GAFA, Geom. funct. anal.
Vol. 8 (1998) 1 – 0
1016-443X/98/030001-0 $ 1.50+0.20/0

GAFA Geometric And Functional Analysis

A NEW PROOF OF SZEMERÉDI'S THEOREM FOR ARITHMETIC PROGRESSIONS OF LENGTH FOUR

W.T. Gowers

1 Introduction

The famous theorem of Szemerédi asserts that, for any positive integer k and any real number $\delta > 0$, there exists N such that every subset of $\{1, 2, \ldots, N\}$ of cardinality at least δN contains an arithmetic progression of length k. The theorem trivially implies van der Waerden's theorem, and was, by the time it was proved by Szemerédi, a renowned and long-standing conjecture of Erdős and Turán [ET].

The first progress towards the theorem was due to Roth [R1], who proved the result in the special case $k = 3$, using exponential sums. Szemerédi later found a different, more combinatorial proof of this case, which he was able to extend to prove the result first for $k = 4$ [Sz1] and then eventually in the general case [Sz2]. There was then a further breakthrough due to Furstenberg [Fu], who showed that techniques of ergodic theory could be used to prove many Ramsey theoretic results, including Szemerédi's theorem and certain extensions of Szemerédi's theorem that were previously unknown.

These results left an obvious avenue unexplored: can Roth's proof for $k = 3$ be generalized to prove the whole theorem? The purpose of this paper is to show that it can, at least for the first "difficult" case $k = 4$. A subsequent paper will give rather more detail and an extension to the general case, which, although based on similar ideas, is significantly more complicated.

The motivation for generalizing Roth's argument is twofold. First, his argument is very natural and beautiful, and it is curious that it should not have an obvious generalization (though there are good reasons for this, as will become clear). Second, the bounds arising from the known proofs of Szemerédi's theorem are very weak, and in general for this sort of problem all the best bounds tend come from the use of exponential sums. For example, Roth shows that when $k = 3$ one can take N to be $\exp\exp(C/\delta)$ for some absolute constant C, which is far better than the bound given by any known combinatorial argument. This estimate has been reduced by Szemerédi [Sz3] and Heath-Brown [H] to $\exp((1/\delta)^C)$, also using exponential sums.

With our new approach, it is possible to show that there is an absolute constant $c > 0$ such that every subset of $\{1, 2, \dots, N\}$ of size at least $N(\log\log N)^{-c}$ contains an arithmetic progression of length four. Equivalently, there is an absolute constant C such that any subset of $\{1, 2, \dots, N\}$ of size at least δN contains an arithmetic progression of length four, as long as $N \geqslant \exp\exp((1/\delta)^C)$. In this paper we obtain instead a bound of $\exp\exp\exp((1/\delta)^C)$, as the argument is simpler. The improved bound will be presented in the later paper dealing with the general case.

Although a bound of this type may seem weak (and is almost certainly far from best possible) it is nevertheless a significant improvement on what went before. Even to state the earlier bounds needs some effort. Let us define the tower function T inductively by $T(1) = 2$ and $T(n+1) = 2^{T(n)}$. Next, define a function W inductively by $W(1) = 2$ and $W(n+1) = T(W(n))$. The previous best known bound for N has not been carefully calculated, but is at least as bad as $W(1/\delta)$. Even the bounds for van der Waerden's theorem are weak: to show that any r-colouring of $\{1, 2, \dots, N\}$ gives a monochromatic arithmetic progression of length four, the proofs need N to be at least as large as $T(T(r))$.

These earlier estimates rely on van der Waerden's theorem in its full generality, for which the best known bounds, due to Shelah [S], involve functions of the same type as the function W above. An important feature of our proof is that we avoid using van der Waerden's theorem, and also have no need for Szemerédi's uniformity lemma, which is known to require a bound similar to the function T [G]. Instead, our main tools are a well known consequence of Weyl's inequality and a deep theorem of Freiman.

It should be mentioned that Roth himself did find a proof for $k = 4$ [R2] which used analytic methods, but these were combined with certain combinatorial arguments of Szemerédi and the proof still used van der Waerden's theorem. The argument of this paper is quite different and more purely analytic, which is why it gives a better bound.

2 Quadratically Uniform Sets

In this section, we shall reduce Szemerédi's theorem for progressions of length four to a question that looks somewhat different. The rough idea is to define a notion of pseudorandomness, which we shall call quadratic uniformity, and show that every pseudorandom set, in the appropriate sense, contains about the same number of arithmetic progressions of length four as a random set of the same size. In later sections, we shall then prove that a set which *fails* to be pseudorandom can be restricted to a large arithmetic progression where its density increases noticeably. These two facts then easily imply the result.

In order to define quadratic uniformity, we shall need to introduce some notation. Given a positive integer N, we shall write $\mathbb{Z}_N$ for the group of integers mod N. When N is clear from the context (which will be always) we shall write ω for the number $\exp(2\pi i/N)$. Given any function $f : \mathbb{Z}_N \to \mathbb{C}$, we shall define its r^{th} Fourier coefficient $\tilde{f}(r)$ to be $\sum_{s\in\mathbb{Z}_N} f(s)\omega^{-rs}$. It would be more standard to write $\sum_{s\in\mathbb{Z}_N} f(s)e(-rs/N)$, where $e(x)$ is the function $\exp(2\pi ix)$. However, we have found the less standard notation convenient.

In our context, we shall often wish to consider convolutions of the form $h(s) = \sum_{t-u=s} f(t)\overline{g(u)}$. Again departing from standard notation, we shall write $f * g$ for this function. The two main properties of the discrete Fourier transform that we shall use are then

$$\sum_{r\in\mathbb{Z}_N} |\tilde{f}(r)|^2 = N \sum_{s\in\mathbb{Z}_N} |f(s)|^2 \tag{1}$$

and

$$(f * g)^{\sim}(r) = \tilde{f}(r)\overline{\tilde{g}(r)} \qquad (r \in \mathbb{Z}_N)\,. \tag{2}$$

There are two classes of functions to which we shall apply Fourier techniques. The first is what we shall call *balanced* functions associated with subsets $A \subset \mathbb{Z}_N$. Given such a set A, of size δN, we define its balanced function $f = f_A$ by

$$f(s) = \begin{cases} 1-\delta & s \in A \\ -\delta & s \notin A\,. \end{cases}$$

This is the characteristic function of A minus the constant function $\delta\mathbf{1}$. Note that $\sum_{s\in\mathbb{Z}_N} f_A(s) = \tilde{f}_A(0) = 0$ and that $\tilde{f}_A(r) = \tilde{A}(r)$ for $r \neq 0$. (Here, we have identified A with its characteristic function. We shall continue to do this.) The second class of functions that will interest us is functions of the form

$$g(s) = \begin{cases} \omega^{\phi(s)} & s \in B \\ 0 & s \notin B\,, \end{cases}$$

where B is a subset of $\mathbb{Z}_N$ and $\phi : B \to \mathbb{Z}_N$.

Another convention we shall adopt from now on is that any sum is over $\mathbb{Z}_N$ if it is not specified as being over another set. The next lemma contains some well known facts about functions on $\mathbb{Z}_N$ with small Fourier coefficients. When we say below that one statement with constant c_i implies another with constant c_j, we mean that the second statement follows from the first provided that $c_j \geqslant \gamma(c_i)$, for some function γ which tends to zero at zero. In fact, $\gamma(c_i)$ will always be some power of c_i.

