Quasirandomness, Counting and Regularity for 3-Uniform Hypergraphs

W. T. Gowers

Abstract. *The main results of this paper are regularity and counting lemmas for 3-uniform hypergraphs. A combination of these two results gives a new proof of a theorem of Frankl and Rödl, of which Szemerédi's theorem for arithmetic progressions of length 4 is a notable consequence. Frankl and Rödl also prove regularity and counting lemmas, but the proofs here, and even the statements, are significantly different. Also included in this paper is a proof of Szemerédi's regularity lemma, some basic facts about quasirandomness for graphs and hypergraphs, and detailed explanations of the motivation for the definitions used.*

§1. Introduction.

One of the most important tools in extremal graph theory is Szemerédi's regularity lemma. Given a graph G with vertex set V, the regularity lemma provides a partition of V into sets $V_1, \dots, V_K$, such that for almost every pair (i, j) the induced bipartite graph $G(V_i, V_j)$ (that is, the restriction of G to the set of edges xy such that $x \in V_i$ and $y \in V_j$) behaves like a typical random graph of the same density. This result has been such an important tool in graph theory and has so many applications that it is very natural to try to find a generalization that will help with extremal problems for hypergraphs. Indeed, there is an application that, even on its own, provides sufficient motivation for a project of this kind. In 1976, Ruzsa and Szemerédi discovered a simple way to deduce Roth's theorem (Szemerédi's theorem for progressions of length 3) from the regularity lemma [RS], and for many years, Vojta Rödl, with several collaborators, has had a very promising programme for obtaining a new proof of the full Szemerédi theorem by generalizing the Ruzsa-Szemerédi argument. In order to do this, one needs not just a regularity lemma but also a so-called "counting lemma", which says, very roughly, that a sufficiently quasirandom hypergraph will contain small subhypergraphs with approximately the correct frequency. (This way of putting it, though it will do for now, is in fact so rough as to be positively misleading - see §5 for an explanation of why.) In 2002, Frankl and Rödl carried out this programme for 3-uniform hypergraphs, thereby giving a new proof of Szemerédi's theorem in the case of progressions of length four [FR2]. Ten years earlier, they had proved a regularity lemma for k-uniform hypergraphs [FR1], but obtaining an appropriate counting lemma to go with it was much harder.

The main purpose of this paper is expository: we shall explain the ideas that lie behind a paper of the author [G3], the main advance of which is to establish a counting lemma for k-uniform hypergraphs, but also a regularity lemma that is not quite the same as that of Frankl and Rödl. These two results imply not just Szemerédi's theorem in full generality, but also, as has already been observed by Solymosi [S2], its multidimensional version. This result had previously been obtained only by Furstenberg's ergodic theory approach [FK]. To explain how these results are proved, we shall concentrate on the case of 3-uniform hypergraphs, since the generalization to k-uniform hypergraphs, though notationally more complicated, is not different in an essential way. The main results of this paper are therefore not new, but the point is that the definitions and proofs are different from those of Frankl and Rödl and more readily generalized. Thus, the principal novelty of this paper occurs at the technical level. However, since this is an area where it is easy to become overwhelmed by technical difficulties, technical simplifications are of more than merely technical interest. That said, Nagle, Rödl and Schacht have also recently generalized the Frankl-Rödl approach to k-uniform hypergraphs [NRS], in work independent of the work discussed here. In the final section of this paper, we explain what the main difference is between the Frankl-Rödl approach and the one given here.

The methods of this paper have their roots in part of the analytic proof of Szemerédi's theorem given by the author in [G1,2]. Another purpose of the paper is to explain this connection. We shall occasionally assume a nodding acquaintance with the ideas of [G1] (which are not used here in any formal way, but which shed light on some of our arguments): otherwise this paper is self-contained.

The reader is urged not to be put off by the length of the paper. The main results - counting and regularity lemmas for 3-uniform hypergraphs - are dealt with in §6 and §8 respectively. The rest of the paper consists of discussion, motivating examples and well-known background results in graph theory. Even the sections with the main results contain quite a bit of discussion, rather than being written as densely as possible. If you are looking for a short proof of Szemerédi's theorem for progressions of length four, then the "true" length of this paper is shorter than that of other papers that establish the same result. Almost all the proofs are straightforward applications of the second-moment method, otherwise known as the Cauchy-Schwarz inequality.

To begin with, here is a brief sketch of a variant of the argument of Ruzsa and Szemerédi that was the starting point for Rödl's programme. Their first step was to prove the following simple-looking statement about graphs.

Theorem 1.1. *For every constant $c > 0$ there exists a constant $a > 0$ with the following property. If G is any graph with n vertices that contains at most an^3 triangles, then it is possible to remove at most cn^2 edges from G to make it triangle-free.*

Sketch Proof. For anybody with experience of the regularity lemma, this is an easy and standard argument (but it was of course a serious achievement of Ruzsa and Szemerédi to notice that this kind of result was both easy and significant). First, apply the regularity lemma to obtain a $c/4$-regular partition of G into vertex sets $V_1, \ldots, V_K$ of almost equal size. Remove all edges that belong to pairs (V_i, V_j) that fail to be $c/4$-regular or that have density at most $c/2$. The result is a subgraph G' of G such that every edge belongs to a pair (V_i, V_j) that is $c/4$ regular and has density at least $c/2$, and the number of edges we have removed from G to achieve this is less than cn^2.

It remains to show that G' is triangle-free. But if xyz is a triangle in G' then we must have $x \in V_i$, $y \in V_j$ and $z \in V_k$ with the pairs (V_i, V_j), (V_j, V_k) and (V_i, V_k) all $c/4$ regular and of density at least $c/2$. It is not hard to deduce from this that the number of triangles in G is at least $c^3N^3/256K^3$. Since K depends on c only, the result is proved. □

Remark. In the above argument we are allowing i, j and k to coincide, so strictly speaking the triangles we obtain should have labellings on their vertices. But this affects the bound by a factor of at most 6. It is more common to insist that K is large, so that there are very few edges joining vertices in the same V_i and one can afford to remove them.

The bound one gets for the dependence of a on c in Theorem 1.1 is extremely weak, and it is a fascinating problem to find a proof that does not use the regularity lemma and therefore, one hopes, gives a better bound. The reason this would be more than a minor curiosity is that Theorem 1.1 implies Roth's theorem - that is, Szemerédi's theorem in the case of progressions of length three. There are several ways to demonstrate this - the way we give here is a small modification of an argument of Solymosi [S1], which is a precursor to the argument mentioned earlier. He was the first to observe that one could obtain the following two-dimensional statement as well.

Corollary 1.2. *For every $\delta > 0$ there exists N such that every subset $A \subset [N]^2$ of size at least δN^2 contains a triple of the form (x, y), $(x + d, y)$, $(x, y + d)$ with $d > 0$.*

Proof. First, note that an easy argument allows us to replace A by a set B that is symmetric about some point. Briefly, if the point (x, y) is chosen at random then the intersection of A with $(x, y) - A$ has expected size $c\delta^2N^2$ for some absolute constant $c > 0$, lives inside the grid $[-N, N]^2$, and has the property that $B = (x, y) - B$. So B is still reasonably dense, and if it contains a subset K then it also contains a translate of $-K$. So we shall not worry about the condition $d > 0$. (I am grateful to Ben Green for bringing this trick to my attention.)

Without loss of generality, the original set A is symmetric in this sense. Let X be the set of all vertical lines through $[N]^2$, that is, subsets of the form $\{(x, y) : x = u\}$ for some $u \in [N]$. Similarly, let Y be the set of all horizontal lines. Define a third set, Z, of diagonal lines, that is, lines of constant $x + y$. These sets form the vertex sets of a tripartite graph, where a line in one set is joined to a line in another if and only if their intersection belongs to A. For example, the line $x = u$ is joined to the line $y = v$ if and only if $(u, v) \in A$ and the line $x = u$ is joined to the line $x + y = w$ if and only if $(u, w - u) \in A$.

Suppose that the resulting graph G contains a triangle of lines $x = u$, $y = v$, $x+y = w$. Then the points (u, v), $(u, w - u)$ and $(w - v, v)$ all lie in A. Setting $d = w - u - v$, we can rewrite them as (u, v), $(u, v + d)$, $(u + d, v)$, which shows that we are done unless $d = 0$. When $d = 0$, we have $u + v = w$, which corresponds to the degenerate case when the vertices of the triangle in G are three lines that intersect in a single point. Clearly, this can happen in at most $|A| = o(N^3)$ ways.

Therefore, if A contains no configuration of the desired kind, then the hypothesis of Theorem 1.1 holds, and we can remove $o(N^2)$ edges from G to make it triangle-free. But this is a contradiction, because there are at least δN^2 degenerate triangles and they are edge-disjoint. □

It is straightforward to deduce Roth's theorem from the above result. Note that for this deduction we do not mind if the d obtained in Corollary 1.2 is negative.

Corollary 1.3. *For every $\delta > 0$ there exists N such that every subset A of $\{1, 2, \ldots, N\}$ of size at least δN contains an arithmetic progression of length 3.*

Proof. Define $B \subset [N]^2$ to be the set of all (x, y) such that $x-y \in A$. It is straightforward to show that B has density at least $\eta > 0$ for some η that depends on δ only. Applying Corollary 1.2 to B we obtain inside it three points (x, y), $(x+d, y)$ and $(x, y+d)$. Then the three numbers $x-y-d$, $x-y$, $x+d-y$ belong to A and form an arithmetic progression. □

Rödl's programme, outlined in [R], was to generalize Theorem 1.1 to hypergraphs, using generalized regularity and counting lemmas to prove it. The various ways of deducing Roth's theorem from Theorem 1.1 can then be straightforwardly modified to give deductions of the full Szemerédi theorem.

To be more precise about this, let H be a 3-uniform hypergraph. By a *simplex* in H we mean a collection of four edges of the form $\{xyz, xyw, xzw, yzw\}$, that is, a complete subhypergraph on four vertices. If one thinks of the edges of H as (two-dimensional) triangles, then the "edges" of a simplex can be thought of as the faces of a tetrahedron. However one looks at it, this is the natural generalization of the notion of a triangle in a graph. The result of Frankl and Rödl mentioned earlier is the following.

Theorem 1.4. *For every constant $c > 0$ there exists a constant $a > 0$ with the following property. If H is any 3-uniform hypergraph with n vertices that contains at most an^4 simplices, then it is possible to remove at most cn^3 edges from H to make it simplex-free.*

As Solymosi demonstrated, it is easy to adapt the proof of Theorem 1.2 and show that Theorem 1.4 has the following consequence.

Theorem 1.5. *For every $\delta > 0$ there exists N such that every subset $A \subset [N]^3$ of size at least δN^3 contains a quadruple of points of the form*

$$\{(x, y, z), (x + d, y, z), (x, y + d, z), (x, y, z + d)\}$$

with $d > 0$.

Similarly, Szemerédi's theorem for progressions of length four is an easy consequence of Theorem 1.5.

Thus, once one has appropriate generalizations of the regularity and counting lemmas to hypergraphs, the rest of the argument goes through quite easily. However, as will become clear over the next three sections, even to come up with the right statements is harder than one might at first think. To begin with, however, we shall provide some essential background by discussing several notions of quasirandomness and the relationships between them.

§2. Quasirandom graphs, hypergraphs and subsets of $\mathbb{Z}_N$: some definitions.

Every known proof of Szemerédi's theorem involves somewhere a notion of quasirandomness and a two-case argument of the following kind: if a certain structure is quasirandom then it contains several configurations of the kind one is looking for (just as one expects from a random structure), and if it is not then one can exploit the non-quasirandomness and pass to the next stage of an iteration. In this section we review certain notions of quasirandomness and point out connections between them (some of which are well known).

The first is the definition of a quasirandom graph, which was introduced by Chung, Graham and Wilson [CGW]. (A similar notion was discovered independently by Thomason [T].) There are in fact several different definitions, and the main purpose of their paper was to show that they are all equivalent. Here we restrict attention to bipartite graphs and focus on just two of the definitions. (Chung and Graham state their results in the case $p = 1/2$ only but the generalization to arbitrary p is not hard.) A proof of a slightly modified result will be given in the next section.

Theorem 2.1. *Let G be a bipartite graph with vertex sets X and Y. Let $|X| = M$ and $|Y| = N$ and suppose that G has pMN edges. Then the following properties of G are equivalent.*

(i) The number of labelled 4-cycles in G that start in X (that is, quadruples $(x_1, x_2, y_1, y_2) \in X^2 \times Y^2$ such that x_1y_1, x_1y_2, x_2y_1 and x_2y_2 are all edges of G) is at most $p^4M^2N^2 + c_1M^2N^2$.

(ii) If X' and Y' are any two subsets of X and Y respectively, then the number of edges from X' to Y' differs from $p|X||Y|$ by at most c_2MN.

A graph that satisfies property (i) is often called *α-quasirandom.* For the time being, we shall adopt this definition, though in the next section we shall choose a different, equivalent definition that is more convenient.

Chung and Graham [CG2] went on to define a notion of quasirandomness for subsets of $\mathbb{Z}_N$. Again, they gave the definition in the form of a theorem that asserted the equivalence of various randomness properties, and once again we shall focus on just two of these and consider the case of an arbitary "probability" p rather than just $p = 1/2$. By a *mod-N progression* we mean a set of the form $(a, a+d, \dots, a+(m-1)d)$, where addition is in the group $\mathbb{Z}_N$.

Theorem 2.2. *Let A be a subset of $\mathbb{Z}_N$ of size pN. Then the following properties of A are equivalent.*

(i) The number of quadruples $(a, b, c, d) \in A^4$ such that $a + b = c + d$ is at most $p^4N^4 + o(N^4)$.

(ii) If X is any mod-N progression then $|A \cap X| = p|X| + o(N)$.

Chung and Graham also pointed out that there is a close connection between quasirandom subsets of $\mathbb{Z}_N$ and quasirandom graphs. To see this, let A be a subset of $\mathbb{Z}_N$ and define a bipartite graph G with vertex sets $X = Y = \mathbb{Z}_N$ by letting $(x, y) \in X \times Y$ be an edge if and only if $x+y\in A$. Then, given a 4-cycles (x_1,x_2,y_1,y_2) of the kind counted in property (ii) of Theorem 2.1, we know that x_1+y_1, x_1+y_2, x_2+y_1 and x_2+y_2 all belong to A, and moreover that

$$(x_1+y_1)+(x_2+y_2)=(x_1+y_2)+(x_2+y_1)\ ,$$

from which it is easy to see that there is an N-to-one correspondence between the 4-cycles from Theorem 2.1 (i) and the quadruples from Theorem 2.2 (i). Thus, the set A is quasirandom if and only if the corresponding graph G is quasirandom.

Why are quasirandom sets useful for Szemerédi's theorem? Briefly, the reason is that a quasirandom set A of density p must contain approximately the same number of arithmetic progressions of length 3 as a typical random set of that density, while if it fails to be quasirandom, then property (ii) of Theorem 2.2 tells us that there are well-structured subsets of $\mathbb{Z}_N$ inside which A has density substantially different from p (and hence, by an averaging argument, sometimes substantially larger). This makes possible an iteration argument of the kind mentioned at the beginning of this section.

In all known proofs of Szemerédi's theorem, the difficulty increases sharply when the length of the progression increases from 3 to 4. In the analytic approach of [G1,2], the reason for this difficulty is that the quasirandomness of Theorem 2.2 is not a sensitive enough property to detect that a set has too many or too few arithmetic progressions of length 4 (compared with a random set of the same density). The idea of [G1] is to define a stronger property, known there as *quadratic uniformity*, which *is* sensitive to progressions of length 4. However, because it is a stronger property, it is much harder to say anything about sets that fail to be quadratically uniform - a point to which we shall return.

The main purpose of this section, however, is to show how the notion of quadratic uniformity leads naturally to a definition, also due to Chung and Graham [CG1] (but arrived at in a different way), of quasirandomness for 3-uniform hypergraphs. First, then, here is the definition of a quadratically uniform subset of $\mathbb{Z}_N$.

Definition 2.3. *Let $\alpha>0$. A subset $A\subset\mathbb{Z}_N$ of size pN is* α-quadratically uniform *if A^8 contains at most $(p^4+\alpha)N^4$ octuples of the form*

$$(x,x+a,x+b,x+c,x+a+b,x+a+c,x+b+c,x+a+b+c)\ .$$

Since quadruples (a,b,c,d) with $a+b=c+d$ are in one-to-one correspondence with quadruples of the form $(x,x+a,x+b,x+a+b)$, this definition is a natural generalization of property (i) of Theorem 2.2.

In order to define quasirandomness for 3-uniform hypergraphs, we shall "complete the square", by finding a property that stands in relation to quadratic uniformity as quasirandomness for graphs does to quasirandomness for subsets of $\mathbb{Z}_N$. To establish that connection we started with a set A and defined the bipartite graph G to consist of all pairs (x,y) such that $x+y\in A$. Now we would like to find an associated 3-uniform hypergraph, and it seems sensible to try the tripartite hypergraph H with vertex sets $X=Y=Z=\mathbb{Z}_N$, with the triple (x,y,z) forming an edge of H if and only if $x+y+z\in A$.

