  and  $x''_i$  have the same label,  $x'_i$  and  $x'''_i$  have the same label,  $y_i$  and  $y''_i$  have the same label,  $y'_i$  and  $y'''_i$  have the same label,  $u_i$  and  $z_i$  have the same label, and  $w_i$  and  $v_{i+1}$  have the same label.*

The reason for this terminology is as follows. Consider a ring decomposition of a label  $2r$ -cycle. It consists of another label  $2r$ -cycle and a collection of label 4-cycles. To obtain a dispersed ring decomposition, we take the various cycles that form the decomposition and ‘disperse’ them by replacing them by other cycles of the same length that have the same label sequences. The conditions above are precisely the ones that are guaranteed to hold for the various cycles after we have done this: a point in one cycle has to have the same label as a point in another cycle if before the ‘dispersing’ they were the same point. Note that to say that a ring decomposition is popular is precisely to say that one can obtain many dispersed ring decompositions from it in this way.

The equivalence mentioned above is given more precisely by the following simple lemma.

**LEMMA 3.5.** *Let  $(X, Y, Z, A, \phi)$  be a partial Latin square with  $|X| = |Y| = |Z| = n$ . Let  $F$  be a  $2r$ -cycle which is  $\theta$ -popularly decomposable in  $A$  in at least  $\gamma n^{2r}$  different ways. Then  $F$  has at least  $\gamma \theta^{2r+1} n^{4r+1}$  different dispersed ring decompositions.*

*Proof.* Without loss of generality  $F$  is a label  $2r$ -cycle. There are at least  $\gamma n^{2r}$  different ring decompositions of  $F$  into cycles that are  $\theta$ -popular. Each of these popular cycles can be replaced with one of  $\theta n$  different cycles with the same label sequence as the original, giving a total of  $(\theta n)^{2r+1}$  further choices, from which the result follows.  $\square$

In the next section we shall see how to describe a dispersed ring decomposition of a  $2r$ -cycle as a triangulated surface of a certain shape that has that  $2r$ -cycle as its boundary, which should make the concept significantly clearer.

## 4 The van Kampen picture

In this section we discuss in more detail the role that van Kampen diagrams play in our proof. Though the material is well known, we include it here partly because our treatment of it is not standard in all respects, and partly for the convenience of readers who may not be familiar with it.

Let  $(X, Y, Z, A, \phi)$  be a partial Latin square. As we have seen, we can think of it as a linear tripartite 3-uniform hypergraph  $H$ . However, it will be very helpful to represent  $H$  in an unusual way as follows.

**DEFINITION 4.1.** *Let  $H$  be a linear tripartite 3-uniform hypergraph with vertex sets  $X, Y$  and  $Z$ . The van Kampen complex  $K(H)$  corresponding to  $H$  is a directed simplicial complex with three vertices  $u, v, w$ , a directed 1-cell joining  $u$  to  $v$  for each  $x \in X$ , a directed 1-cell joining  $v$  to  $w$  for each  $y \in Y$ , a directed 1-cell joining  $u$  to  $w$  for each  $z \in Z$ , and a 2-cell bounded by the 1-cells corresponding to  $x \in X$ ,  $y \in Y$  and  $z \in Z$  if and only if  $xyz$  is a face of  $H$ .*

The directions are chosen to conform with the definition of a van Kampen diagram. A face  $xyz$  of  $H$  corresponds to a point  $(x, y, z)$  in the partial Latin square, which, when thought of as a multiplication table, corresponds to the relation  $xy = z$ . We therefore want the 2-cell corresponding to  $xyz$  to have boundary word  $xyz^{-1}$ .

To put the above definition more loosely,  $K(H)$  is simply the result of replacing the vertex sets of  $H$  by sets of edges joining  $u$  to  $v$ ,  $v$  to  $w$ , and  $u$  to  $w$ , so for each face of  $H$ , its vertices become edges.

Several of the concepts that we have defined up to now become more natural and geometrical when reinterpreted in terms of triangulated surfaces that may or may not live in  $K(H)$ . The vertices  $u, v$  and  $w$  in  $K(H)$  do not play an important role, and in order to represent these reinterpretations pictorially it is convenient to ‘de-identify’ them. For example, to obtain the surface corresponding to what we have called an octahedron, we take two edges between  $u$  and  $v$ , two between  $v$  and  $w$ , and two between  $u$  and  $w$ , and we fill all of the eight resulting triangles with faces, which results not in an octahedron, but in an octahedron with opposite vertices identified. However, there is no harm in drawing it without the identifications and remembering that strictly speaking the identifications are needed if we want to talk

about copies of the octahedron inside  $K(H)$ . The next definitions allow us to do this precisely.

**DEFINITION 4.2.** *An  $n_1 \times n_2 \times n_3$  tripartite simplicial complex is a directed 2-dimensional simplicial complex with three vertex classes  $U, V$  and  $W$ , and  $n_1$  edges directed from  $U$  to  $V$ ,  $n_2$  edges directed from  $V$  to  $W$ , and  $n_3$  edges directed from  $U$  to  $W$ . (Multiple edges are allowed). A homomorphism from a tripartite simplicial complex  $K_1$  to a tripartite simplicial complex  $K_2$  is a map  $\phi$  that takes the vertices, edges and faces of  $K_1$  to the vertices, edges and faces of  $K_2$  and respects incidences, in the sense that if  $u, v$  are the start and end points of an edge  $e$  in  $K_1$ , then  $\phi(u)$  and  $\phi(v)$  are the start and end points of  $\phi(e)$ , and if the edges  $e_1, e_2, e_3$  bound a face  $f$  in  $K_1$ , then  $\phi(e_1), \phi(e_2)$  and  $\phi(e_3)$  bound the face  $\phi(f)$  in  $K_2$ . If in addition  $\phi$  is a bijection on the edges and faces (but not necessarily the vertices), then it is an isomorphism.*

*A tripartite surface is a tripartite simplicial complex such that each edge is contained in at most two faces. The boundary of a tripartite surface is the set of edges that are contained in exactly one face. A copy of a surface  $S$  in a tripartite simplicial complex  $K$  is the image of a homomorphism  $\phi : S \rightarrow K$ . We call the surface a disc if it is homeomorphic to a disc, and a sphere if it is homeomorphic to the 2-sphere. The area of a surface is the number of faces it contains, and the length of the boundary of a surface is the number of edges it contains.*

It may seem strange not to insist that a copy of a surface is *isomorphic* to the surface itself, but for the purposes of counting it is convenient to allow a small proportion of the copies to be degenerate.

Note that the van Kampen complex of a linear tripartite 3-uniform hypergraph is a tripartite simplicial complex with the additional properties that it has just one vertex in each vertex class, and that no two edges are contained in more than one face. We call such a complex *linear* as well. As the next (almost trivial) proposition shows, every linear tripartite simplicial complex arises in this way, so there is a simple equivalence between linear tripartite 3-uniform hypergraphs and linear tripartite simplicial complexes.

**PROPOSITION 4.3.** *Every linear tripartite simplicial complex is the van Kampen complex of a linear tripartite 3-uniform hypergraph.*

*Proof.* Let  $K$  be a linear tripartite simplicial complex with vertices  $u, v$ , and  $w$ , and let  $X$  be the set of edges from  $u$  to  $v$ ,  $Y$  the set of edges from  $v$  to  $w$ , and  $Z$  the set of edges from  $u$  to  $w$ . Let  $H$  be the hypergraph with vertex sets  $X, Y$  and  $Z$ , where  $(x, y, z)$  is a face of  $H$  if and only if the edges  $x, y$  and  $z$  bound a face of  $K$ . Then  $H$  is clearly tripartite. It is also linear, since the condition that  $K$  is linear is precisely the condition that no two vertices of  $H$  are contained in more than one face.  $\square$

The main statement we wish to prove, which for reasons sketched in the introduction (and explained in more detail in Sect. 5.2) will imply our main theorem, is the following.

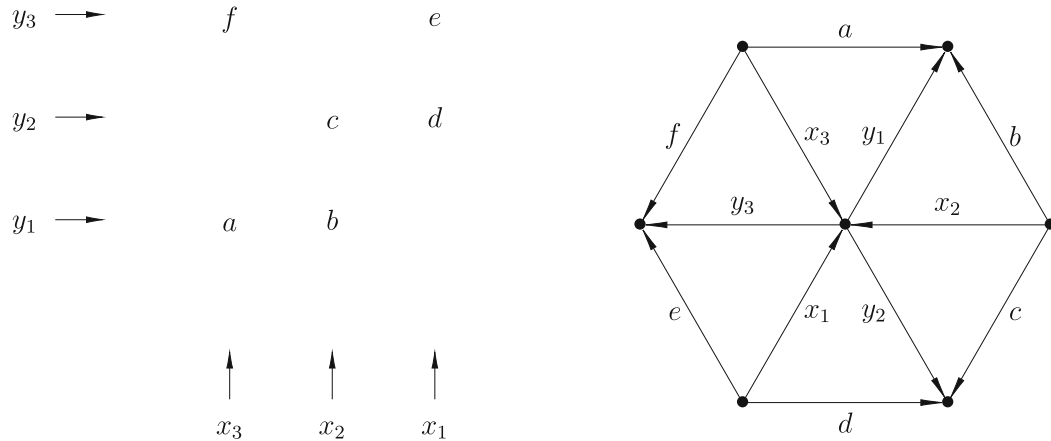


Figure 7: A label 6-cycle in a partial Latin square is shown on the left, with the corresponding rows and columns shown. The corresponding 6-gon (see Definition 4.5) is shown on the right. A copy of this disc would appear in the van Kampen complex of the partial Latin square

**Theorem 4.4.** *Let  $b$  be a positive integer, let  $K$  be linear tripartite simplicial complex  $K$  with  $n$  edges joining each pair of vertices, and suppose that  $K$  contains at least  $\epsilon n^5$  octahedra. Then  $K$  contains a subcomplex  $L$  with at least  $c(\epsilon)n^2$  faces such that  $L$  does not contain a copy of any disc with area less than  $b$  and boundary of length 2.*

One particular kind of surface plays an important role in our arguments.

**DEFINITION 4.5.** *A  $2r$ -gon is the disc with boundary of length  $2r$  formed by  $2r$  triangles that share a central vertex.*

A  $2r$ -gon corresponds to a  $2r$ -cycle in a partial Latin square, as shown in Fig. 7.

We can now give the geometrical interpretation of a dispersed ring decomposition from Definition 3.4 that we promised earlier. A dispersed ring decomposition of a  $2r$ -cycle is obtained by starting with a prism, the two ends of which are  $2r$ -gons. Each face of this prism (that is, two  $2r$ -gons and  $2r$  rectangles) is triangulated by joining a central point to every vertex. That is, the dispersed ring decomposition contains two  $2r$ -gons, one at each end of the prism, with  $2r$  4-gons joining them. One of the  $2r$ -gons corresponds to the  $2r$ -cycle that has been decomposed, and the rest of the prism is a more complicated triangulation of the disc that has the same boundary, which is obtained by taking another  $2r$ -gon and surrounding it with  $2r$  4-gons. Figure 8 below shows this more complicated triangulation in the case  $r = 2$ : that is, it illustrates the disc that comes from a dispersed ring decomposition of a 4-cycle. Note that in this case the prism is a cube.

For convenience, we encapsulate the above remarks in the following definition, which translates the notion of a dispersed ring decomposition (see Definition 3.4) into the language of tripartite surfaces.

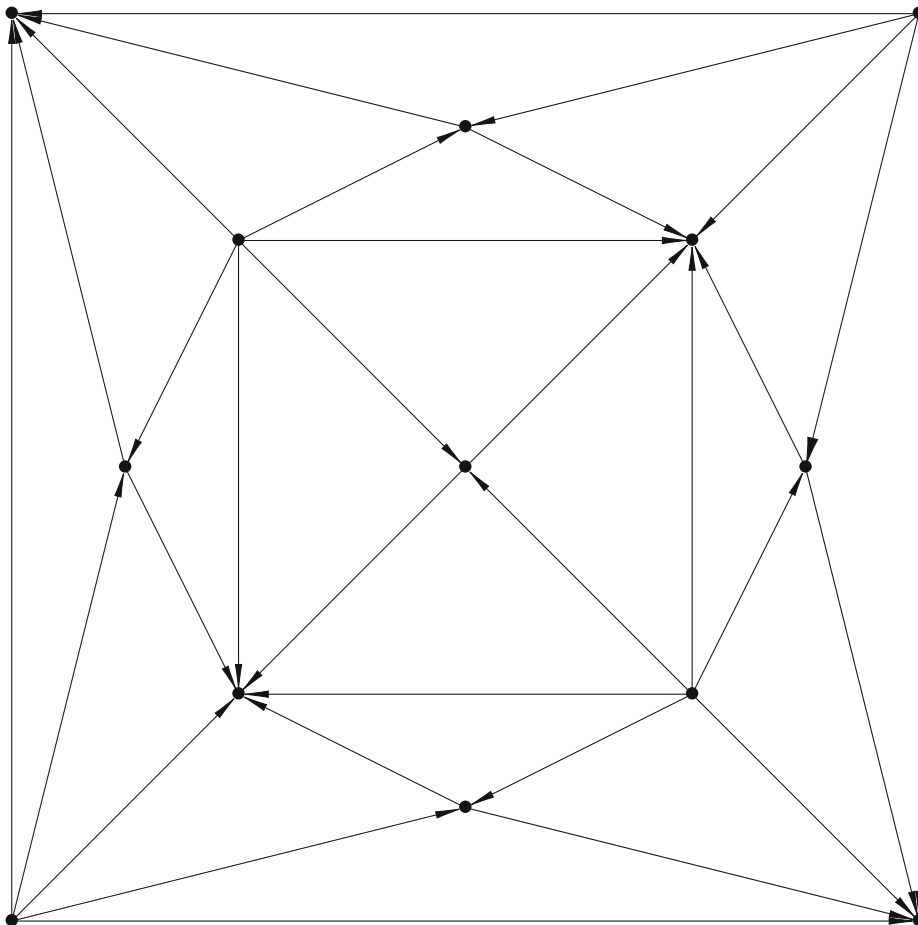


Figure 8: The disc corresponding to the dispersed ring decomposition of a 4-cycle. We have omitted the labels on the edges

DEFINITION 4.6. Given a tripartite simplicial complex  $K$  and a copy  $D$  of a  $2r$ -gon in  $K$ , a dispersed ring decomposition of  $D$  in  $K$  is a copy  $D'$  in  $K$  of the disc obtained by surrounding a  $2r$ -gon with  $2r$  4-gons as described above to give a triangulation of the disc, in which the boundary of  $D'$  is the same as the boundary of  $D$ .

### 5 Popular replacement of discs

We now turn to the proof of Theorem 4.4. Since the details will get somewhat involved, it will be instructive to begin with the case concerning what we call *slit octahedra*. These are tripartite surfaces that are isomorphic to the triangulated disc illustrated in Fig. 1, which can be obtained from an octahedron by cutting along one of its edges and opening up the cut. Since these are precisely the discs that show that the quadrangle condition fails, this special case will prove Theorem 1.11.



The main fact that we shall need to prove is that if we pass to a suitable dense subset, then every copy of a slit octahedron in that subset can be replaced by a certain more complicated triangulated disc with the same boundary in a near-maximal number of ways (where ‘near-maximal’ means within a constant of the trivial maximum). To prove this, we shall start with a fixed copy of the surface, and repeatedly replace  $2r$ -gons by discs with the same boundary that correspond to dispersed ring decompositions (see Definition 4.6). This we shall be able to do in many ways, so we regard it as a kind of ‘unfixing’ process, where little by little we unfix vertices in order to convert the original completely fixed slit octahedron into a variable surface, at each stage ensuring that the number of possibilities for the variable surface is within a constant of the trivial maximum, given the points that are still fixed. This idea will be explained in more detail later in the section.

The structure of the section will be as follows. First, we shall translate the main results of Sect. 3 into the language of linear tripartite simplicial complexes introduced in Sect. 4. Next, we shall give a more detailed overview of the popular replacement argument. This is followed by a brief section containing a technical lemma in which we determine the maximum possible number of copies of a certain tripartite surface that can appear in a linear tripartite simplicial complex. We are then ready to describe the popular replacement argument. We begin with the special case of the slit octahedron, before generalizing the approach to prove Theorem 4.4.

**5.1 The surface picture.** As a first step, we reinterpret Theorem 3.3 as a statement about the van Kampen complex from Definition 4.1. Doing so will allow us to work almost entirely with simplicial complexes in this section of the proof, thereby avoiding the need to think about the same object in two different ways at once.