LEMMA 1. *Let f be a function from $\mathbb{Z}_N$ to the unit disc in $\mathbb{C}$. The following are equivalent.*

(i) $\sum_r |\tilde{f}(r)|^4 \leqslant c_1 N^4$.
(ii) $\max_r |\tilde{f}(r)| \leqslant c_2 N$.
(iii) $\sum_k \big|\sum_s f(s)\overline{f(s-k)}\,\big|^2 \leqslant c_3 N^3$.
(iv) $\sum_k \big|\sum_s f(s)\overline{g(s-k)}\,\big|^2 \leqslant c_4 N^2 \|g\|_2^2$ *for every function $g : \mathbb{Z}_N \to \mathbb{C}$.*

Proof. Using identities (2) and (1) above, we have

$$\begin{aligned}\sum_k \Big|\sum_s f(s)\overline{g(s-k)}\,\Big|^2 &= \sum_k |f * g(k)|^2 \\ &= N^{-1}\sum_r |(f*g)^{\sim}(r)|^2 \\ &= N^{-1}\sum_r |\tilde{f}(r)|^2 |\tilde{g}(r)|^2 \\ &\leqslant \Big(\sum_r |\tilde{f}(r)|^4\Big)^{1/2}\Big(\sum_r |\tilde{g}(r)|^4\Big)^{1/2}\end{aligned}$$

by the Cauchy–Schwarz inequality. If $f = g$, then equality holds above, which gives the equivalence between (i) and (iii) with $c_1 = c_3$. It is obvious that (iv) implies (iii) if $c_3 \geqslant c_4$. Using the additional inequality

$$\Big(\sum_r |\tilde{g}(r)|^4\Big)^{1/2} \leqslant \sum_r |\tilde{g}(r)|^2,$$

we can deduce (iv) from (i) if $c_4 \geqslant c_1^{1/2}$.

Since $\max_r |\tilde{f}(r)| \leqslant \big(\sum_r |\tilde{f}(r)|^4\big)^{1/4}$, one can see that (ii) follows from (i) if $c_2 \geqslant c_1^{1/4}$. For the reverse implication, we use the fact that

$$\sum_r |\tilde{f}(r)|^4 \leqslant \max_r |\tilde{f}(r)|^2 \sum_r |\tilde{f}(r)|^2 .$$

By identity (1) and the restriction on the image of f, we have the estimate $\sum_r |\tilde{f}(r)|^2 \leqslant N^2$, so that (i) follows from (ii) if $c_1 \geqslant c_2^2$. □

If f satisfies condition (i) with $c_1 = \alpha$, then we shall say that f is *α-uniform.* If f is the balanced function of a set A, we shall say also that A is α-uniform. (This definition coincides with the definition made by Chung and Graham of a *quasirandom* subset of $\mathbb{Z}_N$ [CGr].)

Roth's proof can be presented as follows. Let A be a subset of $\mathbb{Z}_N$ of size δN. If A is α-uniform for a suitable α (a power of δ, where $|A| = \delta N$) then A contains roughly the expected number of arithmetic progressions of length three. (This follows easily from Lemma 6 below.) If not, then some non-zero Fourier coefficient of the characteristic function of A is a large fraction of N. It follows easily that there is a subset $I = \{a + d, a + 2d, \dots, a + md\} \subset \mathbb{Z}_N$ such that m is a substantial fraction of N and $|A \cap I| \geqslant (\delta + \epsilon)m$ for some $\epsilon > 0$ which is also a power of δ. It can be shown quite easily (see for example Lemma 17 of this paper) that I can be partitioned into genuine arithmetic progressions (that is, when considered as subsets of $\mathbb{Z}$) of size about $m^{1/2}$. Hence, there is an arithmetic progression P of about this size such that $|A \cap P| \geqslant (\delta + \epsilon)|P|$. Now repeat the argument for P. The number of times it can be repeated depends only on δ, so, provided N is large enough, there must be an arithmetic progression of size three in A.

It turns out that, even if α is extremely small, an α-uniform set need not contain roughly the expected number of arithmetic progressions of length four. (An example will be presented in a future paper.) For this reason, if we wish to have an approach similar to the above one, but for progressions of length four, then we need a stronger notion of pseudorandomness. Given a function $f : \mathbb{Z}_N \to \mathbb{Z}_N$ and $k \in \mathbb{Z}_N$, define a function $\Delta(f; k)$ by $\Delta(f; k)(s) = f(s)\overline{f(s-k)}$. Notice that if $f(s) = \omega^{\phi(s)}$ for some function $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$, then $\Delta(f; k)(s) = \omega^{\phi(s) - \phi(s-k)}$.

LEMMA 2. *Let f be a function from $\mathbb{Z}_N$ to the closed unit disc in $\mathbb{C}$. The following are equivalent.*

(i) $\sum_u \sum_v \left| \sum_s f(s)\overline{f(s-u)f(s-v)}f(s-u-v)\right|^2 \leqslant c_1 N^4$.

(ii) $\sum_k \sum_r |\Delta(f;k)^{\sim}(r)|^4 \leqslant c_2 N^5$.

(iii) $|\Delta(f;k)^{\sim}(r)| \geqslant c_3 N$ *for at most* $c_3^2 N$ *pairs* (k, r).

(iv) *For all but* $c_4 N$ *values of* k *the function* $\Delta(f;k)$ *is* c_4*-uniform.*

Proof. The equivalence of (i) and (ii) with $c_1 = c_2$ follows, as in the proof of the equivalence of (i) and (iii) in Lemma 1, by expanding. Alternatively, it can be deduced by applying that result to each function $\Delta(f;k)$ and adding.

If $|\Delta(f;k)\tilde{}(r)| \geqslant c_3N$ for more than c_3^2N pairs (k,r) then obviously

$$\sum_k\sum_r |\Delta(f;k)\tilde{}(r)|^4 > c_3^6N^5\,,$$

so (ii) implies (iii) provided that $c_2 \leqslant c_3^6$. If (ii) does not hold, then there are more than $c_2N/2$ values of k such that $\sum_r |\Delta(f;k)\tilde{}(r)|^4 \leqslant c_2N^4/2$. By the implication of (i) from (ii) in Lemma 1 this implies that there are more than $c_2N/2$ values of k such that $\max_r |\Delta(f;k)\tilde{}(r)| \geqslant (c_2/2)^{1/2}N$, and hence (iii) implies (ii) as long as $c_2 \geqslant 2c_3^2$. Finally, it is easy to see that (iv) implies (ii) if $c_2 \geqslant 2c_4$ and (ii) implies (iv) if $c_2 \leqslant c_4^2$. □

A function satisfying property (i) above with $c_1 = \alpha$ will be called *quadratically* α-uniform. A set will be called quadratically α-uniform if its balanced function is. Let us define a *square* and a *cube in* $\mathbb{Z}_N$ to be sequences of the form $(s, s+a, s+b, s+a+b)$ and $(s, s+a, s+b, s+c, s+a+b, s+a+c, s+b+c, s+a+b+c)$ respectively. The number of squares in a set A is easily seen to be $N^{-1}\sum_r |\tilde{A}(r)|^4$. It follows that if A has cardinality δN, then it contains at least δ^4N^3 squares and is α-uniform if and only if it contains at most $(\delta^4+\alpha)N^3$ squares. It is not hard to show that A contains at least δ^8N^4 cubes, and that A is quadratically uniform if and only if it contains at most $\delta^8(1+\epsilon)N^4$ cubes for some small ϵ. However, we shall not need this result. The aim of the rest of this section is to show that a quadratically uniform set contains roughly the expected number of arithmetic progressions of length four.

LEMMA 3. *For $1 \leqslant i \leqslant k$ let $f_i : \mathbb{Z}_N \to D$ be an α_i-uniform function. Then $f_1 + \cdots + f_k$ is $(\alpha_1^{1/4} + \cdots + \alpha_k^{1/4})^4$-uniform.*

Proof. This follows immediately from the definition and the fact that $f \mapsto \left(\sum_r |\tilde{f}(r)|^4\right)^{1/4}$ is a norm. □

LEMMA 4. *Let $A \subset \mathbb{Z}_N$ be a quadratically α-uniform set of size δN. Then, for all but at most $\alpha^{1/2}N$ values of k, $A \cap (A+k)$ is $81\alpha^{1/2}$-uniform, and, for all but at most $\alpha^{1/4}N$ values of k, $\big||A\cap(A+k)| - \delta^2N\big| \leqslant \alpha^{1/8}N$.*

Proof. Let f be the balanced function of A. Then

$$A\cap(A+k)(s) = \delta^2 + \delta f(s) + \delta f(s-k) + f(s)f(s-k)\,.$$

The implication of (iv) from (i) in Lemma 2 implies that for all but $\alpha^{1/2}N$ values of k, the function $f(s)f(s-k)$ is $\alpha^{1/2}$-uniform. Expanding condition (iii) of Lemma 1 and then applying the Cauchy–Schwarz inequality shows that if f is quadratically α-uniform, then it is also $\alpha^{1/2}$-uniform. Therefore, by Lemma 3, $A \cap (A+k)$ is $81\alpha^{1/2}$-uniform for at least $(1-\alpha^{1/2})N$ values of k. As for the size of $A\cap(A+k)$, it is $\delta^2+\sum_s f(s)f(s-k)$. Since f is $\alpha^{1/2}$-uniform, condition (iii) of Lemma 1 tells us that $\sum_k\left|\sum_s f(s)f(s-k)\right|^2 \leqslant \alpha^{1/2}N^3$, which implies the assertion. □

Let $f : \mathbb{Z}_N \to \mathbb{R}$. Then the Cauchy–Schwarz inequality implies that $\|f\|_2 \geqslant N^{-1/2}\,\|f\|_1$. At one point in the argument to come, we shall exploit the fact that a function $f : \mathbb{Z}_N \to \mathbb{R}_+$ for which equality almost occurs is close to being constant. A precise statement of what we shall use follows (which is basically Tchebyshev's inequality).