In order to generalize Definition 2.3, we now want to find a structure in H that corresponds to octuples of the given kind. It is natural to think of these octuples as labelling vertices of a cube (as they would if a, b and c denoted three orthogonal vectors of the same length). It is also natural to think of the edges of a 3-uniform hypergraph as triangles. Is there a configuration of triangles naturally associated with a cube?

Indeed there is: the dual of a cube is an octahedron and an octahedron is made of triangles. This suggests that we should regard a quadruple $(x,x+a,x+b,x+a+b)$ in $\mathbb{Z}_N$ as a kind of square, and a 4-cycle (x_1,x_2,y_1,y_2) in a bipartite graph as its dual, which happens, confusingly, to look like a square as well. (This dual square gives rise to the square $(x_1+y_1,x_1+y_2,x_2+y_1,x_2+y_2)$ in $\mathbb{Z}_N$.) The points x_i and y_i of the 4-cycle are vertices, so the vertices of the square in $\mathbb{Z}_N$ correspond to (1-dimensional) faces, and the value attached to a face is the sum of the values at the vertices. This gives us a construction that generalizes easily to all dimensions.

The above remarks provide the justification for the following construction. Given a subset $A\subset\mathbb{Z}_N$, define a tripartite 3-uniform hypergraph H with vertex sets $X=Y=Z=\mathbb{Z}_N$ to be the set of all triples (x,y,z) such that $x+y+z\in A$. Define an *octahedron* in this (or any other) hypergraph to be a set of eight 3-edges of the form $\big\{(x_i,y_j,z_k):i,j,k\in\{1,2\}\big\}$, where $x_1,x_2\in X$, $y_1,y_2\in Y$, $z_1,z_2\in Z$. Equivalently, an octahedron is a complete tripartite subhypergraph with two vertices from each vertex set of H.

It is now reasonable to guess that a good definition of quasirandomness for 3-uniform hypergraphs is the following.

Definition 2.4. *Let H be a tripartite 3-uniform hypergraph with vertex sets X, Y and Z of size L, M and N respectively and suppose that H has $pLMN$ edges. Then H is* α-quasirandom *if it contains at most $(p^8+\alpha)L^2M^2N^2$ octahedra.*

This suggestion can be motivated further (to do so one attempts to generalize arguments about graphs and sees what one needs) but for now it is perhaps enough simply to say that the guess turns out to be correct, and to point out that the number of octahedra must be at least $p^8L^2M^2N^2$. (This last observation follows easily from the Cauchy-Schwarz inequality, as will become clear later in this paper.)

A first sign that the definition is a good one is that the eight numbers $x_i + y_j + z_k$ do form a cube, in the sense discussed earlier. It follows that a subset A of $\mathbb{Z}_N$ gives rise to an α-quasirandom 3-graph H if and only if A is α-quadratically uniform.

As it happens, there is a different definition (Definition 4.3 below) that is even better. However, it is equivalent apart from the precise value of α, and the advantage is a technical one, so for the purposes of the discussion in this section we shall stick with the more obvious Definition 2.4.

Theorem 2.1 stated that two quasirandomness properties of graphs are equivalent. We have just generalized the second of these properties. What about the first? It is here that hypergraph quasirandomness springs a surprise. The property that most obviously generalizes property (ii) of Theorem 2.1 is the following, which we shall call vertex-uniformity.

Definition 2.5. *Let H be a 3-uniform hypergraph with vertex sets X, Y and Z of sizes L, M and N respectively and suppose that H has $pLMN$ edges. Then H is* β-vertex-uniform *if, for any choice of subsets $X' \subset X$, $Y' \subset Y$ and $Z' \subset Z$, the number of triples $(x, y, z) \in X' \times Y' \times Z'$ that belong to H differs from $p|X'||Y'||Z'|$ by at most βLMN.*

In the next section we shall see that an α-quasirandom 3-graph is β-vertex-uniform for some β depending on α only, but the reverse is not true. In §4 we shall give two examples that demonstrate the failure of this implication. We end this section with one last definition, of a property that generalizes property (ii) of Theorem 2.1 in a less naive way. This property *will* turn out to be equivalent to quasirandomness - another result to be proved in the next section.

Definition 2.6. *Let H be a 3-uniform hypergraph with vertex sets X, Y and Z of sizes L, M and N respectively and suppose that H has $pLMN$ edges. Then H is* γ-edge-uniform *if for every $t \in [0, 1]$ and every tripartite graph G with vertex sets X, Y and Z and $tLMN$ triangles, the number of triangles in G that belong to H differs from $ptLMN$ by at most γLMN.*

Here, of course, a triangle in G is said to belong to H if its vertices form a triple that belongs to H.

Why is this a generalization of property (ii)? Well, that property says of a bipartite graph that it doesn't significantly correlate with graphs that are induced by sets of vertices (that is, complete bipartite graphs on subsets of the vertex sets). Edge-uniformity says of a 3-uniform hypergraph that it doesn't correlate with 3-uniform hypergraphs that are induced by sets of edges (as opposed to hyperedges).

Although in this paper we are concentrating on 3-uniform hypergraphs, it should be clear how the definitions of this section are generalized. A k-partite k-uniform hypergraph H of density p with vertex sets X_i of size N_i is quasirandom if it contains at most $(p^{2^k} + c)(N_1 \dots N_k)^2$ k-dimensional octahedra and c is small; this turns out to be equivalent to the assertion that H is $((k-1)$-edge)-uniform, in the sense that H does not significantly correlate with any k-uniform hypergraph induced from a $(k-1)$-uniform hypergraph. Using this kind of language, one could say that a quasirandom graph is a 2-uniform hypergraph that is (1-edge)-uniform. The right generalization of the pair $(2, 1)$ is not $(k, 1)$ as one might at first suppose, but $(k, k-1)$. Thus, Definition 2.6 is more useful to us than Definition 2.5.

§3. Quasirandom functions and a counting lemma.

Let G be a tripartite graph with vertex sets X, Y and Z, of sizes L, M and N respectively, and write $G(X,Y)$, $G(Y,Z)$ and $G(X,Z)$ for the three bipartite parts of G. Suppose that these parts are quasirandom with probabilities p, q and r respectively. How many triangles does G contain?

If we use genuinely random graphs as our guide, we should expect the answer to be about $pqrLMN$, and indeed it is easy to use property (ii) of Theorem 2.1 to prove that this is approximately right: a typical vertex in X has about pM neighbours in Y and rN neighbours in Z; these two neighbourhoods are linked by about $q(pM)(rN)$ edges; summing over all $x \in X$ we obtain the desired estimate.

The argument just sketched can be generalized quite easily from triangles to copies of any small graph, and this generalization is what we shall refer to as the *counting lemma* for graphs. The first aim of this section will be to give it a different and less transparent proof. Why should we wish to do something so apparently perverse? Because the alternative proof has a number of advantages: first, it is closely modelled on the analytic arguments of [G1] and related arguments from Furstenberg's ergodic-theory proof of Szemerédi's theorem (see for example [FKO]); more importantly, this analytic approach is much easier to generalize, an advantage that is very noticeable even for "the first non-trivial case", that of 3-uniform hypergraphs. Thus, the reader who is prepared to make the small effort needed to understand the proof of Theorem 3.5 below will understand the basic structure of the longer arguments of §4 and §6, and also that of [G3]: the main lemmas and arguments there all have their prototypes here.

An important technicality in the analytic approach to these arguments is to think of sets as $\{0,1\}$-valued functions and to generalize set-theoretic arguments to functions taking values in more general sets such as $[0,1]$, $[-1,1]$, $\mathbb{R}_+$, $\mathbb{R}$, $\{z \in \mathbb{C} : |z| \leqslant 1\}$ or $\mathbb{C}$. We shall do that now, proving results about $[-1,1]$-valued functions and only occasionally pausing to deduce results about graphs and hypergraphs. Our first result is a version of Theorem 2.1 for functions. It is not quite a direct generalization of Theorem 2.1 because the conclusion is generalized to functions as well as the hypothesis. However, as we shall see, if G is regular then it is easy to deduce Theorem 2.1 from it. We shall not bother with the irregular case, because we shall base our later arguments on Theorem 3.1 and not use Theorem 2.1.

Theorem 3.1. *Let X and Y be sets of sizes M and N respectively and let $f : X \times Y \to [-1,1]$. Then the following statements are equivalent.*

(i) $\sum_{x,x' \in X} \sum_{y,y' \in Y} f(x,y)f(x',y)f(x,y')f(x',y') \leqslant c_1 M^2 N^2$.

(ii) For any pair of functions $u : X \to [-1,1]$ and $v : Y \to [-1,1]$ we have the inequality $\left|\sum_{x,y} f(x,y)u(x)v(y)\right| \leqslant c_2 MN$.

(iii) For any pair of sets $X' \subset X$ and $Y' \subset Y$ we have the inequality $\left|\sum_{x \in X'} \sum_{y \in Y'} f(x,y)\right| \leqslant c_3 MN$.

Proof. We shall begin with a very simple argument that shows that (ii) implies (i). Let us assume that (i) is false, or in other words that

$$\sum_{x,x' \in X} \sum_{y,y' \in Y} f(x,y)f(x',y)f(x,y')f(x',y') > c_1 M^2 N^2 .$$

Now choose $x' \in X$ and $y' \in Y$ randomly and independently and fix them. Then the average of the sum $\sum_{x,y} f(x,y)f(x',y)f(x,y')f(x',y')$ is greater than $c_1 MN$, so the sum itself is greater than $c_1 MN$ for at least one choice of x' and y'. But then (ii) is false if we set $c_2 = c_1$, $u(x) = f(x,y')$ and $v(y) = f(x',y)f(x',y')$. (What is important here is that fixing x' and y' turns all the terms except for $f(x,y)$ into constants or functions of one variable.)

The reverse implication uses the Cauchy-Schwarz inequality several times. It is this technique that, when suitably generalized, lies at the heart of the proof of the counting lemma for hypergraphs. (To be more precise, what we keep using is not the Cauchy-Schwarz inequality directly, but the inequality $\left(\sum_{i=1}^N a_i\right)^2 \leqslant N \sum_{i=1}^N a_i^2$, which follows from it.) In the expressions that follow, sums involving x and x' are over X and sums involving y and y' are over Y.

$$\begin{aligned}
\Big|\sum_{x,y} f(x,y)u(x)v(y)\Big|^4 &= \Big(\Big(\sum_x\sum_y f(x,y)u(x)v(y)\Big)^2\Big)^2 \\
&\leqslant \Big(M\sum_x\Big(\sum_y f(x,y)u(x)v(y)\Big)^2\Big)^2 \\
&\leqslant \Big(M\sum_x\Big(\sum_y f(x,y)v(y)\Big)^2\Big)^2 \\
&= M^2\Big(\sum_x\sum_{y,y'} f(x,y)f(x,y')v(y)v(y')\Big)^2 \\
&\leqslant M^2N^2\sum_{y,y'}\Big(\sum_x f(x,y)f(x,y')v(y)v(y')\Big)^2 \\
&\leqslant M^2N^2\sum_{y,y'}\Big(\sum_x f(x,y)f(x,y')\Big)^2 \\
&= M^2N^2\sum_{x,x'}\sum_{y,y'} f(x,y)f(x,y')f(x',y)f(x',y')
\end{aligned}$$

These calculations show that if (i) is true then (ii) is true for $c_2 = c_1^{1/4}$.

It is obvious that (ii) implies (iii), since one can take u and v to be the characteristic functions of X' and Y' respectively, so it remains to prove that (iii) implies (ii).

Suppose, then, that (ii) is false and we have functions $u : X \to [-1,1]$ and $v : Y \to [-1,1]$ such that $\big|\sum_{x,y} f(x,y)u(x)v(y)\big| \geqslant c_2MN$. We can write $u = u_+ - u_-$ with u_+ and u_- disjointly supported and taking values in $[0,1]$, and similarly we can write $v = v_+ - v_-$. It follows that there are $[0,1]$-valued functions s and t such that $\big|\sum_{x,y} f(x,y)s(x)t(y)\big| \geqslant c_1MN/4$. Now let X_1 and Y_1 be random subsets of X and Y respectively, with their elements chosen independently with probabilities given by the functions s and t. Then the expectation of $\sum_{x\in X_1}\sum_{y\in Y_1} f(x,y)$ is $\sum_{x,y} f(x,y)s(x)t(y)$, so there must exist a choice of X_1 and Y_1 such that $\big|\sum_{x\in X_1}\sum_{y\in Y_1} f(x,y)\big|$ is at least $c_1MN/4$. If this sum is positive, then we are done. Otherwise, let $X_2 = X \setminus X_1$ and $Y_2 = Y \setminus Y_1$, and for $i,j \in \{1,2\}$ let $S_{ij} = \sum_{x\in X_i}\sum_{y\in Y_j} f(x,y)$. Then $S_{11}+S_{12}+S_{21}+S_{22} = 0$ from which it follows that either there exists a pair $(i,j) \neq (1,1)$ such that $S_{ij} \geqslant c_1MN/12$. Whatever happens, we have found a pair of sets $X' \subset X$ and $Y' \subset Y$ such that $\Big|\sum_{x\in X'}\sum_{y\in Y'} f(x,y)\Big| > c_2MN/12$, contradicting (iii) when $c_3 \leqslant c_2/12$. □

Theorem 3.1 motivates the following definition.

Definition 3.2. *Let X and Y be sets of size M and N. A function f is* α-quasirandom *if* $\sum_{x,x'\in X}\sum_{y,y'\in Y} f(x,y)f(x',y)f(x,y')f(x',y') \leqslant \alpha M^2N^2$.

If G is a bipartite graph with vertex sets X and Y, let us write $G(x,y)$ for the function that is 1 if xy is an edge of G and 0 otherwise. Suppose that $|X| = M$, $|Y| = N$ and every vertex in X has degree pN, and set $f(x,y) = G(x,y) - p$. Then it is easy to verify that

$$\sum_{x,x',y,y'} G(x,y)G(x,y')G(x',y)G(x',y')$$
$$= \sum_{x,x',y,y'} f(x,y)f(x,y')f(x',y)f(x',y') + p^4M^2N^2 \ .$$

It follows that G is α-quasirandom, in the sense of §2, if and only if f is α-quasirandom. From now on we shall adopt this as our definition of quasirandomness even in the non-regular case.

Definition 3.3. *Let G be a bipartite graph with vertex sets X and Y of size M and N and suppose that G has pMN edges. Then G is* α-quasirandom *if the function $f(x,y) = G(x,y) - p$ is α-quasirandom.*

Notice also that, with this definition of f,

$$\sum_{x\in X'}\sum_{y\in Y'} G(x,y) = \sum_{x\in X'}\sum_{y\in Y'} f(x,y) + p|X'||Y'| \ ,$$

so property (iii) of Theorem 3.1 and property (ii) of Theorem 2.1 are trivially equivalent. We have therefore proved Theorem 2.1 in the case of regular graphs.

Now let us state and prove a counting lemma for graphs. We begin with the special case of triangles in order to demonstrate the argument without getting tied up with notation. It is also the case of most immediate interest.

Lemma 3.4. *Let G be a tripartite graph with vertex sets X, Y and Z, of sizes L, M and N respectively. Suppose that the bipartite graphs $G(X,Y)$, $G(Y,Z)$ and $G(X,Z)$ are α-quasirandom with densities p, q and r respectively. Then the number of triangles in G differs from $pqrLMN$ by at most $4\alpha^{1/4}LMN$.*

Proof. Let the variables x, y and z always stand for elements of X, Y and Z respectively, so that we do not keep needing to specify this. Define a function $f : X \times Y \to [-1,1]$ by $f(x,y) = G(x,y) - p$, and similarly let $g(y,z) = G(y,z) - q$ and $h(x,z) = G(x,z) - r$. In terms of this notation, the number of triangles in G is given by the sum

$$\sum_{x,y,z} (p + f(x,y))(q + g(y,z))(r + h(x,z)) \ .$$

This sum splits naturally into eight parts, and the idea of the proof is that if α is small then only the main term $pqrLMN$ makes a significant contribution to it. To see this, let us consider any one of the four terms that involves $f(x,y)$ rather than p. It will have the form $\sum_{x,y,z} f(x,y)u(y,z)v(x,z)$, where u is either q or g and v is either r or h.

If we now fix z, we obtain an expression of the form $\sum_{x,y} f(x,y)u(y)v(z)$. The deduction of property (ii) from property (i) in Theorem 3.1 tells us that this is at most $\alpha^{1/4}LM$, which gives us that $\sum_{x,y,z} f(x,y)u(y,z)v(x,z)$ is at most $\alpha^{1/4}LMN$. There are seven terms other than the main one, of this kind, three of which are easily seen to be zero, so the result follows. □

The general counting lemma is proved in essentially the same way. We shall state it in an equivalent form, as a "probability lemma". Notice that the probability in the conclusion of the lemma is what one would expect in the case of random graphs. The conclusion of the lemma therefore says that there are about as many copies of H in G as one would expect.