Recall that the hypothesis of Theorem 3.3 involves a 3-uniform, linear hypergraph  $A$  that contains at least  $\epsilon n^5$  octahedra. The van Kampen complex  $K(A)$  is therefore a linear tripartite simplicial complex (see Definition 4.2) with  $n$  edges joining each pair of vertices and containing at least  $\epsilon n^5$  octahedra, where in the van Kampen case an octahedron is simply a (tripartite) triangulated surface isomorphic to an octahedron, when an octahedron is also considered as a triangulated sphere.

Given this hypothesis, the conclusion of Theorem 3.3 is that there exists a sequence  $A = A_0 \supset A_1 \supset \dots$  such that each  $A_i$  has density at least  $\alpha_i(\epsilon, k)$ , each  $A_i$  contains at least  $\epsilon_i(\epsilon, k)n^5$  octahedra, and for each  $r = 2, \dots, k$ , every  $2r$ -cycle in  $A_i$  has at least  $\gamma_i(\epsilon, k)n^{2r}$  different  $\theta_i(\epsilon, k)$ -popular ring decompositions in  $A_{i-1}$ . Each  $A_i$  has an associated van Kampen complex  $K(A_i)$  which is a subcomplex of  $K(A)$ , and the theorem tells us that  $K(A_i)$  has at least  $\alpha_i(\epsilon, k)n^2$  faces. Moreover,  $K(A_i)$  contains at least  $\epsilon_i(\epsilon, k)n^5$  octahedra.

Recall that a  $2r$ -cycle corresponds in the surface picture to a  $2r$ -gon (see Definition 4.5). It remains to interpret, in the language of tripartite surfaces, the conclusion that for each  $r = 2, \dots, k$ , every  $2r$ -cycle in  $A_i$  has at least  $\gamma_i(\epsilon, k)n^{2r}$  different  $\theta_i(\epsilon, k)$ -popular ring decompositions in  $A_{i-1}$ . By applying Lemma 3.5, it follows that

every  $2r$ -cycle in  $A_i$  has at least  $\gamma_i(\epsilon, k)\theta_i(\epsilon, k)^{2r+1}n^{4r+1}$  different dispersed ring decompositions (see Definition 3.4). At the end of Sect. 4 we discussed the dispersed ring decomposition in terms of tripartite surfaces (see Definition 4.6), so we are now ready to formulate the van Kampen version of Theorem 3.3.

**Theorem 5.1.** *Let  $\epsilon \leq 10^{-3}$  and let  $k \geq 100$ . Let  $K$  be an  $n \times n \times n$  linear tripartite simplicial complex (see Definition 4.2) containing at least  $\epsilon n^5$  octahedra. Then there exists a sequence  $K = K_0 \supset K_1 \supset \dots$  of subcomplexes of  $K$  such that each  $K_i$  has at least  $\alpha_i(\epsilon, k)n^2$  faces, at least  $\epsilon_i(\epsilon, k)n^5$  octahedra, and also has the property that, for each  $r = 2, \dots, k$ , a given  $2r$ -gon in  $K_i$  has at least  $\gamma_i(\epsilon, k)\theta_i(\epsilon, k)^{2r+1}n^{4r+1}$  different dispersed ring decompositions in  $K_i$ .*

**5.2 Overview.** The structure of our proof of Theorem 4.4 will be as follows. We start with a linear tripartite simplicial complex that contains many octahedra. We then apply Theorem 5.1 to create a sequence  $K_0 \supset K_1 \supset \dots$  of tripartite simplicial complexes with the properties stated. We fix some  $s$  and pick some particular small disc  $D_0$  with boundary of length 2 (see Definition 4.2, or for an illustration see Fig. 1) that we wish to get rid of, and consider the following auxiliary graph on the edges of the 1-skeleton of  $K_s$ : we join two edges  $x$  and  $y$  of  $K_s$  by an edge if there is a copy of  $D_0$  in  $K_s$  such that  $x$  and  $y$  are the boundary edges of that copy. If the maximum degree of this auxiliary graph is bounded then we may pass to a dense independent set, and then we are done, since this independent set corresponds to a dense set of faces in  $K_s$  such that no two faces are the boundary faces of a copy of  $D_0$ , and that implies that there is no copy of  $D_0$ . (Of course, we need to repeat this argument for all the discs we are trying to eliminate).

If the maximum degree is *not* bounded, then we would like to find a contradiction. We are given a vertex of large degree in the auxiliary graph, which corresponds to a face of  $K_s$  that is contained in many different copies of  $D_0$ , each with a different ‘opposite face’. Given one of these discs, we perform our unfixing process. Initially, we say that all edges are *fixed*, meaning that we have specified precisely one copy of  $D_0$ . We then find a  $2r$ -gon in this copy and use dispersed ring decompositions guaranteed by Theorem 5.1 to replace it with a new, more complicated disc, which we can do in many different ways. However we do the replacement,  $D_0$  turns into a copy of a larger disc  $D_1$  that still has a boundary of length 2. The copies of  $D_1$  thus obtained lie in  $K_{s-1} \supset K_s$ , and we obtain  $\Omega(n^{4r+1})$  of them, the trivial maximum being  $n^{4r+1}$ . We say that the internal edges in the chosen  $2r$ -gons are *unfixed*, since they may differ from copy to copy. Note that the number of fixed edges has decreased.

We may continue this process, choosing at each step a  $2r$ -gon with some fixed internal edges from  $D_i$  and using dispersed ring decompositions to generate a larger collection of copies of a disc  $D_{i+1}$  that lies in  $K_{s-i-1}$ , with fewer fixed edges. If  $s$  is chosen sufficiently large relative to the area of  $D_0$  then we may proceed until we obtain a collection  $\mathcal{D}$  of copies of some disc  $D_t$  in which the two boundary edges are fixed but every edge incident to an internal vertex is unfixed. One of the boundary edges corresponds to our initial vertex of high degree in the auxiliary graph. By

repeating this process for each choice of neighbour of our chosen vertex from that auxiliary graph, we obtain many different collections of copies of  $D_t$ , which all share one of the two boundary edges. By taking the union of all of these collections, we end up violating the trivial upper bound on the maximum possible number of copies of  $D_t$  in an  $n \times n \times n$  linear tripartite simplicial complex.

The next sections will expand on the details required for this argument. As promised earlier, we shall begin with a detailed account of the argument when  $D_0$  is the ‘slit octahedron’ illustrated in Fig. 1, and then we shall tackle the necessary generalizations. Before we embark on this it will be necessary to work out the trivial maximum for the number of copies of a given disc with a given set of fixed edges in a linear tripartite simplicial complex. The main task of this section will then be to verify that during the unfixing process, the number of copies we obtain is always within a constant of the appropriate trivial maximum, so that in particular this is the case when we reach the tripartite disc  $D_t$  with all non-boundary edges unfixing. This is essential for obtaining our desired contradiction.

**5.3 The maximum number of copies of a partially fixed disc.** We begin with a definition that formalizes the notion of a partially fixed disc, which will be needed to describe the surfaces  $D_i$  that appear part way through the overview described above.

**DEFINITION 5.2.** *Let  $K$  be an  $n \times n \times n$  linear tripartite simplicial complex. We define a partially fixed disc in  $K$  to be a triple  $(D, E, \gamma)$ , where  $D$  is a disc (see Definition 4.2),  $E$  is a subset of the edges of  $D$ , and  $\gamma$  is a homomorphism from  $E$  to the 1-skeleton of  $K$  that respects the tripartition of the vertices of  $D$ . We call the edges in  $E$  fixed and the other edges unfixing. We call a face unfixing if it contains at least one unfixing edge.*

A copy of  $(D, E, \gamma)$  in  $K$  is a copy of  $D$  in  $K$  that extends  $\gamma$  in the obvious sense. Less formally, it is a copy of  $D$  in  $K$  for which the images of the fixed edges have to be given by  $\gamma$ . By the trivial maximum number of copies of a partially fixed disc  $(D, E, \gamma)$  we mean the maximum possible number of copies of a partially fixed disc  $(D, E, \gamma')$  in an  $n \times n \times n$  linear tripartite simplicial complex  $K$ . Since the trivial maximum does not depend on the complex  $K$  or the map  $\gamma$ , we also define an abstract partially fixed disc to be just a pair  $(D, E)$ , where  $D$  and  $E$  are as above. If no confusion is likely to arise, we shall omit the word ‘abstract’. As above, the edges in  $E$  will be called fixed.

Next, we prove an upper bound on the trivial maximum possible number of copies of a partially fixed disc.

**LEMMA 5.3.** *Let  $D$  be an abstract partially fixed disc obtained by triangulating the disc and fixing the boundary edges. Then the trivial maximum number of copies of  $D$  is at most  $n^{V_I}$  where  $V_I$  is the number of internal vertices—that is, vertices that do not lie on the boundary.*

*Proof.* The proof is by induction on the number of faces of  $D$ . The result is trivial when  $D$  is a single face with all three edges fixed. Now suppose that  $D$  has at least two faces. Suppose first that there is a face  $f$  that has two boundary edges. Then the third edge must be internal. The label of this edge is determined by the labels on the two boundary edges. If we remove the face  $f$  and fix its internal edge, then we obtain a disc that still has  $V_I$  internal vertices, and hence at most  $n^{V_I}$  copies, so we are done.

If  $D$  does not have such a face, then we split into two further cases. Suppose first that  $D$  has an internal vertex: that is, a vertex that does not lie on the boundary. Then there must be an internal vertex that is joined by an edge to a boundary vertex  $w$ . The neighbours of  $w$  form a path from its predecessor along the boundary to its successor. Let  $v$  be the first internal vertex along this path. Then  $v$  is joined to  $w$  and to its predecessor, which gives us a face that has one boundary edge and two internal edges. We can choose the label for one of the internal edges in at most  $n$  ways, and that determines the label for the other. Having done so, if we remove the face and fix the two internal edges, we obtain a simply connected disc  $D'$  with one less internal vertex. For each of the at most  $n$  choices of labelling for the newly fixed edges we get at most  $n^{V_I-1}$  copies of  $D'$ , by the inductive hypothesis, so the number of copies of  $D$  is at most  $n^{V_I}$  as required.

The final case is where  $D$  does not have any internal vertices or any faces with two boundary edges. This case cannot in fact occur. Indeed, if it did, then note that the number of vertices would equal the number of boundary edges, and the number of faces would be at most the number of internal edges (since each face would contain at least two internal edges and each internal edge would be contained in two faces). It would follow that  $V - E + F \leq 0$ , contradicting Euler's formula (which would give  $V - E + F = 1$ , since we are not counting the external face as a face).  $\square$

A simple example that is important for us is that of a  $2r$ -gon: if the boundary is fixed, then we are left with at most  $n$  possibilities. In the grid picture, this corresponds to the fact that if we know the labels of a label  $2r$ -cycle (say), then the first point of the cycle (which can be chosen in at most  $n$  ways) determines the rest of the cycle if it exists.

An even more important example is where  $D$  is taken to be the disc corresponding to the dispersed ring decomposition of a  $2r$ -gon as described in Definition 4.6, again with the boundary cycle fixed. This bounds the maximum possible number dispersed ring decompositions of a given  $2r$ -gon. The number of internal vertices is  $4r+1$ , since the opposite  $2r$ -gon contributes  $2r+1$  vertices, and each of the  $2r$  4-gons has a further internal vertex in the middle. Thus, Lemma 5.3 gives an upper bound of  $n^{4r+1}$  for the number of dispersed ring decompositions of a given  $2r$ -gon. But Lemma 3.5 gives us  $\Omega(n^{4r+1})$  such decompositions, so we see again that our machinery from the previous section gives us within a constant factor of the maximum number of such objects.

**5.4 The slit octahedron case.** The aim of this section is to prove that if we are given a linear tripartite simplicial complex  $K$  with  $\epsilon n^5$  octahedra, then we can pass to a dense subset  $L$  of  $K$  in which there are no slit octahedra.

We begin by defining the auxiliary graph described in Sect. 5.2.

**DEFINITION 5.4.** *Given a linear tripartite simplicial complex, we let the auxiliary graph  $G(K)$  have vertex set given by the edges in the 1-skeleton of  $K$ , with edges  $x$  and  $y$  joined if  $x$  and  $y$  are the boundary edges of some copy of the slit octahedron in  $K$ .*

As discussed in Sect. 5.2, if we can prove that this auxiliary graph is of bounded degree, then we will be able to pass to a dense independent subset of the vertices and thereby eliminate all slit octahedra. We shall achieve this by applying Theorem 5.1 to  $K$  and obtaining a subcomplex  $K_s$  for some appropriately chosen  $s$ . The rough idea is that if we fix a vertex  $z$  of the auxiliary graph  $G(K_s)$  (recall that this is an edge of the complex  $K_s$ ), then each edge  $zz_i$  in  $G(K_s)$  gives rise to a large number of copies of a certain tripartite surface (see Definition 4.2) homeomorphic to a disc with boundary of length 2, which we build from the initial slit-octahedron by unfixing all the interior edges using dispersed ring decompositions (see Definition 4.6). If there are too many edges  $zz_i$  in  $G(K_s)$ , this ends up contradicting Lemma 5.3.

We now give the details.

**LEMMA 5.5.** *Let  $K$  be an  $n \times n \times n$  linear tripartite simplicial complex containing at least  $\epsilon n^5$  octahedra. Then there is a subcomplex  $L$  of  $K$  with at least  $\epsilon^{2400} n^2$  faces such that the maximum degree in the graph  $G(L)$  is at most  $\epsilon^{-2450}$ .*

*Proof.* First we apply Theorem 5.1 with  $k = 100$  (we only need  $k \geq 4$ ) to obtain a sequence  $K = K_0 \supset K_1 \supset \dots \supset K_4$  with the property that  $K_i$  has at least  $\alpha_i(\epsilon)n^2$  faces, and for each  $r = 2, \dots, 4$  we have that every  $2r$ -gon in  $K_i$  has at least  $\gamma_i(\epsilon)\theta_i(\epsilon)^{2r+1}n^{4r+1}$  different dispersed ring decompositions in  $K_{i-1}$ . The parameters  $\alpha_i(\epsilon)$ ,  $\theta_i(\epsilon)$  and  $\gamma_i(\epsilon)$  are all at least  $\epsilon^{100^{15i}} \geq \epsilon^{2^{100^i}}$ .

Now suppose that the auxiliary graph  $G(K_4)$  of  $K_4$  has a vertex of degree at least  $M$ . Without loss of generality, let us assume that the vertex class that contains this vertex is  $Z$ . If the vertex is  $z$ , then we can find a set  $\{z_1, \dots, z_M\}$  of distinct vertices in  $Z$  such that for each  $j$  there exists a copy of the slit octahedron in  $K_4$  for which the boundary has edges  $z$  and  $z_j$  (recall from Definition 5.4 that the vertices of  $G(K_4)$  are edges of  $K_4$ ).

Let us now fix  $j$  and let  $D_0$  be the corresponding copy of the slit octahedron, which we shall think of as a partially fixed disc for which every edge is fixed. Let  $\gamma$  be the inclusion map from the abstract slit octahedron to its copy in  $D_0$ . The slit octahedron is illustrated again in Fig. 9, for ease of reference. We now select a 4-gon in  $D_0$  by choosing some internal vertex and taking the four faces that surround it. For instance, we may select the bottom internal vertex, which is incident to the edges labelled  $x_3, y_3, x_4$  and  $y_4$ . This gives us the triangulated 4-gon represented by

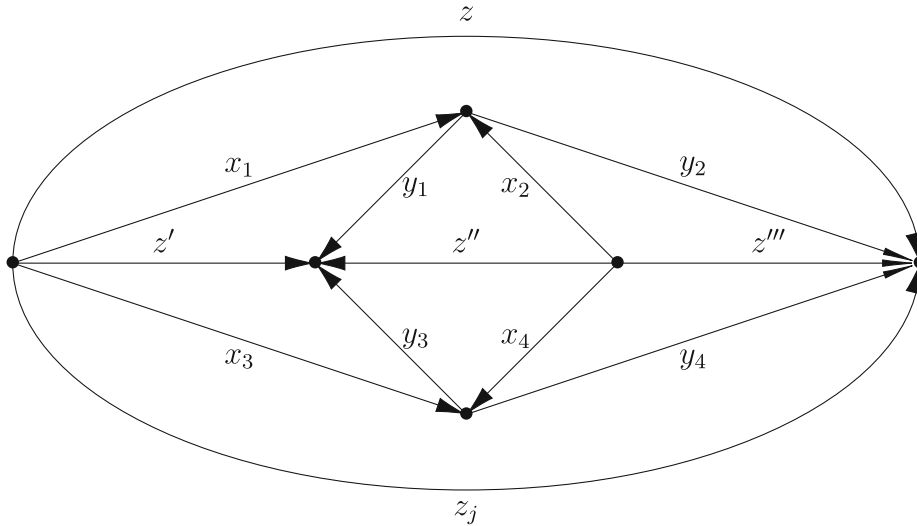


Figure 9: A slit octahedron yielding an edge between  $z$  and  $z_j$  in the auxiliary graph

the four faces in the bottom half of the diagram. We now create a new partially fixed disc  $D_1$  as follows. First we remove this 4-gon from  $D_0$  and replace it by a dispersed ring decomposition of the 4-gon (see Definition 4.6). Then we declare all the internal edges of the dispersed ring decomposition to be unfixed, and the map  $\gamma$  takes the same values as before, but is applied only to the fixed edges. The disc  $D_1$  is illustrated in Fig. 10, with the unfixed edges in red.