LEMMA 5. *Let $f : \mathbb{Z}_N \to \mathbb{R}_+$ be a function with $\|f\|_1 = wN$ and suppose that $\|f\|_2^2 \leqslant (1+\epsilon)w^2N$. Let A be a subset of $\mathbb{Z}_N$. Then $|\sum_{s\in A} f(s) - w|A|| \leqslant \epsilon^{1/2}wN^{1/2}|A|^{1/2}$.*

Proof. The mean of f is w and the variance is ϵw^2. Therefore

$$\Big|\sum_{s\in A} f(s) - w|A|\Big| \leqslant \sum_{s\in A}|f(s)-w| \leqslant |A|^{1/2}\Big(\sum_{s\in A}(f(s)-w)^2\Big)^{1/2}$$
$$\leqslant \epsilon^{1/2}wN^{1/2}|A|^{1/2}\,.$$ □

The proof of the next lemma gives a better bound than the one we shall actually state. However, the improvement is less tidy to use and does not make a significant difference to our eventual bound.

LEMMA 6. *Let A, B and C be subsets of $\mathbb{Z}_N$ of cardinalities $\alpha N, \beta N$ and γN respectively. Suppose that C is η-uniform. Then*

$$\Big|\sum_r |A\cap(B+r)\cap(C+2r)| - \alpha\beta\gamma N^2\Big| \leqslant \eta N^2\,.$$

Proof. Let us identify A, B and C with their characteristic functions. Then

$$\begin{aligned}\sum_r |A\cap(B+r)\cap(C+2r)| &= \sum_r\sum_s A(s)B(s-r)C(s-2r)\\ &= N^{-1}\sum_p\sum_{x,y,z} A(x)B(y)C(z)\omega^{-p(x-2y+z)}\\ &= N^{-1}\sum_{p\neq 0}\tilde{A}(p)\tilde{B}(-2p)\tilde{C}(p) + N^{-1}|A|\,|B|\,|C|\,.\end{aligned}$$

However, by the η-uniformity of C and the Cauchy–Schwarz inequality,

$$\Big|\sum_{p\neq 0}\tilde{A}(p)\tilde{B}(-2p)\tilde{C}(p)\Big| \leqslant \eta N\|\tilde{A}\|_2\|\tilde{B}\|_2 \leqslant \eta N^2\,,$$

which proves the lemma. □

Lemma 7. *Let A,B,C and D be subsets of $\mathbb{Z}_N$ of cardinality $\alpha N,\beta N,\gamma N$ and δN respectively. Suppose that C is η-uniform and D is quadratically η-uniform for some $\eta\leqslant 2^{-20}$. Then*

$$\Big|\sum_r \big|A\cap(B+r)\cap(C+2r)\cap(D+3r)\big|-\alpha\beta\gamma\delta N^2\Big|\leqslant 3\eta^{1/16}N^2/\beta\gamma\delta\,.$$

Proof. Once again, identify sets with their characteristic functions and let $f(s)=\sum_r B(s-r)C(s-2r)D(s-3r)$. We shall estimate the norms $\|f\|_1$ and $\|f\|_2$. The proof of Lemma 4 tells us that D is $\eta^{1/2}$-uniform. Hence, by Lemma 6,

$$\begin{aligned}\|f\|_1&=\sum_s\sum_r B(s-r)C(s-2r)D(s-3r)\\&=\sum_r\big|B\cap(C+r)\cap(D+2r)\big|\\&\geqslant N^2(\beta\gamma\delta-\eta^{1/2})\,.\end{aligned}$$

Lemma 6 also tells us that $\|f\|_1\leqslant N^2(\beta\gamma\delta+\eta^{1/2})$, which we shall need to know later. As for $\|f\|_2$, we have that

$$\|f\|_2^2=\sum_s\sum_{r,q}B(s-r)B(s-q)C(s-2r)C(s-2q)D(s-3r)D(s-3q)\,.$$

If we substitute $p=q-r$, then this becomes

$$\begin{aligned}&\sum_s\sum_{r,p}B(s-r)B(s-r-p)C(s-2r)C(s-2r-2p)D(s-3r)D(s-3r-3p)\\&=\sum_{r,p}\big|(B+r)\cap(B+r+p)\cap(C+2r)\cap(C+2r+2p)\cap(D+3r)\cap(D+3r+3p)\big|\\&=\sum_{r,p}\big|(B\cap(B+p))\cap(C\cap(C+2p)+r)\cap(D\cap(D+3p)+2r)\big|\,.\end{aligned}$$

By Lemma 4, $D\cap(D+3p)$ is $\eta^{1/2}$-uniform for all but at most $81\eta^{1/2}N$ values of p. When $D\cap(D+3p)$ is $\eta^{1/2}$-uniform, Lemma 6 implies that

$$\sum_r\big|(B\cap(B+p))\cap(C\cap(C+2p)+r)\cap(D\cap(D+3p)+2r)\big|$$

is at most

$$N^{-1}\big|B\cap(B+p)\big|\big|C\cap(C+2p)\big|\big|D\cap(D+3p)\big|+\eta^{1/2}N^2\,.$$

Summing over p, this tells us that

$$\|f\|_2^2\leqslant N^{-1}\sum_p\big|B\cap(B+p)\big|\big|C\cap(C+2p)\big|\big|D\cap(D+3p)\big|+82\eta^{1/2}N^3\,.$$

Because C and D are quadratically η-uniform, Lemma 4 implies that

$$\left|C\cap(C+2p)\right|\leqslant\gamma^2N+\eta^{1/8}N$$

and

$$\left|D\cap(D+3p)\right|\leqslant\delta^2N+\eta^{1/8}N$$

except for at most $2\eta^{1/4}N$ values of p. Therefore,

$$\begin{aligned}\|f\|_2^2&\leqslant N^{-1}\sum_p\left|B\cap(B+p)\right|(\gamma^2\delta^2N+2\eta^{1/8}N)+2\eta^{1/4}N^3+82\eta^{1/2}N^3\\&\leqslant N^3(\beta^2\gamma^2\delta^2+3\eta^{1/8})\end{aligned}$$

because of our restriction on the size of η. We have now shown that

$$\|f\|_2^2\leqslant N^{-1}\|f\|_1^2\left(1+\frac{3\eta^{1/8}}{\beta^2\gamma^2\delta^2}\right)\left(1-\frac{\eta^{1/2}}{\beta\gamma\delta}\right)^{-2}\leqslant N^{-1}\|f\|_1^2\left(1+4\frac{\eta^{1/8}}{\beta^2\gamma^2\delta^2}\right).$$

We now apply Lemma 5 with $\epsilon=4\eta^{1/8}/\beta^2\gamma^2\delta^2$ and $|w-\beta\gamma\delta N|\leqslant\eta^{1/2}N$, to deduce that

$$\left|\sum_{s\in A}f(s)-\alpha\beta\gamma\delta N^2\right|\leqslant\eta^{1/2}N^2+2\alpha^{1/2}\eta^{1/16}N^2/\beta\gamma\delta\leqslant3\eta^{1/16}N^2/\beta\gamma\delta$$

which is equivalent to the assertion of the lemma. □

COROLLARY 8. *Let $A_0\subset\mathbb{Z}_N$ be a quadratically η-uniform set of size δN, where $\eta\leqslant2^{-208}\delta^{112}$ and $N>200\delta^{-3}$. Then A_0 contains an arithmetic progression of length four.*