Theorem 3.5. *Let G be an m-partite graph with vertex sets $X_1, \ldots, X_m$ and write N_i for the size of X_i. Suppose that for each pair (i,j) the induced bipartite graph $G(X_i, X_j)$ is α-quasirandom with density p_{ij}. Let H be any graph with vertex set $\{1, 2, \ldots, m\}$ and let $(x_1, \ldots, x_m)$ be a random element of $X_1 \times \ldots \times X_m$. Then the probability that the function $i \mapsto x_i$ is an isomorphic embedding of H into G differs from $\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij})$ by at most $2^{\binom{m}{2}} \alpha^{1/4}$. The probability that $x_i x_j$ is an edge of G whenever ij is an edge of H (but not necessarily conversely) differs from $\prod_{ij \in E(H)} p_{ij}$ by at most $2^{|E(H)|} \alpha^{1/4}$.*

Proof. This time, let x_i always stand for an element of X_i. Let $f_{ij}(x,y) = G(x,y) - p_{ij}$ for each pair (i,j). Then the probability in question is

$$(N_1 \ldots N_m)^{-1} \sum_{x_1, \ldots, x_m} \prod_{ij \in E(H)} (p_{ij} + f_{ij}(x_i, x_j)) \prod_{ij \notin E(H)} (1 - p_{ij} - f_{ij}(x_i, x_j)) \ .$$

Once again, we have a sum that splits up into several terms. The main term is $\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij})$ and it remains to show that all other terms are small. But any other term must choose $f_{ij}(x_i, x_j)$ from at least one bracket, and since only one bracket involves both x_i and x_j, if we fix all the other x_k we obtain an expression of the form $\sum_{x_i, x_j} f_{ij}(x_i, x_j) u(x_i) v(x_j)$, with u and v taking values in the interval $[-1, 1]$. It follows from the α-quasirandomness of f_{ij} and Theorem 3.1 that this is at most $\alpha^{1/4} N_i N_j$. Summing over the other $m-2$ variables and multiplying by $(N_1 \ldots N_m)^{-1}$ we find that each term apart from the main one has size at most $\alpha^{1/4}$. Since there are $2^{\binom{m}{2}}$ terms, the result follows (and could be improved slightly since some of the terms are zero).

The proof of the second assertion is similar, but slightly simpler. $\square$

§4. A counting lemma for quasirandom 3-graphs.

The virtue of the arguments in §3 is that they generalize easily. Our first demonstration of this takes the form of very similar proofs of corresponding results for quasirandom 3-uniform hypergraphs. As in §3, we begin with a result about functions that serves as a definition of quasirandomness. The statement is very similar to that of Theorem 3.1, and the proofs of the implications are also very similar to the corresponding proofs in Theorem 3.1. Properties (i) and (iii) below are functional versions of Definitions 2.4 and 2.6 respectively.

During part of the proof, we shall make use of a non-standard but very convenient "product convention". If g is any function of k variables $x_1, \ldots, x_k$, then $g_{x,x'}(x_2, \ldots, x_k)$ will be shorthand for $g_x(x_2, \ldots, x_k)g_{x'}(x_2, \ldots, x_k)$. What's more, we shall iterate this, writing $g_{x,x',y,y'}$ for $(g_{x,x'})_{y,y'}$ and so on. For instance, if g is a function of three variables, then

$$g_{x,x',y,y'}(z) = g(x, y, z)g(x', y, z)g(x, y', z)g(x', y', z) .$$

If we iterate k times, then the resulting function is a function of no variables, that is, a constant. To be precise, $g_{x_1,x'_1,\ldots,x_k,x'_k}$ is the number $\prod_{\epsilon\in\{0,1\}} g\big(u_1(\epsilon), \ldots, u_k(\epsilon)\big)$, where $u_i(\epsilon) = x_i$ if $\epsilon_i = 0$ and $u_i(\epsilon) = x'_i$ if $\epsilon_i = 1$.

Theorem 4.1. *Let X, Y and Z be sets of sizes L, M and N and let $f : X\times Y\times Z \to [-1, 1]$. Then the following statements are equivalent.*

(i) $\sum_{x_0,x_1\in X} \sum_{y_0,y_1\in Y} \sum_{z_0,z_1\in Z} \prod_{(i,j,k)\in\{0,1\}^3} f(x_i, y_j, z_k) \leqslant c_1L^2M^2N^2$.

(ii) For any three functions $u : X \times Y \to [-1, 1]$, $v : Y \times Z \to [-1, 1]$ and $w : X \times Z \to [-1, 1]$ we have the inequality $\left|\sum_{x,y,z} f(x, y, z)u(x, y)v(y, z)w(x, z)\right| \leqslant c_2LMN$.

(iii) For any tripartite graph G with vertex sets X, Y and Z, the sum of $f(x, y, z)$ over all triangles xyz of G is at most c_3LMN in magnitude.

Proof. Assume that (i) is false, so that

$$\sum_{x_0,x_1\in X} \sum_{y_0,y_1\in Y} \sum_{z_0,z_1\in Z} \prod_{(i,j,k)\in\{0,1\}^3} f(x_i, y_j, z_k) > c_1L^2M^2N^2 .$$

Choose x_1, y_1 and z_1 randomly and independently and fix them. Then the average of the sum

$$\sum_{x_0,y_0,z_0} \prod_{(i,j,k)\in\{0,1\}^3} f(x_i, y_j, z_k)$$

is greater than c_1LMN, so the sum itself is greater than c_1LMN for at least one choice of x_1, y_1 and z_1. But then (ii) is false if we set $c_2 = c_1$, $u(x,y) = f(x,y,z_1)$, $v(y,z) = f(x_1,y,z)f(x_1,y_1,z)$ and $w(x,z) = f(x,y_1,z)f(x,y_1,z_1)f(x_1,y_1,z)f(x_1,y_1,z_1)$.

Again, the details do not matter here: the point is that if we write x, y and z for x_0, y_0 and z_0, then $f(x,y,z)$ is the only term in the product that depends on all of x, y and z, so fixing x_1, y_1 and z_1 results in an expression of the form $\sum_{x,y,z} f(x,y,z)u(x,y)v(y,z)w(x,z)$.

Now let us prove the reverse inequality by making repeated use of the Cauchy-Schwarz inequality. This is where we shall use the notation introduced before the statement of the theorem.

$$\begin{aligned}
&\Big(\sum_{x,y,z} f(x,y,z)u(x,y)v(y,z)w(x,z)\Big)^8\\
&\qquad\leqslant \Big(MN\sum_{y,z}\Big(\sum_x f(x,y,z)u(x,y)v(y,z)w(x,z)\Big)^2\Big)^4\\
&\qquad\leqslant M^4N^4\Big(\sum_{y,z}\Big(\sum_x f(x,y,z)u(x,y)w(x,z)\Big)^2\Big)^4\\
&\qquad= M^4N^4\Big(\sum_{x,x'}\sum_{y,z} f(x,y,z)f(x',y,z)u(x,y)u(x',y)w(x,z)w(x',z)\Big)^4\\
&\qquad= M^4N^4\Big(\sum_{x,x'}\sum_{y,z} f_{x,x'}(y,z)u_{x,x'}(y)w_{x,x'}(z)\Big)^4\\
&\qquad\leqslant M^4N^4L^6\sum_{x,x'}\Big(\sum_{y,z} f_{x,x'}(y,z)u_{x,x'}(y)w_{x,x'}(z)\Big)^4 .
\end{aligned}$$

We now perform on the inner sum the steps from the corresponding part of the proof of

Theorem 3.1. The last expression is at most

$$
\begin{aligned}
M^4N^4L^6\sum_{x,x'}\Big(\sum_z\Big(\sum_y f_{x,x'}(y,z)u_{x,x'}(y)w_{x,x'}(z)\Big)^2\Big)^2 \\
&\leqslant M^4N^4L^6\sum_{x,x'}\Big(N\sum_z\Big(\sum_y f_{x,x'}(y,z)u_{x,x'}(y)\Big)^2\Big)^2 \\
&= M^4N^6L^6\sum_{x,x'}\Big(\sum_{y,y'}\sum_z f_{x,x'}(y,z)f_{x,x'}(y',z)u_{x,x'}(y)u_{x,x'}(y')\Big)^2 \\
&= M^4N^6L^6\sum_{x,x'}\Big(\sum_{y,y'}\sum_z f_{x,x',y,y'}(z)u_{x,x',y,y'}\Big)^2 \\
&\leqslant M^4N^6L^6\sum_{x,x'}M^2\sum_{y,y'}\Big(\sum_z f_{x,x',y,y'}(z)u_{x,x',y,y'}\Big)^2 \\
&\leqslant M^6N^6L^6\sum_{x,x'}\sum_{y,y'}\Big(\sum_z f_{x,x',y,y'}(z)\Big)^2 \\
&= M^6N^6L^6\sum_{x,x'}\sum_{y,y'}\sum_{z,z'} f_{x,x',y,y',z,z'}\ .
\end{aligned}
$$

This is another way of writing

$$M^6N^6L^6\sum_{x_0,x_1\in X}\sum_{y_0,y_1\in Y}\sum_{z_0,z_1\in Z}\prod_{(i,j,k)\in\{0,1\}^3} f(x_i,y_j,z_k)\ .$$

It follows that if (i) is true then (ii) is true with $c_2=c_1^{1/8}$.

It is obvious that (ii) implies (iii), since one can take u, v and w to be the characteristic functions of the bipartite graphs $G(X,Y)$, $G(Y,Z)$ and $G(X,Z)$. It therefore remains to prove that (iii) implies (ii).

Suppose, then, that (ii) is false and we have functions $u:X\times Y\to[-1,1]$, $v:Y\times Z\to[-1,1]$ and $w:X\times Z\to[-1,1]$ such that

$$\Big|\sum_{x,y,z} f(x,y,z)u(x,y)v(y,z)w(x,z)\Big| > c_2LMN\ .$$

One can write u as u_+-u_- with u_+ and u_- disjointly supported and taking values in $[0,1]$, and one can do the same for v and w. It follows that there are functions a, b and c taking values in $[0,1]$ such that

$$\Big|\sum_{x,y,z} f(x,y,z)a(x,y)b(y,z)c(x,z)\Big| > c_2LMN/8\ .$$

Now let A_1, B_1 and C_1 be random subsets of $X \times Y$, $Y \times Z$ and $X \times Z$, their elements chosen randomly and independently with probabilities given by the functions a, b and c. Writing A_1, B_1 and C_1 for the characteristic functions of the sets as well, we have that the expectation of

$$\Big|\sum_{x,y,z} f(x,y,z)A_1(x,y)B_1(y,z)C_1(x,z)\Big|$$

is at least $c_2LMN/8$. Choose A_1, B_1 and C_1 such that the absolute value of the sum is at least this big and let the complements of A_1, B_1 and C_1 be A_2, B_2 and C_2. Write S_{ijk} for the sum $\sum_{x,y,z} f(x,y,z)A_i(x,y)B_j(y,z)C_k(x,z)$. Since f sums to zero, we know that the S_{ijk} add up to zero. Since S_{111} has absolute value at least $c_2LMN/8$, it follows that at least one S_{ijk} exceeds $c_2LMN/56$. If we let G be the tripartite graph with edge sets A_i, B_j and C_k, then we have disproved (iii) for any $c_3 \leqslant c_2/56$. □

It is now very natural to make the following pair of definitions.

Definition 4.2. *Let X, Y and Z be sets of sizes L, M and N. A function $f : X \times Y \times Z$ is* α-quasirandom *if*

$$\sum_{x_0,x_1\in X}\sum_{y_0,y_1\in Y}\sum_{z_0,z_1\in Z}\prod_{(i,j,k)\in\{0,1\}^3} f(x_i,y_j,z_k) \leqslant \alpha L^2M^2N^2 \ .$$

Definition 4.3. *Let H be a tripartite 3-uniform hypergraph with vertex sets X, Y and Z of sizes L, M and N. Let the number of edges of H be $pLMN$ and let $f(x,y,z) = H(x,y,z) - p$. Then H is* α-quasirandom *if f is α-quasirandom.*

As we commented earlier, this definition is not identical to Definition 2.4, but it is equivalent (give or take the precise value of α) and it is the one we shall use. The advantage it has is that it is easier to use when proving a counting lemma. Note that if H fails to be quasirandom in this sense, then we have the easy deduction that f fails property (ii) of Theorem 4.1, which in turn shows easily that H is not α-edge uniform (in the sense of Definition 2.6).

We come now to a counting lemma for quasirandom 3-uniform hypergraphs. For simplicity, we prove it only in one special case, that of simplices (these were defined just before the statement of Theorem 1.4), but it is an easy exercise to generalize this case to a full counting lemma, just as we generalized Lemma 3.4 to Theorem 3.5.

Lemma 4.4. *Let H be a quadripartite 3-uniform hypergraph with vertex sets X, Y, Z and W, of sizes L, M and N and P respectively. Suppose that the induced subhypergraphs*

$H(X,Y,Z)$, $H(X,Y,W)$, $H(X,Z,W)$ and $H(Y,Z,W)$ are α-quasirandom with densities p, q, r and s respectively. Then the number of simplices in H differs from $pqrsLMNP$ by at most $15\alpha^{1/8}LMNP$.

Proof. Let the variables x, y, z and w stand for elements of X, Y, Z and W. Let the letter H stand for the characteristic function of the hypergraph H as well as the hypergraph itself. Define functions f, g, h and k (with obvious domains) by $f(x,y,z) = H(x,y,z) - p$, $g(x,y,w) = H(x,y,w) - q$, $h(x,z,w) = H(x,z,w) - r$ and $k(y,z,w) = H(y,z,w) - s$. Then the number of simplices in H is

$$\sum_{x,y,z,w} (p + f(x,y,z))(q + g(x,y,w))(r + h(x,z,w))(s + k(y,z,w)) \ .$$

The main term in this sum is $pqrsLMNP$. We shall now show that all other terms are significantly smaller. Consider, for example, any term that chooses $f(x,y,z)$ rather than p from the first bracket. For each fixed w this results in a sum of the form $\sum_{x,y,z} f(x,y,z)t(x,y)u(y,z)v(z,x)$, with t, u and v taking values in the interval $[-1,1]$. Therefore, by Theorem 4.1 (and the bound obtained in the proof), the entire sum comes to at most $\alpha^{1/8}LMNP$. The same argument works for g, h and k, and that is enough to show that all terms apart from the main term have modulus at most $\alpha^{1/8}LMNP$, which proves the result. □

§5. Why quasirandom 3-graphs are not enough.

To prove Roth's theorem (that is, Szemerédi's theorem for progressions of length 3), one uses a combination of Szemerédi's regularity lemma and Lemma 3.4. The reader who has followed the paper so far may be disappointed to learn that Lemma 4.4 is not very useful when it comes to generalizing that argument. However, any effort spent on understanding it will pay dividends later, since the result that *is* useful is a further generalization, proved by a similar technique, and the steps of that result will make much more sense if they are compared with the steps in the proofs of Theorem 4.1 and Lemma 4.4.

What, then, is inadequate about quasirandomness of 3-uniform hypergraphs? The answer is not that the property is too weak - as Lemma 4.4 demonstrates - but rather that it is too strong. In other words, we are delighted if we are lucky enough to be presented with a quasirandom hypergraph, but in general it is too much to hope for. If we wish to generalize the proof for graphs, triangles and progressions of length 3, then we shall need two components: a regularity lemma and an associated counting lemma. The regularity lemma will tell us that we can divide any 3-uniform hypergraph H into random-like pieces (whatever this turns out to mean, it should somehow be analogous to the statement of the usual regularity lemma for graphs) and the counting lemma will allow us to use this information to approximate the number of simplices in H. One might think that "random-like pieces" should simply be quasirandom sub-hypergraphs, but any sensible statement along these lines turns out to be false.

Here is a simple example of a tripartite 3-uniform hypergraph that has no large quasirandom subhypergraph, and which therefore cannot be decomposed into a small number of them. Let X, Y and Z be three sets of size N and let G be a random tripartite graph with vertex sets X, Y and Z. Let H be the hypergraph consisting of all triangles in G, that is, all triples (x, y, z) such that xy, yz and xz are edges of G. Then the density of H is $1/8$, but the number of octahedra in H is about $2^{-12}N^6$ (because an octahedron, considered as a graph, has 12 edges) rather than $8^{-8}N^6$ as it should have if H is quasirandom.

Now, given any large subsets $X' \subset X$, $Y' \subset Y$ and $Z' \subset Z$, the graphs $G(X', Y')$, $G(Y', Z')$ and $G(X', Z')$ are (with high probability) quasirandom, and therefore the same reasoning shows that the induced subhypergraph $H(X', Y', Z')$ still fails to be quasirandom.

Indeed, the situation is even worse, as it is not just induced subhypergraphs that fail to be quasirandom. Let H' be *any* subhypergraph of H and let the density of H' be p. Since H' is a subhypergraph of H,

$$\sum_{x,y,z} H'(x, y, z)G(x, y)G(y, z)G(x, z) = pN^3 \ .$$

If we set $f(x, y, z) = H'(x, y, z) - p$, then we can deduce that

$$\sum_{x,y,z} f(x, y, z)G(x, y)G(y, z)G(x, z) = pN^3 - pN^3/8 \ ,$$

which is a clear violation of property (iii) of Theorem 4.1. This argument remains valid even if one starts by restricting to large subsets X', Y' and Z' of X, Y and Z.