Any 4-gon in  $K_4$  has at least  $\gamma_4(\epsilon)\theta_4^5 n^9$  different dispersed ring decompositions in  $K_3$ . Since the trivial maximum number of these dispersed ring decompositions is  $n^9$ , by Lemma 5.3 (because the number of internal vertices is 9), the number of copies of  $D_1$  in the complex  $K_3$  is within a constant of its trivial maximum, as we wanted.

The next step is to select another  $2r$ -gon by choosing another internal vertex, this time of  $D_1$ . We can do this by picking all the faces of  $D_1$  that contain some given internal vertex that is incident to at least one unfixed edge. For instance, we might take the leftmost internal black vertex in Fig. 10.

This gives us a 6-gon  $F$ , since this vertex is contained in six faces of  $D_1$ . Let  $D_2$  be the partially fixed disc obtained by replacing  $F$  with a dispersed ring decomposition and declaring all its internal edges to be unfixed. It is already challenging to draw  $D_2$  in detail, and we shall see shortly that it is not important to track the precise structure of the discs that we obtain at each step. Nevertheless, we include an illustration of  $D_2$  in Fig. 11 to help clarify the process.

Since any given 6-gon in  $K_3$  has at least  $\gamma_3\theta_3^7 n^{13}$  different dispersed ring decompositions in  $K_2$ , we may obtain a copy of  $D_2$  in  $K_2$  by taking any one of the  $\gamma_4\theta_4^5 n^9$  copies of  $K_1$  and then replacing the image of the 6-gon  $F$  in that copy by any one

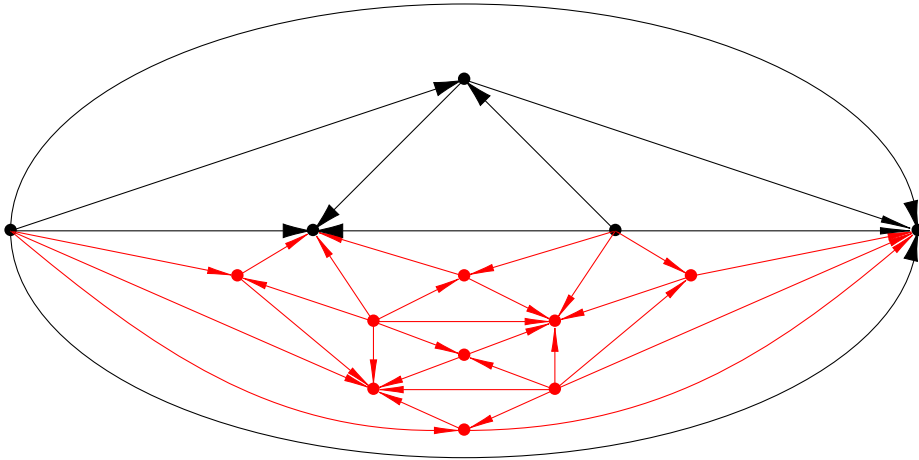


Figure 10: The disc  $D_1$  obtained after the first popular replacement in a slit octahedron. The dispersed ring decomposition is represented with the red part of the diagram. All labels have been omitted for simplicity

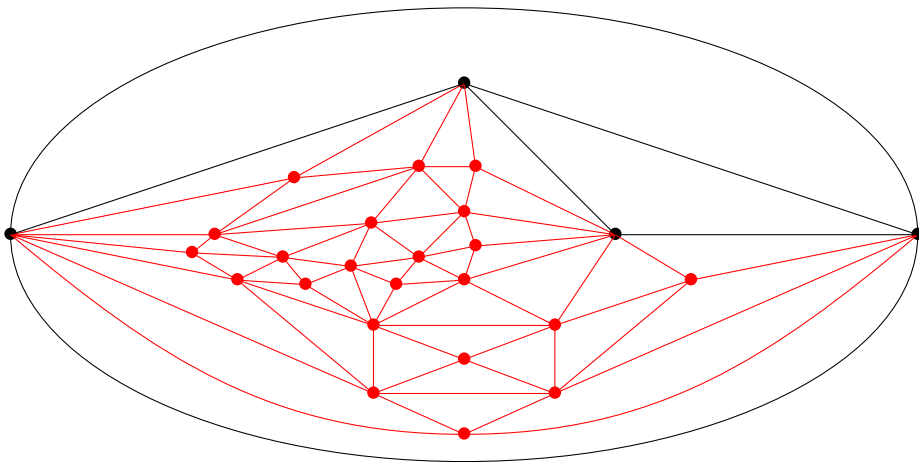


Figure 11: The partially fixed disc  $D_2$  obtained after the second replacement. The fixed part is shown in black, and the unfixed part in red. All labels and directions have been omitted for simplicity

of its  $\gamma_3\theta_3^7n^{13}$  dispersed ring decompositions. We now claim that this gives us

$$(\gamma_4\theta_4^5n^9)(\gamma_3\theta_3^7n^{13}) = \gamma_3\gamma_4\theta_3^7\theta_4^5n^{22}$$

different copies of  $D_2$  in  $K_2$ , but to verify this we must ensure that each copy we have just described is counted at most once.

Suppose that  $D$  and  $D'$  are copies of  $D_1$  in  $K_3$  that, following replacements of their respective copies of  $F$ , both give the same copy of  $D_2$  in  $K_2$ . Then  $D$  and  $D'$  must agree on all but the internal edges of  $F$ . However, we chose  $F$  in such a way that one of the internal edges of  $F$  is fixed and thus shared between  $D$  and  $D'$ . But

since a  $2r$ -gon is fully determined by the boundary edges and a single internal edge (since two edges of a face in the linear simplicial complex uniquely determine the third), we see that  $D = D'$ .

Therefore we do not overcount, and the number of copies of  $D_2$  in  $K_2$  is indeed  $\gamma_3\gamma_4\theta_3^7\theta_4^5n^{22}$ . Again, it is easy to see that this is within a constant of the trivial maximum, since a dispersed ring decomposition of a 6-gon has thirteen internal vertices, so the number of internal red vertices after the second unfixing is 22 (as the sceptical reader can verify from Fig. 11).

The remaining two steps are similar. At the next step, we can replace the 8-gon around the rightmost, internal black vertex in Fig. 11 by its dispersed ring decomposition, with all the internal edges unfixed, to create a partially fixed disc  $D_3$ .

By Lemma 3.5, the number of dispersed ring decompositions in  $A_1$  is at least  $\gamma_2\theta_2^9n^{17}$ , so that the number of copies of  $D_3$  in  $K_1$  is at least  $\gamma_2\theta_2^9n^{17}$  times the number of copies of  $D_2$  in  $K_2$ . But we will also have added  $8 + 8 + 1 = 17$  new internal red vertices, so the trivial maximum increases by a factor of  $n^{17}$ . Therefore, the number of copies of  $D_3$  is within a factor  $\gamma_2\gamma_3\gamma_4\theta_2^9\theta_3^7\theta_4^5$  of the maximum possible.

In  $D_3$  there is one remaining internal vertex that is incident to fixed edges. This vertex is the internal vertex of an 8-gon in  $K_3$ , so we may finish by replacing this 8-gon with a dispersed ring decomposition to obtain a partially fixed disc  $D_4$ , for which only the two boundary edges are fixed. As before, Lemma 3.5 gives us at least  $\gamma_1\theta_1^9n^{17}$  dispersed ring decompositions in  $K_0$ , and therefore Lemma 5.3 tells us that the number of copies of  $D_4$  is within the constant factor  $\gamma_1\gamma_2\gamma_3\gamma_4\theta_1^9\theta_2^9\theta_3^7\theta_4^5$  of the trivial maximum.

Drawings of the full structure of  $D_3$  and  $D_4$  would be too complicated to be illuminating, but we include Fig. 12, which gives a global view of the replacement sequence we have performed. In this figure we show  $D_1, D_2, D_3$  and  $D_4$  but instead of drawing all the unfixed edges, we simply indicate where they are with red hatching.

Recall that this entire collection of copies of  $D_4$  in  $K_0$  was obtained by starting with a given slit octahedron, which yielded a disc with boundary edges  $z$  and  $z_j$ . By performing this sequence of popular replacements for each choice of  $j \in \{1, \dots, M\}$  we obtain  $M$  different collections of copies of the same partially fixed disc. Each of these collections has a fixed boundary, but one of the two fixed boundary edges differs from collection to collection. By taking the union over all these collections, we obtain a final collection  $\mathcal{D}$  of copies of  $D_4$  in which only the label on one of the two boundary edges is fixed.

Now we need an upper bound for the maximum number of copies of the partially fixed disc  $D'_4$ , which is the same as  $D_4$  except that only one of the two boundary edges is fixed. We cannot immediately apply Lemma 5.3 since the entire boundary is not fixed. But we can modify  $D'_4$  by attaching one new triangular face onto the unfixed boundary edge and fixing the other two edges of this face. We thus obtain a new partially fixed disc  $D''_4$  with a boundary consisting of three fixed edges, and



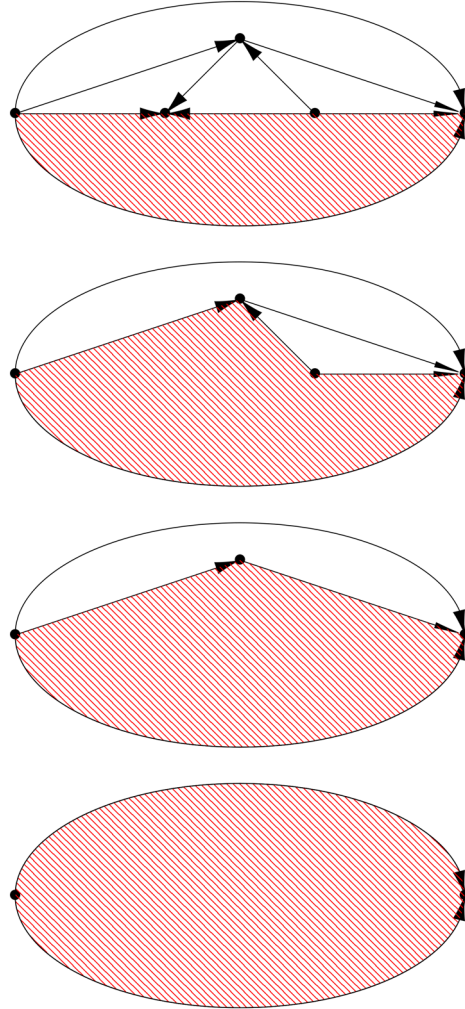


Figure 12: The sequence of four popular replacements from the proof of Lemma 5.5. Starting with a fixed disc corresponding to a slit octahedron, we progressively unfix all but the two boundary edges. Our unfixing process modifies the triangulation, and we represent the modified part with the red hatching (for example, the top figure represents  $D_1$ , shown in full detail in Fig. 10). All edges in the triangulation represented by the red hatching are unfix

every internal edge of  $D_4''$  is unfix. The maximum number of copies of  $D_4''$  is at most the maximum number of copies of  $D_4'$ , since adding extra fixed edges cannot increase the number. We can now apply Lemma 5.3 to  $D_4''$ , which has the same number of internal vertices as  $D_4$ . Therefore the maximum number of copies of  $D_4''$  is the same as that of  $D_4$ , and hence the maximum number of copies of  $D_4'$  is at most that of  $D_4$ .

But the size of the collection  $\mathcal{D}$  is at least  $M$  times the number of copies of  $D_4$  found in the popular replacement process just described, which is within a constant

factor  $\gamma_1\gamma_2\gamma_3\gamma_4\theta_1^9\theta_2^9\theta_3^7\theta_4^5$  of the maximum possible. Therefore if

$$\begin{aligned} M\gamma_1\gamma_2\gamma_3\gamma_4\theta_1^9\theta_2^9\theta_3^7\theta_4^5 \\ \geq M\epsilon^{2^{450}} > 1 \end{aligned}$$

then we have our contradiction. Therefore we may take  $L = K_4$ , which has at least  $\alpha_4 n^2 \geq \epsilon^{2^{400}} n^2$  faces. □

Our ‘removal lemma’ for slit octahedra follows from this lemma.

**Theorem 5.6.** *Let  $K$  be an  $n \times n \times n$  linear tripartite simplicial complex that contains at least  $\epsilon n^5$  octahedra. Then there is a subcomplex  $L$  of  $K$  with at least  $\epsilon^{O(1)} n^2$  faces that contains no slit octahedra.*

*Proof.* We apply Lemma 5.5. This gives us a subcomplex  $L$  of  $K$  with at least  $\epsilon^{2^{400}} n^2$  faces such that the associated graph  $G(K)$  has maximum degree at most  $\epsilon^{-2^{450}}$ .

Let the vertices of  $K$  and  $L$  be  $u, v$  and  $w$ , and let  $X$  be the set of edges from  $u$  to  $v$ , let  $Y$  the set of edges from  $v$  to  $w$ , and let  $Z$  the set of edges from  $u$  to  $w$ . To avoid confusion, it is important to keep in mind that these edges are the *vertex* sets of the auxiliary graph  $G(K)$ .

We now pick a maximal independent subset  $I_X$  of  $X$  in  $G(K)$ . We first pick a vertex  $x \in X$  that (when considered as an edge of  $K$ ) belongs to the largest number of faces of  $K$  and add it to  $I_X$ . Then we discard all vertices in the neighbourhood of  $x$  in the graph  $G(K)$  and repeat, picking at each stage the remaining vertex that belongs to the largest number of faces of  $K$ . Since the maximum degree of  $G(K)$  is at most  $\epsilon^{-2^{450}}$ , we end up picking at least  $\epsilon^{2^{450}} n$  vertices from  $X$ , and these vertices when thought of as edges belong to at least a fraction  $\epsilon^{2^{450}}$  of the faces of  $K$  since at each stage we chose an edge that belonged to an above average number of faces. Let  $K_1$  be the subcomplex of  $K$  induced by  $I_X$ ,  $Y$  and  $Z$ . Then  $K_1$  has at least  $\epsilon^{2^{451}} n^2$  faces, and inside  $K_1$  there is no slit octahedron with its boundary edges belonging to  $X$ .

Since  $G(K_1)$  is a subgraph of  $G(K)$ , it also has maximum degree at most  $\epsilon^{-2^{450}}$ , and we may similarly choose an independent set  $I_Y$  in the graph  $G(K_1)$  of at least  $\epsilon^{2^{450}} n$  vertices from  $Y$ , accounting for at least a fraction  $\epsilon^{2^{450}}$  of the faces of  $K_1$ . This gives us a subcomplex  $K_2$  with at least  $\epsilon^{2^{452}} n^2$  faces with no slit octahedron with its boundary edges belonging to either  $X$  or  $Y$ .

Finally, we choose an independent set  $I_Z$  in the graph  $G(K_2)$  of at least  $\epsilon^{2^{450}} n$  vertices from  $Z$ , accounting for the greatest fraction of faces of  $K_2$ . This gives us a subcomplex  $K_3$  with at least  $\epsilon^{2^{453}} n^2$  faces with no slit octahedra, which we take as our subcomplex  $L$ . □

Theorem 1.11, stated in the introduction, follows immediately from Theorem 5.6.

*Proof of Theorem 1.11.* The result follows by applying Theorem 5.6 to the partial Latin square  $A$  (viewed as a linear hypergraph as explained in Lemma 1.6) and

noting that the resulting subcomplex  $L$  of  $K(A)$  corresponds to a subset  $B$  of the partial Latin square  $A$  of positive density (depending only on  $\epsilon$ ). Moreover, since  $L$  contains no slit octahedra, it follows that  $B$  satisfies the quadrangle condition (Definition 1.10).  $\square$

**5.5 The general cases.** Almost all of the complexity of the general case is contained in the detailed account given for the slit octahedron in the previous section. What remains is to describe how the replacement steps work in general, so that we can see that the argument for the slit octahedron generalizes straightforwardly to arbitrary discs with boundaries of length 2.

The outline of the approach is as above. Given a disc  $D$  with boundary of length 2 and a linear tripartite simplicial complex  $K$ , we shall define the auxiliary graph  $G(K, D)$  on the set of edges in the 1-skeleton of  $K$  by joining  $K$ -edges  $x$  and  $y$  by a  $G(K, D)$ -edge if there is a copy of  $D$  in  $K$  with  $x$  and  $y$  as its two boundary edges.