Proof. In Lemma 7, take A and B to be $A_0\cap[2N/5,3N/5)$ and take C and D to be A_0. Since A_0 is $\eta^{1/2}$-uniform, the upper bound on η implies that A and B have cardinality at least $\delta N/10$. (Otherwise, it can easily be shown, there would be at least one non-trivial large Fourier coefficient.) The lemma and the bound on η then imply that there are at least $\delta^4N^2/200$ sequences of the form $(a,a+d,a+2d,a+3d)$ in $A\times B\times C\times D$. Of these, at most δN can have $d=0$. Therefore, there is at least one with $d\neq0$. Since a and $a+d$ belong to the interval $[2N/5,3N/5)$, we have $a+2d$ in the interval $[N/5,4N/5)$ and $a+3d$ in $[0,N)$, even when these numbers are considered as elements of $\mathbb{Z}$. That is, the sequence $(a,a+d,a+2d,a+3d)$ is a genuine arithmetic progression and not just an arithmetic progression mod N. □

3 Finding Many Additive Quadruples

We have just seen that a quadratically uniform set must contain an arithmetic progression of length four. We now begin an argument of several steps, which will eventually show that if A is a subset of $\mathbb{Z}_N$ of cardinality δN which *fails* to be quadratically α-uniform, then there is an arithmetic progression $P \subset \mathbb{Z}_N$ (which is still an arithmetic progression when regarded as a subset of $\{1, 2, \ldots, N\}$) of size N^β such that $|A \cap P| \geqslant (\delta+\epsilon)|P|$, where β and ϵ depend on α and δ only.

If A fails to be quadratically α-uniform, then so does its balanced function f (by definition). This tells us that there are many values of k for which the function $\Delta(f;k)$ has a large (meaning proportional to N) Fourier coefficient r. In the next result, we shall show that the set of pairs (k,r) for which $\Delta(f;k)^\sim(r)$ is large is far from arbitrary.

PROPOSITION 9. *Let $\alpha > 0$, let $f : \mathbb{Z}_N \to D$, let $B \subset \mathbb{Z}_N$ and let $\phi : B \to \mathbb{Z}_N$ be a function such that*

$$\sum_{k\in B} \left|\Delta(f;k)^\sim(\phi(k))\right|^2 \geqslant \alpha N^3 .$$

Then there are at least $\alpha^4 N^3$ quadruples $(a,b,c,d) \in B^4$ such that $a+b = c+d$ and $\phi(a)+\phi(b) = \phi(c)+\phi(d)$.

Proof. Expanding the left hand side of the inequality in the statement tells us that

$$\sum_{k\in B}\sum_{s,t} f(s)\overline{f(s-k)f(t)}f(t-k)\omega^{-\phi(k)(s-t)} \geqslant \alpha N^3 .$$

If we now introduce the variable $u = s-t$ we can rewrite this as

$$\sum_{k}\sum_{s,u} f(s)\overline{f(s-k)f(s-u)}f(s-k-u)\omega^{-\phi(k)u} \geqslant \alpha N^3 .$$

Since $|f(s)| \leqslant 1$ for every s, it follows that

$$\sum_{u}\sum_{s}\left|\sum_{k\in B}\overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u}\right| \geqslant \alpha N^3$$

which implies that

$$\sum_{u}\sum_{s}\left|\sum_{k\in B}\overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u}\right|^2 \geqslant \alpha^2 N^4 . \qquad (*)$$

For fixed u, let $\gamma(u)$ be defined by the equation

$$\sum_{s}\left|\sum_{k\in B}\overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u}\right|^2 = \gamma(u)N^3 .$$

This shows that the function $\overline{B(k)\omega^{\phi(k)u}}$ has a large inner product with many translates of the function $\overline{\Delta(f;u)}$ (both considered as functions of k). Lemma 1 implies that both functions have at least one large Fourier coefficient. To be precise, if we apply the implication of (iv) from (i) in Lemma 1 to these functions, then we can deduce that

$$\sum_r \Big|\sum_{k\in B}\omega^{\phi(k)u-rk}\Big|^4 \geqslant \gamma(u)^2N^4\,. \qquad (**)$$

Inequality $(*)$ implies that $\sum_u \gamma(u) \geqslant \alpha^2 N$, which implies that $\sum_u \gamma(u)^2 \geqslant \alpha^4 N$. Hence, taking inequality $(**)$ and summing over u, we obtain

$$\sum_u\sum_r \Big|\sum_{k\in B}\omega^{\phi(k)u-rk}\Big|^4 \geqslant \alpha^4 N^5\,.$$

Expanding the left hand side we find that

$$\sum_{u,r}\ \sum_{a,b,c,d\in B} \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))}\omega^{-r(a+b-c-d)} \geqslant \alpha^4 N^5\,.$$

But now the left hand side is exactly N^2 times the number of quadruples $(a,b,c,d) \in B^4$ for which $a+b=c+d$ and $\phi(a)+\phi(b)=\phi(c)+\phi(d)$. This proves the proposition. □

We shall call a quadruple with the above property *additive*. In the next section, we shall show that functions with many additive quadruples have a very interesting structure.

4 An Application of Freiman's Theorem

There is a wonderful theorem due to Freiman about the structure of finite sets $A \subset \mathbb{Z}$ with the property that $A + A = \{x + y : x, y \in A\}$ is not much larger than A. Let us define a d-dimensional arithmetic progression to be a set of the form $P_1 + \cdots + P_d$, where the P_i are ordinary arithmetic progressions. It is not hard to see that if $|A| = m$ and A is a subset of a d-dimensional arithmetic progression of size Cm, then $|A + A| \leqslant 2^d Cm$. Freiman's theorem [F1,2] tells us that these are the *only* examples of sets with small double set.

Theorem 10. *Let C be a constant. There exist constants d and K, depending only on C, such that, whenever A is a subset of $\mathbb{Z}$ with $|A| = m$ and $|A + A| \leqslant Cm$, there exists an arithmetic progression Q of dimension at most d such that $|Q| \leqslant Km$ and $A \subset Q$.*

In fact, we wish to apply Freiman's theorem to subsets of $\mathbb{Z}^2$, but it is an easy exercise to embed such a subset "isomorphically" into $\mathbb{Z}$ and deduce the appropriate result from Theorem 10. Freiman's proof of his theorem did not give a bound for d and K, but recently an extremely elegant proof was discovered by Ruzsa which gives quite a good bound [Ru]. A better bound for Szemerédi's theorem can be obtained by modifying the statement of Freiman's theorem, and modifying Ruzsa's proof accordingly. However, this modification will be presented in a future paper - the priority here is to keep the argument as simple as possible, given known results.

We shall be applying Freiman's theorem to graphs of functions with many additive quadruples. If ? is such a graph, then we can regard ? as a subset of $\mathbb{Z}^2$. To every additive quadruple we can associate a quadruple of points $(x,y,z,w) \in$? such that $x+y=z+w$, where the addition is in $\mathbb{Z}^2$. It turns out to be convenient to consider instead quadruples with $x-y=z-w$ but they are clearly in one-to-one correspondence with the other kind.

The assumption that A is a subset of $\mathbb{Z}^2$ containing many quadruples (x,y,z,w) with $x-y=z-w$ tells us virtually nothing about the size of $A+A$, since half of A might be very nice and the remainder arbitrary. Even the stronger property that all large subsets of A contain many such quadruples (which comes out of Proposition 9) is not enough. For example, A could be the union of a horizontal line and a vertical line. What we shall show is that A has a reasonably large *subset* B such that $|B+B|$ is reasonably small. We will then be able to apply Freiman's theorem to the set B. This result, in its qualitative form, is due to Balog and Szemerédi [BSz]. However, they use Szemerédi's uniformity lemma, which, as we mentioned in the introduction, produces a very weak bound. We therefore need a different argument, which will be the main task of this section. We begin with a combinatorial lemma.