This looks like bad news, and in a way it is, because it makes life more complicated, but it is not as bad as all that. To see why not, just look back at the discussion of the example above. Although the hypergraph H was not quasirandom, we had absolutely no difficulty calculating roughly how many octahedra it should have, and the reason was that we were able to use the quasirandomness of the graphs $G(X, Y)$, $G(Y, Z)$ and $G(Z, X)$. More generally, suppose we construct a hypergraph H as follows. First, we take random graphs $G(X,Y)$, $G(Y,Z)$ and $G(Z,X)$ with densities p, q and r and let G be the tripartite graph formed by their union. Next, we define H_0 to be the hypergraph consisting of all triangles of G. Finally, we let H be a random subhypergraph of H_0, choosing each edge of H_0 with probability s and making all choices independently.

How many octahedra do we expect H to contain? Well, if we choose $x_0, x_1 \in X$, $y_0, y_1 \in Y$ and $z_0, z_1 \in Z$ at random, then the probability that a pair $x_i y_j$ belongs to G is p, and these probabilities are more or less independent, so the probability that all four pairs belong to G is almost exactly p^4. Similar statements hold for the $y_i z_j$ and the $x_i z_j$, so the probability that all the 2-edges of the octahedron belong to G is almost exactly $(pqr)^4$. If this happens, then there are eight 3-edges, or faces, each of which has a probability s of lying in H. We therefore expect the number of octahedra in H to be about $(pqr)^4 s^8 (|X||Y||Z|)^2$.

It is clear from that calculation that quasirandom hypergraphs are not the only ones for which it ought to be possible to prove a counting lemma. That is, one ought to be able to relax the assumptions of Lemma 4.4 so that the induced subhypergraphs are not necessarily quasirandom, but are built rather like the hypergraph considered in the last paragraph. It is a lemma of this kind that we shall state and prove in the next section.

This will deal with the difficulty that we have just discussed. If we have proved a counting lemma for a wider class of hypergraphs than just the quasirandom ones, then it is enough, when proving a regularity lemma, to show that a hypergraph can be decomposed into subgraphs from this wider class. And this assertion is weak enough to be true, which is of course a huge advantage.

The situation we have just encountered occurs in other parts of mathematics - indeed, something like it seems to happen for almost any class of mathematical objects that do not have too rigid a structure but are well-endowed with subobjects. In such a situation, it is very useful to find, for any object X in the class, a subobject $Y \subset X$ that is in some way “stable”, in the sense that any further subobject $Z \subset Y$ does not differ interestingly from Y. To do this one must first identify the stable objects and then prove that every object contains a stable subobject. Here the structure we have been talking about is an approximate one (it is not hard to define a notion of *approximate isomorphism* to make it precise).

More exact instances are usually called canonical Ramsey theorems, of which the most famous example concerns arbitrary colourings of the edges of the complete graph on $\mathbb{N}$. Here, one cannot expect to find a monochromatic infinite clique, but one can find an infinite set X such that the restriction of the colouring to the clique $X^{(2)}$ has one of four simple forms. Write all edges as xy with $x < y$ and write $xy \sim zw$ if xy and zw have the same colour. Then one of the following four statements is true for all pairs xy, zw of edges with $x, y, z, w \in X$: (i) $xy \sim zw$ if and only if $x = z$ and $y = w$; (ii) $xy \sim zw$ if and only if $x = z$; (iii) $xy \sim zw$ if and only if $y = w$; (iv) $xy \sim zw$. It is easy to check that if one of these statements holds for X then it holds for all subsets $Y \subset X$, so X (together with its colouring) is stable.

A second class of examples arises in Banach space theory. There are several theorems in the subject that allow one to pass from a Banach space X, perhaps with some extra properties, to a subspace Y that is in some way easier to handle. And in many cases, the property that Y has is a stability property, in the sense that all its subspaces are in some important way similar to the space itself.

Before we embark on the main results of this paper, here is a second hypergraph example to consider. It is not completely obvious that the first one matters, since it is not derived from a subset of $\mathbb{Z}_N$ in the way shown after Definition 2.3. Perhaps hypergraphs that come from sets have some extra property that makes them decomposable into quasirandom pieces.

It turns out that they don't. Rather than show precisely this, we shall briefly discuss a similar result for functions, because it is much easier technically. It is derived in a simple way from an example that plays a similar role in the analytic proof of Szemerédi's theorem [G1].

Let N be an odd positive integer and let $\omega = e^{2\pi i/N}$. Define a function $f : \mathbb{Z}_N^3 \to \mathbb{C}$ by $f(x, y, z) = \omega^{(x+y+z)^2}$. This can be decomposed as $\omega^{(x+y)^2}\omega^{(y+z)^2}\omega^{(x+z)^2}\omega^{-x^2}\omega^{-y^2}\omega^{-z^2}$, or alternatively as $g(x, y)g(y, z)g(x, z)$, where $g(x, y) = \omega^{2^{-1}(x^2+y^2)+2xy}$. It is a straightforward exercise to deduce from the fact that the function ω^{x^2} has very small Fourier coefficients that g is a quasirandom function. So once again, we have a function of three variables that is a product of three quasirandom functions of two variables and therefore not quasirandom itself, even after restriction to any large set.

§6. A counting lemma for two-dimensional quasirandom simplicial complexes.

What the previous section shows is that we should consider objects that are slightly more complicated than 3-uniform hypergraphs. We need to look instead at 3-uniform hypergraphs that are obtained as subhypergraphs H of the set of all triangles in some tripartite graph G, paying attention to both H and G. Let us write $\Delta(G)$ for the set of triangles of G. Then one of these objects can be defined more formally as an ordered pair (G, H) such that $H \subset \Delta(G)$. It should be considered as quasirandom if the three bipartite parts of the graph G are quasirandom and H in some way "sits quasirandomly" inside $\Delta(G)$. Our first task is to make this idea precise. Once we have done that, we shall prove another sequence of results, again following the scheme of §3 and §4.

A slightly better way to think of our objects (G, H) is as two-dimensional simplicial complexes: that is, as collections Σ of sets of size at most 3 with the property that if $A \in \Sigma$ and $B \subset A$ then $B \in \Sigma$. Of course, to do this we need to take not just a graph and a hypergraph, but also a set of vertices (and, to be strictly correct, the empty set). What makes this a better way to think about it is partly that it is more natural when one comes to generalize to k-uniform hypergraphs, and partly that the regularity lemma we shall eventually prove involves restricting vertex sets. Despite all this, it will be simpler to stick to pairs (G, H) for now and bear in mind that our results will later be applied to pairs with restricted vertex sets. Let us make a formal definition.

Definition 6.1. *An r-partite* chain *is a pair (G, H), where G is an r-partite graph, H is an r-partite hypergraph with the same vertex sets as G, and $H \subset \Delta(G)$.*

Suppose, then, that we have a chain (G, H). We know what it means for the bipartite parts of G to be quasirandom, but must now say what it means for H to sit quasirandomly inside $\Delta(G)$. As before, we shall define this in terms of functions. For convenience, let us make another definition.

Definition 6.2. *Let X, Y and Z be three sets and let $f : X \times Y \times Z \to \mathbb{R}$ be a function. Then $\mathrm{oct}(f)$ is defined to be $f_{x,x',y,y',z,z'}$, which equals the sum*

$$\sum f(x,y,z)f(x,y,z')f(x,y',z)f(x,y',z')f(x',y,z)f(x',y,z')f(x',y',z)f(x',y',z')$$

taken over all $x, x' \in X$, $y, y' \in Y$ and $z, z' \in Z$.

Recall that the notation $f_{x,x',y,y',z,z'}$ was introduced before the statement of Theorem 4.1. We shall use it again when proving Theorem 6.5 below.

If f is the characteristic function of a hypergraph H then $\mathrm{oct}(f)$ is the number of octahedra in H. We shall write $\mathrm{oct}(H)$ for this quantity. With this notation it is easy to express precisely what it means for H to sit quasirandomly in G.

Definition 6.3. *Let G be a tripartite graph with vertex sets X, Y and Z of sizes L, M and N, and let $f : X \times Y \times Z \to [-1, 1]$ be a function such that f is supported in $\Delta(G)$.*

Let the densities of $G(X,Y)$*,* $G(Y,Z)$ *and* $G(X,Z)$ *be* p*,* q *and* r *respectively. Then* f *is* α-quasirandom relative to G *if* $\mathrm{oct}(f) \leqslant \alpha(pqr)^4(LMN)^2$*. Now let* H *be a tripartite 3-uniform hypergraph with vertex sets* X*,* Y *and* Z *and suppose that* $H \subset \Delta(G)$ *and* $|H| = \gamma|\Delta(G)|$*. Let* $f(x,y,z) = H(x,y,z) - \gamma$ *for* $(x,y,z) \in \Delta(G)$ *and* 0 *otherwise. Then* H *is* α-quasirandom relative to G *if* f *is* α*-quasirandom relative to* G*.*

It might be more natural to say that f is α-quasirandom relative to G if $\mathrm{oct}(f) \leqslant \alpha\,\mathrm{oct}(G)$. If G is quasirandom, then this is roughly what the definition does say, and we shall apply the definition only to quasirandom graphs, where the formulation we have given turns out to be slightly more convenient for technical reasons.

These definitions are very similar to those of §4, but now everything takes place inside $\Delta(G)$. What we shall show is that if the bipartite parts of G are sufficiently quasirandom, then relative quasirandomness has consequences that are also similar to those of §4, though for reasons that will be explained more fully later, some of them are a bit more complicated.

The main result of this section is a counting lemma for simplices in quadripartite chains. The reader who follows the proof will see that it can be generalized easily to a counting lemma for arbitrary subchains. However, the case of simplices is easier to present, and is enough for the application to arithmetic progressions of length four.

Before we prove this counting lemma, we need to prepare for it with a technical lemma (Lemma 6.6 below), which itself needs small amount of preparation.

Definition 6.4. *Let* G *and* H *be* k*-partite graphs with vertex sets* $X = X_1 \cup \ldots \cup X_k$ *and* $A = A_1 \cup \ldots \cup A_k$ *respectively. A* homomorphism *from* H *to* G *is a map* $\phi : A \to X$ *such that* $\phi(A_i) \subset X_i$ *for each* i *and such that* $\phi(v)\phi(w)$ *is an edge of* G *whenever* vw *is an edge of* H*.*

Note that in the above definition we say nothing about what happens if vw is not an edge of H. Nor do we insist that ϕ is an injection.

Let G be a quadripartite graph with vertex sets X, Y, Z and W of sizes L, M, N and P respectively. Let H be another quadripartite graph, with vertex sets A, B, C and D of sizes q, r, s and t, and let $a, a' \in A$, $b, b' \in B$ and $c, c' \in C$ be six vertices and suppose that the set $\{a, a', b, b', c, c'\}$ is independent. For any $x, x' \in X$, $y, y' \in Y$ and $z, z' \in Z$ let $h(x, x', y, y', z, z')$ be the number of homomorphisms ϕ from H to G such that $\phi(a) = x$, $\phi(a') = x'$, $\phi(b) = y$, $\phi(b') = y'$, $\phi(c) = z$ and $\phi(c') = z'$. Lemma 6.6 will tell us that if G is α-quasirandom for a sufficiently small α, then the function h is approximately constant. This is a simple application of the counting lemma for graphs (Theorem 3.5) and the second-moment method.

Before we embark on the lemma, let us think about the constant we expect to obtain. For each pair i, j let the density of the graph $G(X_i, X_j)$ be δ_{ij}. Given any edge e of H, let us set $\delta(e)$ to be the δ_{ij} for which A_i and A_j are the vertex sets containing the two vertices joined by e. The number of ways of choosing a function ϕ that respects the partitions of G and H is $L^q M^r N^s P^t$. If we fix the images of a, a', b, b', c, c' then the number of possible extensions is $L^{q-2}M^{r-2}N^{s-2}P^t$. Given an edge e of H, it joins A_i to A_j for some $1 \leqslant i < j \leqslant 4$. The probability that $\phi(e)$ is an edge of G is the density δ_{ij} of the bipartite subgraph $G(X_i, X_j)$. If G behaves like a random graph, then the probability that ϕ is a homomorphism will be roughly the product, δ, of all these individual edge-probabilities. We shall call this the *expected H-density* of G. (The assumption that $\{a, a', b, b', c, c'\}$ is an independent set means that we do not have to worry about whether there are edges joining vertices in the set $\{x, x', y, y', z, z'\}$.) We expect the approximately constant value of h to be about $\delta L^{q-2}M^{r-2}N^{s-2}P^t$.

We shall be using second moments, so for convenience here first is an easy technical lemma that encapsulates what we need for the main lemma.

Lemma 6.5. *Let $\alpha, \delta \in [0,1]$, let R be a real number and let $a_1, \dots, a_n$ be real numbers such that $\sum_{i=1}^n a_i \geqslant (\delta - \alpha)Rn$ and $\sum_{i=1}^n a_i^2 \leqslant (\delta^2 + \alpha)R^2 n$. Then $|a_i - \delta R| \leqslant R\alpha^{1/4}$ for all but at most $3n\sqrt{\alpha}$ values of i.*

Proof. Using our hypotheses, we find that

$$\begin{aligned}\sum_{i=1}^n (a_i - \delta R)^2 &= \sum_{i=1}^n a_i^2 - 2R\sum_{i=1}^n a_i\delta + n\delta^2 R^2 \\ &\leqslant R^2 n\big(\delta^2 + \alpha - 2\delta(\delta - \alpha) + \delta^2\big) \\ &= R^2 n\alpha(1 + 2\delta) \leqslant 3R^2 n\alpha\ .\end{aligned}$$

The result follows immediately. □

The statement that follows will look somewhat peculiar: we are giving the particular case that happens to arise later of a more general statement which, though more natural, is perhaps harder to digest. Some readers may wish to jump to Lemma 6.7 and then come back to this point of the paper when the motivation for it has become clear.

Lemma 6.6. *Let G, H and δ be as in the remarks preceding Lemma 6.5. Let $0 < \epsilon \leqslant 1$, let $\alpha > 0$ be such that $2^m \alpha^{1/16} \leqslant \epsilon\delta/3$ and suppose that G is α-quasirandom. Let m be the number of edges of H and let δ be the expected H-density of G. Then the number of* *sextuples* (x,x',y,y',z,z') *for which* $h(x,x',y,y',z,z')$ *differs from* $\delta L^{q-2}M^{r-2}N^{s-2}P^t$ *by more than* $\epsilon\delta L^{q-2}M^{r-2}N^{s-2}P^t$ *is at most* $\epsilon\delta L^2M^2N^2$.

Proof. First, we estimate $\sum_{x,x',y,y',z,z'} h(x,x',y,y',z,z')$. This is the number of homomorphisms ϕ from H to G, which, by Theorem 3.5, is at least $(\delta - 2^m\alpha^{1/4})L^qM^rN^sP^t$.

Next, we estimate $\sum_{x,x',y,y',z,z'} h(x,x',y,y',z,z')^2$. For this we can use the counting lemma again, but first we must define an auxiliary graph J. For each vertex v of H apart from a, a', b, b', c and c', let v_1 be a copy of v. Let γ be a function defined on the vertex set V of H that takes the vertices a, a', b, b', c and c' to themselves and takes any other vertex v to its copy v_1. The vertex set of J is $V \cup \gamma(V)$ and the edges of J are the edges of H together with all pairs $\gamma(u)\gamma(v)$ such that uv is an edge of H. It is easy to see that $\sum_{x,x',y,y',z,z'} h(x,x',y,y',z,z')^2$ is the number of homomorphisms from J to G, which is of course why we defined J. Since every edge of H has been doubled up in J, the product of the edge-probabilities of J is δ^2, so Theorem 3.5 tells us that the number of homomorphisms from J to G is at most $(\delta^2 + 2^{2m}\alpha^{1/4})L^{2q-2}M^{2r-2}N^{2s-2}P^t$.

Let us now apply the previous lemma, with $n = L^2M^2N^2$, $R = L^{q-2}M^{r-2}N^{s-2}P^t$, δ as it is and α replaced by $2^{2m}\alpha^{1/4}$. Then the hypotheses of the lemma are satisfied and it tells us that the number of sextuples (x,x',y,y',z,z') for which $h(x,x',y,y',z,z')$ differs from $\delta L^{q-2}M^{r-2}N^{s-2}P^t$ by more than $2^{m/2}\alpha^{1/16}L^{q-2}M^{r-2}N^{s-2}P^t$ is at most $3.2^m\alpha^{1/8}L^2M^2N^2$. Our upper bound on α then implies the result. □

A few words of explanation are needed before the next lemma, to draw attention to how it differs from the implication of (ii) from (i) in Theorem 4.1. As in that proof we have a sum and we wish to show that it is small, subject to a quasirandomness assumption about the function f. As in that proof we shall use the Cauchy-Schwarz inequality several times, and the manipulations will be very similar. However, this time the functions we look at are supported in the set of triangles of a quasirandom quadripartite graph G, and the bound we obtain is stronger because it depends on the densities of the six bipartite parts of G. The significance of this is that the theorem says something even when the quasirandomness parameter η below is much larger than any of these densities. This extra strength is very significant when G is sparse, as it will be if it is one of the graphs given to us by the hypergraph regularity lemma proved later. To obtain the extra strength, we shall be very careful to use the full strength of the Cauchy-Schwarz inequality whenever we apply it: if the number of i such that $a_i \neq 0$ is m, then we shall bound $\left(\sum_{i=1}^n a_i\right)$ above by $m\sum_{i=1}^n a_i^2$ rather than by $n\sum_{i=1}^n a_i^2$.