The main lemma will show that we may pass to a dense subcomplex  $L$  of  $K$  such that the auxiliary graph  $G(L, D)$  has bounded degree for each  $D$  of size below some chosen bound. If this is the case, then the elimination of small discs of boundary length 2 is straightforward—as in the proof of Theorem 5.6, we will simply pass down to independent sets in the graphs  $G(L, D)$  in such a way that we avoid discarding too much of  $L$ .

The proof of the main lemma is similar to that of Lemma 5.5. Given  $M$  fixed copies of the disc  $D$  with boundary edges  $z$  and  $z_j$  (for  $j = 1, \dots, M$ ), we shall unfix the edges by using popular decompositions of constituent  $2r$ -gons that surround internal vertices. At each stage we have, for each  $j$ , a collection of almost maximal size of copies of a partially fixed disc with boundary edges  $z$  and  $z_j$ . We aim to show that once all edges incident to internal vertices are unfixed, we will have more than the trivial maximum number of copies of a certain partially fixed disc in  $K'$  unless  $M$  is bounded above by some constant that is independent of  $n$  (which will have a power dependence on  $\epsilon$ , with the exponent depending on the number of faces of  $D$ ).

The condition that  $b \geq 100$  in the next lemma is purely for convenience, as when  $b \geq 100$  it makes certain calculations easier. Of course, the result for  $b \geq 100$  implies the result for  $b < 100$ .

**LEMMA 5.7.** *Let  $K$  be an  $n \times n \times n$  linear tripartite simplicial complex containing at least  $\epsilon n^5$  octahedra, and  $b \geq 100$ . Then we can pass to a subcomplex  $L$  of  $K$  with at least  $\epsilon^{b^{20b}} n^2$  faces such that for each disc  $D$  with at most  $b$  faces and a boundary of length 2, the maximum degree in the graph  $G(L, D)$  is at most  $\epsilon^{-b^{20b}}$ .*

*Proof.* We begin the proof, as we began the proof of Lemma 5.5, by applying Theorem 5.1, which we do with  $k = 2b$ . We obtain a sequence  $K = K_0 \supset K_1 \supset \dots$  with the property that  $K_i$  has at least  $\alpha_i(\epsilon, 2b)n^2$  faces and for each  $r = 2, \dots, k$  we have that every  $2r$ -gon in  $A_i$  has at least  $\gamma_i(\epsilon, 2b)\theta_i(\epsilon, 2b)^{2r+1}n^{4r+1}$  different dispersed ring decompositions, where each of  $\alpha_i, \gamma_i$  and  $\theta_i$  are at least  $\epsilon^{(2b)^{15i}} \geq \epsilon^{b^{20i}}$ . Our tripartite complex  $L$  will be  $K_b$ , which has at least  $\epsilon^{b^{20b}} n^2$  faces. Note that  $L$  has been chosen

independently of any particular disc  $D$ : it will in fact serve for all discs with boundary of length 2 and at most  $b$  faces. Note that the number of internal vertices of any such disc is at most  $3b/4 < b$ , since each internal vertex is contained in at least four faces and each face contains at most three internal vertices.

Now let  $D$  be any disc with at most  $b$  faces and with all its edges fixed. Our goal is to unfix all edges except the boundary edges. As before, our unfixing steps involve picking vertices from the diagram, removing all of their incident faces and re-triangulating the resulting  $2r$ -gon using the dispersed ring decomposition, taking all internal edges of this dispersed ring decomposition to be unfixing. Starting with  $D = D_0$ , this process will lead us to construct a sequence  $D = D_0, D_1, D_2, \dots$  of partially fixed discs and associated collections  $\mathcal{D}_i$  of copies of these discs, where the copies in the family  $\mathcal{D}_i$  live in the complex  $K_{s-i}$ .

In the previous section, we performed the replacements one by one and ensured at each stage that the size of  $\mathcal{D}_i$  was within a constant of the maximum possible. For the general case, it will be simplest to perform the latter check at the end, once all replacements have been made and we have reached a partially fixed disc  $D_s$  in which all edges incident to internal vertices are unfixing.

At each stage, we pick any vertex  $v$  inside  $D_i$  (not on the boundary) such that  $v$  is incident to fixed edges. We then consider the faces containing  $v$ —there are  $2r_i$  of them giving a  $2r_i$ -gon. We replace this  $2r_i$ -gon with a dispersed ring decomposition with unfixing internal edges, giving us  $D_{i+1}$ . As before, the collection of copies  $\mathcal{D}_{i+1}$  is obtained from  $\mathcal{D}_i$  by choosing each possible replacement for each member of  $\mathcal{D}_i$ . As in the slit octahedron case, we will have that the size of  $\mathcal{D}_{i+1}$  is equal to at least the size of  $\mathcal{D}_i$  times the minimum number of different dispersed ring decompositions of the  $2r_i$ -gon in the complex  $K_{s-i-1}$ . We do not overcount, since if two copies of  $D_i$  agree on all edges apart from those incident to  $v$  then, since  $v$  is also incident to a fixed edge, they must agree everywhere.

At each stage we reduce the number of internal vertices incident to fixed edges by exactly one, so the number of unfixing steps that we need to perform is equal to the number of internal vertices of the disc  $D$ , which is at most  $3b/4$ . Moreover, the maximum degree of a vertex in  $D$  is bounded above by  $b$  and this increases by at most two with each popular replacement. Thus, the maximum value of  $r$  for which we ever perform a popular replacement of a  $2r$ -gon is bounded above by  $(b + 2(3b/4))/2 \leq 2b = k$ .

We now consider the disc  $D_s$  that we get at the end of this process. Each time we do a popular replacement of a  $4r_i$ -gon, we increase the size of the family by a factor  $\gamma_{k+1-i} \theta_{k+1-i}^{2r_i-1} n^{4r_i+1}$ . So at the end of the process, the size of the collection  $\mathcal{D}_s$  is at least

$$\gamma_b^b \theta_b^{4b^2} \prod_{i=1}^s n^{4r_i+1} \geq \epsilon^{b^{20b}} \prod_{i=1}^s n^{4r_i+1}.$$

The number of internal vertices of  $D_s$  is  $\sum_{i=1}^s (4r_i + 1)$ , since at each step of the unfixing process we replace one internal vertex by the  $4r_i + 1$  internal vertices of a

dispersed ring decomposition. So, by Lemma 5.3, the maximum possible size of a collection of copies of  $D_s$  that agree on the boundary edges is  $\prod_{i=1}^s n^{4r_i+1}$ .

Therefore  $|\mathcal{D}_s|$  is within a constant factor of the maximum possible. Indeed the constant factor  $\eta$  is bounded by

$$\eta \geq \epsilon^{b^{20b}}.$$

As before, we may perform the same unfixing process (in the same order) for each vertex  $z_i$  ( $i = 1, \dots, M$ ). Each one gives us a collection of discs with fixed boundary edges. The union of these collections is  $\mathcal{D}$ , a collection of copies of the partially fixed disc  $D'$  obtained by unfixing the appropriate boundary edge of  $D_s$ . By the same trick as in the previous section, we can apply Lemma 5.3 to deduce that the maximum possible number of copies of  $D'$  is in fact the same as the maximum possible number of copies of  $D_s$ , and therefore we obtain a contradiction if  $M\eta > 1$ . Therefore  $M \leq \epsilon^{-b^{20b}}$ , which proves the lemma.  $\square$

We are finally ready to prove Theorem 4.4, which we restate here with explicit bounds.

**Theorem 5.8.** *Let  $b$  be a positive integer, let  $K$  be an  $n \times n \times n$  linear tripartite simplicial complex, and suppose that  $K$  contains at least  $\epsilon n^5$  octahedra. Then  $K$  contains a subcomplex  $L$  with at least  $\epsilon^{b^{25b}} n^2$  faces such that  $L$  does not contain a copy of any disc with area less than  $b$  and boundary of length 2.*

*Proof.* Let  $u, v$  and  $w$  be the vertices of  $K$ , let  $X_1$  be the set of edges from  $u$  to  $v$ , let  $X_2$  be the set of edges from  $v$  to  $w$ , and let  $X_3$  be the set of edges from  $u$  to  $w$ . We apply Lemma 5.7 to obtain a subcomplex  $L$  of  $K$  such that the graph  $G(L, D)$  has maximum degree at most  $\epsilon^{-b^{20b}}$  for any tripartite disc  $D$  homeomorphic to a disc with fewer than  $b$  faces and with boundary of length 2. The goal is now to pass to subsets  $X'_1 \subset X$ ,  $X'_2 \subset X_2$ , and  $X'_3 \subset X_3$  such that for  $i = 1, 2, 3$ ,  $G_i(L_X, D)$  contains no edges for any choice of  $D$  as above, where  $L_X$  is the subcomplex of  $L$  induced by the edges in  $X_1, X_2$  and  $X_3$  and  $G_i(L_X, D)$  is the auxiliary graph with respect to  $L_X$  and  $D$  with vertex set  $X_i$ .

In order to do this, we introduce the graph  $G(L, b)$  which is the union of all graphs  $G_i(L, D)$  where  $D$  is as above and  $i = 1, 2$  or  $3$ . Since a tripartite disc with boundary of length 2 has  $3b/2 + 1$  vertices, the number of different  $D$  is at most  $(3b/2 + 1)^{b+1}$ , so  $G(L, b)$  has maximum degree at most  $(3b/2 + 1)^{b+1} \epsilon^{-b^{20b}}$ .

Now, as in the proof of Theorem 5.6, we select our subsets  $X_i$  by passing to independent sets in the  $G(L, b)$  in such a way that the number of faces in the induced subcomplex  $L_V$  is maximized. Doing this gives us a subcomplex  $L$  which is guaranteed to have at least

$$\left( (3b/2 + 1)^{-(b+1)} \epsilon^{b^{20b}} \right)^3 n^2 \geq \epsilon^{b^{25b}} n^2$$

faces, and which contains no copy of any disc with area less than  $b$  and boundary of length 2.  $\square$

REMARK 5.9. Of course, Theorem 5.8 implies a version of Theorem 5.6, although the bound is somewhat worse because Theorem 5.8 uses crude estimates for the number of replacements required (whereas in the proof of Theorem 5.6 we determine an exact sequence of four replacements for the slit octahedron, and determine each  $r_i$  required).

## 6 Obtaining the approximate isomorphism

We sketched the rest of the proof of our main theorem in Sect. 1.2. In this short section we give a detailed proof. The previous sketch is probably sufficient for a reader with a background in geometric group theory, so this section is mainly for the benefit of combinatorialists who may not have such a background.

Our aim in this section is to prove the following statement, the hypotheses of which come from the conclusion of Theorem 5.8.

PROPOSITION 6.1. *Let  $b$  be a positive integer, let  $K$  be the van Kampen complex of a partial Latin square  $(X, Y, Z, A, \lambda)$  and suppose that  $K$  does not contain a disc with area less than  $b$  and boundary of length 2. Then there is a metric group  $G$  and maps  $\phi : X \rightarrow G$ ,  $\psi : Y \rightarrow G$  and  $\omega : Z \rightarrow G$  such that the images  $\phi(X)$ ,  $\psi(Y)$  and  $\omega(Z)$  are 1-separated sets, and  $d(\phi(x)\psi(y), \omega(z)) \leq b^{-1}$  whenever  $(x, y) \in A$  and  $\lambda(x, y) = z$ .*

More informally, the conclusion of the proposition is saying (for large  $b$ ) that the partial Latin square is approximately isomorphic to part of the multiplication table of a metric group.

The definition of the metric group is given by a simple universal construction, as we said in Sect. 1.2. Assume, as we clearly may, that  $X, Y$  and  $Z$  are disjoint sets. Then  $G$  is simply the free group generated by  $X \cup Y \cup Z$ . As for the metric on  $G$ , it is the largest metric that is compatible with the ‘approximate relations’  $d(xy, z) \leq b^{-1}$ , where we have such a relation for every triple  $(x, y, z)$  with  $(x, y) \in A$  and  $\lambda(x, y) = z$  (or equivalently for every face of the associated linear tripartite 3-uniform hypergraph). Recall that we allow infinite distances.

However, if we want to prove Proposition 6.1, it is simpler to use the more explicit description of this metric that we also mentioned in Sect. 1.2. For the main result, we do not need to know that the two metrics coincide, but we shall briefly indicate the (standard) proof once Proposition 6.1 is established.

DEFINITION 6.2. *Let  $v_1$  and  $v_2$  be two elements of the free group on  $X \cup Y \cup Z$ , let  $w_1$  and  $w_2$  be two words representing  $v_1$  and  $v_2$  and let  $R$  be a set of relations of the form  $xy = z$ . The van Kampen distance  $d(v_1, v_2)$  is the smallest area of a van Kampen diagram with relations from  $R$  and boundary word  $w_1w_2^{-1}$ . If no such diagram exists, then we set  $d(v_1, v_2)$  to be infinite.*

Of course, the above definition can be generalized to any set of relations on any free group, but we content ourselves with the case that concerns us in this paper.

It is not hard to check that the distance above is well-defined. To prove this, it is enough to show that if we insert an inverse pair into either  $w_1$  or  $w_2$ , the smallest area of a van Kampen diagram with boundary word  $w_1w_2^{-1}$  does not change. This is obvious, provided one allows van Kampen diagrams to contain degenerate parts that consist of an edge traversed forwards and then immediately backwards (appropriately directed, and with a label  $u$  that counts as  $u$  in the forwards direction and  $u^{-1}$  in the backwards direction).

A similar argument shows that it is translation invariant. To see, for example, that it is left-translation invariant, note that any van Kampen diagram with boundary word  $w_1w_2^{-1}$  can be converted into a van Kampen diagram with boundary word  $uw_1w_2^{-1}u^{-1}$  by adding an edge labelled  $u$  and directed into the vertex at the beginning of  $w_1$  and the end of  $w_2$ , and conversely, given any van Kampen diagram with that boundary word, we can simply remove the edge labelled  $u$  and obtain a van Kampen diagram for  $w_1w_2^{-1}$ .

The fact that it satisfies the triangle inequality is again straightforward. Given van Kampen diagrams with boundary words  $w_1w_2^{-1}$  and  $w_2w_3^{-1}$ , one can create a van Kampen diagram with boundary word  $w_1w_3^{-1}$  by gluing them together along their common section of boundary  $w_2$ . The area of this new word is the sum of the areas of the first two words, so we are done.

Symmetry is obvious. To see that  $d(v_1, v_2) = 0$  only if  $v_1 = v_2$ , observe that a van Kampen diagram of area zero is simply a labelled tree. The boundary word is obtained by following a path round the tree in the standard way, and a simple induction, removing one isolated vertex at a time, then shows that this word reduces to the empty word when one cancels inverse pairs.

*Proof of Proposition 6.1.* Let  $G$  be the free group on  $X \cup Y \cup Z$  with  $b^{-1}$  times the van Kampen metric. If  $(x, y) \in A$  and  $\lambda(x, y) = z$ , then the triangle with directed edges  $ab$  labelled  $x$ ,  $bc$  labelled  $y$ , and  $ac$  labelled  $z$  is a van Kampen diagram of area 1 with boundary word  $xyz^{-1}$ , so  $d(xy, z) \leq b^{-1}$ .

If  $u$  and  $v$  are distinct elements of  $X \cup Y \cup Z$ , then by hypothesis the smallest van Kampen diagram with boundary word  $uv^{-1}$  has area at least  $b$ , so  $d(u, v) \geq 1$ . It follows that the images of  $X, Y$  and  $Z$  are 1-separated, as claimed.  $\square$

REMARK 6.3. It is also straightforward to show that  $d(u, v) = \infty$  unless  $u$  and  $v$  belong to the same one of the sets  $X, Y$  and  $Z$ . Indeed, since every relation derived from the partial Latin square is of the form  $x_iy_j = z_k$ , there is a homomorphism from the group with generators  $X \cup Y \cup Z$  and all those relations to the group with presentation  $\langle x, y, z | xy = z \rangle$ , which is isomorphic to the free group generated by any two of  $x, y$  and  $z$ . Therefore, it is not possible to prove using the relations that two generators from different classes are equal. Since a van Kampen diagram with boundary word  $uv^{-1}$  does prove that  $x = y$ , the claim follows.

REMARK 6.4. The fact that ( $b^{-1}$  times) the van Kampen metric really is the largest that is compatible with the ‘approximate relations’  $d(x_i y_j, z_k) \leq b^{-1}$  comes from an inductive argument. Suppose that  $D$  is another metric compatible with these approximate relations. We need to prove that if  $w_1, w_2$  are two words and there is a van Kampen diagram of area  $s$  and boundary word  $w_1 w_2^{-1}$ , then  $D(w_1, w_2) \leq b^{-1} s$ . If  $s = 1$ , this is trivial. For larger  $s$ , we can split the van Kampen diagram into two parts, both of positive area, as follows.