LEMMA 11. *Let X be a set of size m, let $\delta > 0$ and let $A_1, \dots, A_n$ be subsets of X such that $\sum_{x=1}^n \sum_{y=1}^n |A_x \cap A_y| \geqslant \delta^2 mn^2$. There is a subset $K \subset [n]$ of cardinality at least $2^{-1/2}\delta^5 n$ such that for at least 90% of the pairs $(x,y) \in K^2$ the intersection $A_x \cap A_y$ has cardinality at least $\delta^2 m/2$. In particular, the result holds if $|A_x| \geqslant \delta m$ for every x.*

Proof. For every $j \leqslant m$ let $B_j = \{i : j \in A_i\}$ and let $E_j = B_j^2$. Choose five numbers $j_1, \dots, j_5 \leqslant m$ at random (uniformly and independently), and let $X = E_{j_1} \cap \dots \cap E_{j_5}$. The probability p_{xy} that a given pair $(x,y) \in [n]^2$ belongs to E_{j_r} is $m^{-1}|A_x \cap A_y|$, so the probability that it belongs to X

is p_{xy}^5. By our assumption we have that $\sum_{x,y=1}^n p_{xy} \geqslant \delta^2 n^2$, which implies (by Hölder's inequality) that $\sum_{x,y=1}^n p_{xy}^5 \geqslant \delta^{10} n^2$. In other words, the expected size of X is at least $\delta^{10} n^2$.

Let Y be the set of pairs $(x,y) \in X$ such that $|A_x \cap A_y| < \delta^2 m/2$, or equivalently $p_{xy} < \delta^2/2$. Because of the bound on p_{xy}, the probability that $(x,y) \in Y$ is at most $(\delta^2/2)^5$, so the expected size of Y is at most $\delta^{10} n^2/32$.

It follows that the expectation of $|X| - 16|Y|$ is at least $\delta^{10} n^2/2$. Hence, there exist $j_1, \dots, j_5$ such that $|X| \geqslant 16|Y|$ and $|X| \geqslant \delta^{10} n^2/2$. This proves the lemma, with $X = K^2$ (so $K = B_{j_1} \cap \dots \cap B_{j_5}$). □

Let A be a subset of $\mathbb{Z}^D$ and identify A with its characteristic function. Then $A * A(x)$ is the number of pairs $(y,z) \in A^2$ such that $y - z = x$. (Recall that we have a non-standard use for the symbol "$*$".) Hence, the number of quadruples $(x,y,z,w) \in A^4$ with $x - y = z - w$ is $\|A * A\|_2^2$. The next result is a precise statement of the Balog–Szemerédi theorem, but, as we have mentioned, the bounds obtained in the proof are new.

Proposition 12. *Let A be a subset of $\mathbb{Z}^D$ of cardinality m such that $\|A * A\|_2^2 \geqslant c_0 m^3$. There are constants c and C depending only on c_0 and a subset $A'' \subset A$ of cardinality at least cm such that $|A'' - A''| \leqslant Cm$.*

Proof. The function $f(x) = A * A(x)$ (from $\mathbb{Z}^D$ to $\mathbb{Z}$) is non-negative and satisfies $\|f\|_\infty \leqslant m$, $\|f\|_2^2 \geqslant c_0 m^3$ and $\|f\|_1 = m^2$. This implies that $f(x) \geqslant c_0 m/2$ for at least $c_0 m/2$ values of x, since otherwise we would have

$$\|f\|_2^2 < (c_0/2)m.m^2 + (c_0 m/2).m^2 = c_0 m^3 .$$

Let us call a value of x for which $f(x) \geqslant c_0 m/2$ a *popular difference* and let us define a graph G with vertex set A by joining a to b if $b - a$ (and hence $a - b$) is a popular difference. The average degree in G is at least $c_0^2 m/4$, so there must be at least $c_0^2 m/8$ vertices of degree at least $c_0^2 m/8$. Let $\delta = c_0^2/8$, let $a_1, \dots, a_n$ be vertices of degree at least $c_0^2 m/8$, with $n \geqslant \delta m$, and let $A_1, \dots, A_n$ be the neighbourhoods of the vertices $a_1, \dots, a_n$. By Lemma 11 we can find a subset $A' \subset \{a_1, \dots, a_n\}$ of cardinality at least $\delta^5 n/\sqrt{2}$ such that at least 90% of the intersections $A_i \cap A_j$ with $a_i, a_j \in A'$ are of size at least $\delta^2 m/2$. Set $\alpha = \delta^6/\sqrt{2}$ so that $|A'| \geqslant \alpha m$.

Now define a graph H with vertex set A', joining a_i to a_j if and only if $|A_i \cap A_j| \geqslant \delta^2 m/2$. The average degree of the vertices in H is at least $(9/10)|A'|$, so at least $(4/5)|A'|$ vertices have degree at least $(4/5)|A'|$. Define A'' to be the set of all such vertices.

We claim now that A'' has a small difference set. To see this, consider any two elements $a_i, a_j \in A''$. Since the degrees of a_i and a_j are at least

$(4/5)|A'|$ in H, there are at least $(3/5)|A'|$ points $a_k \in A'$ joined to both a_i and a_j. For every such k we have $|A_i \cap A_k|$ and $|A_j \cap A_k|$ both of size at least $\delta^2 m/2$. If $b \in A_i \cap A_k$, then both $a_i - b$ and $a_k - b$ are popular differences. It follows that there are at least $c_0^2 m^2/4$ ways of writing $a_i - a_k$ as $(p-q)-(r-s)$, where $p,q,r,s \in A$, $p-q = a_i - b$ and $r - s = a_k - b$. Summing over all $b \in A_i \cap A_k$, we find that there are at least $\delta^2 c_0^2 m^3/8$ ways of writing $a_i - a_k$ as $(p-q)-(r-s)$ with $p,q,r,s \in A$. The same is true of $a_j - a_k$. Finally, summing over all k such that a_k is joined in H to both a_i and a_j, we find that there are at least $(3/5)|A'|\delta^4 c_0^4 m^6/64 \geqslant \alpha\delta^4 c_0^4 m^7/120$ ways of writing $a_i - a_j$ in the form $(p-q)-(r-s)-((t-u)-(v-w))$ with $p,q,\dots,w \in A$.

Since there are at most m^8 elements in A^8, the number of differences of elements of A'' is at most $120m/\alpha\delta^4 c_0^4 \leqslant 2^{38} m/c_0^{24}$. Note also that the cardinality of A'' is at least $(4/5)\alpha m \geqslant c_0^{12} m/2^{19}$. The proposition is proved. □

Combining Theorem 10 and Proposition 12 gives us the following consequence of Freiman's theorem.

Corollary 13. *Let A be a subset of $\mathbb{Z}^D$ of cardinality m such that $\|A * A\|_2^2 \geqslant c_0 m^3$. There is an arithmetic progression Q of cardinality at most Cm and dimension at most d such that $|A \cap Q| \geqslant cm$, where C, d and c are constants depending only on c_0.* □

It turns out that a small step from Ruzsa's proof of Freiman's theorem allows one to make the reverse deduction: in other words, Freiman's theorem and Corollary 13 can be seen to be equivalent.

Ruzsa's proof also allows us to make a small but convenient modification to Corollary 13, and it provides us with some bounds. A d-dimensional arithmetic progression $Q = P_1 + \dots + P_d$ is said to be *proper* if every $x \in Q$ has a unique representation of the form $x_1 + \dots + x_d$ with $x_i \in P_i$. Ruzsa showed that if A is any set such that $|A - A| \leqslant C|A|$, then there is a proper arithmetic progression Q of dimension $d \leqslant 2^{18}C^{32}$ and size at least $(2^{20}C^{32})^{-2^{18}C^{32}}|A|$, such that $|A \cap Q| \geqslant C^{-5}2^{-d}|Q|$ (which of course implies that $|Q| \leqslant C^5 2^d |A|$). Applying this result to the set A'' arising from Proposition 12, we find that we can ask for the progression Q in Corollary 13 to be proper.