We shall use the following notation. The graph G will have vertex sets X, Y, Z and W, but we shall also think of them as X_1, X_2, X_3 and X_4 respectively. (Sometimes one notation is easier to handle, sometimes the other.) The density of G_{ij} will again be denoted δ_{ij}. We shall write h_i for $|X_i|$, h_{ij} for $|G(X_i, X_j)|$ and h_{ijk} for the number of triangles in $G(X_i, X_j) \cup G(X_j, X_k) \cup G(X_i, X_k)$. If $x, x' \in X_1$ and $i > 1$ then we shall also write $h_i(x, x')$ for the number of vertices in X_i that are joined to both x and x'. Similarly, if $1 < i < j$ then we shall write $h_{ij}(x, x')$ for the set of all edges yz in $G(X_i, X_j)$ such that y and z are both joined to both of x and x'. Similarly, if $x, x' \in X_1$, $y, y' \in X_2$ and $i > 2$, then we shall write $h_i(x, x', y, y)$ for the number of $z \in X_i$ that are joined to all of x, x', y and y'. It is numbers such as these that will appear when we make our more efficient uses of the Cauchy-Schwarz inequality.

Lemma 6.7. *Let G be as just described, and let $f : X \times Y \times Z \to [-1, 1]$, $g : X \times Y \times W \to [-1, 1]$, $h : X \times Z \times W \to [-1, 1]$ and $k : Y \times Z \times W \to [-1, 1]$ be functions that are non-zero only at triples that form triangles in G. Suppose that G is α-quasirandom and that f is η-quasirandom relative to the tripartite graph $G(X, Y, Z)$. Suppose also that $2^{36}\alpha^{1/16} \leqslant \eta(\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34})^8/6$. Then*

$$\Big|\sum_{x,y,z,w} f(x,y,z)g(x,y,w)h(x,z,w)k(y,z,w)\Big| \leqslant (2\eta)^{1/8}\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP\ .$$

Proof. We begin with several applications of the Cauchy-Schwarz inequality, of a similar kind to ones that we have seen already.

$$\begin{aligned}
\Big(\sum_{x,y,z,w} & f(x,y,z)g(x,y,w)h(x,z,w)k(y,z,w)\Big)^8 \\
&\leqslant \Big(h_{234}\sum_{y,z,w}\Big(\sum_x f(x,y,z)g(x,y,w)h(x,z,w)k(y,z,w)\Big)^2\Big)^4 \\
&\leqslant \Big(h_{234}\sum_{y,z,w}\Big(\sum_x f(x,y,z)g(x,y,w)h(x,z,w)\Big)^2\Big)^4 \\
&= h_{234}^4\Big(\sum_{x,x'}\sum_{y,z,w} f_{x,x'}(y,z)g_{x,x'}(y,w)h_{x,x'}(z,w)\Big)^4 \\
&\leqslant h_{234}^4 h_1^6\sum_{x,x'}\Big(\sum_{y,z,w} f_{x,x'}(y,z)g_{x,x'}(y,w)h_{x,x'}(z,w)\Big)^4 \\
&\leqslant h_{234}^4 h_1^6\sum_{x,x'}\Big(h_{34}(x,x')\sum_{z,w}\Big(\sum_y f_{x,x'}(y,z)g_{x,x'}(y,w)h_{x,x'}(z,w)\Big)^2\Big)^2
\end{aligned}$$

$$
\begin{aligned}
&\leqslant h_{234}^4 h_1^6 \sum_{x,x'} \Big(h_{34}(x,x') \sum_{z,w} G(z,w) \Big(\sum_y f_{x,x'}(y,z) g_{x,x'}(y,w) \Big)^2 \Big)^2 \\
&= h_{234}^4 h_1^6 \sum_{x,x'} h_{34}(x,x')^2 \Big(\sum_{y,y'} \sum_{z,w} f_{x,x',y,y'}(z) g_{x,x',y,y'}(w) G(z,w) \Big)^2 \\
&\leqslant h_{234}^4 h_1^6 \sum_{x,x'} h_{34}(x,x')^2 h_2(x,x')^2 \sum_{y,y'} \Big(\sum_{z,w} f_{x,x',y,y'}(z) g_{x,x',y,y'}(w) G(z,w) \Big)^2 \\
&\leqslant h_{234}^4 h_1^6 \sum_{x,x'} h_{34}(x,x')^2 h_2(x,x')^2 \sum_{y,y'} h_4(x,x',y,y') \\
&\qquad\qquad \sum_w \Big(\sum_z f_{x,x',y,y'}(z) g_{x,x',y,y'}(w) G(z,w) \Big)^2 \\
&\leqslant h_{234}^4 h_1^6 \sum_{x,x'} h_{34}(x,x')^2 h_2(x,x')^2 \sum_{y,y'} h_4(x,x',y,y') \\
&\qquad\qquad \sum_w \Big(\sum_z f_{x,x',y,y'}(z) G_{x,x'} G_{y,y'}(w) G(z,w) \Big)^2 \\
&= h_{234}^4 h_1^6 \sum_{x,x'} h_{34}(x,x')^2 h_2(x,x')^2 \sum_{y,y'} h_4(x,x',y,y') \\
&\qquad\qquad \sum_{z,z'} f_{x,x',y,y',z,z'} \sum_w G_{x,x'} G_{y,y'} G_{z,z'}(w) \\
&= h_{234}^4 h_1^6 \sum_{x,x'} h_{34}(x,x')^2 h_2(x,x')^2 \sum_{y,y'} h_4(x,x',y,y') \sum_{z,z'} f_{x,x',y,y',z,z'} h_4(x,x',y,y',z,z')
\end{aligned}
$$

The main idea of the proof has now been given. The rest of the argument consists in showing that the h-terms are all approximately constant and calculating what the result would be if they *were* constant. For this we use the lemma about graphs proved earlier.

The final line above can be written as

$$\sum_{x,x',y,y',z,z'} h(x,x',y,y',z,z') f_{x,x',y,y',z,z'} = \langle h, F \rangle \ ,$$

where $F(x,x',y,y',z,z') = f_{x,x',y,y',z,z'}$ and

$$h(x,x',y,y',z,z') = h_{234}^4 h_1^6 h_{34}(x,x')^2 h_2(x,x')^2 h_4(x,x',y,y') h_4(x,x',y,y',z,z') \ .$$

This last quantity is the number of homomorphisms from a certain quadripartite graph H to G, given that a particular six of its vertices (none of which are joined to each other) map to x, x', y, y', z and z'. To see this, note first that it is true of each individual term in the product. For example, to understand the term $h_{34}(x,x')^2$ in this way, take the graph J with vertex set $\{a, a', b, b', c, c', d, e\}$ and edges ad, ae, $a'd$, $a'e$ and de. The number of homomorphisms ϕ from J to G such that $\phi(a) = x$, $\phi(a') = x'$, $\phi(b) = y$, $\phi(b') = y'$, $\phi(c) = z$, $\phi(c') = z'$, $\phi(d) \in Z$ and $\phi(e) \in W$ is the number of pairs $(z, w) \in Z \times W$ such that z is joined to w and both z and w are joined to both x and x', which is the definition of $h_{34}(x, x')$. To obtain the product, one takes disjoint copies of all the graphs J constructed in this way and identifies the vertices a, a', b, b', c and c' from each one. (That is, the a in one graph is the same as the a in another, and so on.)

Let the vertex set of H be $A \cup B \cup C \cup D$, and let us look at the sizes of A, B, C and D. The set A contains a and a', and receives an additional six (isolated) vertices from the term h_1^6. B contains b and b', and receives in addition four vertices from h_{234}^4 and two from $h_2(x, x')$. C contains c and c' and receives four vertices from h_{234}^4 and two from $h_{34}(x, x')^2$. Finally, D receives four vertices from h_{234}^4, two from $h_{34}(x, x')$ and one each from $h_4(x, x', y, y')$ and $h_4(x, x', y, y', z, z')$.

In a similar way one can work out how many edges there are between each pair from A, B, C and D. For example, the number of edges between B and D is $4+2+2 = 8$, since h_{234}^4 contributes four, $h_4(x, x', y, y')$ contributes two and $h_4(x, x', y, y', z, z')$ contributes two. As another example, the number of edges between A and C is 4, all coming from $h_{34}(x, x')^2$. It turns out that there are eight edges between any pair of sets that includes D and four between any other pair.

We wish to apply Lemma 6.6. It follows from the simple calculations we have just made that the expected H-density of G is

$$\delta = (\delta_{12}\delta_{23}\delta_{13})^4(\delta_{14}\delta_{24}\delta_{34})^8 \ .$$

We also have $q = r = s = t = 8$ and $m = 36$. Let $\epsilon = \eta(\delta_{12}\delta_{23}\delta_{13})^4/2$ and call a sextuple (x, x', y, y', z, z') *bad* if $\left|h(x, x', y, y', z, z') - \delta L^6M^6N^6P^8\right| > \epsilon\delta L^6M^6N^6P^8$, and *good* otherwise.

By Lemma 6.6, the number of bad sextuples is at most $\epsilon\delta L^2M^2N^2$, and a trivial upper bound for each $h(x, x', y, y', z, z')$ is $L^6M^6N^6P^8$. Therefore, if we let h' be a new function that equals h for every good sextuple and takes the value $\delta L^6M^6N^6P^8$ otherwise, then $\|h - h'\|_1 \leqslant \epsilon\delta(LMNP)^8$. Writing d for the constant function $\delta L^6M^6N^6P^8$, we also have that $\|h' - d\|_\infty \leqslant \epsilon\delta L^6M^6N^6P^8$. As for F, we know that $\|F\|_1$ is at most the number of octahedra in the graph $G(X, Y, Z)$. By Theorem 3.5, this is at most $2(\delta_{12}\delta_{23}\delta_{13})^4(LMN)^2$. Finally, we are assuming also the bound $\|F\|_\infty \leqslant 1$. Putting all these facts together, we

find that
$$\big|\langle h,F\rangle-\langle d,F\rangle\big| \leqslant |\langle h-h',F\rangle|+|\langle h'-d,F\rangle| \\ \leqslant \|h-h'\|_1\|F\|_\infty+\|h'-d\|_\infty\|F\|_1 \\ \leqslant \epsilon\delta(LMNP)^8+\epsilon\delta(LMNP)^8 \\ =\eta(\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP)^8 \ .$$
But the the relative quasirandomness assumption on f tells us that
$$\langle d,F\rangle = \delta L^6M^6N^6P^8\,\mathrm{oct}(f) \\ \leqslant \eta\delta(\delta_{12}\delta_{23}\delta_{13})^4(LMNP)^8 \\ =\eta(\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP)^8 \ .$$
The result follows. □

We are now ready to prove a generalization of Lemma 4.4 from quasirandom hypergraphs to quasirandom chains. Notice the dependence of parameters in the statement. The graph G is α-quasirandom and the hypergraph H is relatively η-quasirandom. Both α and η need to be small for the conclusion to hold and be useful, but whereas the condition on α depends on η, the density of G and the relative density of H, the smallness of η depends only on the last of these. In particular, as we have already mentioned, η can be much larger than the density of G. This is critically important, since it is all that can be guaranteed by the regularity lemma later.

Theorem 6.8. *Let X, Y, Z and W be sets of size L, M, N and P respectively. Let G be a quadripartite graph with vertex sets X, Y, Z and W and suppose that the six bipartite parts of G are α-quasirandom. Write δ_{12} for the density of the graph $G(X,Y)$, and similarly for the other parts, and suppose that all these graphs are α-quasirandom. Let H_{123} be a tripartite hypergraph with vertex sets X, Y and Z that is η-quasirandom relative to $\Delta\big(G(X,Y,Z)\big)$ and similarly for H_{124}, H_{134} and H_{234}. For each triple ijk Let the relative density of H_{ijk} be δ_{ijk}. Let H be the union of the hypergraphs H_{ijk}. Suppose that α satisfies the condition $2^{36}\alpha^{1/16}\leqslant\eta(\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34})^8/6$. Then the number of simplices in H differs from $\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}\delta_{123}\delta_{124}\delta_{134}\delta_{234}LMNP$ by at most $8\eta^{1/8}\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP$.*

Proof. We wish to estimate the sum
$$\sum_{x,y,z,w}H(x,y,z)H(x,y,w)H(x,z,w)H(y,z,w)\ .$$

For each triple $1 \leqslant i < j < k \leqslant 4$ let $d_{ijk}(x,y,z) = \delta_{ijk} G(x,y)G(y,z)G(x,z)$. Then

$$\sum_{x,y,z,w} d_{123}(x,y,z)d_{124}(x,y,w)d_{134}(x,z,w)d_{234}(y,z,w)$$

is $\delta_{123}\delta_{124}\delta_{134}\delta_{234}$ times the number of simplices in G. Since G is α-quasirandom, Theorem 3.5 tells us that the number of simplices in G is $\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP$, to within an error of at most $64\alpha^{1/4}LMNP$, which is certainly at most $\eta^{1/8}LMNP$.

If we let $f(x,y,z) = H(x,y,z) - d_{123}(x,y,z)$, then our hypothesis implies that f is η-quasirandom relative to G. If we take the sum we wish to estimate and change $H(x,y,z)$ into $d_{123}(x,y,z)$, then the difference we make to the sum is

$$\Big|\sum_{x,y,z,w} f(x,y,z)H(x,y,w)H(x,z,w)H(y,z,w)\Big| \ .$$

By Lemma 6.7, this is at most $(2\eta)^{1/8}\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP$. By a similar argument we can replace $H(x,y,w)$ by $d_{124}(x,y,w)$, again making a difference of at most $(2\eta)^{1/8}\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP$. Repeating this process twice more, we find that

$$\Big|\sum_{x,y,z,w} H(x,y,z)H(x,y,w)H(x,z,w)H(y,z,w) - \sum_{x,y,z,w} d_{123}(x,y,z)d_{124}(x,y,w)d_{134}(x,z,w)d_{234}(y,z,w)\Big|$$

is at most $4(2\eta)^{1/8}\delta_{12}\delta_{23}\delta_{13}\delta_{14}\delta_{24}\delta_{34}LMNP$. Combining this with the estimate of the previous paragraph, we obtain the desired result. □

We have now finished the hardest part of the proof, by identifying a class of stable hypergraphs and proving a counting lemma for them. It remains to prove a regularity lemma, which says, roughly speaking, that every dense 3-uniform hypergraph can be decomposed into stable subhypergraphs.

§7. A proof of Szemerédi's regularity lemma.

The statement of Szemerédi's regularity lemma given in the introduction is not quite standard, but it can be proved more cleanly and is better suited for generalizing to hypergraphs. To demonstrate the first of these assertions, to keep this paper self-contained and to illuminate the proof of hypergraph regularity, we shall now prove it in full.

Let G be a bipartite graph with vertex sets X and Y of sizes M and N respectively. Let $X_1 \cup \ldots \cup X_m$ and $Y_1 \cup \ldots \cup Y_n$ be partitions of X and Y, with $|X_i| = \alpha_i M$ and $Y_j = \beta_j N$. Write $d(X_i, Y_j)$ for the *density* of the induced subgraph $G(X_i, Y_j)$, that is, $|X_i|^{-1}|Y_j|^{-1}$ times the number of edges from X_i to Y_j. The *mean-square density* of G with respect to the partitions is defined to be $\sum_{i,j} \alpha_i \beta_j d(X_i, Y_j)^2$. Sometimes, when G is clear from the context, we shall call this the mean-square density of the partitions. This concept can also be viewed probabilistically. Choose a random $x \in X$ and $y \in Y$. Then x belongs to some X_i and y to some Y_j and the mean-square density is the expectation of the square of the density $d(X_i, Y_j)$.

Our first steps are very simple - all they say is that certain projections on certain Hilbert spaces have norm at most 1. However, let us quickly establish them in our particular context.