Without loss of generality at least one edge  $ab$  on the part of the boundary corresponding to  $w_1$  bounds a face  $F$ . Let  $w_3$  be the word obtained by following the boundary until it gets to  $a$ , then going round  $F$  the other way to  $b$ , and proceeding to the end of  $w_1$ . Then  $d(w_1, w_3) = 1$  and  $d(w_3, w_2) = s - 1$ . By induction it follows that  $D(w_1, w_3) \leq b^{-1}$  and  $D(w_3, w_2) \leq b^{-1}(s - 1)$ , and then by the triangle inequality we deduce that  $D(w_1, w_2) \leq b^{-1} s$ .

REMARK 6.5. Another way of thinking about the van Kampen distance, and one of the main reasons van Kampen diagrams were defined, is that it corresponds to the length of the shortest proof that two words are equal given a certain set of relations. Roughly speaking, the length of the proof is defined to be the number of times the relations need to be used in order to transform one word into another. (Inserting or cancelling inverse pairs is not regarded as contributing to the proof length). Thus, we can think of Theorem 5.8 as saying that a partial Latin square with many octahedra contains a dense partial Latin square for which there is no short proof that it is not isomorphic to part of the multiplication table of a group.

## 7 Concluding remarks

It is important to stress that although algebraically the group  $G$  introduced in the proof of Proposition 6.1 is just a free group, the metric gives it a much more interesting structure. Indeed, one can think of this metric as an approximate group presentation: instead of declaring that certain words are equal to the identity, we declare that they are *close* to the identity, and then we take the distance to be the largest one that is compatible with these ‘approximate relations’. (Note that this should be read as ‘approximate group-presentation’ and not ‘approximate-group presentation’).

Theorem 1.9 gives us in particular a metric group  $G$  and three 1-separated subsets  $\phi(X), \psi(Y), \omega(Z)$  of  $G$  of comparable size with the property that for a constant proportion of pairs  $(x, y) \in \phi(X) \times \psi(Y)$  there exists  $z \in \omega(Z)$  such that  $d(xy, z) \leq \delta$ , where  $\delta = b^{-1}$ . If we replace the condition  $d(xy, z) \leq \delta$  by the condition that  $xy = z$ , we obtain a condition that is very closely related to the definition of an approximate group. In particular, we can conclude that there is an approximate group  $H$  of size not much larger than  $|\phi(X)|$  and translates  $xH$  and  $Hy$  of  $H$  such that a constant proportion of the points of  $\phi(X)$  belong to  $xH$  and a constant proportion of the points of  $\psi(Y)$  belong to  $Hy$ . In the first appendix we show that a suitable



‘metric entropy version’ of this result holds, which allows us to replace equality by approximate equality and obtain an appropriate conclusion, where the notion of an approximate group is replaced by that of an approximate group that is also approximate in a metric sense. We call these structures ‘rough approximate groups’. (To the best of our knowledge, this concept was first formulated by Tao [Taoa], and a slight adaptation of it was introduced and studied by Hrushovski [Hru], who called it a metrically approximate subgroup).

It would be very interesting to go further and describe in a more concrete way the structure of rough approximate groups, ideally obtaining an analogue of the results of Breuillard, Green and Tao on approximate groups [BGT12]. We have not attempted to formulate a conjecture along these lines, but examples such as taking a maximal  $\delta$ -separated subset of a small ball about the identity in  $\mathrm{SO}(3)$ , where the size of the ball tends to zero with  $\delta$  but much more slowly than  $\delta$ , suggest that Lie groups of bounded rank are likely to play a role, and also that the part played by nilpotency may be significantly different.

It is natural to ask whether there is an analogue of the results of this paper for Abelian groups. In a forthcoming paper we address this question, identifying a structure that plays the role that the octahedron plays for general groups, in the sense that if the number of copies of that structure in a partial Latin square is within a constant of maximal, then the partial Latin square has Abelian-group-like behaviour. The proof turns out to be quite a lot harder, because it is necessary to consider tripartite surfaces of higher genus, and that leads to significant complications.

## Acknowledgments

We would like to thank an anonymous referee, whose suggestions led us to rewrite the paper substantially and greatly improve its presentation.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Appendix

**A. Rough approximate groups.** Let  $G$  be a group. A subset  $H$  of  $G$  is a  $k$ -*approximate subgroup* if it contains the identity, it is closed under taking inverses, and there exists a set  $K$  of size at most  $k$  such that  $HH \subset KH$ —that is, if the product set  $HH$  can be covered by a bounded number of (left) translates of  $H$ . If  $G$  is a metric group, we shall say that a subset  $H$  is a  $(k, \delta)$ -*rough approximate subgroup* if there is a set  $K$  of size at most  $k$  such that  $HH \subset (KH)_\delta$ , where for any subset  $U$  we write  $U_\delta$  denotes the  $\delta$ -expansion  $\{x : d(x, U) \leq \delta\}$  of  $U$ . Thus,  $H$  is a rough approximate subgroup if every point in  $HH$  can be approximated by a point in one of a bounded number of translates of  $H$ . By a rough approximate group, we mean simply a rough approximate subgroup of some metric group. (As with approximate groups themselves, it is possible to define rough approximate groups more intrinsically, but since ours arise naturally as subsets of an ambient group, we shall not do this).

From now on, we will take the mappings  $\phi, \psi$  and  $\omega$  from Theorem 1.9 as given, and refer to  $\phi(X)$ ,  $\psi(Y)$  and  $\omega(Z)$  as  $X$ ,  $Y$  and  $Z$  instead.

With this convention, Theorem 1.9 yields for us three 1-separated subsets  $X, Y, Z$  of a metric group  $G$ , all of roughly the same size, and a small positive number  $\delta$ , such that  $d(xy, Z) \leq \delta$  for a positive proportion of pairs  $(x, y) \in X \times Y$ . In this appendix we shall deduce that there is a rough approximate subgroup  $H$  of  $G$  such that  $X$  has substantial overlap with a left translate of  $H$ ,  $Y$  has substantial overlap with a right translate, and  $Z$  has substantial overlap with a two-sided translate. The (slightly stronger) precise statement is Theorem A.14 below. The arguments are mostly contained in either [Taoa] or [Tao08], and those that are not are fairly straightforward modifications or extensions of those arguments. It is for that reason, and because the result is something of an optional extra to our main result, that we present it in an appendix rather than in the main body of the paper.

**A.1 Metric entropy definitions and some basic observations.** Given a subset  $X$  of a metric space, and another subset  $\Delta$ , we say that  $\Delta$  is an  $\epsilon$ -*net* of  $X$  if for every  $x \in X$  there exists  $y \in \Delta$  such that  $d(x, y) < \epsilon$ . An  $\epsilon$ -*separated subset* of  $X$  is a subset  $\Gamma$  such that  $d(x, x') \geq \epsilon$  for every pair of distinct elements  $x, x' \in \Gamma$ . Write  $\nu_\epsilon(X)$  for the smallest size of an  $\epsilon$ -net of  $X$ , and  $\sigma_\epsilon(X)$  for the largest size of an  $\epsilon$ -separated subset. We begin with three very basic lemmas.

LEMMA A.1. *Let  $X$  be a subset of a metric space and let  $\epsilon > 0$ . Then  $\nu_\epsilon(X) \leq \sigma_\epsilon(X) \leq \nu_{\epsilon/2}(X)$ .*

*Proof.* Let  $\Gamma$  be an  $\epsilon$ -separated set of maximal size. Then in particular it is maximal. It follows that it is an  $\epsilon$ -net. This proves the first inequality.

Now let  $\Delta$  be an  $(\epsilon/2)$ -net. Then the balls of radius  $\epsilon/2$  about the points of  $\Delta$  cover  $X$ , and no  $\epsilon$ -separated set can contain more than one element in any of these balls. This proves the second inequality.  $\square$

LEMMA A.2. *Let  $X$  and  $Y$  be subsets of metric spaces and let  $d$  be the metric on  $X \times Y$  defined by  $d((x, y), (x', y')) = d(x, x') \vee d(y, y')$ . Then  $\nu_\epsilon(X \times Y) \leq \sigma_{\epsilon/2}(X)\sigma_{\epsilon/2}(Y)$ .*

*Proof.* By Lemma A.1, we have that

$$\nu_\epsilon(X \times Y) \leq \nu_{\epsilon/2}(X)\nu_{\epsilon/2}(Y) \leq \sigma_{\epsilon/2}(X)\sigma_{\epsilon/2}(Y).$$

□

LEMMA A.3. *Let  $X$  be a subset of a metric group and let  $\epsilon > 0$ . Then  $\nu_\epsilon(X) = \nu_\epsilon(X^{-1})$  and  $\sigma_\epsilon(X) = \sigma_\epsilon(X^{-1})$ .*

*Proof.* This is an immediate consequence of the fact that

$$d(x, y) = d(y, x) = d(e, y^{-1}x) = d(x^{-1}, y^{-1})$$

for any two elements  $x, y$  of a metric group. □

We shall write  $\bar{\nu}_\epsilon(X)$  for the size of the smallest non-strict  $\epsilon$ -net of  $X$ —that is, of the smallest set  $\Delta$  such that for every  $x \in X$  there exists  $y \in \Delta$  with  $d(x, y) \leq \epsilon$ .

LEMMA A.4. *Let  $X, Y, Z$  be 1-separated subsets of a metric group  $G$ , let  $\delta < \frac{1}{100}$ , let  $\epsilon < 1/6$ , and suppose that  $|Z| \leq \delta^{-1}|X|^{1/2}|Y|^{1/2}$  and that  $d(xy, Z) \leq \epsilon$  for at least  $\delta|X||Y|$  pairs  $(x, y) \in X \times Y$ . Then there are subsets  $X' \subset X$  and  $Y' \subset Y$  with  $|X'| \geq \delta^7|X|$  and  $|Y'| \geq \delta^7|Y|$  such that  $\bar{\nu}_{6\epsilon}(X'Y') \leq \delta^{-16}|X|^{1/2}|Y|^{1/2}$  and such that  $d(xy, Z) \leq \epsilon$  for at least  $\delta|X'||Y'|/4$  pairs  $(x, y) \in X' \times Y'$ .*

*Proof.* Form a bipartite graph  $G$  with vertex sets  $X, Y$  by joining  $x$  to  $y$  if and only if  $d(xy, z) \leq \epsilon$ . Then by hypothesis  $G$  has density  $\delta$ .

We shall apply Lemma 2.11, but in order to do so we must first balance the sizes of the vertex sets. Suppose without loss of generality that  $|X| \leq |Y|$ . From the above discussion, we recall that  $|Y| \leq \delta^{-4}|X|$ . We now discard vertices of minimal degree from  $Y$  one by one, until we arrive at a subset  $Y_1 \subset Y$  with  $|Y_1| = |X|$ . The edge density of the graph  $G|_{X \times Y_1}$  is still at least  $\delta$ .

Applying Lemma 2.11 with  $k = 1$ , we can find  $X' \subset X$  and  $Y' \subset Y_1$  with  $|X'| \geq \delta^2|X|/16 \geq \delta^7|X|$  and  $|Y'| \geq \delta^2|Y_1|/16 \geq \delta^6|Y|/16 \geq \delta^7|Y|$  such that between any  $x \in X'$  and  $y \in Y'$  there are at least  $\delta^9|X||Y|$  paths of length 3 (with the two vertices in between not required to live in  $X'$  and  $Y'$ ) and such that the graph  $G|_{X' \times Y'}$  has density at least  $\delta/4$ .

For each  $x \in X'$  and  $y \in Y'$ , let  $T(x, y)$  be the set of triples  $(z_1, z_2, z_3) \in Z^3$  such that there exist  $x_1 \in X$  and  $y_1 \in Y$  with  $d(xy_1, z_1), d(x_1y_1, z_2)$  and  $d(x_1y, z_3)$  all at most  $\epsilon$ . Since  $X, Y$  and  $Z$  are all 1-separated, there is a bijection between triples in  $T(x, y)$  and paths of length 3 from  $x$  to  $y$  in the graph, so each set  $T(x, y)$  has size at least  $\delta^9|X||Y_1| \geq \delta^{13}|X||Y|$ .

Suppose now that  $(z_1, z_2, z_3)$  belongs to  $T(x, y)$  and  $x_1, y_1$  are as above. Then from the three approximate relations and the fact that

$$xy = xy_1(x_1y_1)^{-1}x_1y,$$

it follows that

$$d(xy, z_1 z_2^{-1} z_3) \leq 3\epsilon.$$

Now let  $\Gamma = \{(x_1 y_1), \dots, (x_m y_m)\}$  be a  $6\epsilon$ -separated subset of  $X'Y'$ . Then the balls of radius  $3\epsilon$  about the  $x_i y_i$  are disjoint, from which it follows that the sets  $T(x_i, y_i)$  are disjoint. But each one has size at least  $\delta^{13}|X||Y|$  and their union has size at most  $|Z|^3$ , so  $m \leq \delta^{-13}|Z|^3|X|^{-1}|Y|^{-1} \leq \delta^{-16}|X'|^{1/2}|Y'|^{1/2}$ . This bound holds for all  $6\epsilon$ -separated subsets, so the result now follows from Lemma A.1.  $\square$

We remark that since  $X$  and  $Y$  are 1-separated sets, we could if we wanted replace the cardinalities  $|X'|$  and  $|Y'|$  in the statement above by the quantities  $\sigma_1(X')$  and  $\sigma_1(Y')$ .

One of the main results of [Tao08] is that if  $X, Y$  are finite subsets of a group and  $|XY| \leq C|X|^{1/2}|Y|^{1/2}$ , then there exists an approximate group  $H$  and sets  $K, L$  of bounded size such that  $X \subset KH$  and  $Y \subset HL$ . (One can of course take  $K$  and  $L$  to be the same by taking their union). In the next subsection, we shall prove an analogous statement for our metric-entropy context.

**A.2 Products with small metric entropy come from rough approximate groups.** The main theorem we prove in this subsection is the following metric-entropy variant of Theorem 4.6 of [Tao08].

**Theorem A.5.** *Let  $G$  be a metric group, let  $\beta \geq 2048\epsilon$ , and let  $X, Y \subset G$  be subsets such that  $\nu_\epsilon(XY) \leq C\sigma_\beta(X)^{1/2}\sigma_\beta(Y)^{1/2}$ . Then there exists a  $(16C^{16}, 256\epsilon)$ -rough approximate group  $H \subset G$  and sets  $K, L$  of sizes at most  $256C^{32}$  and  $2048C^{48}$ , respectively, such that  $KH$  is a  $584\epsilon$ -net of  $X$ ,  $HL$  is a  $2304\epsilon$ -net of  $Y$ , and  $\nu_{128\epsilon}(H) \leq 8C^{15}\sigma_\beta(X)^{1/2}\sigma_\beta(Y)^{1/2}$ .*

We begin with an analogue of the Ruzsa triangle inequality (which can also be found in [Taoa]).

LEMMA A.6. *Let  $G$  be a metric group and let  $U, V, W$  be subsets of  $G$ . Then  $\nu_\epsilon(U)\nu_\epsilon(VW^{-1}) \leq \sigma_{\epsilon/4}(UV^{-1})\sigma_{\epsilon/4}(UW^{-1})$ .*

*Proof.* Let  $\Gamma_1$  be an  $\epsilon$ -separated subset of  $U$  and let  $\Gamma_2$  be an  $\epsilon$ -separated subset of  $VW^{-1}$ . Define  $\phi : \Gamma_1 \times \Gamma_2 \rightarrow UV^{-1} \times UW^{-1}$  by choosing for each  $x \in \Gamma_2$  a pair of elements  $(v(x), w(x)) \in V \times W$  such that  $v(x)w(x)^{-1} = x$ , and then for each  $(u, x) \in \Gamma_1 \times \Gamma_2$  defining  $\phi(u, x)$  to be  $(uv(x)^{-1}, uw(x)^{-1})$ .

Suppose now that  $(u_1, x_1)$  and  $(u_2, x_2)$  are elements of  $\Gamma_1 \times \Gamma_2$  such that  $d(\phi(u_1, x_1), \phi(u_2, x_2)) < \delta$ , where for our product metric we take the maximum of the metrics on  $UV^{-1}$  and  $UW^{-1}$ . Then  $d(u_1 v(x_1)^{-1}, u_2 v(x_2)^{-1}) < \delta$  and  $d(u_1 w(x_1)^{-1}, u_2 w(x_2)^{-1}) < \delta$ . Since  $G$  is a metric group, it follows that

$$\begin{aligned} d(x_1, x_2) &= d(v(x_1)w(x_1)^{-1}, v(x_2)w(x_2)^{-1}) \\ &= d(v(x_1)u_1^{-1}u_1w(x_1)^{-1}, v(x_2)u_2^{-1}u_2w(x_2)^{-1}) \end{aligned}$$

$$< \delta + \delta = 2\delta.$$

Therefore, if  $\delta \leq \epsilon/2$  we can deduce that  $x_1 = x_2$ , since they are both elements of  $\Gamma_2$ . But then  $d(u_1, u_2) = d(u_1v(x_1)^{-1}, u_2v(x_1)^{-1}) < \delta$ , which implies that  $u_1 = u_2$  as well.