Corollary 14. *Let $B \subset \mathbb{Z}_N$ be a set of cardinality βN, and let $\phi : B \to \mathbb{Z}_N$ be a function with at least $c_0 N^3$ additive quadruples. Then there are constants γ and η depending on β and c_0 only, a mod-N arithmetic* *progression $P \subset \mathbb{Z}_N$ of cardinality at least N^γ and a linear function $\psi : P \to \mathbb{Z}_N$ such that $\phi(s)$ is defined and equal to $\psi(s)$ for at least $\eta|P|$ values of $s \in P$.*

Proof. Let ? be the graph of ϕ, embedded in the obvious way into $\mathbb{Z}^2$. By Corollary 13 with the modification mentioned above, we may find a proper d-dimensional arithmetic progression Q of cardinality at most CN, with $|? \cap Q| \geqslant cN$, where d, C and c depend on β and c_0 only. Let $Q = P_1 + \cdots + P_d$. Then at least one P_i has cardinality at least $(CN)^{1/d} \geqslant (cN)^{1/d}$, so Q can be partitioned into (one-dimensional) arithmetic progressions of at least this cardinality. Hence, by averaging, there is an arithmetic progression $R \subset \mathbb{Z}^2$ of cardinality at least $(CN)^{1/d} \geqslant (cN)^{1/d}$ such that $|R \cap ?| \geqslant cC^{-1}|R|$. Because ? is the graph of a function, we know that R is not vertical (unless $|R \cap ?| = 1$ in which case the result we wish to prove is true anyway). Hence, there is an arithmetic progression $P \subset \mathbb{Z}$ with $|P| = |R|$ and a linear function $\psi : P \to \mathbb{Z}$ such that ? contains at least $cC^{-1}|P|$ pairs $(s, \psi(s))$. Reducing mod N now proves the result stated. □

It can be checked that Ruzsa's bounds imply that there is an absolute constant K such that, in the above corollary, we may take γ to be c_0^K and η to be $\exp(-(1/c_0)^K)$. As mentioned earlier, the use of Freiman's theorem and these bounds is somewhat uneconomical when it comes to proving the main result. That is because all we need is Corollary 14, which forgets most of the structure guaranteed by the theorem. It turns out that there is a weakening of Freiman's theorem with a better bound and a strong enough statement for Corollary 14 still to follow.

5 Obtaining Quadratic Bias

Let $A \subset \mathbb{Z}_N$ be a set which fails to be quadratically α-uniform and let f be the balanced function of A. Then there is a subset $B \subset \mathbb{Z}_N$ of cardinality at least αN, and a function $\phi : B \to \mathbb{Z}_N$ such that $|\Delta(f;k)^{\sim}(\phi(k))| \geqslant \alpha N$ for every $k \in B$. From section 3 we know that B contains at least $\alpha^{12}N^3$ additive quadruples for the function ϕ. The last section then implies that ϕ can be restricted to a large arithmetic progression P where it often agrees with a linear function $s \mapsto as + b$. We shall now use this fact to show that $\mathbb{Z}_N$ can be uniformly covered by large arithmetic progressions $P_1, \dots, P_N$ such that, for every s we can choose a quadratic function $\psi_s : P_s \to \mathbb{Z}_N$ such that $\sum_{z \in P_s} f(z)\omega^{-\psi_s(z)}$ is on average large in modulus (meaning an appreciable fraction of $|P_s|$). In the next section we shall use this result to find an arithmetic progression where the density of A increases.

Proposition 15. *Let $A \subset \mathbb{Z}_N$ have balanced function f. Let P be an arithmetic progression (in $\mathbb{Z}_N$) of cardinality T. Suppose that there exist λ and μ such that $\sum_{k\in P} |\Delta(f;k)^{\sim}(\lambda k+\mu)|^2 \geqslant \beta N^2 T$. Then there exist quadratic polynomials $\psi_0, \psi_1, \dots, \psi_{N-1}$ such that*

$$\sum_s \Big| \sum_{z\in P+s} f(z)\omega^{-\psi_s(z)} \Big| \geqslant \beta NT/\sqrt{2}\,.$$

Proof. Expanding the assumption we are given, we obtain the inequality

$$\sum_{k\in P}\sum_{s,t} f(s)f(s-k)f(t)f(t-k)\omega^{-(\lambda k+\mu)(s-t)} \geqslant \beta N^2 T\,.$$

Substituting $u = s-t$, we deduce that

$$\sum_{k\in P}\sum_{s,u} f(s)f(s-k)f(s-u)f(s-k-u)\omega^{-(\lambda k+\mu)u} \geqslant \beta N^2 T\,.$$

Let $P = \{x+d, x+2d, \dots, x+td\}$. Then we can rewrite the above inequality as

$$\sum_{i=1}^{T}\sum_{s,u} f(s)f(s-x-id)f(s-u)f(s-k-id-u)\omega^{-(\lambda x+\lambda id+\mu)u} \geqslant \beta N^2 T. \tag{$*$}$$

Since there are exactly T ways of writing $u = y+jd$ with $y \in \mathbb{Z}_N$ and $1 \leqslant j \leqslant T$, we can rewrite the left-hand side above as

$$\frac{1}{T}\sum_s\sum_{i=1}^{T}\sum_y\sum_{j=1}^{T} f(s)f(s-x-id)f(s-y-jd) \cdot f(s-x-id-y-jd)\omega^{-(\lambda x+id+mu)(y+jd)}\,.$$

Let us define $\gamma(s,y)$ by the equation

$$\Big|\sum_{i=1}^{T}\sum_{j=1}^{T} f(s-x-id)f(s-y-jd)f(s-x-id-y-jd)\omega^{-(\phi(x)+i\mu)(y+jd)}\Big| = \gamma(s,y)T^2\,.$$

Since $|f(s)| \leqslant 1$, $(*)$ tells us that the average value of $\gamma(s,y)$ is at least β.

In general, suppose we have real functions f_1, f_2 and f_3 such that

$$\Big|\sum_{i=1}^{T}\sum_{j=1}^{T} f_1(i)f_2(j)f_3(i+j)\omega^{-(ai+bj-2cij)}\Big| \geqslant cT^2\,.$$

Since $2cij = c((i+j)^2 - i^2 - j^2)$, we can rewrite this as

$$\left|\sum_{i=1}^{T}\sum_{j=1}^{T} f_1(i)\omega^{-(ai+ci^2)} f_2(i)\omega^{-(bj+cj^2)} f_3(i+j)\omega^{c(i+j)^2}\right| \geqslant cT^2$$

and then replace the left hand side by

$$\frac{1}{N}\left|\sum_{r}\sum_{i=1}^{T}\sum_{j=1}^{T}\sum_{k=1}^{2T} f_1(i)\omega^{-(ai+ci^2)} f_2(j)\omega^{-(bj+cj^2)} f_3(k)\omega^{ck^2}\omega^{-r(i+j-k)}\right| .$$

If we now set $g_1(r)=\sum_{i=1}^{T} f_1(i)\omega^{-(ai+ci^2)}\omega^{-ri}$, $g_2(r)=\sum_{j=1}^{T} f_2(j)\omega^{-(bj+cj^2)}\omega^{-rj}$ and $g_3(r) = \sum_{k=1}^{2T} f_3(k)\omega^{-ck^2}\omega^{-rk}$, then we have

$$\left|\sum_{r} g_1(r)g_2(r)g_3(r)\right| \geqslant cT^2N ,$$

which implies, by the Cauchy–Schwarz inequality, that $\|g_1\|_\infty \|g_2\|_2 \|g_3\|_2 \geqslant cT^2N$. Since $\|g_2\|_2^2 \leqslant NT$ and $\|g_3\|_2^2 \leqslant 2NT$ (by identity (1) of section 2), this tells us that $|g_1(r)| \geqslant cT/\sqrt{2}$ for some r. In particular, there exists a quadratic polynomial ψ such that $\left|\sum_{i=1}^{T} f_1(i)\omega^{-\psi(i)}\right| \geqslant cT/\sqrt{2}$.