Lemma 7.1. *Let U be a finite set and let $f : U \to \mathbb{R}$ be a function with mean d. Let $U = U_1 \cup \ldots \cup U_s$ with $|U_i| = \gamma_i |U|$, and let d_i be the mean of f restricted to U_i. Then $d^2 \leqslant \sum_{i=1}^r \gamma_i d_i^2$.*

Proof. By the Cauchy-Schwarz inequality,

$$\Big(\sum_{i=1}^r \gamma_i d_i\Big)^2 \leqslant \Big(\sum_{i=1}^r \gamma_i\Big)\Big(\sum_{i=1}^r \gamma_i d_i^2\Big) .$$

The left-hand side is d^2 and the first bracket on the right-hand side is 1, so the result is proved. □

Lemma 7.2. *Let U, f and $U_1, \ldots, U_r$ be as in Lemma 7.1. Suppose that each U_i is partitioned further into sets U_{ij}, let $|U_{ij}| = \gamma_{ij}|U|$ and let d_{ij} be the mean of f restricted to U_{ij}. Then $\sum_i \gamma_i d_i^2 \leqslant \sum_{ij} \gamma_{ij} d_{ij}^2$.*

Proof. By Lemma 7.1, we have for each i the inequality

$$d_i^2 \leqslant \sum_j \frac{\gamma_{ij}}{\gamma_i} d_{ij}^2 .$$

Multiplying both sides by γ_i and summing over i gives the result. □

Corollary 7.3. *Let G be a bipartite graph of density d with vertex sets X and Y of sizes M and N respectively. Let $X_1 \cup \ldots \cup X_m$ and $Y_1 \cup \ldots \cup Y_n$ be partitions of X and Y. Let each X_i be partitioned further into sets X_{ik} and each Y_j into sets Y_{jl}. Then the* *mean-square density of G with respect to the partitions $\{X_{ik}\}$ and $\{Y_{jl}\}$ is at least the mean-square density of G with respect to the partitions $\{X_i\}$ and $\{Y_j\}$.*

Proof. Let U be the set $X \times Y$ and let f be the characteristic function of G. Then the result follows from Lemma 7.2 if we take as our cruder partition of U all sets of the form $X_i \times Y_j$ and as our finer one all sets of the form $X_{ik} \times Y_{jl}$, since the quantities compared in that lemma are the mean-square densities of G with respect to the two sets of partitions. □

Lemma 7.4. *Let G be a bipartite graph of density d with vertex sets X and Y of sizes M and N, and suppose that G fails to be ϵ-quasirandom. Then there are partitions $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ of the vertex sets that have mean-square density at least $d^2 + \epsilon^2/16$.*

Proof. Let $f(x,y) = G(x,y) - d$. The proof of Theorem 3.1 provides us with subsets $X_1 \subset X$ and $Y_1 \subset Y$ such that $\left|\sum_{x\in X_1}\sum_{y\in Y_1} f(x,y)\right| \geqslant \epsilon MN/4$. Let us write $\phi(X_i, Y_j)$ for the "density" of f when restricted to $X_i \times Y_j$, that is, for $|X_i|^{-1}|Y_j|^{-1}\sum_{x\in X_i}\sum_{y\in Y_j} f(x,y)$. Then the mean-square density of the partitions is

$$\sum_{i,j=1}^{2} \alpha_i\beta_j\big(d + \phi(X_i,Y_j)\big)^2 = \sum_{i,j=1}^{2} \alpha_i\beta_j\big(d^2 + 2d\phi(X_i,Y_j) + \phi(X_i,Y_j)^2\big)\ .$$

The first term adds up to d^2. The second adds up to zero, since the average of f is zero. The third adds up to at least $\alpha_1\beta_1(\alpha_1 M)^{-2}(\beta_1 N)^{-2}(\epsilon MN/4)^2$, which is at least $\epsilon^2/16$. □

Lemma 7.5. *Let G be a bipartite graph with vertex sets X and Y of sizes M and N respectively. Let $X_1 \cup \ldots \cup X_m$ and $Y_1 \cup \ldots \cup Y_n$ be partitions of X and Y, with $|X_i| = \alpha_i M$ and $Y_j = \beta_j N$. Suppose that the mean-square density of these partitions is d^2. Let B be the set of all pairs (i,j) such that the subgraph $G(X_i, Y_j)$ fails to be ϵ-quasirandom, and suppose that $\sum_{(i,j)\in B}\alpha_i\beta_j > \epsilon$. Then one can find partitions $X_i = X_{i1} \cup \ldots \cup X_{is}$ and $Y_j = Y_{j1} \cup \ldots \cup Y_{jt}$ such that the refined partitions $\{X_{ik}\}$ and $\{Y_{jl}\}$ have mean-square density at least $d^2 + \epsilon^3/16$. Moreover, s and t are uniformly bounded above by 2^n and 2^m respectively.*

Proof. For every pair (X_i, Y_j) that fails to be ϵ-quasirandom Lemma 7.4 provides us with partitions $X_i = X_{i1}^{(j)} \cup X_{i2}^{(j)}$ and $Y_j = Y_{j1}^{(i)} \cup Y_{j2}^{(i)}$ of mean-square density at least $d(X_i,Y_j)^2 + \epsilon^2/16$. For each i let $X_i = X_{i1} \cup \ldots \cup X_{is}$ be a partition with $s \leqslant 2^n$ that simultaneously refines all the partitions $X_{i1}^{(j)} \cup X_{i2}^{(j)}$, and for each j let $Y_j = Y_{j1} \cup \ldots \cup Y_{jt}$ be a partition with $t \leqslant 2^m$ that simultaneously refines all the partitions $Y_j = Y_{j1}^{(i)} \cup Y_{j2}^{(i)}$.

For each $(i,j) \in B$, Lemma 7.3 implies that the mean-square density of G with respect to the partitions $X_{i1} \cup \ldots \cup X_{is}$ and $Y_{j1} \cup \ldots \cup Y_{jt}$ is at least $d(X_i, Y_j)^2 + \epsilon^2/16$. For every other (i,j), Lemma 7.3 implies that it is at least $d(X_i, Y_j)^2$. Multiplying by $\alpha_i \beta_j$ and summing over all i, j tells us that the mean-square density of the partitions $\{X_{ik}\}$ and $\{Y_{jl}\}$ is at least

$$\sum_{(i,j)\notin B} \alpha_i \beta_j d(X_i, Y_j)^2 + \sum_{(i,j)\in B} \alpha_i \beta_j \big(d(X_i, X_j)^2 + \epsilon^2/16\big) ,$$

which is at least $d^2 + \epsilon^3/16$, by our hypothesis on the size of B. □

The next result is our non-standard statement of Szemerédi's regularity lemma.

Theorem 7.6. *Let $\epsilon > 0$ and let G be any bipartite graph with vertex sets X and Y. Then there are partitions $X = X_1 \cup \ldots \cup X_m$ and $Y = Y_1 \cup \ldots \cup Y_n$ with m and n bounded above by functions of ϵ, with the following property. Let B be the set of all pairs (i,j) such that the subgraph $G(X_i, Y_j)$ fails to be ϵ-quasirandom. Then $\sum_{(i,j)\in B} \alpha_i \beta_j \leqslant \epsilon$. Equivalently, the probability that a random pair $(x,y) \in X \times Y$ belongs to an $X_i \times Y_j$ for which $G(X_i, Y_j)$ fails to be ϵ-quasirandom is at most ϵ.*

Proof. If G itself is ϵ-quasirandom then we are done. Otherwise, let d be the density of G. Then Lemma 7.4 gives us partitions $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ of mean-square density at least $d^2 + \epsilon^3/16$. In general, given any pair of partitions $X = X_1 \cup \ldots \cup X_m$ and $Y = Y_1 \cup \ldots \cup Y_n$, either $\sum_{(i,j)\in B} \alpha_i \beta_j \leqslant \epsilon$ and we are done (where B is defined as in the statement of the theorem) or we can find refinements for which the mean-square density is greater by at least $\epsilon^3/16$. This allows us to construct a sequence of partitions of ever-increasing mean-square density, each refining the one before. Since mean-square density is bounded above by 1, this sequence must terminate in at most $16\epsilon^{-3}$ steps, and it terminates at a pair of partitions that satisfy the conclusion of the theorem. By the bound in Lemma 7.5, the number of sets in these partitions is bounded above by a function of ϵ only. (This function is given by a tower of 2s of height proportional to ϵ^{-3}.) □

To end this section, we prove a (known and easy) generalization of Szemerédi's regularity lemma, which we shall need later.

Lemma 7.7. *Let X and Y be sets and let $G_1, \ldots, G_r$ be bipartite graphs that form a partition of the edges of the complete bipartite graph $K(X,Y)$. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be partitions of X and Y respectively and suppose that the sum of the mean-square densities of all the graphs G_u with respect to these partitions is D. Choose an* *element (x,y) of the set $X \times Y$ uniformly at random, and suppose that with probability at least ϵ it lies in some $X_i \times Y_j$ for which not all the graphs $G_u(X_i \times Y_j)$ are ϵ-quasirandom. Then one can find partitions $X_i = X_{i1} \cup \ldots \cup X_{is_i}$ and $Y_j = Y_{j1} \cup \ldots \cup Y_{jt_j}$ such that the sum of the mean-square densities of the G_u with respect to the refined partitions $\{X_{ik}\}$ and $\{Y_{jl}\}$ is at least $D + \epsilon^3/16$. Moreover, all the s_i are bounded above by 2^n and all the t_j are bounded above by 2^m.*

Proof. This is very similar to the proof of Lemma 7.5 so we shall be brisk. Let (i,j) be a pair such that some graph $G_u(X_i \times Y_j)$, of density d, say, fails to be ϵ-quasirandom. Then by Lemma 7.4 we can find partitions of X_i and Y_j into two sets each in such a way that the mean-square density of $G_u(X_i \times Y_j)$ with respect to these partition is at least $d^2 + \epsilon^2/16$, and that will be true of any refinements of them, by Corollary 7.3.

Now let us find such partitions for every pair (i,j) for which a suitable u exists. Then each set X_i has been partitioned into two in at most n ways and each Y_j has been partitioned into two in at most m ways. Let $X_i = X_{i1} \cup \ldots \cup X_{is_i}$ and $Y_j = Y_{j1} \cup \ldots \cup Y_{jt_j}$ be common refinements of these partitions, into at most 2^n sets and 2^m sets respectively.

For each pair (i,j) for which there was a non-ϵ-quasirandom graph $G_u(X_i, X_j)$ of density d the mean-square density of $G_u(X_i, X_j)$ with respect to the partitions $X_i = X_{i1} \cup \ldots \cup X_{is_i}$ and $Y_j = Y_{j1} \cup \ldots \cup Y_{jt_j}$ is at least $d^2 + \epsilon^2/3$. Since a random pair (x,y) has a probability of at least ϵ of belonging to $X_i \times X_j$ for such a pair (i,j), a calculation similar to that of Lemma 7.5 shows that the sum of the mean-square densities of all the graphs G_u with respect to the partitions $\{X_{ik}\}$ and $\{Y_{jl}\}$ is at least $D + \epsilon^3/16$, as claimed. □

Theorem 7.8. *Let $\epsilon > 0$, let $X_1, \ldots, X_k$ be finite sets. For each i let $X'_{i1}, \ldots, X'_{im_i}$ be a partition of X_i, and for each i,j let $G_{ij}(1), \ldots, G_{ij}(r_{ij})$ be bipartite graphs that form a partition of the complete bipartite graph $K(X_i, X_j)$. Then for each i one can find a partition $X_{i1}, \ldots, X_{in_i}$ of X_i that refines the partition $X'_{i1}, \ldots, X'_{im_i}$, and this can be done in such a way that, for every i and j, if a random pair (x,y) is chosen from $X_i \times X_j$, then with probability at least $1 - \epsilon$ it lies in some set $X_{is} \times X_{jt}$ for which all the r_{ij} induced subgraphs $G_{ij}(u)(X_{is}, X_{jt})$ are ϵ-quasirandom. Moreover, all the n_i are bounded above by a function that depends on ϵ, the m_i and the r_{ij} only.*

Proof. Suppose that we have partitions $X_{i1}, \ldots, X_{in_i}$ of each X_i, and suppose that for these partitions the conclusion of the theorem is false. Then there exist i and j such that at least $\epsilon|X_i||X_j|$ of the pairs $(x,y) \in X_i \times X_j$ lie in sets $X_{is} \times X_{jt}$ for which at least one of the graphs $G_{ij}(u)(X_{is}, X_{jt})$ is not ϵ-quasirandom. By Lemma 7.7 we can refine the partitions $X_{i1}, \ldots, X_{in_i}$ and $X_{j1}, \ldots, X_{jn_j}$ in such a way that the sum of the mean-square densities of the graphs $G_{ij}(u)$ with respect to the refined partitions is greater by at least $\epsilon^3/16$ than it was for the original ones. Moreover, the numbers of sets in the new partitions are bounded above by an exponential function of the numbers in the old ones.

Since the sum of the mean-square densities of the graphs $G_{ij}(u)$ (over all i, j and u) cannot exceed $\sum_{i<j} r_{ij}$, this procedure must terminate after at most $16\epsilon^{-3}\sum_{i<j} r_{ij}$ steps. At that point we have a partition with the desired properties. If we start the iteration with the partitions given in the first place, then we end up proving the theorem. □

§8. Regularity for 3-uniform hypergraphs.

As ever, the picture for hypergraphs is more complicated. One of the reasons for this we have already met - we shall split our hypergraphs into "stable" subhypergraphs (see the discussion at the end of §5) rather than quasirandom ones, and this forces us to discuss chains (G, H) as well as hypergraphs. We shall find ourselves partitioning not just the vertex sets of H (and G) but also the edge sets of the graphs G, so the statements we prove are rather more elaborate.

A more technical complication, but nevertheless a fundamental one, arises out of the fact that we must consider chains (G, H) for which G is very sparse. This makes it hard to generalize Lemma 7.4 adequately. To see why, let G be an α-quasirandom tripartite graph of density p with vertex sets X, Y and Z, and suppose that p is very small. Let $H \subset \Delta(G)$ be a hypergraph that fails to be η-quasirandom relative to $\Delta(G)$. If α is small enough, then an averaging argument similar to that of Theorem 4.1 allows us to find sets $X' \subset X$, $Y' \subset Y$ and $Z' \subset Z$ such that the restriction of H to $X' \times Y' \times Z'$ is of significantly greater density than H itself. (This statement is true both relative to G and, since G is quasirandom, in absolute terms.) Unfortunately, the sets X', Y' and Z' that this argument gives are contained in certain neighbourhoods of vertices of G, and therefore their sizes depend not just on η but on p as well. Therefore, any increase in mean-square density that we can hope to get from the dense hypergraph $H(X', Y', Z')$ will also depend on p.

This matters a lot, because as the iteration proceeds in the hypergraph regularity lemma, we are forced to consider a sequence of graphs G_i with rapidly decreasing densities p_i, so if the increase in mean-square density at stage i depends on p_i, there is no guarantee that the iteration will come to an end.

The solution to this problem is to squeeze a bit more out of the proof of Theorem 4.1. Instead of choosing just one triple (X', Y', Z'), we shall choose several, and prove that they are sufficiently spread out to provide us with an increase in mean-square density that is strong enough to use.

The statement of the hypergraph regularity lemma is somewhat complicated, so we shall postpone it until after we have made the above remarks precise in Lemma 8.4 below. At that point, the formulation of the regularity lemma will be better motivated and the rest of the proof quite easy.

To begin with, here is a simple and general criterion that we can use when we are trying to establish that a partition gives us an increase in mean-square density.

Lemma 8.1. *Let U be a set of size n and let f and g be functions from U to the interval $[-1,1]$. Let $B_1,\dots,B_r$ be a partition of U and suppose that g is constant on each B_i. Then the mean-square density of f with respect to the partition $B_1,\dots,B_r$ is at least $\langle f,g\rangle^2/n\|g\|_2^2$.*

Proof. For each j Let a_j be the value taken by g on the set B_j. Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\langle f,g\rangle &= \sum_j a_j \sum_{x\in B_j} f(x)\\ &\leqslant \Big(\sum_j |B_j| a_j^2\Big)^{1/2}\Big(\sum_j |B_j|^{-1}\sum_{x\in B_j} f(x)\Big)^{1/2}\\ &= \|g\|_2\Big(\sum_j |B_j|\Big(|B_j|^{-1}\sum_{x\in B_j} f(x)\Big)^2\Big)^{1/2}.\end{aligned}$$

But $\sum_j |B_j|\big(|B_j|^{-1}\sum_{x\in B_j} f(x)\big)^2$ is n times the mean-square density of f (by definition), so the lemma follows. □

So that it does not clutter up the proof of Lemma 8.4, here is a second very simple technical lemma.