Since  $\Gamma_1$  and  $\Gamma_2$  were arbitrary  $\epsilon$ -separated subsets, it follows that

$$\sigma_\epsilon(U)\sigma_\epsilon(VW^{-1}) \leq \sigma_{\epsilon/2}(UV^{-1} \times UW^{-1}),$$

and hence by Lemmas A.1 and A.2, that

$$\nu_\epsilon(U)\nu_\epsilon(VW^{-1}) \leq \sigma_{\epsilon/4}(UV^{-1})\sigma_{\epsilon/4}(UW^{-1}).$$

□

**COROLLARY A.7.** *Let  $\epsilon, \delta > 0$  and let  $X, Y$  be a subsets of a metric group such that  $\nu_\epsilon(XY) \leq C\sigma_\delta(X)^{1/2}\sigma_{16\epsilon}(Y)^{1/2}$ . Then  $\nu_{8\epsilon}(XX^{-1}) \leq C^2\sigma_\delta(X)$ .*

*Proof.* By Lemma A.6, Lemma A.3 and our hypothesis, we have that

$$\nu_{8\epsilon}(Y^{-1})\nu_{8\epsilon}(XX^{-1}) \leq \sigma_{2\epsilon}(Y^{-1}X^{-1})^2 = \sigma_{2\epsilon}(XY)^2 \leq \nu_\epsilon(XY)^2 \leq C^2\sigma_\delta(X)\sigma_{16\epsilon}(Y).$$

By Lemmas A.3 and A.1,  $\nu_{8\epsilon}(Y^{-1}) = \nu_{8\epsilon}(Y) \geq \sigma_{16\epsilon}(Y)$ , so the result follows. □

Our next lemma is a version of the Ruzsa covering lemma.

**LEMMA A.8.** *Let  $\epsilon > 0$  and let  $A, B$  be subsets of a metric group such that  $\nu_\epsilon(AB) \leq C\sigma_{2\epsilon}(B)$ . Then there exists a set  $K$  of size at most  $C$  such that  $KBB^{-1}$  is a  $2\epsilon$ -net of  $A$ .*

*Proof.* Let  $K \subset A$  be maximal such that for any two distinct elements  $x, x' \in K$  the distance between the sets  $xB$  and  $x'B$  is at least  $2\epsilon$ . Then if  $y \in A$  there must be some  $x \in K$  such that  $d(xB, yB) < 2\epsilon$ , by maximality, from which it follows that  $d(y, xBB^{-1}) < 2\epsilon$ . Therefore,  $KBB^{-1}$  is a  $2\epsilon$ -net of  $A$ .

Now let  $\Gamma$  be a  $2\epsilon$ -separated subset of  $B$ . Then  $K\Gamma$  is a  $2\epsilon$ -separated subset of  $KB$ , which is contained in  $AB$ . It follows that  $K\sigma_{2\epsilon}(B) \leq \sigma_{2\epsilon}(AB)$ , which by Lemma A.1 is at most  $\nu_\epsilon(AB)$ . By hypothesis this is at most  $C\sigma_{2\epsilon}(B)$  and the result follows. □

Next we need a notion of ‘popular differences’ that will be suitable for this metric-entropy context.

**DEFINITION A.9.** *Let  $A$  be a subset of a metric group. We say that an element  $d \in A^2$  is  $(\epsilon, \delta, m)$ -popular if there are  $m$  pairs  $(x_i, y_i) \in A^2$  such that the sets  $\{x_1, \dots, x_m\}$  and  $\{y_1, \dots, y_m\}$  are  $\delta$ -separated and  $d(y_i^{-1}x_i, d) < \epsilon$  for every  $i$ ,*

**LEMMA A.10.** *Let  $\delta \geq 2\epsilon$ , let  $A$  be a subset of a metric group such that  $\nu_\epsilon(AA^{-1}) \leq C\sigma_\delta(A)$  and let  $S$  be the set of  $(2\epsilon, \delta, \sigma_\delta(A)/2C)$ -popular elements of  $A^{-1}A$ . Then  $\sigma_\delta(S) \geq \sigma_\delta(A)/2C$ .*

*Proof.* Let  $\Gamma$  be a  $\delta$ -separated subset of  $A$  of size  $\sigma_\delta(A)$ . Choose a partition of  $AA^{-1}$  into  $\nu_\epsilon(AA^{-1})$  sets, each contained in an open ball of radius  $\epsilon$ , and write  $z \sim w$  if  $z$  and  $w$  belong to the same cell of the partition.

If we choose a random cell from the partition, then the expected number of pairs  $(x_1, x_2) \in \Gamma^2$  with  $x_1x_2^{-1}$  in that cell is at least  $\sigma_\delta(A)^2/\nu_\epsilon(AA^{-1})$ . It follows that there are at least  $\sigma_\delta(A)^4/\nu_\epsilon(AA^{-1}) \geq \sigma_\delta(A)^3/C$  quadruples  $(x_1, x_2, x_3, x_4) \in \Gamma^4$  such that  $x_1x_2^{-1} \sim x_3x_4^{-1}$ , and hence, since the cells are contained in balls of radius  $\epsilon$ , such that  $d(x_3^{-1}x_1, x_4^{-1}x_2) < 2\epsilon$ . It follows that for a randomly chosen  $(x_1, x_3) \in \Gamma^2$  the expected number of pairs  $(x_2, x_4) \in \Gamma^2$  such that  $d(x_3^{-1}x_1, x_4^{-1}x_2) < 2\epsilon$  is at least  $\sigma_\delta(A)/C$ . Since  $\delta \geq 2\epsilon$ , it is not possible to find  $x, y, z \in \Gamma$  such that  $x^{-1}y = x^{-1}z$  or such that  $x^{-1}z = y^{-1}z$ . It follows that the maximum number of pairs  $(x_2, x_4)$  with  $d(x_3^{-1}x_1, x_4^{-1}x_2) < 2\epsilon$  is at most  $\sigma_\delta(A)$ . Therefore, there are at least  $\sigma_\delta(A)^2/2C$  pairs  $(x_1, x_3) \in \Gamma$  such that  $x_3^{-1}x_1$  is  $(2\epsilon, \delta, \sigma_\delta(A)/2C)$ -popular.

By averaging we can find some  $x_i$  for which there are at least  $\sigma_\delta(A)/2C$  popular pairs  $(x_i, x_j)$ . If  $(x_i, x_j)$  and  $(x_i, x_k)$  are two distinct such pairs, then  $d(x_j^{-1}x_i, x_k^{-1}x_i) \geq \delta$ . It follows that there is a  $\delta$ -separated subset of  $S$  of size at least  $\sigma_\delta(A)/2C$ , as claimed. □

LEMMA A.11. *Let  $\delta \geq 4\epsilon$ , let  $A$  be a subset of a metric group such that  $\nu_\epsilon(AA^{-1}) \leq C\sigma_\delta(A)$  and let  $S$  be the set of  $(2\epsilon, \delta, \sigma_\delta(A)/2C)$ -popular elements of  $A^{-1}A$ . Then  $\nu_{16\epsilon}(AS^3A^{-1}) \leq 8C^7\sigma_\delta(A)$ .*

*Proof.* Let  $x_0, x_7$  be elements of  $A$  and let  $d_1, d_2, d_3 \in S$ . Since each  $d_i$  is popular, we can approximate  $x_0d_1d_2d_3x_7^{-1}$  as  $x_0x_1^{-1}x_2x_3^{-1}x_4x_5^{-1}x_6x_7^{-1}$  in several ways. More precisely, for each  $i = 1, 3, 5$  we have at least  $\sigma_\delta(A)/2C$  independent choices for the pair  $(x_i, x_{i+1})$ , and the individual coordinates of these choices form  $\delta$ -separated sets.

Each such product gives us an element  $(x_0x_1^{-1}, x_2x_3^{-1}, x_4x_5^{-1}, x_6x_7^{-1})$  of the set  $(AA^{-1})^4$ . If  $(x_0x_1^{-1}, x_2x_3^{-1}, x_4x_5^{-1}, x_6x_7^{-1})$  and  $(x_0x_1'^{-1}, x_2'x_3'^{-1}, x_4'x_5'^{-1}, x_6'x_7'^{-1})$  are two different such quadruples, then if their first  $i$  coordinates agree and the  $(i + 1)$ st coordinate is different, then  $x_j = x_j'$  for  $0 \leq j < 2i$ , and hence for  $j = 2i$  as well, so we find that the two  $(i + 1)$ st coordinates are  $x_{2i}x_{2i+1}^{-1}$  and  $x_{2i}x_{2i+1}'^{-1}$ , which are separated by at least  $\delta \geq 4\epsilon$ .

We also have that if two elements of  $AS^3A^{-1}$  are separated by at least  $16\epsilon$  and for each one we choose a quadruple as above, then at least one coordinate of the two quadruples will be separated by at least  $4\epsilon$ , since the products of the two quadruples give the two elements.

It follows that

$$\sigma_{16\epsilon}(AS^3A^{-1})(\sigma_\delta(A)/2C)^3 \leq \sigma_{4\epsilon}((AA^{-1})^4) \leq \nu_\epsilon(AA^{-1})^4.$$

Since  $\nu_\epsilon(AA^{-1}) \leq C\sigma_\delta(A)$ , this implies the result. □

LEMMA A.12. *Let  $\delta \geq 2\epsilon$  and let  $S$  be a subset of a metric group such that  $S = S^{-1}$  and  $\nu_\epsilon(S^3) \leq C\sigma_\delta(S)$ . Then  $S^2$  is a  $(C, 2\epsilon)$ -rough approximate group.*

*Proof.* By Lemma A.8 with  $A = S^2$  and  $B = S$  there is a set  $K$  of size at most  $C$  such that  $KS^2$  is a  $2\epsilon$ -net of  $S^2$ .  $\square$

LEMMA A.13. *Let  $\delta \geq 2\epsilon_1$ , let  $A$  be a subset of a metric group, let  $H$  be a  $(C_2, \epsilon_2)$ -rough approximate group, and suppose that  $\nu_{\epsilon_1}(AH) \leq C_1\sigma_\delta(H)$ . Then there is a set  $K$  of size at most  $C_1C_2$  such that  $KH$  is a  $(2\epsilon_1 + \epsilon_2)$ -net of  $A$ .*

*Proof.* By Lemma A.8 there is a set  $K_1$  of size at most  $C_1$  such that  $K_1H^2$  is a  $2\epsilon_1$ -net of  $A$ . By the definition of an approximate group there is also a set  $K_2$  of size at most  $C_2$  such that  $K_2H$  is an  $\epsilon_2$ -net of  $H$ . But then  $K_1K_2H$  is a  $(2\epsilon_1 + \epsilon_2)$ -net of  $A$ .  $\square$

*Proof of Theorem A.5.* If  $X, Y$  are subsets of a metric group and  $\nu_\epsilon(XY) \leq C\sigma_\beta(X)^{1/2}\sigma_\beta(Y)^{1/2}$ , then by Corollary A.7 we have the inequality  $\nu_{8\epsilon}(XX^{-1}) \leq C^2\sigma_\beta(X)$ . By Lemmas A.10 and A.11 we obtain a set  $S$  with  $S = S^{-1}$  and  $\sigma_\beta(S) \geq \sigma_\beta(X)/2C^2$  such that  $\nu_{128\epsilon}(XS^3X^{-1}) \leq 8C^{14}\sigma_\beta(X)$ .

It follows that  $\nu_{128\epsilon}(S^3) \leq 16C^{16}\sigma_\beta(S)$ . Therefore, by Lemma A.12,  $S^2$  is a  $(16C^{16}, 256\epsilon)$ -rough approximate group.

We also have that  $\nu_{128\epsilon}(XS^2) \leq 16C^{16}\sigma_\beta(S^2)$ . Therefore, by Lemma A.13 there is a set  $K$  of size at most  $256C^{32}$  such that  $KS^2$  is a  $512\epsilon$ -net of  $X$ .

By Lemma A.6,

$$\begin{aligned} \nu_{1024\epsilon}(X)\nu_{1024\epsilon}(S^2Y) &\leq \sigma_{256\epsilon}(XS^2)\sigma_{256\epsilon}(XY) \\ &\leq \nu_{128\epsilon}(XS^2)\nu_{128\epsilon}(XY) \\ &\leq 16C^{16}\sigma_\beta(S^2).C\sigma_\beta(X)^{1/2}\sigma_\beta(Y)^{1/2}. \end{aligned}$$

But

$$\sigma_\beta(X) \leq \sigma_\beta(XY) \leq \nu_{\beta/2}(XY) \leq C\sigma_\beta(X)^{1/2}\sigma_\beta(Y)^{1/2},$$

so  $\sigma_\beta(X) \leq C^2\sigma_\beta(Y)$  and therefore  $\sigma_\beta(X)^{1/2}\sigma_\beta(Y)^{1/2} \leq C\sigma_\beta(Y)$ . Also, since  $\beta \geq 2048\epsilon$ ,

$$\sigma_\beta(S^2) \leq \nu_{128\epsilon}(XS^3X) \leq 8C^{14}\sigma_\beta(X) \leq 8C^{14}\nu_{1024\epsilon}(X).$$

It follows that  $\nu_{1024\epsilon}(Y^{-1}S^2) = \nu_{1024\epsilon}(S^2Y) \leq 128C^{32}\sigma_\beta(Y)$ .

Therefore, by Lemma A.13 again it follows that there is a set  $L$  of size at most  $2048C^{48}$  such that  $LS^2$  is a  $2304\epsilon$ -net of  $Y^{-1}$ , which implies that  $S^2L^{-1}$  is a  $2304\epsilon$ -net of  $Y$ .  $\square$

We conclude this appendix by combining Lemma A.4 and Theorem A.5. We shall present the result (mostly) without explicit constants, but it is not hard to obtain them.

**Theorem A.14.** *Let  $X, Y, Z$  be 1-separated subsets of a metric group  $G$ , let  $0 < \delta < 1/100$ , let  $\epsilon > 0$  be sufficiently small, and suppose that  $|Z| \leq \delta^{-1}|X|^{1/2}|Y|^{1/2}$  and that  $d(xy, Z) \leq \epsilon$  for at least  $\delta|X||Y|$  pairs  $(x, y) \in X \times Y$ . Then there exist subsets  $X'' \subset X$ ,  $Y'' \subset Y$  and  $Z'' \subset Z$  with  $|X''| = \delta^{O(1)}|X|$ ,  $|Y''| = \delta^{O(1)}|Y|$  and  $|Z''| = \delta^{O(1)}|Z|$ , a  $(\delta^{-O(1)}, O(\epsilon))$ -rough approximate group  $H \subset G$ , and elements  $u, v, w$  of  $G$  such that  $\nu_{O(\epsilon)}(H) = \delta^{-O(1)}|X|^{1/2}|Y|^{1/2}$ ,  $X'' \subset (uH)_{O(\epsilon)}$ ,  $Y'' \subset (Hv)_{O(\epsilon)}$ ,  $Z'' \subset (X''Y'')_\epsilon \cap (uwHv)_{O(\epsilon)}$  and  $d(xy, Z'') \leq \epsilon$  for  $\delta^{O(1)}|X''||Y''|$  pairs  $(x, y) \in X'' \times Y''$ .*

*Proof.* Lemma A.4 gives us  $X' \subset X$  and  $Y' \subset Y$  with  $|X'| \geq \delta^7|X|$  and  $|Y'| \geq \delta^7|Y|$  such that  $\nu_{O(\epsilon)}(X'Y') = \delta^{-O(1)}|X|^{1/2}|Y|^{1/2}$  and such that  $d(xy, Z) \leq \epsilon$  for  $\delta^{O(1)}|X'||Y'|$  pairs  $(x, y) \in X' \times Y'$ . Applying Theorem A.5 (with  $\beta = 1$ ), we obtain a  $(\delta^{-O(1)}, O(\epsilon))$ -rough approximate group  $H \subset G$  and sets  $K, L$  of sizes  $\delta^{-O(1)}$  such that  $X' \subset (KH)_{O(\epsilon)}$  and  $Y' \subset (HL)_{O(\epsilon)}$ .

We will pick  $u \in K$  and  $v \in L$  at random, and let  $X'' = X' \cap (uH)_{O(\epsilon)}$  and  $Y'' = Y' \cap (Hv)_{O(\epsilon)}$ . By averaging there are choices  $u \in K$  and  $v \in L$  such that  $|X''| = \delta^{O(1)}|X|$ ,  $|Y''| = \delta^{O(1)}|Y|$  and  $d(xy, Z) \leq \epsilon$  for  $\delta^{O(1)}|X''||Y''|$  pairs  $(x, y) \in X'' \times Y''$ .

Observe that since  $X'' \subset (uH)_{O(\epsilon)}$  and  $Y'' \subset (Hv)_{O(\epsilon)}$ , we have that  $X''Y'' \subset (uHHv)_{O(\epsilon)}$ . Since  $H$  is a  $(\delta^{-O(1)}, O(\epsilon))$ -rough approximate subgroup of  $G$ , this means that there exists a set  $M \subset G$  of size  $\delta^{-O(1)}$  such that  $X''Y'' \subset (uMHv)_{O(\epsilon)}$ .

Since  $X''Y'' \subset (uMHv)_{O(\epsilon)}$ , we have that  $(X''Y'')_\epsilon \subset (uMHv)_{O(\epsilon)}$ . Let

$$Z' = Z \cap (X''Y'')_\epsilon \subset (uMHv)_{O(\epsilon)}$$

and observe that  $d(xy, Z') \leq \epsilon$  for  $\delta^{O(1)}|X''||Y''|$  pairs  $(x, y) \in X'' \times Y''$ .

Now we choose  $w \in M$  uniformly at random, and let  $Z'' = Z' \cap (uwHv)_{O(\epsilon)}$ . Since  $|M| = \delta^{-O(1)}$ , we have in expectation that  $d(xy, Z'') \leq \epsilon$  for  $\delta^{O(1)}|X''||Y''|$  pairs  $(x, y) \in X'' \times Y''$ . Suppose without loss of generality that  $|X| \geq |Y|$ . If  $d(xy, Z'') \leq \epsilon$  for at least  $\delta^{O(1)}|X''||Y''|$  pairs  $(x, y) \in X'' \times Y''$ , then there exists a choice of  $y \in Y''$  such that  $d(xy, Z'') \leq \epsilon$  for  $\delta^{O(1)}|X''|$  choices of  $x \in X''$ . Since  $X''$  is 1-separated, this implies that  $|Z''| = \delta^{O(1)}|X''| = \delta^{O(1)}|Z|$ . Therefore there is some choice of  $w \in M$  satisfying our requirements.  $\square$

### B. A Bogolyubov-type lemma for $\text{SO}(3)$ .

In this section we look at properties of product sets of dense subsets of  $\text{SO}(3)$ . The main result we shall prove is the following lemma. Once we have it, it will enable us to prove that the partial binary operation on a maximal  $\delta$ -separated subset of  $\text{SO}(3)$  described in the introduction is defined for a constant proportion of pairs and gives rise to within a constant of the maximum possible number of associative triples.

LEMMA B.1. *For every  $\theta > 0$  and  $\epsilon > 0$  there exists  $\eta > 0$  such that if  $A$  is any subset of  $\text{SO}(3)$  of Haar measure at least  $\theta$ , then  $AA^{-1}$  contains a proportion  $1 - \epsilon$  of a ball of radius  $\eta$  about the identity.*



It follows straightforwardly that  $AA^{-1}AA^{-1}$  contains the whole of a ball of some radius  $\eta(\theta)$ . Thus, this result can be thought of as a Bogolyubov lemma [Bog39] for  $\mathrm{SO}(3)$ , with balls about the identity playing the role of Bohr sets. Note that unlike in the Abelian case, there is an extra uniformity here: the ball we obtain depends only on the measure of  $A$  and not on  $A$  itself. This fact, which can be thought of as saying that the only departure from quasirandomness of the group  $\mathrm{SO}(3)$  is the obvious one that a product of two small balls is contained in a small ball, will be essential to our argument. A corollary of this result will be a statement that we claimed earlier in the paper: that if  $\Gamma$  is a maximal  $\delta$ -separated subset of  $\mathrm{SO}(3)$  and  $\circ : \Gamma \times \Gamma \rightarrow \Gamma$  is a partially defined operation where  $x \circ y = z$  if and only if  $xy$  is close to  $z$ , then  $\circ$  is defined for a dense set of pairs. (We give a precise formulation later).

To prove Lemma B.1 we shall begin, as one might expect, by imitating the proof of Bogolyubov's lemma, using non-Abelian Fourier analysis. This will show that the structure of the convolution  $\mathbb{1}_A * \mathbb{1}_{A^{-1}}$  essentially depends on the large Fourier coefficients of  $\mathbb{1}_A$ . As is well known, these all come from the low-dimensional representations of  $\mathrm{SO}(3)$ : the further uniformity mentioned above comes from the fact that the number of low-dimensional representations is bounded. To prove the assertion about the ball, we use the fact that the low-dimensional representations can be described explicitly as follows. Every irreducible representation has odd dimension, and for each odd dimension there is exactly one irreducible representation, which is given by the action of  $\mathrm{SO}(3)$  on the space of spherical harmonics of degree  $d$ . (See for example [KS10]).

Let us briefly recall the basic facts about non-Abelian Fourier analysis that we shall need. Given an integrable function  $f : \mathrm{SO}(3) \rightarrow \mathbb{C}$  and an irreducible representation  $\rho$  of  $\mathrm{SO}(3)$ , we define the Fourier coefficient  $\hat{f}(\rho)$  by the formula

$$\hat{f}(\rho) = \mathbb{E}_x f(x) \overline{\rho(x)},$$

where we are writing  $\mathbb{E}_x$  for the average with respect to Haar measure on  $\mathrm{SO}(3)$ . Note that if  $\rho$  is a  $k$ -dimensional representation, then  $\hat{f}(\rho)$  is a  $k \times k$  matrix.

The non-Abelian versions of Parseval's identity, the convolution identity, and the inversion formula are as follows. Parseval's identity states that for any two square-integrable functions  $f, g : \mathrm{SO}(3) \rightarrow \mathbb{C}$ ,

$$\int_{\mathrm{SO}(3)} f(x) \overline{g(x)} dx = \sum_{\rho} n_{\rho} \mathrm{tr}(\hat{f}(\rho) \hat{g}(\rho)^*),$$

where the sum is over all irreducible representations and  $n_{\rho}$  is the dimension of  $\rho$ . The left-hand side is the obvious definition of the inner product of  $f$  and  $g$ . As for the right-hand side, the matrix inner product  $\langle A, B \rangle$  of two  $k \times k$  matrices  $A$  and  $B$  is  $\mathrm{tr}(AB^*) = \sum_{ij} A_{ij} B_{ij}^*$ , so we can rewrite it as  $\sum_{\rho} n_{\rho} \langle \hat{f}(\rho), \hat{g}(\rho) \rangle$ , which is a natural way of defining the inner product  $\langle \hat{f}, \hat{g} \rangle$ . So, suitably interpreted, Parseval's identity is just the usual identity  $\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle$ .

The convolution identity is also the same as it is in the Abelian case:  $\widehat{f * g}(\rho) = \widehat{f}(\rho)\widehat{g}(\rho)$ . Of course, here the product  $\widehat{f}(\rho)\widehat{g}(\rho)$  is a matrix product.

Finally, the inversion formula is

$$f(x) = \sum_{\rho} n_{\rho} \text{tr}(\widehat{f}(\rho)\overline{\rho(x)^*}),$$

where the equality is valid almost everywhere.

The *Hilbert-Schmidt norm* of a complex matrix  $A$  is defined by the formula

$$\|A\|_{HS}^2 = \text{tr}(AA^*) = \sum_{x,y} |A(x,y)|^2.$$

The *box norm* is defined by the formula

$$\|A\|_{\square}^4 = \sum_{x,x',y,y'} A(x,y)\overline{A(x,y')}\overline{A(x',y)}A(x',y').$$

It is also equal to  $\text{tr}(AA^*AA^*) = \langle AA^*, AA^* \rangle = \|AA^*\|$ .

A Cauchy-Schwarz argument shows that  $\|AB\|_{HS} \leq \|A\|_{\square}\|B\|_{\square}$ , and it is also not hard to prove that  $\|A\|_{\square} \leq \|A\|_{HS}$ .

Now let us apply these facts to say something about the convolution  $f * g$  of two bounded measurable functions on  $\text{SO}(3)$ . (By ‘bounded’ we mean that  $\|f\|_{\infty}, \|g\|_{\infty} \leq 1$ ). By Parseval’s identity we have that

$$\|f * g\|_2^2 = \sum_{\rho} n_{\rho} \|\widehat{f}(\rho)\widehat{g}(\rho)^*\|_{HS}^2$$

The generalized Cauchy-Schwarz inequality for the box norm implies that  $\|AB\|_{HS}^2 \leq \|A\|_{\square}^2\|B\|_{\square}^2$ , so the right-hand side is at most

$$\sum_{\rho} n_{\rho} \|\widehat{f}(\rho)\|_{\square}^2 \|\widehat{g}(\rho)\|_{\square}^2 \leq \sum_{\rho} n_{\rho} \|\widehat{f}(\rho)\|_{HS}^2 \|\widehat{g}(\rho)\|_{HS}^2.$$

Also, the convolution identity and inversion formula together imply that

$$f * g(x) = \sum_{\rho} n_{\rho} \text{tr}(\widehat{f}(\rho)\widehat{g}(\rho)\overline{\rho(x)^*}).$$

Let us fix a constant  $C$  and split the right-hand side into the two functions

$$u(x) = \sum_{\rho: n_{\rho} \leq C} n_{\rho} \text{tr}(\widehat{f}(\rho)\widehat{g}(\rho)\overline{\rho(x)^*}).$$

and

$$v(x) = \sum_{\rho: n_{\rho} > C} n_{\rho} \text{tr}(\widehat{f}(\rho)\widehat{g}(\rho)\overline{\rho(x)^*}).$$

By Parseval’s identity,

$$\|v\|_2 = \sum_{\rho:n_\rho>C} n_\rho \|\hat{f}(\rho)\hat{g}(\rho)\|_{HS}^2.$$

Also, Parseval’s identity implies that  $\|\hat{f}(\rho)\|_{HS}^2 \leq n_\rho^{-1}\|f\|_2^2 \leq n_\rho^{-1}$  for any bounded function  $f$ , so

$$\begin{aligned} \sum_{\rho:n_\rho>C} n_\rho \|\hat{f}(\rho)\hat{g}(\rho)\|_{HS}^2 &= \sum_{\rho:n_\rho>C} \|\hat{f}(\rho)\|_{\square}^2 \|\hat{g}(\rho)\|_{\square}^2 \\ &\leq \left( \sum_{\rho:n_\rho>C} n_\rho \|\hat{f}(\rho)\|_{\square}^4 \right)^{1/2} \left( \sum_{\rho:n_\rho>C} n_\rho \|\hat{g}(\rho)\|_{\square}^4 \right)^{1/2} \\ &\leq \left( \sum_{\rho:n_\rho>C} n_\rho \|\hat{f}(\rho)\|_{HS}^4 \right)^{1/2} \left( \sum_{\rho:n_\rho>C} n_\rho \|\hat{g}(\rho)\|_{HS}^4 \right)^{1/2} \\ &\leq C^{-2} \left( \sum_{\rho:n_\rho>C} n_\rho \|\hat{f}(\rho)\|_{HS}^2 \right)^{1/2} \left( \sum_{\rho:n_\rho>C} n_\rho \|\hat{g}(\rho)\|_{HS}^2 \right)^{1/2} \\ &\leq C^{-2} \|f\|_2 \|g\|_2. \end{aligned}$$

It follows that if  $C$  is large, then the function  $f * g$  is well approximated in  $L_2(\text{SO}(3))$  by the function  $u$  defined above, which was the part that comes from the representations of dimension at most  $C$ .

Now let  $B$  be a ball of radius  $\eta$ , where  $\eta > 0$  is a constant to be chosen later, and let  $\mu_B$  be the characteristic measure of  $B$ . That is, if  $B$  has Haar measure  $\beta$ , then  $\mu_B(x) = \beta^{-1}$  for  $x \in B$  and  $\mu_B(x) = 0$  otherwise. We shall show that if  $\eta$  is sufficiently small, then  $\|u - u * \mu_B\|_\infty$ , and hence  $\|u - u * \mu_B\|_2$ , is small.

By the convolution law and the inversion formula,

$$u(x) - u * \mu_B(x) = \sum_{\rho:n_\rho \leq C} n_\rho \text{tr}(\hat{f}(\rho)\hat{g}(\rho)(I_{n_\rho} - \widehat{\mu}_B(\rho))\overline{\rho(x)^*}). \tag{1}$$

In order to bound the size of the right-hand side, we shall show that if  $\eta$  is small enough, then  $\widehat{\mu}_B(\rho)$  is close to the identity (on  $\mathbb{C}^{n_\rho}$ ) for all irreducible representations  $\rho$  of dimension at most  $C$ .

By definition,

$$\widehat{\mu}_B(\rho) = \mathbb{E}_{x \in \text{SO}(3)} \mu_B(x) \overline{\rho(x)} = \mathbb{E}_{x \in B} \overline{\rho(x)}.$$

One can show easily that the  $d$ -dimensional spherical harmonics are equicontinuous: for instance, it follows from the fact that the space of  $d$ -dimensional spherical harmonics is compact when considered as a subset of  $C(\text{SO}(3))$ . (This is the easy direction of the Arz\`ela-Ascoli theorem. If one wants, one can obtain estimates for

the equicontinuity by using explicit formulae for the spherical harmonics, but we shall content ourselves with a qualitative statement here). Therefore, for every  $\epsilon > 0$  and every irreducible representation  $\rho$  of dimension  $n_\rho = 2d + 1$  there exists  $\eta > 0$  such that if  $x$  is sufficiently close to the identity in  $\text{SO}(3)$ , then  $\langle \rho(x)p, p \rangle \geq 1 - \eta$  for every spherical harmonic  $p$  of dimension  $d$ . It follows by averaging over all  $p$  that  $n_\rho^{-1} \text{tr} \rho(x) \geq 1 - \eta$ , which implies that  $\|\rho(x) - I_{n_\rho}\|_{HS}^2 \leq 2\eta n_\rho$ , and therefore that  $\|\rho(x) - I_{n_\rho}\|_{\text{op}}^2 \leq 2\eta n_\rho$  as well.

It follows that for every  $\delta > 0$  and every  $C$  we may choose  $\eta > 0$  such that  $\|\rho(x) - I_{n_\rho}\|_{\text{op}} \leq \delta$  for every  $x \in B$  and every irreducible representation  $\rho$  of  $\text{SO}(3)$  of dimension at most  $C$ . This in turn implies by averaging that  $\|\widehat{\mu}_B(\rho) - I_{n_\rho}\|_{\text{op}} \leq \delta$  for every such  $\rho$ , where  $B$  is the ball of radius  $\eta$  about the identity. But then in the right-hand side of (1) we are taking the trace of a product of four matrices of which three have operator norm at most 1 and one has operator norm at most  $\delta$ . It follows that the trace is at most  $\delta n_\rho$ , and therefore that the right-hand side is in total at most  $\delta \sum_{\rho: n_\rho \leq C} n_\rho^2$ . Choosing  $\delta$  in such a way that this sum is at most  $\epsilon/2$  (which we can do with  $\delta$  depending on  $\epsilon$  and  $C$  only) we obtain that  $|u(x) - u * \mu_B(x)| \leq \epsilon$  for every  $x \in \text{SO}(3)$ , which implies that  $\|u - u * \mu_B\|_2 \leq \epsilon/2$ .

If we choose  $C$  such that  $C^{-2} \leq \epsilon/4$ , then  $\|v - v * \mu_B\|_2 \leq 2\|v\|_2 \leq \epsilon/4$ . Since  $f * g = u + v$ , it follows that

$$\|f * g - f * g * \mu_B\|_2 \leq \epsilon.$$

Let us state this result formally for later reference. When we say ‘the ball of radius  $\eta$ ’ this is to be understood to be the ball with respect to any reasonable metric, such as the one coming from the operator norm or the Hilbert-Schmidt norm—the result is true for all of them.

**LEMMA B.2.** *For every  $\epsilon > 0$  there exists  $\eta > 0$  such that the following statement holds. Let  $f$  and  $g$  be two bounded measurable complex-valued functions defined on  $\text{SO}(3)$ , let  $B$  be the ball of radius  $\eta$  around the identity in  $\text{SO}(3)$  and let  $\mu_B$  be the characteristic measure of  $B$ . Then*

$$\|f * g - f * g * \mu_B\|_2 \leq \epsilon.$$

We can interpret this result as a kind of partial quasirandomness property of  $\text{SO}(3)$ . For a fully quasirandom group (as defined in [Gow11]), we can replace  $\mu_B$  by the constant function that takes value 1 everywhere and the lemma above holds. Thus, when two bounded functions are convolved, the resulting function depends, up to a small  $L_2$  error, only on the averages of those functions. In  $\text{SO}(3)$  we cannot say that, but we can say that the resulting function does not depend on the fine structure of  $f$  and  $g$  and only on the averages over balls of radius  $\eta$ , since one can replace  $f$  by  $f * \mu_B$  or  $g$  by  $g * \mu_B$  without having much effect on the answer. (Strictly speaking, we have not proved that  $f * g$  is close to  $f * \mu_B * g$ , since  $\text{SO}(3)$  is non-Abelian, but a very minor modification of the above argument will do this as well).

Thus,  $\mathrm{SO}(3)$  is ‘quasirandom at fine scales’. This observation will play an important role in our argument.

Now we give the promised proof that the approximate multiplication we defined earlier is densely defined.

**LEMMA B.3.** *For every  $\theta \in (0, 1)$  there exists  $\delta > 0$  such that for any maximal  $\delta$ -separated subset  $\Gamma$  of  $\mathrm{SO}(3)$ , the proportion of  $(x, y) \in \Gamma^2$  with  $d(xy, \Gamma) \leq \theta\delta$  is at least  $(\theta/3)^9/16$ .*

*Proof.* Let us write  $\Gamma_\theta$  for the set of all points  $x$  within distance  $\theta\delta$  of a point in  $\Gamma$ . Since  $\Gamma$  is a maximal  $\delta$ -separated set, it is also a  $\delta$ -net, so the union of the balls of radius  $\delta$  about each point is all of  $\mathrm{SO}(3)$ . Since the balls of radius  $\theta\delta$  are disjoint and each one occupies at least a proportion  $\theta^3/2$  of the ball of radius  $\delta$  with the same centre (proving this is one detail that we omit, but it uses the fact that  $\mathrm{SO}(3)$  is a three-dimensional manifold), it follows that  $\Gamma_\theta$  has Haar measure at least  $\theta^3/2$ .

Furthermore, given any  $\eta > 0$ , if  $\delta$  is sufficiently small, we have for every ball  $B$  of radius  $\eta$  in  $\mathrm{SO}(3)$  that  $|\Gamma_\theta \cap B| \geq (\theta^3/2)|B|$ , where we write  $|A|$  for the Haar measure of a subset  $A$ , except that if  $A$  is a finite set then  $|A|$  will denote its cardinality.

Let  $f$  be the characteristic function of  $\Gamma_\theta$ . Then the last assertion is equivalent to the statement that  $f * \mu(B)(x)$  is at least  $\theta^3/2$  for every  $x \in \mathrm{SO}(3)$ . But that implies that  $f * f * \mu_B(x)$  is at least  $(\theta^3/2)\mathbb{E}_x f(x) \geq \theta^6/4$  for every  $x \in \mathrm{SO}(3)$ .

Now we apply Lemma B.2. For any  $\gamma > 0$  we can choose  $\eta > 0$  (depending on  $\theta$  and  $\epsilon$  but not on  $\delta$ ) such that  $\|f * f - f * f * \mu_B\|_2 \leq \gamma$ . From this it follows that

$$|\langle f * f, f \rangle - \langle f * f * \mu_B, f \rangle| \leq \gamma \|f\|_2.$$

But  $\langle f * f * \mu_B, f \rangle \geq \theta^9/8$  by the estimates above, so for suitable choice of  $\gamma$  we can ensure that  $\langle f * f, f \rangle \geq \theta^9/16$ .

The left-hand side of this inequality is the quantity  $\mathbb{E}_{x,y} f(x)f(y)f(xy)$ . It is non-zero if and only if all of  $x, y$  and  $xy$  are within  $\theta\delta$  of points  $x', y', z'$  of  $\Gamma$ . Since all balls of radius  $\theta\delta$  have the same measure, it follows that the proportion of  $(x', y') \in \Gamma^2$  such that  $x'y'$  is within  $3\theta\delta$  of some point  $z' \in \Gamma$  is at least  $\theta^9/16$ . Replacing  $\theta$  by  $\theta/3$  gives the lemma as stated.  $\square$

We make another observation that uses part of the proof above.

**LEMMA B.4.** *For every  $\epsilon, \theta \in (0, 1)$  there exists  $\delta > 0$  such that for any maximal  $\delta$ -separated subset  $\Gamma$  of  $\mathrm{SO}(3)$ , the proportion of  $z \in \Gamma$  such that  $d(xy, z) \leq \theta\delta$  for at least  $(\theta/3)^6|\Gamma|/8$  pairs  $(x, y) \in \Gamma^2$  is at least  $1 - \epsilon$ .*

*Proof.* Choose  $\eta$ , and therefore  $B$ , as in the proof of the previous lemma. Then the measure of the set of  $x$  such that  $|f * f(x) - f * f * \mu_B(x)| < \theta^6/8$  is at least  $1 - 64\gamma^2\theta^{-12}$ . For each  $x$  in this set, we have  $f * f(x) \geq \theta^6/8$  (since  $f * f * \mu_B$  is always at least  $\theta^6/4$ ). Let us denote this set of ‘popular products’ by  $W$ .

Choosing  $\gamma$  appropriately, we can ensure that the measure of  $W$  is at least  $1 - \epsilon\theta^3/2$ . Since  $|\Gamma_\theta| \geq \theta^3/2$ , it follows that  $|W \cap \Gamma_\theta| \geq (1 - \epsilon)|\Gamma_\theta|$ . From this it follows

that the proportion of  $z \in \Gamma$  such that there exists  $z' \in W$  with  $d(z, z') \leq \theta\delta$  is at least  $1 - \epsilon$ . But if  $z' \in W$ , then  $f * f(z') \geq \theta^6/8$ , which implies that the proportion of  $x \in \Gamma$  such that there exist  $x', y, y'$  with  $d(x, x') \leq \theta\delta$ ,  $y \in \Gamma$ ,  $d(y, y') \leq \theta\delta$  and  $d(x'y', z') \leq \theta\delta$  is at least  $\theta^6/8$ . But that implies that the number of pairs  $(x, y) \in \Gamma$  such that  $d(xy, z) \leq 3\theta\delta$  is at least  $(\theta^6/8)|\Gamma|$ . Replacing  $\theta$  by  $\theta/3$  gives the result as stated.  $\square$

Note that the power  $\theta^6$  makes sense above. Since  $SO(3)$  is three-dimensional, the probability that two random points will be within  $\theta$  of points in  $\Gamma$  should be around  $\theta^3 \cdot \theta^3$ , and we have shown that most of the time we are within a constant of what this random model would predict.

For the partially defined operation  $\circ$  to satisfy the hypothesis of Theorem 1.2, we need in particular that there should be many associative triples—that is, triples  $x, y, z$  such that both  $(x \circ y) \circ z$  and  $x \circ (y \circ z)$  are defined (in which case, as we have noted, they must be equal). This can also be deduced from Lemma B.2, as we now show.

LEMMA B.5. *Let  $\delta, \theta > 0$ , let  $\Gamma$  be a maximal  $\delta$ -separated subset of  $SO(3)$ , let  $U$  and  $V$  be  $\delta$ -separated subsets of  $SO(3)$  and suppose that  $U$  and  $V$  have cardinalities at least  $\alpha\delta^{-3}$  and  $\beta\delta^{-3}$ , respectively. Then*

$$|\{(u, v) \in U \times V : d(uv, \Gamma) \leq 3\theta\delta\}| \geq \alpha\beta\theta^3\delta^{-6}/64.$$

*Proof.* Write  $U_\theta, V_\theta$  and  $\Gamma_\theta$  for the  $\theta\delta$ -expansions of  $U, V$  and  $\Gamma$ . Then  $\Gamma_\theta$  has density at least  $\theta^3/2$ , while  $U_\theta$  and  $V_\theta$  have densities at least  $\alpha\theta^3/2$  and  $\beta\theta^3/2$ , respectively.

Let  $\epsilon = \alpha\beta\theta^{15/2}/32$ . By Lemma B.2, there exists  $\eta > 0$  such that, writing  $\mu_B$  for the characteristic measure of the ball of radius  $\eta$  about the identity, we have that

$$\|U_\theta * V_\theta - U_\theta * V_\theta * \mu_B\| \leq \epsilon,$$

where we have written  $U_\theta$  and  $V_\theta$  for the characteristic functions of the sets  $U_\theta$  and  $V_\theta$ .

Since  $|\Gamma_\theta| \geq \theta^3/2$ , it follows that its characteristic function, which again we write  $\Gamma_\theta$ , has the property that  $\|\Gamma_\theta\|_2 \leq 2\theta^{3/2}$ . Writing  $x \approx_\eta y$  as an abbreviation for  $|x - y| \leq \eta$ , we therefore have

$$\langle U_\theta * V_\theta, \Gamma_\theta \rangle \approx_{2\epsilon\theta^{3/2}} \langle U_\theta * V_\theta * \mu_B, \Gamma_\theta \rangle = \langle U_\theta * V_\theta, \Gamma_\theta * \mu_B \rangle.$$

Recall from the proof of Lemma B.3 that  $\Gamma_\theta * \mu_B$  is bounded below by  $\theta^3/2$  everywhere. It follows that the inner product on the right-hand side is at least  $\alpha\beta\theta^9/8$ , and therefore that the inner product on the left-hand side is at least  $\alpha\beta\theta^9/8 - 2\epsilon\theta^{3/2}$ , which is at least  $\alpha\beta\theta^9/16$ .

Now let  $(u, v) \in U \times V$ . If  $d(uv, \Gamma) > 3\theta$ , then by the triangle inequality the product of the balls of radius  $\theta\delta$  about  $u$  and  $v$  does not intersect  $\Gamma_\theta$ , so the pair  $(u, v)$  contributes nothing to the inner product. And otherwise, since the balls have volume at most  $2\theta^3\delta^3$  and the product of the balls intersects at most one ball of

radius  $\theta\delta$  about a point of  $\Gamma$ , the contribution is at most  $4\theta^6\delta^6$ . It follows that the number of pairs  $(u, v) \in U \times V$  such that  $d(uv, \Gamma) \leq 3\theta$  is at least  $\alpha\beta\theta^3\delta^{-6}/64$ , as claimed.  $\square$

**COROLLARY B.6.** *Let  $\Gamma$  be a maximal  $\delta$ -separated subset of  $\text{SO}(3)$ . Then for at least half of the elements  $y \in \Gamma$  there are at least  $\theta^3\delta^{-3}/128$  elements  $x \in \Gamma$  such that  $d(xy, \Gamma) \leq 3\theta\delta$  and at least  $\theta^3\delta^{-3}/128$  elements  $z \in \Gamma$  such that  $d(yz, \Gamma) \leq 3\theta\delta$ .*

*Proof.* Let  $A$  be the set of all  $y$  for which there are fewer than  $\xi\delta^{-3}$  elements  $x \in \Gamma$  with  $d(xy, \Gamma) \leq 3\theta$ . Let  $|A| = \alpha\delta^{-3}$ . Since  $|\Gamma| \geq \delta^{-3}/2$ , we have by Lemma B.5 that  $\xi\alpha\delta^{-6} \geq \alpha\theta^3\delta^{-6}/128$ .

This appears to place no restriction on  $\alpha$ , but that is because the restriction is hidden. We need  $\delta$  to be small compared with a parameter  $\eta$  in the previous lemma which depends on  $\epsilon$ , which in turn depends on  $\alpha, \beta$  and  $\theta$ . However, for any fixed  $\alpha, \theta$  we obtain the above result for sufficiently small  $\delta$ . In particular, we can obtain it for  $\alpha = 1/4$  and deduce that  $\xi \geq \theta^3/128$ .

A similar argument proves that at most quarter of all  $y$  fail the other property, and we are done.  $\square$

We are now ready to obtain many associative triples.

**COROLLARY B.7.** *Let  $\Gamma$  be a maximal  $\delta$ -separated subset of  $\text{SO}(3)$ . Then there are at least  $\theta^9\delta^{-9}/2^{22}$  triples  $(x, y, z) \in \Gamma^3$  such that both  $x \circ (y \circ z)$  and  $(x \circ y) \circ z$  are defined.*

*Proof.* Let  $y$  satisfy the conclusion of Corollary B.6. Now let  $U = \{x : d(xy, \Gamma) \leq 3\theta\}$  and let  $V = \{yz : d(yz, \Gamma) \leq 3\theta\}$ . (The lack of symmetry between those two definitions is deliberate). Then  $U$  and  $V$  satisfy the assumptions of Lemma B.5 with  $\alpha = \beta = \theta^3/128$ . It follows that there are at least  $\theta^9\delta^{-6}/2^{20}$  elements  $(x, yz)$  of  $U \times V$  such that  $d(xyz, \Gamma) \leq 3\theta\delta$ . But in that case, if we define the operation  $\circ$  using the parameter  $6\theta$  in place of  $\theta$ , then for each such pair  $(x, yz)$  we have that  $z \in V$ , so  $d(yz, \Gamma) \leq 3\theta$ , so  $y \circ z$  is defined, and then

$$d(x(y \circ z), \Gamma) \approx_{3\theta} d(xyz, \Gamma) \leq 3\theta,$$

so  $x \circ (y \circ z)$  is also defined.

Since  $x \in V$ , we have that  $d(xy, \Gamma) \leq 3\theta$ , so  $x \circ y$  is defined, and finally

$$d((x \circ y)z, \Gamma) \approx_{3\theta} d(xyz, \Gamma) \leq 3\theta,$$

so  $(x \circ y) \circ z$  is also defined and equal to  $x \circ (y \circ z)$ .

We can do this for at least  $\delta^{-3}/4$  elements  $y$ , so it follows that there are at least  $\theta^9\delta^{-9}/2^{22}$  associative triples, as claimed.  $\square$

## References

- [ACK17] V. ARTAMONOV, S. CHAKRABARTI and S. KUMAR PAL. Characterizations of highly non-associative quasigroups and associative triples. *Quasigroups and Related Systems* 25 (2017), 1–19
- [Bog39] N. BOGOLIOÛBOFF. Sur quelques propriétés arithmétiques des presque-périodes. *Ann. Chaire Phys. Math. Kiev* 4 (1939), 185–205
- [Bra27] H. BRANDT. Verallgemeinerung des Gruppenbegriffs. *Math. Ann.* 96 (1927), 360–366
- [BGT12] E. BREUILLARD, B.J. GREEN and T.C. TAO. The structure of approximate groups. *Publ. Math. IHES* 116 (2012), 115–221
- [CHPS11] D. CONLON, H. HÀN, Y. PERSON and M. SCHACHT. Weak quasi-randomness for uniform hypergraphs. *Random Struct. Algor.* 40 (2011), 1–38
- [DK74] J. DENÉS and A.D. KEEDWELL. *Latin Squares and their Applications*, Akadémiai Kiadó, Budapest, Hungary, (1974).
- [Dra83] A. DRAPAL. On quasigroups rich in associative triples. *Discrete Math* (3)44 (1983), 251–265
- [FS11] J. FOX, B. SUDAKOV. Dependent random choice. *Random Structures and Algorithms* 38 (2011), 68–99
- [Gow01] W.T. GOWERS. A new proof of Szemerédi’s theorem. *GAFA, Geom. Funct. Anal.* 11 (2001), 465–588
- [Gow11] W.T. GOWERS. Quasirandom groups. *Combin. Probab. Comp.* (3)17 (2011), 363–387
- [Gre20] B. GREEN. *On a conjecture of Gowers and Long*, B. Lond. Math. Soc., to appear, 2020.
- [Hru] E. HRUSHOVSKI. *Metrically approximate subgroups*, preprint. <http://www.ma.huji.ac.il/~ehud/metric-stabilizer.pdf>
- [KS10] P.Y. KOSMANN-SCHWARZBACH and S.F. SINGER. *Spherical Harmonics*, In: Groups and Symmetries. Universitext. Springer, New York, NY, (2010).
- [Lev] E. LEVI. *Symmetric abstract independence relations and the group configuration theorem*, Masters thesis, private communication.
- [Taoa] T. TAO. *Metric entropy analogues of sum set theory*, blog post. <https://terrytao.wordpress.com/2014/03/19/metric-entropy-analogues-of-sum-set-theory/>
- [Tao08] T. TAO. Product set estimates for non-commutative groups. *Combinatorica* (5)28 (2008), 547–594
- [Taob] T. TAO. *1% quasimorphisms and group cohomology*, blog post. <https://terrytao.wordpress.com/2018/07/07/1-quasimorphisms-and-group-cohomology/>

W. T. GOWERS

Collège de France, 11, place Marcelin-Berthelot, 75231 Paris cedex 05, France.

[timothy.gowers@college-de-france.fr](mailto:timothy.gowers@college-de-france.fr)



J. LONG

Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter Woodstock Road, Oxford OX2 6GG, UK.

[jasonlong272@gmail.com](mailto:jasonlong272@gmail.com)

Received: November 5, 2019

Revised: August 21, 2020

Accepted: October 2, 2020