Let us apply this general fact to the functions $f_1(i) = f(x-s-id)$, $f_2(j) = f(s-y-jd)$ and $f_3(k) = f(s-x-y-kd)$. It gives us a quadratic polynomial $\psi_{s,y}$ such that

$$\left|\sum_{i=1}^{T} f(s-x-id)\omega^{-\psi_{s,y}(i)}\right| \geqslant \gamma(s,y)T/\sqrt{2}\,.$$

Let $\gamma(s)$ be the average of $\gamma(s,y)$, and choose ψ_s to be one of the $\psi_{s,y}$ in such a way that

$$\left|\sum_{i=1}^{T} f(s-x-id)\omega^{-\psi_s(i)}\right| \geqslant \gamma(s)T/\sqrt{2}\,.$$

If we now sum over s, we have the required statement (after a small change to the definition of the ψ_s). □

Combining the above result with the results of the previous section, we obtain a statement of the following kind. If A fails to be quadratically uniform, then $\mathbb{Z}_N$ can be uniformly covered by large arithmetic progressions, on each of which the balanced function of A exhibits "quadratic bias". It is not immediately obvious that this should enable us to find a progression where the restriction of A has an increased density. That is a task for the next section.

6 An Application of Weyl's Inequality.

A famous result of Weyl asserts that, if α is an irrational number and k is an integer, then the sequence $\alpha, 2^k\alpha, 3^k\alpha, \dots$ is equidistributed mod 1. As an immediate consequence, if α is any real number and $\epsilon > 0$, then there exists n such that the distance from $n^2\alpha$ to the nearest integer is at most ϵ. This is the result we need to finish the proof. For the purposes of a bound, we need an estimate for n in terms of ϵ. It is not particularly easy to find an appropriate statement in the literature. In the longer paper to come, we shall give full details of the deduction of the statement we need, with estimates, from Weyl's inequality. Here we shall merely state the result in a convenient form, almost certainly not with the best known bound.

Theorem 16. *Let N be sufficiently large and let $a \in \mathbb{Z}_N$. For any $t \leqslant N$ there exists $p \leqslant t$ such that $|p^2a| \leqslant Ct^{-1/8}N$, where C is an absolute constant.*

Before we apply Theorem 16, we need a standard lemma (essentially due to Dirichlet).

Lemma 17. *Let $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ be linear (i.e., of the form $\phi(x) = ax+b$) and let $r, s \leqslant N$. For some $m \leqslant (2rN/s)^{1/2}$ the set $\{0, 1, 2, \dots, r-1\}$ can be partitioned into arithmetic progressions $P_1, \dots, P_m$ such that the diameter of $\phi(P_j)$ is at most s for every j. Moreover, the sizes of the P_j differ by at most 1.*

Proof. Let t be an integer greater than or equal to $(2rN/s)^{1/2}$ and note that this is at least $r^{1/2}$. Of the numbers $\phi(0), \phi(1), \dots, \phi(t)$, at least two must be within N/t and hence there exists $u \leqslant t$ such that $|\phi(u) - \phi(0)| \leqslant N/t$. Split $\{0, 1, \dots, r-1\}$ into u congruence classes mod u, each of size at most $\lceil r/u \rceil$. Each congruence class is an arithmetic progression. If P is a set of at most st/N consecutive elements of a congruence class, then P is an arithmetic progression with $\phi(P)$ of diameter at most s. Hence, each congruence class can be divided into at most $2rN/ust$ sub-progressions P with $\phi(P)$ of diameter at most s and with different Ps differing in size by at most 1. Since the congruence classes themselves differ in size by at most 1, it is not too hard to see that the whole of $\{0, 1, \dots, r\}$ can be thus partitioned. Hence, the total number of subprogressions is at most $2rN/st \leqslant (2rN/s)^{1/2}$. (Note that we cannot make t larger because we needed the estimate $r/u \geqslant st/N$ above.) □

Proposition 18. *There is an absolute constant C with the following* *property. Let $\psi : \mathbb{Z}_N \to \mathbb{Z}_N$ be any quadratic polynomial and let $r \in \mathbb{N}$. For some $m \leqslant Cr^{1-1/128}$ the set $\{0, 1, 2, \dots, r-1\}$ can be partitioned into arithmetic progressions $P_1, \dots, P_m$ such that the diameter of $\psi(P_j)$ is at most $Cr^{-1/128}N$ for every j. The lengths of any two P_j differ by at most 1.*

Proof. Let us write $\psi(x) = ax^2 + bx + c$. By Theorem 18 we can find $p \leqslant r^{1/2}$ such that $|ap^2| \leqslant C_1 r^{-1/8}N$ for some absolute constant C_1. Then for any s we have

$$\begin{aligned}\psi(x+sp) &= a(x+sp)^2 + b(x+sp) + c \\ &= s^2(ap^2) + \theta(x,p)\end{aligned}$$

where θ is a bilinear function of x and p. (Throughout this paper, we use the word "linear" where "affine" is, strictly speaking, more accurate.)

For any u, the diameter of the set $\{s^2(ap^2) : 0 \leqslant s < u\}$ is at most $u^2|ap^2| \leqslant C_1u^2r^{-1/8}N$. Therefore, for any $u \leqslant r^{1/4}$, we can partition the set $\{0, 1, \dots, r-1\}$ into arithmetic progressions of the form

$$Q_j = \{x_j, x_j + p, \dots, x_j + (u_j - 1)p\},$$

such that, for every j, $u - 1 \leqslant u_j \leqslant u$ and there exists a linear function ϕ_j such that, for any subset $P \subset Q_j$,

$$\mathrm{diam}(\psi(P)) \leqslant C_1u^2r^{-1/8}N + \mathrm{diam}(\phi_j(P)).$$

Let us choose $u = r^{1/64}$, with the result that $u^2r^{-1/16} = r^{-1/32}$. By Lemma 17, if $v \leqslant u^{1/2}/2$, then every Q_j can be partitioned into arithmetic progressions P_{jt} of length $v-1$ or v in such a way that $\mathrm{diam}(\phi_j(P_{jt})) \leqslant 2u^{-1/2}N$ for every t. This, with our choice of u above, gives us the result. □

COROLLARY 19. *Let $\psi : \mathbb{Z}_N \to \mathbb{Z}_N$ be a quadratic polynomial and let $r \leqslant N$. There exists $m \leqslant Cr^{1-1/128}$ (where C is an absolute constant) and a partition of the set $\{0, 1, \dots, r-1\}$ into arithmetic progressions $P_1, \dots, P_m$ such that the sizes of the P_j differ by at most one, and if $f : \mathbb{Z}_N \to D$ is any function such that*

$$\left|\sum_{x=0}^{r-1} f(x)\omega^{-\psi(x)}\right| \geqslant \alpha r,$$

then

$$\sum_{j=1}^{m}\left|\sum_{x\in P_j} f(x)\right| \geqslant \alpha r/2.$$

Proof. By Proposition 18 we can choose $P_1, \dots, P_m$ such that $\mathrm{diam}(\phi(P_j)) \leqslant CNr^{-1/128}$ for every j. For sufficiently large r this is at most $\alpha N/4\pi$. By the triangle inequality,

$$\sum_{j=1}^{m}\Big|\sum_{x\in P_j} f(x)\omega^{-\psi(x)}\Big| \geqslant \alpha r\ .$$

Let $x_j \in P_j$. The estimate on the diameter of $\psi(P_j)$ implies that $|\omega^{-\psi(x)} - \omega^{-\psi(x_j)}|$ is at most $\alpha/2$ for every $x \in P_j$. Therefore

$$\begin{aligned}\sum_{j=1}^{m}\Big|\sum_{x\in P_j} f(x)\Big| &= \sum_{j=1}^{m}\Big|\sum_{x\in P_j} f(x)\omega^{-\psi(x_j)}\Big| \\ &\geqslant \sum_{j=1}^{m}\Big|\sum_{x\in P_j} f(x)\omega^{-\psi(x)}\Big| - \sum_{j=1}^{m}(\alpha/2)|P_j| \\ &\geqslant \alpha r/2\,.\end{aligned}$$

The statement about the sizes of the P_j follows easily from our construction. □

7 Putting Everything Together

Theorem 20. *There is an absolute constant C with the following property. Let A be a subset of $\mathbb{Z}_N$ with cardinality δN. If $N \geqslant \exp\exp\exp((1/\delta)^C)$, then A contains an arithmetic progression of length four.*