Lemma 8.2. *Let $0<\delta<1$ and let r be an integer greater than or equal to δ^{-1}. Let $v_1,\dots,v_n$ be vectors in ℓ_2^n such that $\|v_i\|^2\leqslant n$ for each i and such that $\|\sum v_i\|_2^2\leqslant \delta n^3$. Let r vectors $w_1,\dots,w_r$ be chosen uniformly and independently from the v_i. (To be precise, for each w_j an index i is chosen randomly between 1 and n and w_j is set equal to v_i.) Then the expectation of $\|\sum w_j\|_2^2$ is at most $2\delta r^2 n$.*

Proof. The expectation of $\|\sum w_j\|_2^2$ is the expectation of $\sum_{i,j}\langle w_i,w_j\rangle$. If $i\neq j$ then the expectation of $\langle w_i,w_j\rangle$ is $n^{-2}\|\sum v_i\|_2^2$ which, by hypothesis, is at most δn. If $i=j$, then

$\langle w_i, w_j \rangle$ is at most n, again by hypothesis. Therefore, the expectation we are trying to bound is at most $(\delta r(r-1)+r)n$. Since $\delta r \geqslant 1$, this is at most $2\delta r^2 n$, as claimed. $\square$

Definition 8.3. *Let G be a tripartite graph with vertex sets X, Y and Z and let the bipartite graphs $G(X,Y)$, $G(Y,Z)$ and $G(X,Z)$ be partitioned into subgraphs $G_i(X,Y)$, $G_j(Y,Z)$ and $G_k(X,Z)$ respectively. For each triangle $(x,y,z) \in \Delta(G)$, define its* index *to be the triple (i,j,k) such that $xy \in G_i(X,Y)$, $yz \in G_j(Y,Z)$ and $xz \in G_k(X,Z)$. The* induced partition *of $\Delta(G)$ is the partition of the triples of $\Delta(G)$ according to their index. If $f : X \times Y \times Z \to [-1,1]$, then the* mean-square density *of f* relative to *the partitions $(G_i(X,Y))$, $(G_j(Y,Z))$ and $(G_k(X,Z))$ is defined to be the mean-square density of f relative to the induced partition of $\Delta(G)$.*

Note that a typical cell of the induced partition is of the form

$$\Delta\big(G_i(X,Y) \cup G_j(Y,Z) \cup G_k(X,Z)\big) \ .$$

Lemma 8.4. *Let G be a tripartite graph with vertex sets X, Y and Z of sizes L, M and N respectively, let the densities $G(X,Y)$, $G(Y,Z)$ and $G(X,Z)$ be δ_{12}, δ_{23} and δ_{13}, let $\delta = \delta_{12}\delta_{23}\delta_{13}$ and suppose that these three graphs are α-quasirandom. Let H be a tripartite 3-uniform hypergraph with the same vertex sets as G, let the relative density $|H|/|\Delta(G)|$ of H in G be d, and suppose that H is not η-quasirandom relative to $\Delta(G)$. Suppose also that $2^{21}\alpha^{1/4} \leqslant \delta^7$. Then there are partitions $G_1(X,Y) \cup \ldots \cup G_l(X,Y) = G(X,Y)$, $G_1(Y,Z) \cup \ldots \cup G_m(Y,Z) = G(Y,Z)$ and $G_1(X,Z) \cup \ldots \cup G_n(X,Z) = G(X,Z)$, relative to which the mean-square density of H is at least $d^2 + 2^{-10}\eta^2$. Moreover, l, m and n are all at most $3^{\delta^{-4}}$.*

Proof. Let $f(x,y,z) = H(x,y,z) - d$ whenever $(x,y,z) \in \Delta(G)$ and let it be zero otherwise. Then the hypothesis of the lemma is that

$$\sum_{x,x',y,y',z,z'} f_{x,x',y,y',z,z'} \geqslant \eta\delta^4 L^2 M^2 N^2 \ .$$

Let $U = \Delta(G)$ and for each triple $(x,y,z) \in U$ define $F_{xyz}(x',y',z')$ to be $f_{x,x',y,y',z,z'}/f(x',y',z')$ and let $F(x',y',z') = \sum_{x,y,z} G_{xyz}(x',y',z')$. With this notation, the hypothesis can be rewritten $\langle f, F \rangle \geqslant \eta\delta^4(LMN)^2$.

Let us now build some new functions E_{xyz}. These will have similar properties to the F_{xyz} but will take values 0, 1 and -1 only. A vital property of each function F_{xyz} is that it can be written in the form $u(x',y')v(y',z')w(x',z')$, since of the eight terms in the product

$f_{x,x',y,y',z,z'}$ the only one that depends on all three of x', y' and z' is $f(x',y',z')$, which is absent from $F_{xyz}(x',y',z')$. Moreover, we can do this with u, v and w taking values in the interval $[-1,1]$.

Fix a triple (x,y,z) as above, and for each pair (x',y') define $u'(x,y)$ randomly according to the following simple rule. If $u(x',y')$ is positive then $u'(x',y')$ is 1 with probability $u(x',y')$ and 0 otherwise. If $u(x',y')$ is negative then $u'(x',y')$ is -1 with probability $|u(x',y')|$ and 0 otherwise. If $u(x',y')=0$ then $u'(x',y')=0$ as well. Construct functions v' and w' similarly, and let all the random choices that have been made be independent. Finally, let $E_{xyz}(x',y',z')=u'(x',y')v'(y',z')w'(x',z')$, and note that the expectation of $E_{xyz}(x',y',z')$ is $F_{xyz}(x',y',z')$.

Let E be the sum of all the functions E_{xyz}. The expectations of $\langle f,F_{xyz}\rangle$ and $\langle f,E_{xyz}\rangle$ are the same, so we can make the random choices in such a way that $\langle f,E\rangle\geqslant\eta\delta^4(LMN)^2$. The other property we shall use is that $E_{xyz}(x',y',z')$ is non-zero only if x, x', y, y', z and z' are the vertices of an octahedron in the graph G. (This is true despite the fact that we have divided by $f(x',y',z')$: for each pair there is still a triple containing it for which f is required to be non-zero.) So that our notation will be reasonably concise, Let us set $G_{x,x',y,y',z,z'}$ to be 1 if x, x', y, y', z and z' are the vertices of an octahedron and 0 otherwise. (Strictly speaking, it would be more accurate to write $\Delta(G)_{x,x',y,y',z,z'}$.)

Our next aim is to obtain an upper bound for $\|E\|_2^2$. To start with, we have

$$\begin{aligned}\|E\|_2^2 &= \sum_{x',y',z'}\Big(\sum_{x,y,z}E_{xyz}(x',y',z')\Big)^2\\ &\leqslant \sum_{x',y',z'}\Big(\sum_{x,y,z}G_{x,x',y,y',z,z'}\Big)^2\\ &= \sum_{x',y',z'}\sum_{x_1,y_1,z_1,x_2,y_2,z_2}G_{x_1,x',y_1,y',z_1,z'}G_{x_2,x',y_2,y',z_2,z'}\end{aligned}$$

This counts the number of 9-tuples $(x',y',z',x_1,y_1,z_1,x_2,y_2,z_2)$ such that the sextuples (x',x_1,y',y_1,z',z_1) and (x',x_2,y',y_2,z',z_2) both form octahedra in G. That is, it counts the number of copies in G of a certain graph with nine vertices and seven edges between each pair of vertex sets (four for each octahedron but intersecting in one). By Theorem 3.5 and our upper bound for α, there are at most $2\delta^7(LMN)^3$ of these.

We are not yet in a position to apply Lemma 8.1: E is the sum of $|U|$ functions E_{xyz} and we are unlikely to find a partition into just a few sets on which it is constant. The next stage of the argument is to make a random selection of a small number of E_{xyz} in such a way that their sum preserves the good properties of E. Let $r\geqslant\delta^{-4}$, choose

$E_1,\dots,E_r$ randomly from the E_{xyz} and let $D=\sum_{i=1}^r E_i$. Since the expectation of $\langle f,E_i\rangle$ is $|U|^{-1}\langle f,E\rangle$ and we know that $\langle f,E\rangle \geqslant \eta\delta^4(LMN)^2$, the expectation of $\langle f,D\rangle$ is at least $\eta\delta^4 r|U|^{-1}(LMN)^2$. By Theorem 3.5 and our upper bound for α, we know that $|U|$ lies between $\delta LMN/2$ and $2\delta LMN$, so this is at least $\eta r\delta^2|U|/4$.

We shall now apply Lemma 8.2, with $n=|U|$, to the functions E_{xyz}. We have shown that $\|E\|_2^2 \leqslant 2\delta^7(LMN)^3$, which is at most $16\delta^4|U|^3$. Moreover, each E_{xyz} is supported in U and takes values in $[-1,1]$, so $\|E_{xyz}\|_2^2$ is at most $|U|$. Therefore, Lemma 8.2 implies that the expectation of $\|D\|_2^2$ is at most $32\delta^4r^2|U|$. Therefore, the expectation of $256\delta^2r\langle f,D\rangle-\eta\delta\|D\|_2^2$ is at least $64\eta\delta^4r^2|U|$. From this it follows that we can choose D in such a way that $\langle f,D\rangle \geqslant \eta\delta^2r|U|/4$ and $\|D\|_2^2\leqslant 256\delta^2r\langle f,D\rangle\eta^{-1}$. For such a D, we have

$$\frac{\langle f,D\rangle}{\|D\|_2^2}\geqslant\frac{\langle f,D\rangle}{256\delta^2r\eta^{-1}}\geqslant\frac{\eta\delta^2r|U|}{2^{10}\delta^2r\eta^{-1}}=\frac{\eta^2|U|}{2^{10}}\ .$$

We finish the proof by applying Lemma 8.1 to a suitable partition of $U=\Delta(G)$. Each of the r functions E_{xyz} that we added up to make D is of the form $u'(x',y')v'(y',z')w'(x',z')$, where u', v' and w' take values in the set $\{-1,0,1\}$. We can partition the bipartite graph $G(X,Y)$ into at most 3^r subgraphs $G_i(X,Y)$ such that every u' is constant on each $G_i(X,Y)$. In a similar way we can partition $G(Y,Z)$ into subgraphs $G_j(Y,Z)$ and $G(X,Z)$ into subgraphs $G_k(X,Z)$. For each triangle $(x,y,z)\in\Delta(G)$, define its *index* to be the triple (i,j,k) such that $xy\in G_i(X,Y)$, $yz\in G_j(Y,Z)$ and $xz\in G_k(X,Z)$. Then partition $\Delta(G)$ into triples according to their index. This partition has at most 3^{3r} cells, each of the form $\Delta\big(G_i(X,Y)\cup G_j(Y,Z)\cup G_k(X,Z)\big)$.

The function D is constant on each cell, so Lemma 8.1 and our estimate for $\langle f,D\rangle/\|D\|_2^2$ imply that the mean-square density of f with respect to this partition is at least $2^{-10}\eta^2$. It follows that the mean-square density of H is at least $d^2+2^{-10}\eta^2$, which proves the lemma. □

Now let us prepare for the statement of our regularity lemma for 3-uniform hypergraphs. As with the counting lemma, we will keep things simple by restricting to the case of quadripartite hypergraphs, but the result can easily be generalized. The idea will be to take a quadripartite 3-uniform hypergraph H and "decompose it into quasirandom chains". Before we say exactly what this means, let us try to motivate the slightly technical definition we shall give in a moment of a quasirandom chain.

What we ultimately want from a chain is that it should satisfy the conditions for Theorem 6.8, the counting lemma for chains that we proved earlier. There we had a quadripartite graph G with vertex sets of sizes L, M, N and P and a quadripartite 3-uniform hypergraph $H \subset \Delta(G)$. Writing δ for the product of the densities of the six bipartite parts of G, the assumptions of the theorem were that each tripartite part of H was η-quasirandom relative to G, each bipartite part of G was α-quasirandom and $2^{36}\alpha^{1/16} \leqslant \eta\delta^8/6$. Writing γ for the product of the relative densities of the four tripartite parts of H, the conclusion was that the number of simplices in H differed from $\gamma\delta LMNP$ by at most $8\eta^{1/8}\delta LMNP$. We therefore consider this to be a small error if $\eta^{1/8}$ is small compared with γ.

Definition 8.5. *Let $\psi(\eta, \delta)$ be a polynomial in η and δ that vanishes when either η or δ is zero. A quadripartite chain (G, H) with δ defined as above is* (η, ψ)-quasirandom *if all six parts of G are $\psi(\eta, \delta)$-quasirandom and all four parts of H are η-quasirandom relative to G.*

Later we shall use this definition in the case where $\psi(\eta, \delta) = (2^{-40}\eta\delta^8)^{16}$. Then, if $\alpha = \psi(\eta, \delta)$ it will satisfy the condition $2^{36}\alpha^{1/16} \leqslant \eta\delta^8/6$ discussed above.

In Szemerédi's regularity lemma one decomposes a graph into quasirandom pieces using partitions of its vertex sets. As we have already mentioned, the decompositions we shall consider of hypergraphs are more complicated, so let us say precisely what they are. We shall use the following notation for this purpose and throughout the rest of the paper. If G is a quadripartite graph, X and Y are two of its vertex sets and $X' \subset X$ and $Y' \subset Y$, then $G(X', Y')$ stands for the induced bipartite subgraph of G with vertex sets X' and Y'. Similarly, if H is a quadripartite 3-uniform hypergraph and X', Y' and Z' are subsets of three of its vertex sets then $H(X', Y', Z')$ stands for the induced tripartite subhypergraph with vertex sets X', Y' and Z'. The complete k-partite graph with vertex sets $X_1, \dots, X_k$ will be denoted $K(X_1, \dots, X_k)$.

Definition 8.6. *Let X_1, X_2, X_3 and X_4 be four sets. A* decomposition *of the complete quadripartite graph $K(X_1, X_2, X_3, X_4)$ consists of the following:*

(a) for each vertex set X_i a partition into subsets $X_{i1}, \dots, X_{in_i}$;

(b) for each bipartite graph $K(X_i, X_j)$ a partition into subgraphs $G_{ij}(1), \dots, G_{ij}(m_{ij})$.

Such a decomposition provides us with a collection of tripartite graphs. A typical one of these graphs has vertex sets of the form X_{ir}, X_{js} and X_{kt}, and an edge set of the form

$$G_{ij}(u)(X_{ir}, X_{js}) \cup G_{jk}(v)(X_{js}, X_{kt}) \cup G_{ik}(w)(X_{ir}, X_{kt}) \ .$$

If H is a quadripartite 3-uniform hypergraph with vertex sets X_1, X_2, X_3 and X_4, then the decomposition also provides us with a collection of tripartite hypergraphs. A typical one of these has vertex sets as above and edge set $H \cap \Delta(G)$, where G is one of the tripartite graphs of the form the collection just defined. The pair (G, H) will then be a chain.

Definition 8.7. *Let H be a quadripartite 3-uniform hypergraph with vertex sets $X_1, \ldots, X_4$ and suppose that we have a decomposition of $K(X_1, X_2, X_3, X_4)$. Then the associated* chain decomposition *of H is the set of all chains formed in the way explained above.*

Every triple $(x, y, z) \in (X_i, X_j, X_k)$ belongs to exactly one of the tripartite graphs defined above, and hence we can associate with it exactly one of the chains. Hence, to each (x, y, z, w) we can associate four chains, one for each triple.

Definition 8.8. *Let H be as above. Then $H(X_i, X_j, X_k)$ is* (ϵ, η, ψ)-quasirandom *if for all but $\epsilon|X_i||X_j||X_k|$ of its edges (x, y, z) the associated tripartite chain is (η, ψ)-quasirandom. H itself is* (ϵ, η, ψ)-quasirandom *if all its four parts are.*

Just before we prove the regularity lemma, here is a generalization of Corollary 7.3 that we will need in the proof.

Lemma 8.9. *Let G be a tripartite graph with vertex sets X, Y and Z, let H be a tripartite 3-uniform hypergraph with the same vertex sets, and let d be the relative density $|H \cap \Delta(G)|/|\Delta(G)|$ of H in $\Delta(G)$. Let $E_1 \cup \ldots \cup E_l$, $F_1 \cup \ldots \cup F_m$ and $G_1 \cup \ldots \cup G_n$ be partitions of $G(X, Y)$, $G(Y, Z)$ and $G(X, Z)$ respectively. Let each E_i, F_j and G_k be partitioned further into sets E_{ir}, F_{js} and G_{kt}. Then the mean-square density of H relative to the partitions $\{E_{ir}\}$, $\{F_{js}\}$ and $\{G_{kt}\}$ is at least as big as the mean-square density of H relative to $\{E_i\}$, $\{F_j\}$ and $\{G_k\}$.*

Proof. Like Corollary 7.3, this result is an immediate consequence of Lemma 7.2. This time, let $U = \Delta(G)$ and let f be the characteristic function of H. The partition of U induced by the partitions $\{E_{ir}\}$, $\{F_{js}\}$ and $\{G_{kt}\}$ is a refinement of the partition induced by the partitions $\{E_i\}$, $\{F_j\}$ and $\{G_k\}$, and again the quantities compared in Lemma 7.2 are the mean-square densities we wish to compare here. □

We are now ready for the main result of this section.

Theorem 8.10. *Let H be a quadripartite hypergraph with vertex sets X, Y, Z and W, and let $\epsilon, \eta > 0$. Let the densities of $H(X, Y, Z)$, $H(X, Y, W)$, $H(X, Z, W)$ and $H(Y, Z, W)$*

be δ_{123}, δ_{124}, δ_{134} and δ_{234} respectively . Then there is a decomposition of the complete quadripartite graph $K(X,Y,Z,W)$ such that the associated chain decomposition of H is (ϵ,η,ψ)-quasirandom. Moreover, the number of bipartite graphs in the decomposition is bounded above by a number that depends only on ϵ and η, while the number of sets in the partitions of X, Y, Z and W is bounded above by a function of ϵ, η, ψ and the densities δ_{123}, δ_{124}, δ_{134} and δ_{234}.

Proof. Suppose that we have a decomposition of the vertices and edges of $K(X,Y,Z,W)$, with each vertex set partitioned into at most n sets and each complete bipartite graph formed from two of the vertex sets partitioned into at most m bipartite subgraphs.

Let $\alpha = \psi(\eta,\epsilon/48m)$. By Theorem 7.8 we can refine the partitions of X, Y, Z and W so that if a random quadruple (x,y,z,w) is chosen from $X\times Y\times Z\times W$, then, with probability at least $1-\alpha$, xy lies in a product $X_i\times Y_j$ of cells such that all the induced subgraphs $G(X_i,Y_j)$, where G is a graph from the decomposition of $K(X,Y)$, are α-quasirandom, and similarly for the other five pairs from (x,y,z,w). Suppose that we have passed to such a refinement. Then each vertex set is now partitioned into at most N sets, where N depends on ϵ, m and n only.