Proof. Suppose that the result is false. Then Corollary 8 implies that A is not quadratically $2^{-208}\delta^{112}$-uniform. Let $\alpha = 2^{-208}\delta^{112}$ and let f be the balanced function of A. The implication of (iii) from (ii) in Lemma 2 then implies that there is a set $B \subset \mathbb{Z}_N$ of cardinality at least $\alpha N/2$ together with a function $\phi : B \to \mathbb{Z}_N$, such that $|\Delta(f;k)^{\sim}(\phi(k))| \geqslant (\alpha/2)^{1/2}N$ for every $k \in B$. In particular,

$$\sum_{k\in B}\big|\Delta(f;k)^{\sim}(\phi(k))\big|^2 \geqslant (\alpha/2)^2N^3\,.$$

Hence, by Proposition 9, ϕ has at least $(\alpha/2)^8N^3$ additive quadruples. Corollary 14 and the discussion of bounds immediately after it imply that there is an arithmetic progression P satisfying the hypotheses of Proposition 15, with $T \geqslant N^{\gamma}$, where $\gamma = \delta^K$ and $\beta \geqslant \exp(-(1/\delta)^K)$. (We have changed the absolute constant K, allowing us to write δ instead of $(\alpha^2/2)^8$.) We therefore have quadratic polynomials $\psi_0, \psi_1, \ldots, \psi_{N-1}$ such that

$$\sum_{s}\Big|\sum_{z\in P+s} f(z)\omega^{-\psi_s(z)}\Big| \geqslant \beta NT/\sqrt{2}$$

with these values of β and T. Corollary 19 implies that we can partition each $P+s$ into further progressions $P_{s1},\dots,P_{sm}$ (mod $\mathbb{Z}_N$) of cardinalities differing by at most one and all at least $cT^{1/128}$, where c is another absolute constant, such that

$$\sum_s \sum_{j=1}^m \Big| \sum_{x\in P_{sj}} f(x) \Big| \geqslant \beta NT/2\sqrt{2}\,.$$

It is an easy consequence of Lemma 17 that we can also insist that the P_{sm} are genuine arithmetic progressions (in $\{0,1,\dots,N-1\}$ and not just in $\mathbb{Z}_N$), except that now the condition on the sizes is that the average length of a P_{sj} is $cT^{1/256}$ (for a slightly different c) and no P_{sj} has more than twice this length. With such a choice of P_{sj}, let p_{sj} equal $\sum_{x\in P_{sj}} f(x)$, and let q_{sj} be p_{sj} if this is positive, and zero otherwise. Then $\sum_s \sum_{j=1}^m p_{sj} = T\sum_x f(x) = 0$, which implies that $\sum_s \sum_{j=1}^m q_{sj} \geqslant \beta NT/4\sqrt{2}$. Hence, there exists a choice of s and j such that $\sum_{x\in P_{sj}} f(x) \geqslant \beta T/4m\sqrt{2} = c_1\beta T^{1/256}$, where c_1 is another absolute constant. Then $|P_{sj}|$ is at least $c_1\beta T^{1/256}$ and $|A\cap P_{sj}|$ is at least $(\delta + c_2\beta)|P_{sj}|$.

We now repeat the argument, replacing A and $\{0,1,2,\dots,N\}$ by $A\cap P_{sj}$ and P_{sj}. The function $\delta \mapsto c_2\beta = c_2\exp(-(1/\delta)^K)$ is increasing, so that after each run of the argument, the density of the restriction of A goes up by a factor of at least $1+c_2\beta$. Hence, it can be repeated at most $\exp((1/\delta)^K)$ times. The function $\delta \mapsto \gamma$ is also increasing, so at each stage of the argument we replace the current N with a new one which is at least N^{δ^K} (where K is changed a little to allow for the 256th root taken above). Setting $r = \exp((1/\delta)^K)$ and $\theta = \delta^K$, this tells us that the theorem is proved, provided that N^{θ^r} is sufficiently large. The restriction comes in Corollary 8, which tells us that we must have $N^{\theta^r} \geqslant 200\delta^{-3}$. A small calculation now gives the result stated. □

An alternative formulation of the condition on N and δ is that δ should be at least $(\log\log\log N)^{-c}$ for some absolute constant $c>0$. We have the following immediate corollary.

COROLLARY 21. *There is an absolute constant $c>0$ with the following property. If the set $\{1,2,\dots,N\}$ is coloured with at most $(\log\log\log N)^c$ colours, then there is a monochromatic arithmetic progression of length four.* □

8 Concluding Remarks

Most of the above proof generalizes reasonably easily, with the result that it is not hard to guess the basic outline of a proof of Szemerédi's complete theorem. To be more precise, the results of sections 2 and 6 have straightforward generalizations, and the result of section 5 can also be generalized appropriately, although not in quite as obvious a manner. The main difficulty with the general case is in proving a suitable generalization of Corollary 14. What is needed, which is the main result of our forthcoming paper, is a statement of the following kind. Call a function ψ from $C \subset \mathbb{Z}_N$ to $\mathbb{Z}_N$ strongly additive if every restriction of ψ to a large subset of C has many additive quadruples. If $B \subset \mathbb{Z}_N^k$ is a set of size proportional to N^k and if $\phi : B \to \mathbb{Z}_N$ is a function such that, whenever $k-1$ of the variables are fixed, the resulting function is strongly additive in the remaining variable, then there is a large arithmetic progression $P \subset \mathbb{Z}_N$ and a set of the form $Q = (P+r_1) \times \cdots \times (P+r_k)$ such that ϕ agrees with a multilinear function γ for many points in Q. Even the case $k = 2$ is not at all easy.

The bounds obtained for Theorem 20 and Corollary 21 improve enormously on any that were previously known. However, as was mentioned earlier, it is possible to avoid using Freiman's theorem directly and obtain a further improvement. Doing so removes one exponential from the lower bound for N in terms of δ, or equivalently one logarithm from the lower bound for δ in terms of N. That is, a small modification of our approach shows that it is enough for δ to be at least $(\log\log N)^{-c}$. It might be possible to improve the bound further still to $\delta \geqslant (\log N)^{-c}$ by using ideas from the papers of Szemerédi [Sz3] and Heath-Brown [H].

References

[BSz] A. Balog, E. Szemerédi, A statistical theorem of set addition, Combinatorica 14 (1994), 263–268.

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

[ET] P. Erdős, P. Turán, On some sequences of integers, J. London Math. Soc. 11 (1936), 261–264.

[F1] G.A. Freiman, Foundations of a Structural Theory of Set Addition (in Russian), Kazan Gos. Ped. Inst., Kazan, 1966.

[F2] G.A. Freiman, Foundations of a Structural Theory of Set Addition, Translations of Mathematical Monographs 37, Amer. Math. Soc., Providence, R.I., USA, 1973.

[Fu] H. Furstenberg, Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions, J. Analyse Math. 31 (1977), 204–256.

[G] W.T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, Geometric And Functional Analysis 7 (1997), 322–337.

[H] D.R. Heath-Brown, Integer sets containing no arithmetic progressions, J. London Math. Soc. (2) 35 (1987), 385–394.

[R1] K.F. Roth, On certain sets of integers, J. London Math. Soc. 28 (1953), 245–252.

[R2] K.F. Roth, Irregularities of sequences relative to arithmetic progressions, IV, Period. Math. Hungar. 2 (1972), 301–326.

[Ru] I. Ruzsa, Generalized arithmetic progressions and sumsets, Acta Math. Hungar. 65 (1994), 379–388.

[S] S. Shelah, Primitive recursive bounds for van der Waerden numbers, J. Amer. Math. Soc. 1 (1988), 683–697.

[Sz1] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, Acta Math. Acad. Sci. Hungar. 20 (1969), 89–104.

[Sz2] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 299–345.

[Sz3] E. Szemerédi, Integer sets containing no arithmetic progressions, Acta Math. Hungar. 56 (1990), 155–158.

[W] H. Weyl, Über die Gleichverteilung von Zahlen mod Eins, Math. Annalen 77 (1913), 313–352.

W.T. Gowers
Department of Pure Mathematics
and Mathematical Statistics
16 Mill Lane
Cambridge CB2 1SB
England
E-mail: W.T.Gowers@pmms.cam.ac.uk

Submitted: November 1997
Final Version: May 1998