We do not want to consider chains that are too sparse, so let us show that these do not occur very often. To any quadruple $(x,y,z,w)\in X\times Y\times Z\times W$ there is an associated quadripartite chain. Let its vertex sets be A_1, A_2, A_3 and A_4 and let its edge-sets be G_{ij} for $1\leqslant i<j\leqslant 4$. Here, A_4 is the cell that contains w from the partition of W, G_{23} is the bipartite graph from the decomposition that contains the edge yz, and so on. The graphs G_{ij} form the edges of the quadripartite chain. The hyperedges come from the union of the four tripartite hypergraphs

$$H\cap\Delta\big(G_{ij}(A_i,A_j)\cup G_{jk}(A_j,A_k)\cup G_{ik}(A_i,A_k)\big)\ ,$$

where $1\leqslant i<j<k\leqslant 4$.

Suppose we choose (x,y,z,w) at random. Then there are at most m possibilities for each G_{ij}. It follows that if we condition on the set $A_1\times A_2\times A_3\times A_4$ in which (x,y,z,w) lies, then for each ij the probability that G_{ij} has density less than $\epsilon/48m$ in $A_i\times A_j$ is at most $\epsilon/48$. Therefore, with probability at least $1-\epsilon/8$, each G_{ij} has density at least $\epsilon/48m$ in $A_i\times A_j$.

Suppose that the chain decomposition we now have of H is not (ϵ,η,ψ)-quasirandom. Then, without loss of generality, for at least $\epsilon|X||Y||Z|/4$ edges of $H(X,Y,Z)$ the associated chain is not (η,ψ)-quasirandom. It follows that there are at least $\epsilon|X||Y||Z|/8$ edges

(x, y, z) of $H(X, Y, Z)$ such that the density conditions above are satisfied for the associated chains, and such that either $H(X, Y, Z)$ is not η-quasirandom relative to the associated tripartite graph or the three parts of the tripartite graph are not all α-quasirandom.

However, we have arranged for the second possibility to apply to at most $\alpha|X||Y||Z|$ triples, so there are at least $\epsilon|X||Y||Z|/16$ edges such that $H(X, Y, Z)$ is not η-quasirandom relative to the associated tripartite graph.

Whenever this happens we can use Lemma 8.4 to partition the three bipartite parts of the graph in such a way that the mean-square density of H relative to these partitions is greater by at least $2^{-10}\eta$ than the square of the density of H relative to the tripartite graph. Moreover, by Lemma 8.9 this property is maintained if we refine these partitions. Hence, we can choose a common refinement of all the partitions of all the bipartite parts of all the graphs, preserving all the mean-square density increases. Since at least $\epsilon|X||Y||Z|/16$ triples belong to tripartite graphs where such increases have taken place, we find that the mean-square density of H relative to the new decomposition is at least $2^{-14}\epsilon\eta$ greater than it was relative to the old one. The number of bipartite graphs in the new partitions is bounded above by a function of ϵ, η and m only.

This procedure can be iterated. When we refine the partitions of the vertex sets, the mean-square density of H relative to the decomposition does not decrease, so the iteration must eventually come to an end and the theorem is proved. $\square$

Remark. An examination of the above proof shows that the bound for the numbers of cells in the eventual partition is of "wowzer" type. The wowzer function W is defined in two steps as follows. First let the tower function T be defined by $T(1) = 2$, $T(n) = 2^{n-1}$. Next, let $W(1) = 2$ and $W(n) = T(W(n-1))$. In general, the hypergraph regularity lemma for k-uniform hypergraphs iterates the bound for $(k-1)$-uniform hypergraphs, and thus advances one level in the Ackermann hierarchy. This means that the bounds that it gives for the theorems of van der Waerden and Szemerédi are of Ackermann type. Similar bounds have recently been achieved for Szemerédi's theorem by Tao, who has produced a discretization of Furstenberg's ergodic-theory proof. This answered a question that many people had asked, and the insights gained played an important role in his spectacular result with Green that the primes contain arbitrarily long arithmetic progressions.

We finish the section with a corollary that is designed to make the regularity lemma easy and convenient to use. As with our earlier results, the result we give is not the most general possible, but more general versions can be proved with only small adaptations.

Let H be a quadripartite 3-uniform hypergraph with vertex sets X_1, X_2, X_3 and X_4, and for each i let the size of X_i be N_i. Suppose that for each i we have a subset $A_i \subset X_i$ of size $\delta_i|X_i|$ and that G is a quadripartite graph with vertex sets A_1, A_2, A_3 and A_4. For each pair $i < j$ let δ_{ij} be the density $|G(A_i, A_j)|/|A_i||A_j|$ and for each $i < j < k$ let δ_{ijk} be the relative density of H inside $\Delta\big(G(A_i, A_j) \cup G(A_j, A_k) \cup G(A_i, A_k)\big)$. Let us call the product of the densities δ_i, δ_{ij} and δ_{ijk} the *expected simplex density of H in G*.

Corollary 8.11. *Let $\epsilon > 0$, let H be a quadripartite graph with vertex sets X_1, X_2, X_3 and X_4, and for each i let the size of X_i be N_i. Then for each triple $1 \leqslant i < j < k \leqslant 4$ one can remove at most $\epsilon|H(X_i, X_j, X_k)|$ edges of $H(X_i, X_j, X_k)$, and one can find a decomposition of the complete quadripartite graph $K = K(X_1, X_2, X_3, X_4)$, with the following property. Let each X_i be partitioned into at most n parts and let the number of bipartite graphs in any of the six parts of K be m. Let H' equal H after the edges have been removed and let G be any quadripartite graph arising from the decomposition with vertex sets A_i. Let σ be the expected simplex density of H' in G. Then either $\sigma = 0$ or $\sigma \geqslant (\epsilon/8n)^4(\epsilon/48m)^6(\epsilon/8)^4$ and the number of simplices in the hypergraph $H' \cap \Delta(G)$ differs from $\sigma N_1N_2N_3N_4$ by at most $\epsilon\sigma N_1N_2N_3N_4$.*

Proof. Let γ be $(\epsilon/8)^3$, let $\eta = (\epsilon\gamma/16)^8$ and for any $\delta > 0$ let $\psi(\eta, \delta) = 2^{-40}\eta\delta^8)^{16}$. Using Theorem 8.10, let us take a decomposition of the graph $K = K(X_1, X_2, X_3, X_4)$ such that the associated chain decomposition of H is $(\epsilon/2, \eta, \psi)$-quasirandom, and let m be the maximum number of bipartite graphs in any of the six parts of K. For each $i < j < k$ we shall remove at most $\epsilon|X_i||X_j||X_k|$ edges of H, using the conclusion of Theorem 8.10 and simple averaging arguments. The result will be a subhypergraph H' such that each edge (x_i, x_j, x_k) has several good properties. To describe these properties, let us take an arbitrary such edge, let $A_i \subset X_i$, $A_j \subset X_j$ and $A_k \subset X_k$ be the vertex sets from the decomposition that contain x_i, x_j and x_k respectively and let G_{ij}, G_{jk} let G_{ik} be the bipartite graphs containing x_ix_j, x_jx_k and x_ix_k. The properties are then as follows.

(i) The densities of A_i, A_j and A_k inside X_i, X_j and X_k are all at least $\epsilon/24n$.

(ii) The densities of G_{ij}, G_{jk} and G_{ik} inside $K(A_i, A_j)$, $K(A_j, A_k)$ and $K(A_i, A_k)$ are all at least $\epsilon/48m$.

(iii) The relative density of H' inside the tripartite graph

$$G = G_{ij}(A_i, A_j) \cup G_{jk}(A_j, A_k) \cup G_{ik}(A_i, A_k)$$

is at least $\epsilon/8$.

(iv) The chain $(G, H' \cap \Delta(G))$ is $(\epsilon/2, \eta, \psi)$-quasirandom.

Now let us see how we can get these properties.

(i) Since X_i is partitioned into at most n sets A_i, the number of vertices that belong to an A_i of density less than $e/24n$ is at most $\epsilon|X_i|/24$. Therefore, the number of triples (x_i, x_j, x_k) in $X_i \times X_j \times X_k$ such that at least one of A_i, A_j and A_k has density less than $\epsilon/24n$ is at most $\epsilon|X_i||X_j|X_k|/8$.

(ii) This is obtained by an argument similar to that for (i). In fact, we gave the argument as part of the proof of Theorem 8.10, which implies that the number of triples (x_i, x_j, x_k) in $X_i \times X_j \times X_k$ for which this density condition fails is again at most $\epsilon|X_i||X_j|X_k|/8$.

(iii) Since $K(X_i, X_j, X_k)$ is partitioned into sets $\Delta(G)$, where G is a tripartite graph of the given form, at most $\epsilon|X_i||X_j|X_k|/8$ edges of $H(X_i, X_j, X_k)$ can live in a $\Delta(G)$ where the relative density of H is less than $\epsilon/8$.

(iv) The conclusion of Theorem 8.10 (with ϵ replaced by $\epsilon/2$) and Definition 8.8 together tell us that this is true for all but at most $\epsilon|X_i||X_j|X_k|/2$ edges of $H(X_i, X_j, X_k)$.

It follows that we may remove at most $\epsilon|X_i||X_j|X_k|$ from each part $H(X_i, X_j, X_k)$ of H and obtain a hypergraph H' such that properties (i), (ii), (iii) and (iv) hold for every single edge of H'.

We now apply Theorem 6.8. Let G be any quadripartite graph arising from the decomposition, let its vertex sets be A_1, A_2, A_3 and A_4 and let δ be the product of the densities of its edge sets. If all four hypergraphs of the form $\Delta\big(G(A_i, A_j, A_k)\big)$ contain edges of H' then the densities of all the $G(A_i, A_j)$ inside $K(A_i, A_j)$ are at least $\epsilon/48m$, the relative density of H' inside each $G(A_i, A_j) \cup G(A_j, A_k) \cup G(A_i, A_k)$ is at least $\epsilon/8$ and H' is $(\epsilon/2, \eta, \psi)$-quasirandom there. By our choices of η, ψ and the lower bounds for the densities, this means that the conditions for Theorem 6.8 are satisfied and we may conclude that the number of simplices in $H \cap \Delta(G)$ differs from $\sigma N_1 N_2 N_3 N_4$ by at most $8\eta^{1/8}\delta N_1 N_2 N_3 N_4$. By our choice of η, this is at most $\epsilon\sigma N_1 N_2 N_3 N_4$. □

§9. Szemerédi's theorem for progressions of length 4.

It is now straightforward to generalize the proof of Theorem 1.1 to give a proof of Theorem 1.4. Let us modify the statement a little bit so that it fits better with the statements of the last section.

Theorem 9.1. *For every $a > 0$ there exists $c > 0$ with the following property. Let H be any quadripartite hypergraph H with vertex sets X_1, X_2, X_3 and X_4 of sizes N_1,*

N_2, N_3 and N_4 respectively, and suppose that the number of simplices in H is at most $cN_1N_2N_3N_4$. Then it is possible to remove at most $aN_iN_jN_k$ triples from each of the four induced subhypergraphs $H(X_i, X_j, X_k)$ in such a way that the resulting subhypergraph of H is simplex-free.

Proof. Apply Corollary 8.11 with $\epsilon = a$ and suppose that the resulting hypergraph H' contains a simplex with vertices x_1, x_2, x_3 and x_4. Let G be the associated quadripartite graph coming from the decomposition. Then each tripartite part of G contains an edge of H', so the expected simplex density σ of H in G is non-zero. Corollary 8.11 tells us that it is therefore at least $(1-a)(a/8n)^4(a/48m)^6(a/8)^4$. If c is less than this, then we have obtained a contradiction, which shows that H' could not after all have contained a simplex. Since the numbers m and n depend on ϵ only, the theorem is proved. □

To conclude, let us deduce Szemerédi's theorem for progressions of length four. First we prove a tiny weakening of Theorem 1.5. (The difference is that we prove that $d \neq 0$ rather than that $d > 0$. Ben Green's trick mentioned in the proof of Corollary 1.2 works here as well, but Szemerédi's theorem does not need the positivity of d.)

Theorem 9.2. *For every $\delta > 0$ there exists N such that every subset $A \subset [N]^3$ of size at least δN^3 contains a quadruple of points of the form*

$$\{(x, y, z), (x + d, y, z), (x, y + d, z), (x, y, z + d)\}$$

with $d \neq 0$.

Proof. Define a quadripartite hypergraph with vertex sets $X = Y = Z = [N]$ and $W = [3N]$ as follows. A triple $(x, y, z) \in X \times Y \times Z$ belongs to H if and only if $(x, y, z) \in A$. A triple $(x, y, w) \in X \times Y \times W$ belongs to H if and only if $(x, y, w - x - y) \in A$, and similarly for triples (x, z, w) and (y, z, w).

Suppose now that H contains a simplex (x, y, z, w) and let $d = w - x - y - z$. Then the points (x, y, z), $(x, y, z + d)$, $(x, y + d, z)$ and $(x + d, y, z)$ all belong to A. This proves the theorem unless d is always 0. But in that case, there are at most N^3 simplices in H: in fact, there are exactly $|A|$ of them. Now we apply Theorem 9.1 with $a = \delta/20$. It gives us a $c > 0$ such that, if H contains at most cN^4 simplices, then we can remove at most aN^3 edges from each part of H and remove all of them. Since our hypergraph contains at most N^3 simplices, if $N^{-1} < c$ we may apply the theorem. However, the number of simplices with $d = 0$ is δN^3, and no two of these share a face, since a simplex with $d = 0$ is uniquely determined by any one of its four faces. Therefore, if we remove at most a proportion a of the faces from each of the four parts of H, we end up removing at most $10aN^3 = \delta N^3/2$ simplices, which is not all of them. (The number 10 comes from the fact that $|W| = 3N$.) That is a contradiction, and the theorem is proved. □

Corollary 9.3. *For every $\delta > 0$ there exists N such that every subset $A \subset [N]$ of size at least δN contains an arithmetic progression of length four.*

Proof. Define a subset $B \subset [N]^3$ to consist of all triples (x, y, z) such that $x+2y+3z \in A$. It is a straightforward exercise to show that B has density bounded below by a function of δ, so Theorem 9.3 yields the arithmetic progression

$$x + 2y + 3z, \quad (x + d) + 2y + 3z, \quad x + 2(y + d) + 3z, \quad x + 2y + 3(z + d) \ .$$

□

§10. Concluding Remarks.

The technical details in the papers of Rödl and his coauthors are very different from those here and it is possible and instructive to pinpoint where the difference arises. The answer is that the two approaches use different notions of quasirandomness that are equivalent for graphs and dense hypergraphs but *not* equivalent for the sparse hypergraphs we are forced to consider here.

Briefly, whereas in this paper we have focused on "octahedral quasirandomness", they concentrate their attention on edge-uniformity (see Definition 2.6). These two definitions are not equivalent for chains (G, H), because, as we noted at the beginning of Section 8, if G is sparse, then the failure of H to be relatively quasirandom does not imply that there is a significant increase in mean-square density.

This presented us with a technical problem (dealt with in Lemma 8.4) that does not arise if one uses edge-uniformity instead: that property is weaker, so its denial has stronger consequences. However, one pays the price for this weakness when proving a counting lemma. (In fact, this is an oversimplification: edge-uniformity does not seem to be enough on its own, but a more complicated variant of it can be used instead, more complicated in a way that is rather similar to the way that Lemma 8.4 is more complicated than Lemma 7.4.) Octahedral quasirandomness appears to be exactly the right concept for proving a counting lemma in an analytic style modelled on arguments that have already been used for proving Szemerédi's theorem; edge-uniformity is the concept one naturally comes up with if one wishes to model one's arguments on the usual proof of Szemerédi's regularity lemma. It is just a pity that one cannot use both!

Thus, even though the two approaches share many features, in that they both prove regularity and counting lemmas for chains, they are also genuinely different: for Rödl and his coauthors the counting lemma is harder than the regularity lemma, whereas for us it is the other way round.

References.

[CG1] F. R. K. Chung and R. L. Graham, *Quasi-random hypergraphs*, Random Structures Algorithms **1** (1990), 105–124.

[CG2] F.R.K. Chung, R.L. Graham, *Quasi-random subsets of* $\mathbb{Z}_n$, J. Comb. Th. A **61** (1992), 64–86.

[CGW] F.R.K. Chung, R.L. Graham, R.M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), 345–362.

[FKO] H. Furstenberg, Y. Katznelson, D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. **7** (1982), 527–552.

[FR1] P. Frankl and V. Rödl, *The uniformity lemma for hypergraphs*, Graphs Combin. **8** (1992), 309-312.

[FR2] P. Frankl and V. Rödl, *Extremal problems on set systems*, Random Structures and Algorithms **20** (2002), 131-164.

[FK] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275-291.

[G1] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **8** (1998), 529-551.

[G2] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465-588.

[G3] W. T. Gowers, *Hypergraph regularity and Szemerédi's theorem*, preprint.

[NRS] B. Nagle, V. Rödl and M. Schacht, *The counting lemma for regular* k*-uniform hypergraphs*, Random Structures and Algorithms (to appear).

[RS] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, 939-945.

[S1] J. Solymosi, *Note on a generalization of Roth's theorem*, Discrete and Computational Geometry, 825-827, Algorithms Combin. **25**, Springer, Berlin 2003.

[S2] J. Solymosi, *A note on a question of Erdős and Graham*, Combin. Probab. Comput. **13** (2004), 263-267.

[T] A.G. Thomason, *Pseudo-random graphs*, Proceedings of Random Graphs, Poznán 1985 (M. Karonski, ed.), Annals of Discrete Mathematics **33**, 307–331.

Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK.