Small linear dependencies for binary vectors of low weight

Uriel Feige
Department of Computer Science and Applied Mathematics
Weizmann Institute
Rehovot, Israel
uriel.feige@weizmann.ac.il

June 3, 2008

Abstract

We show that every set of $m \simeq cn\sqrt{n \log \log n}$ vectors in $\{0,1\}^n$ in which every vector has Hamming weight 3 contains a subset of $O(\log n)$ vectors that form a linear dependency. Our proof is based on showing that in every graph of average degree at least $c \log \log n$, every legal edge coloring produces a cycle in which one of the colors appears either once or twice. (In both results, c is some constant.) The results proved are used (in a companion work) in refutation algorithms for semirandom 3CNF formulas.

1 Introduction

The problem studied in this paper can be viewed either as a problem involving linear dependencies among binary vectors, or as a problem on hypergraphs. We present here the hypergraph formulation.

Definition 1.1 *An* even cover *in a hypergraph is a nonempty set of hyperedges that contains each vertex an even number of times (either not at all, or twice, or four times, etc.). The* size *of an even cover is the number of hyperedges in the even cover.*

It is not hard to see that every hypergraph on n vertices with more than n hyperedges has an even cover of size at most $n+1$. This follows by viewing each hyperedge as an indicator vector for its variables, noting that this gives a vector space of dimension at most n, and that every minimal set of linearly dependent binary vectors (addition performed modulo 2) corresponds to an even cover. As the number of hyperedges in a hypergraph increases, smaller even covers may appear. For r-uniform hypergraphs with $r \geq 3$, it is reasonable to conjecture the following relation between number of hyperedges and size of even covers. (The $\tilde{O}$ notation is meant to suppress an $O(\log n)$ multiplicative term, though the author would be happy to settle also for somewhat larger low order multiplicative terms.)

Conjecture 1.2 *Let c be sufficiently large. Then every r-uniform hypergraph on n vertices and $m = c\beta n$ hyperedges (with $1 < \beta \leq O(n^{(r-2)/2})$) has an even cover of size at most $\tilde{O}(n/\beta^{2/(r-2)})$.*

For graphs ($r = 2$), minimal even covers are simply cycles. The natural analog of the conjecture for graphs would be that every graph of sufficiently high constant average degree has a cycle of length $O(\log n)$, which is well known to be true. For general r, the conjecture is not known to hold, except of course at the very low density case when $\beta < (\log n)^{(r-2)/2}$, in which case the conjecture is trivially true. When r is even, the conjecture is known to be true also at the very high density case, say, when $\beta = 2n^{(r-2)/2}$ (see Proposition 2.2). The current work addresses the very high density case of Conjecture 1.2 when r is odd, and comes closer to proving the conjecture in this case.

The case that motivates the current work is that of $r = 3$. In this case, β can range from 1 to $\sqrt{n}$. As we shall explain in Section 1.1, this case comes up in refutation of random 3CNF formulas. Some of the results in this work are stated only for this special case, but easily extend to all odd r.

The following theorem is implicit in the work of Naor and Vastreate [11]. If not for the term $\log n$ in the value of β, it would prove the conjecture for the very high density case.

Theorem 1.3 *Every 3-uniform hypergraph with n vertices and βn hyperedges contains an even cover of size at most $\log n$. Here $\beta = c \log n \sqrt{n}$ for some sufficiently large universal constant c.*

In our work, we improve over the value of β and show:

Theorem 1.4 *Let H be an arbitrary 3-uniform hypergraph with n vertices and $m = \beta n$ hyperedges, and let c be a sufficiently large universal constant. Then:*

1. *If $\beta \geq c\sqrt{n \log n / \log\log n}$ then H contains an even cover of size $O(\log n / \log\log n)$.*
2. *If $\beta \geq c\sqrt{n \log\log n}$ then H contains an even cover of size $O(\log n)$.*

Moreover, in both cases there is a polynomial time algorithm that finds the respective even cover.

Our proof of Theorem 1.4 produces even covers which are of a special form (correspond to linear dependencies over any field). See Corollary 4.5. Our proof technique reduces the problem of even covers in hypergraphs to an extremal problem in graphs.

Definition 1.5 *Given a graph G and an arbitrary legal coloring of its edges (incident edges have different colors), a simple cycle (namely, a cycle that does not visit any vertex more than once) is called a 1-cycle (2-cycle, respectively) if some color is used in order to color exactly one (two, respectively) of its edges. We say that a cycle in an edge colored graph is* distinguished *if it is either a 1-cycle or a 2-cycle.*

The extremal problem is as follows: what is the maximum number of edges that a legally colored n-vertex graph can have and still not contain a distinguished cycle of length at most t? Observe that a graph may have arbitrarily many edges and still not have a 2-cycle (if every edge is colored by a different color). It may have $\Omega(n \log n)$ edges and still not have a 1-cycle (e.g, color edges of the hypercube by the name of the coordinate that is flipped). However, once both 1-cycles and 2-cycles are forbidden, we show that the number of edges is at most $O(n \log\log n)$.

Theorem 1.6 *For a sufficiently large constant c,*

1. *For every graph on n vertices and average degree at least $c\frac{\log n}{\log\log n}$, every legal edge coloring creates a distinguished cycle of length $O(\frac{\log n}{\log\log n})$.*
2. *For every graph on n vertices and average degree at least $c\log\log n$, every legal edge coloring creates a distinguished cycle of length $O(\log n \log\log n)$.*

Moreover, in both cases there is a polynomial time algorithm that finds the respective distinguished cycle.

As we shall see in the proof of Corollary 2.12, the correspondence between Theorem 1.6 and Theorem 1.4 is as follows. A degree of d in Theorem 1.6 gives a density of $\beta = O(\sqrt{dn})$ in Theorem 1.4 for even covers that are twice as large as the corresponding distinguished cycles. Hence item 1 of Theorem 1.4 follows from item 1 of Theorem 1.6. Likewise, item 2 of Theorem 1.4 almost follows from item 2 of Theorem 1.6, except for a $\log\log n$ term in the size of the even cover. To remove this $\log\log n$ term, we appeal to some elementary properties of cycle bases in graphs.

1.1 Motivation and related work

The author's motivation for studying small even covers comes from a sequence of works on refuting random 3CNF formulas. The goal of these works is to design algorithms that when given a nonsatisfiable 3CNF Boolean formula (conjunction of clauses, where each clause is a disjunction of three literals such as $x_1 \vee \bar{x_2} \vee x_3$) certifies that no satisfying assignment exists. In general, this problem is coNP-hard, but it turns out that for sufficiently dense random (or semirandom) formulas efficient refutation algorithms exist (with high probability over the choice of the input formula). The methodology developed in [8, 4, 7, 6, 5] to design such algorithms reduces the problem of refuting 3CNF formulas to a stronger form of refutation but for an easier problem, max-3LIN2. Namely, given an inconsistent system of linear equations with three Boolean variables per equation (such as $x_1 + x_2 + x_3 = 1$ modulo 2), certify that the system is "far" from being satisfiable (in the sense that every assignment leaves "many" equations not satisfied). We call this *strong* refutation (though in a sense not as strong as that of [3]). Refuting satisfiability of 3LIN2 is easy (by Gaussian elimination), but strong refutation is in general NP-hard (by a rephrasing of the known hardness of approximation results [9] for max-3LIN2). However, the max-3LIN2 systems that are obtained from the reduction from random 3CNF formulas are random rather than worst case instances, and hence there is hope for strongly refuting them. Here is the approach developed in the above works.

Given a 3LIN2 system ϕ, let H_ϕ be the following 3-uniform hypergraph. The vertices of H_ϕ are the variables of ϕ. The hyperedges are the equations of ϕ. For example, the clause $x_1 + x_2 + x_3 = 0$ gives rise to the hyperedge (x_1, x_2, x_3). The hypergraph does not represent the right hand side of the equations. Assume that H_ϕ has an even cover of size 2ℓ (observe that an even cover always has an even number of hyperedges, because every hyperedge contains three vertices, and every vertex appears an even number of times). Consider the 2ℓ linear equations that correspond to the hyperedges of the even cover. Summing up all equations, the left hand side gives 0 (since every variable appears an even number of times and addition is performed modulo 2). As to the right hand side, if there is some randomness in the equation in the sense that for every equation independently there is some small probability δ that its right hand side is random, then with probability $\Omega(\delta\ell)$ (or at least 1/4 if $\delta\ell > 1$) the right hand side will sum up to 1, leading to a contradiction. Moreover, if H_ϕ has many disjoint even covers that can be found efficiently, this gives many disjoint subsystems, and if the original 3LIN system is somewhat random in the above sense, then many of them are likely not to be satisfiable. This is exactly what we want to achieve by strong refutation. Observe that a theorem such as Theorem 1.4 implies the existence of many disjoint even covers (and not just one), because after a small even cover is found, it can be removed from the hypergraph without substantially changing the number of hyperedges, and then the theorem can be applied again.

Theorem 1.4 plays a central role in the refutation algorithm presented in [5]. Its proof was only sketched in [5] and is presented in full in the current paper. (Remark: the bounds proved in the current paper are stronger than the corresponding bounds claimed in [5].)

Results in [6] support Conjecture 1.2. There, the special case of $r = 3$ and $\beta = n^{0.4}$ was considered. If was shown that if the hypergraph is random, then indeed it is likely to have even covers of size $O(n/\beta^2) = O(n^{0.2})$. The proof works for other values of β as well (except possibly for very small values of β – this needs to be checked). We remark that random hypergraphs serve as examples showing that Conjecture 1.2 cannot be improved upon. This follows from a simple expectation computation. See for example [6].

The "super-high" density case of $\beta = n^{\delta+(r-2)/2}$ for some $\delta > 0$ was studied in [11]. The motivation there comes from studying the Hamming distance of codes that have low density parity check matrices (see more details in [11]). When $\delta > 0$ there are even covers of constant size, and the goal is to figure out how this constant depends on δ. In our work we slightly improve over the bounds proven in [11] when r is odd and large (see the end of Section 2). More importantly, our proof technique, though sharing some features with that of [11], works for a wider range of parameters than the proof technique of [11]. Hence we start getting meaningful results when $\beta \geq n^{(r-2)/2}\sqrt{\log\log n}$, whereas the techniques of [11] require $\beta \geq n^{(r-2)/2}\log n$.

Some nontrivial upper bounds on the size of even covers can be obtained by proving the existence of subhypergraphs that contain more hyperedges than vertices, and on such a hypergraph invoking the argument that follows Definition 1.1. For example, in [1] it is shown that every 3-uniform hypergraph with βn hyperedges has a set of $\ell = O(n \log n/\beta)$ vertices that induce at least $\ell + 1$ hyperedges. It follows that even covers of size $O(n \log n/\beta)$ always exist. This line of work suggest the following conjecture as an intermediate step towards proving Conjecture 1.2.

Conjecture 1.7 *Let c be sufficiently large. Then every 3-uniform hypergraph on n vertices and $m = c\beta n$ hyperedges (with $1 < \beta \leq O(\sqrt{n})$) has a set of $n' \leq \tilde{O}(n/\beta^2)$ vertices that induce at least $2n'/3$ hyperedges.*

2 Distinguished cycles shorter than $\log n$

This section contains the proof of item (1) of Theorem 1.4. It is based on a simplification of the proof technique of [11], and works when $\beta > \sqrt{n \log n / \log \log n}$. In passing, we also improve some other results of [11]. For this reason, the presentation will be for r-uniform hypergraphs for general r, even though 3-uniform hypergraphs suffice for Theorem 1.4.

It will be convenient for us to view hyperedges of H as r-tuples of vertices, rather than as sets of vertices. Hence we shall use the convention that vertices of H are sorted in some arbitrary order, and likewise, the r vertices in a hyperedge are sorted according to the same order. The r-tuple corresponding to a hyperedge is this sorted list of vertices.

The following lemma is a result taken from [2]. (The result in [2] is slightly stronger. Weaker results that would also suffice for the purpose of our paper were known previous to [2].)

Lemma 2.1 *In any n-vertex graph of average degree $d > 2$ there is a cycle of length no longer than 2ℓ, if $d(d-1)^{\ell-1} > n$.*

Proof: If the graph is d-regular, the proof follows easily by performing breadth first search, starting from an arbitrary vertex. The (known) proof for nonregular graphs is not as simple. See [2] for details. □

The difficulty in proving Theorem 1.4 stems from the fact that it deals with r-uniform hypergraphs with odd $r = 3$. It is instructive to first see how a corresponding (in fact, stronger) theorem can be proved when r is even. The following proposition is taken from [11].

Proposition 2.2 *For even r and $d > 1$, every r-uniform hypergraph H with $m = dn^{r/2}$ hyperedges contains an even cover of size $O(\log n)$.*

Proof: Construct the following auxiliary graph G. It has $\binom{n}{r/2} \leq n^{r/2}$ vertices, labelled by all possible sets of $r/2$ vertices of H. Every hyperedge e of H contributes one edge e' to G, connecting the vertex in G that is labelled by the set of $r/2$ vertices that make the prefix of the r-tuple that labels e and the vertex labelled by the set of $r/2$ vertices that make the suffix of the r-tuple that labels e. The average degree of G is (slightly larger than) $2d > 2$, and hence Lemma 2.1 implies that G has a cycle of length $O(\log n)$. The hyperedges of H that correspond to the edges of G along this cycle form an even cover in H. □

We now return to the case that r is odd (in our case, $r = 3$), and consider an r uniform hypergraph with $m = dn^{r/2}$ edges.

For given n, r and d, let h be maximal such that h and s are positive integers satisfying $n^h < dn^{r/2}$ and $h + 2s = r$. For example, when $r = 3$ we have that $h = s = 1$, and for $r = 5$ we have that $h = 1$ when $d < \sqrt{n}$.

The following notion (used also in [11]) helps simplify later proofs.

Definition 2.3 *For h and s as defined above, an r-uniform hypergraph satisfies the* small overlap *condition if no two hyperedges share $h + s$ vertices.*

The following lemma (similar to [11]) shows that up to a negligible effect on d, we may assume that the small overlap condition holds.

Lemma 2.4 *Let H be an r-uniform hypergraph with $dn^{r/2}$ hyperedges, let h and s be as above, and let $\epsilon > 0$ satisfy $\epsilon dn^{r/2-s} > 1$. Then either*

- *H has a subhypergraph with $(d - 2\epsilon)n^{r/2}$ hyperedges that satisfies the small overlap condition,*

or

- *H has an even cover of size 4ℓ, where ℓ is the smallest integer satisfying $(2\epsilon dn^{r/2-s})^\ell > n^s$. For example, when $r = 3$ and $\epsilon > 1/d$ this corresponds to an even cover of size 8.*

Proof: Given an r-uniform hypergraph H, consider an auxiliary graph G whose vertices are the hyperedges of H, and two vertices of G are connected by an edge if the respective hyperedges share at least $h+s$ vertices (in H). Consider an arbitrary maximal matching M in G. If the matching contains less than $\epsilon dn^{r/2}$ edges, then remove the corresponding matched hyperedges from H. The number of hyperedges in H remains essentially unchanged, and H now satisfies the *small overlap* condition.

If the matching M contains $\epsilon dn^{r/2}$ edges, then consider an auxiliary multigraph F (it may have parallel edges). The vertices of F are s-tuples of vertices of H. Every edge of the matching M contributes one edge to F as follows. Let the matching edge correspond to two hyperedges e_1 and e_2 in H, and without loss of generality, assume that e_1 and e_2 share their first $h + s$ vertices (in H). Then in F add an edge between the vertex that is labelled by the last s vertices of e_1 and the vertex that is labelled by the last s vertices of e_2. Observe that now any cycle in F corresponds in a natural way to an even cover in H (with twice as many hyperedges in H than edges if F). The average degree in F is at least $2\epsilon dn^{r/2-s}$, which is greater than 2 by the conditions in the statement of the lemma. Hence Lemma 2.1 implies that H has an even cover with 4ℓ edges, where ℓ is the smallest value satisfying $(2\epsilon dn^{r/2-s})^\ell > n^s$. □

We shall assume that the hypergraph H satisfies the small overlap condition. This assumption can be made essentially without loss of generality, because results proved under this assumption easily generalize to arbitrary hypergraphs with almost the same parameters, by Lemma 2.4.

Now is our main point of departure from [11]. We construct an auxiliary graph G that is different from the one constructed in [11], and this leads both to a considerable simplification in the proofs, and to a strengthening of the results. The graph that we construct is similar to the one constructed in [8] in their refutation algorithm for random 3SAT.

Each vertex of G corresponds to a set of $2s$ vertices of H. By our convention that vertices of H are sorted, a vertex of G will be labelled by a $2s$-tuple of vertices of H, for which the prefix of size s is sorted and the suffix of size s is sorted. The same vertex of H may appear both in the prefix and in the suffix. Hence G has $\binom{n}{s}^2$ vertices. The edges of G are derived from hyperedges of H as follows. Every hyperedge of H is an r-tuple. Every two hyperedges e_1 and e_2 of H whose r-tuples agree on the last h vertices contribute one edge to G. This edge connects the vertices v_1 and v_2 in G, if the labels of v_1 and v_2 satisfy the following conditions. The tuple labelling v_1 agrees on its first s coordinates with the first s coordinates of the tuple labelling e_1, and agrees in its last s coordinates with coordinates $s+1$ up to $2s$ of the tuple labelling e_2. The tuple labelling v_2 agrees on its first

s coordinates with the first s coordinates of the tuple labelling e_2, and agrees in its last s coordinates with the coordinates $s+1$ up to $2s$ of the tuple labelling e_1. Moreover, we color this edge by the color c, where c is a tuple containing the last h vertices (the overlap vertices) in the tuples e_1 and e_2. Hence every edge of G (together with its color and the labels of its endpoints) uniquely determines which two hyperedges in H generated it.

Here are a few examples to illustrate the construction. When $r = 3$ we have that $h = s = 1$. Hence G contains n^2 vertices. Two hyperedges $e_1 = (1, 2, 5)$ and $e_2 = (3, 4, 5)$ in H would contribute the edge $((1, 4)(3, 2))$ to G, and this edge would be colored (5). When $r = 5$ and d is small, then $h = 1$ and $s = 2$. Hence G contains $\binom{n}{2}^2 < n^4$ vertices. Two hyperedges $e_1 = (1, 3, 4, 6, 8)$ and $e_2 = (2, 4, 5, 7, 8)$ in H would contribute the edge $((1, 3, 5, 7)(2, 4, 4, 6))$ to G, and this edge would be colored (8).

The following proposition is the key reason for introducing the *small overlap* property.

Proposition 2.5 *If H satisfies the small overlap condition, then the coloring of the edges of G is a legal coloring (no two edges of the same color are incident with the same vertex).*

Proof: Otherwise there would be two hyperedges in H whose overlap is at least $h + s$. □

The key to finding even covers in H is by using cycles in G. Observe that for every cycle in G, every vertex of H appears an even number of times on this cycle (counting all its appearances in hyperedges of H that generated the cycle in G). Hence a cycle in G corresponds to an even cover in H. However, there is one potential problem in this correspondence. A hyperedge in H may generate several edges in G. Hence it might be the case that in a given cycle of G, some hyperedges of H appear more than once. Two appearances of the same hyperedge in an even cover can (and should) be removed – this still results in an even cover. Continuing in this fashion, if it happens that every hyperedge of H appears on the cycle an even number of times (say, either twice or not at all), all hyperedges are removed and one remains with the empty even cover. In this case we say that the cycle is *trivial*: it corresponds to the trivial even cover that contains no hyperedges. Hence a cycle in G corresponds to an even cover in H if and only if the cycle is not trivial. In this work, we shall consider one particular class of nontrivial cycles in G, namely, the class of distinguished cycles, as defined in Definition 1.5.

Lemma 2.6 *A distinguished cycle in G of length ℓ corresponds to an even cover in H with at most 2ℓ hyperedges.*

Proof: Every edge of G is generated by two hyperedges of H. Consider one appearance of the color c that appears either once or twice in the distinguished cycle, and the two hyperedges that generated this appearance. If c appears only once, then these two hyperedges each appear only once on the cycle (because any edge that any of them generates will be colored c). If c appears twice, then in can not be that both appearances were generated by the same pair of hyperedges, because then both appearances would correspond to the same edge, contradicting the requirement that the cycle is simple. Hence in a distinguished cycle there is at least one hyperedge that appears exactly once on the cycle, and hence the cycle cannot be trivial. A nontrivial cycle corresponds to an even cover, and the hyperedges of the even cover are those hyperedges that generated the edges of the cycle. Hence the corresponding even cover has at most 2ℓ hyperedges (and possible less, if some hyperedges generated more than one edge along the cycle). □

To show that G has short distinguished cycles, we first bound from below its average degree.

Lemma 2.7 *The graph G has average degree at least (roughly) d^2.*

Proof: Recall that G has $\binom{n}{s}^2$ vertices, that H has $dn^{r/2}$ hyperedges, and that $2s+h=r$. Group the hyperedges into $\binom{n}{h}$ disjoint groups, one for every possible h-suffix of a label of a hyperedge. A simple shifting argument implies that the number of edges in G is minimized when all groups are of the same size. Hence we assume that every group contains $dn^{r/2}/\binom{n}{h}$ hyperedges (ignoring rounding issues). Each group then generates $\binom{dn^{r/2}/\binom{n}{h}}{2}$ edges, and the total number of edges in G is roughly $d^2n^r/2\binom{n}{h}$. The average degree is at least as claimed because $\binom{n}{s}^2\binom{n}{h} \leq n^r$. □

The discussion so far leads to the problem of providing bounds for $L_{avg}(N, D)$ as defined below.

Definition 2.8 *Let $L_{avg}(N, D)$ denote the minimum value of ℓ such that for every graph G with N vertices (in our case $N \simeq n^{r-h}$) and average degree at least D (in our case $D \geq d^2$), and for every legal coloring of its edges, G must contain a distinguished cycle of length at most 2ℓ.*

In our application G need not be a simple graph. It may have parallel edges. But if it does, then it contains a distinguished cycle of length two.

When $D > c\frac{\log N}{\log\log N}$ (here c is some sufficiently large constant) then the existence of short distinguished cycles can be analyzed using the same approach as that used for the existence of short cycles in general. As we do not care for constant factors in the degree D, the analysis can be simplified by the following proposition (also used in [11] for the same purpose).

Proposition 2.9 *Every graph of average degree D has a subgraph of minimum degree $D/2$.*

Proof: Iteratively remove vertices of degree less than $D/2$ together with their incident edges. The total number of edges that can be removed in this process is strictly less than $nD/2$, and hence some subgraph remains. □

We define $L_{min}(N, D)$ in a way similar to Definition 2.8, but with minimum degree replacing average degree. Proposition 2.9 implies that $L_{avg}(N, D) \leq \max_{N' \leq N}[L_{min}(N', D/2)]$. (Our upper bounds on $L_{min}(N, D)$ will be nondecreasing in N, and hence the value of N' to be used in the above inequality will be N.) The following lemma proves item 1 of Theorem 1.6.

Lemma 2.10 *For $L_{min}(N, D)$ as defined above, $L_{min}(N, D) \leq \ell$ where ℓ is the smallest integer for which $D!/(D-\ell)! \geq N$, if such an integer $\ell < D$ exists.*

Proof: Pick an arbitrary vertex r in G as the root, and develop a *colorful* version of a breadth first search tree from r. The root r is at level 0. All neighbors of r (there are at least D of them) are at level 1. Having developed level i, level $i+1$ is developed as follows. For every vertex v of level i, consider all edges incident with it that have colors different from the colors of the tree edges along the path from r to v. There are at least $D-i$ such edges. If any such edge connects to a different vertex v' at level i, then this closes a distinguished cycle (going through v, v' and their least common ancestor). Hence we may assume that all these edges go to level $i+1$. It follows that at level ℓ at the latest (with ℓ as in the lemma), some vertex has two different ancestors at one level below. This closes a cycle. As no color can appear more than twice on this cycle, it is a distinguished cycle. $\Box$

Remark 2.11 *The proof of Lemma 2.10 shows the existence of cycles in which all colors on the cycle appear either once or twice, whereas for a cycle to be distinguished it suffices that one color appears either once or twice.*

Corollary 2.12 *Given n, r, d and s as defined above, let $D = (d-1)^2/2$ and let $\ell < D$ be such that $D!/(D-\ell)! \geq n^{2s}$. Then every r-uniform hypergraph H on n vertices with $m = dn^{r/2}$ hyperedges has an even cover with no more than 4ℓ hyperedges. Moreover, such an even cover can be found in time polynomial in n and m.*

Proof: Given a hypergraph H, use Lemma 2.4 with $\epsilon = 1$ to transform it to a hypergraph satisfying the small overlap condition, with the value of d replaced by $d-1$. (The other alternative in Lemma 2.4, if it holds, already implies an even cover of the desired size.) Construct from the resulting hypergraph the graph G. By lemma 2.7 the average degree in G is at least $(d-1)^2$. By Proposition 2.9 G has a subgraph of minimum degree at least D. By Lemma 2.10, this subgraph has a distinguished cycle of length at most 2ℓ. By Lemma 2.6 this corresponds to an even cover of size at most 4ℓ in H. By inspection one can verify that all parts of the proof are algorithmic, leading to the desired polynomial time algorithm. $\Box$

When D is much larger than ℓ, then the degree condition in Corollary 2.12 is essentially $d^{2\ell} \geq n^{2s}$ for the existence of an even cover of size $k = 4\ell$. Hence, to have an even cover of size k a value of $d = O(n^{4s/k})$ suffices. We remark that in [11] a somewhat different value is proved for d. Namely, for r divisible by 3, the bound in [11] is $d = O(n^{4r/3k})$ (and an error term is introduced in the exponent when k is not divisible by 3). For large r, we can choose $s \leq (r+3)/4$ (or smaller, if $r+1$ is divisible by 4) and we get a better bound of $d = O(n^{r/k+3/k})$.

Setting $r = 3$ and $d \simeq \sqrt{\log n / \log\log n}$ in Corollary 2.12 proves item 1 of Theorem 1.4.

3 Distinguished cycles longer than $\log n$

This section contains the proof of item (2) of Theorem 1.6. Only parts of this section (Proposition 3.3, Theorem 3.4 and Lemma 3.5) will be used in the proof of item 2 of Theorem 1.4, which will appear in section 4.

The proof of Lemma 2.10 assumes the minimum degree to be $D = \Omega(\log n/\log\log n)$. The purpose of this section is to prove the existence of distinguished cycles when the minimum degree is much lower. We note that principles used in the analysis of [11] fail to work already when the minimum degree drops below $\log n$, because (translating their proof technique to our notation) they are using a stricter notion of distinguished cycle in which some color needs to appear exactly once. A hypercube in which edges are colored according to the coordinate of the bit that they flip is an example of a graph of degree $\log n$ that is legally colored and does not have any distinguished cycle in this stronger sense.

3.1 A digression

For the purpose of explaining our proof technique, let us temporarily change the setting in which we seek to find a distinguished cycle. Rather than having a legally-colored graph of minimum degree D, we shall assume that we have a graph in which edges are colored (not necessarily legally) by D colors, and every vertex is incident with at least one edge of every color.

Definition 3.1 *Let $L_{col}(N, D)$ denote the minimum value of ℓ such that for every graph G with N vertices and any coloring of its edges by D colors, if every vertex is incident with edges of all colors, then G contains a distinguished cycle of length at most 2ℓ.*

It may be useful to notice that $L_{col}(N, D)$, $L_{avg}(N, D)$ and $L_{min}(N, D)$ have a common special case, namely D-regular graphs with a legal coloring by D colors. Such colorings exist for all bipartite D-regular graphs.

Observe that the proof of Lemma 2.10 applies to $L_{col}(N, D)$ as well. Hence for $D > \frac{2\log n}{\log\log n}$ we have that $L_{col}(N, D) \leq \frac{2\log n}{\log\log n}$ (the leading constant 2 is for illustrative purposes only and is not meant to be best possible). The following proposition improves the value of D.

Proposition 3.2 *For $D > \log\log N$, $L_{col}(N, D) \leq O(\log^3 N)$. Moreover, a distinguished cycle of this length can be found in polynomial time.*

Proof: Consider an arbitrary graph on N vertices, and an arbitrary coloring of its edges by D colors such that every vertex is incident with all colors. We will show that a distinguished cycle exists. Our proof also provides a polynomial time algorithm for finding such a cycle.

Remove all edges from G and put them back in, one color class at a time. We shall consider the minimum size of connected components that are formed at various steps of this process. We shall show that the assumption that there are no distinguished cycles implies the existence of a connected component of size larger than N, which is a contradiction. Moreover, throughout our proof we shall control the diameter of connected components, and this will lead to a proof that there is a distinguished cycle of length $O(\log^3 N)$.

Initially, all vertices are isolated and there are N components. After adding edges of the first color class, every vertex has degree at least one. We partition the graph into connected components as follows.

1. Iteratively, pick an arbitrary vertex that has not yet been marked. Mark it as a center vertex, and mark all its neighbors and all their neighbors as noncenter vertices.

2. Every center vertex will correspond to exactly one connected component. It will be connected to all its neighbors. All other noncenter vertices (those are at distant two from the set of center vertices) connect to the center vertex that originally marked them (other choices would work as well).

It is not hard to see that every connected component has size at least $s_1 = 2$ and diameter at most $d_1 = 5$, where for convenience diameter is measured here as the number of vertices (including endpoints) on the shortest path between the two most distant vertices in a connected component. (Hence for example, the diameter of C_4, the cycle on four vertices, is 3.)

Consider now what happens when edges of the second color class are added. If any such edge lies in an existing connected component, then this component must contain a distinguished cycle with this edge being the only edge of its color. Hence all second color edges join different components. Moreover, if there are two components that are joined by two edges of the second color, this leads to a distinguished cycle in which the second color appears twice. Hence we may assume that there is at most one edge of the second color joining any two components.

Consider now a graph for which the components after the first phase are the vertices, and edges of the second color are the edges. This must be a simple graph, and moreover, its minimum degree is 2 (because every vertex in every component is incident with at least one edge of the second color). Partitioning this new graph into connected components as described above we get components of size at least $s_2 = s_1(s_1 + 1) = 6$, and diameter at most $5d_1 = 25$.

Likewise, after adding edges of the third color, all components are of size at least $s_3 = s_2(s_2 + 1) = 42$ and the diameter is at most $5d_2 = 125$. By induction, after the $i + 1$th color is added, all components are of size larger than 2^{2^i}. Hence if there are more than $\log \log N$ colors there must be a distinguished cycle. The length of the distinguished cycle is at most twice the size of the maximum diameter reached (plus two, for the two edges of the last color connecting two components), and can readily be seen to be at most essentially $5^{\log \log N}$, and hence $O(\log^3 N)$. □

3.2 Back to the main proof

We now return to the proof of item 2 in Theorem 1.6. The following technical proposition will be used in the proof of our next theorem.

Proposition 3.3 *For every integers $0 < b < a$:*

1. $\log a + \frac{b}{a} < \log(a + b)$.

2. $\frac{b}{a} < \frac{\log(b+1)}{\log(a+1)}$.

3. $\log \log a + \frac{b}{a} < \log \log(a(b + 1))$.

(All logarithms are in base 2.)

Proof:

1. For $b = 0$ and for $b = a$, $\log a + \frac{b}{a} = \log(a+b)$. Hence the result for $0 < b < a$ follows by concavity of the logarithm function.

2. For $b = 0$ and for $b = a$, $\frac{b}{a} = \frac{\log(b+1)}{\log(a+1)}$. Again, the result for $0 < b < a$ follows by concavity of the logarithm function.

3. In the derivation below, the first inequality follows from item (2) and the third inequality follows from item (1) (using $\log a$ as a).

$$\log\log a + \frac{b}{a} < \log\log a + \frac{\log(b+1)}{\log(a+1)} < \log\log a + \frac{\log(b+1)}{\log a}$$

$$< \log(\log a + \log(b+1)) = \log\log(a(b+1))$$

□

We now want to prove a result similar to that of Proposition 3.2 also for $L_{avg}(N, D)$. We first do so without providing any bounds on the length of the distinguished cycle.

Theorem 3.4 *For every graph on n vertices and average degree $d > 4\log\log 2n$, every legal coloring of its edges creates at least one distinguished cycle.*

Proof: Given n and d, start with the empty graph on n vertices, and add in the edges one color class at a time, under the assumption that there is no distinguished cycle. We shall prove the following inductive hypothesis.

Inductive hypothesis. At no stage during the process there is a connected component with n' vertices and average degree larger than $d' = 4\log\log 2n'$.

Base case. All connected components are of size 1, with average degree $4\log\log 2 = 0$.

Inductive step. Assume that the theorem is true before adding color class c. When adding color class c, no new edge lies within an existing connected component, as this edge could be used to close a distinguished cycle with edges of previous colors. Likewise, no two previous components are connected by two new edges, as again these two new edges can be used to close a distinguished cycle with edges of previous colors. Hence any two previous components are connected by at most one new edge.

For every new edge connecting two components, charge both endpoints of the edge to the smaller of the two components (breaking ties arbitrarily), and there spread the charge evenly among all vertices of the smaller component. We show by induction that the total charge of a vertex v does not exceed $4\log\log 2n'$, where n' is the size of the component to which v belongs. Observe that for every connected component, the sum of the charges of all vertices is equal to the sum of the degrees. Hence the fact that in a component of size n' no charge exceeds $4\log\log 2n'$ implies the same with respect to average degree.

For a vertex v, let s be the size of its component before edges of color c are added. By the induction hypothesis, its charge at this point is at most $4\log\log 2s$. Assume that when adding edges of color c, the number of edges that are charged to the component of v is b. Hence the new charge for v is $4\log\log 2s + \frac{2b}{s}$. On the other hand, v must belong now to a component of size at least $s(b+1)$, because each one of the b edges must connect to a distinct component at least as large as s. Hence to establish the inductive step, it remains to see that

$$4\log\log 2s + \frac{2b}{s} \;\leq\; 4(\log\log 2s + \frac{b}{2s}) \;\leq\; 4\log\log(2s(b+1))$$

where the last inequality follows from item (3) in Proposition 3.3 (replacing a by $2s$). □

We shall now show that when the degree is $c\log\log 2n$ (for large enough c) there is a distinguished cycle which is not too long, and that such a cycle can be found efficiently. We shall use the following known lemma.

Lemma 3.5 *There is a polynomial time algorithm that given any graph on n vertices and m edges and a value $k > 1$, removes at most m/k edges and produces a graph in which every connected component has diameter $O(k\log n)$.*

Proof: Pick an arbitrary vertex and grow a ball of radius r around it, where r is the minimum value for which the number of boundary edges (that exit the ball) is smaller than $1/k$ times the number of internal edges (within the ball). Then discard the boundary edges (if any boundary edges exist). Repeat the process starting at an arbitrary vertex not already within a ball, as long as such vertices exist.

For every edge discarded, we keep at least k internal edges, and hence at most a $1/k$ fraction of the edges are discarded. The radius of a component cannot exceed $k\ln m$ because every new layer increases the number of edges by a factor of $(1+1/k)$, and we need to have $(1+1/k)^r \leq m$. □

Our plan is to use Lemma 3.5, multiple times, once for each color class used in the legal coloring. For this reason, we first show that it suffices to consider legal colorings with only few colors.

Lemma 3.6 *Consider an arbitrary N-vertex graph of average degree D and an arbitrary legal coloring of it. Then there is a polynomial time procedure that generates a new graph G with at most $2N$ vertices and average degree at least $D/4$ together with a legal coloring of the edges of G using at most $4D$ colors, such that every distinguished cycle in G can be mapped back to a distinguished cycle of the same length in the original graph.*

Proof: Consider an arbitrary legally colored graph on N vertices with average degree D. Modify the input graph to be nearly regular, with all degrees between D and $D/2$. This can be done as follows. First, iteratively removing vertices of degree at most $D/2$, resulting (as in Proposition 2.9) in a graph of minimum degree at least $D/2$. Thereafter, iteratively split any vertex of degree $D' > D$ into two vertices, one of degree $\lceil D'/2\rceil$ and the other of degree $\lfloor D'/2\rfloor$. The splitting operation preserves the number of edges. Hence the resulting graph G' has minimum degree $D/2$, maximum degree D, and at most $2N$ vertices. It is legally colored and every distinguished cycle in G' corresponds to a distinguished cycle of at most the same length in the original graph.

The legal coloring of G' is arbitrary and there is no a-priori bound on the number of colors that it uses (other than not being larger than the number of edges). We now describe a procedure for replacing this legal coloring by a new legal coloring of the edges with only $4D$ new colors. This is done by going over the original colors one by one. For each of the original colors C_{old} pick one of the new colors C_{new} and recolor all edges of original color C_{old} by the color C_{new}. If this new coloring introduces conflicts (this can happen if an edge of C_{old} is incident with a vertex that already has some other incident edge colored C_{new}), then drop the edge of C_{old} that leads to the conflict. To avoid dropping too many edges, we use the following rule when mapping a color C_{new} to C_{old}: we choose the new color that will result in the smallest number of dropped edges from C_{old} (breaking ties arbitrarily). As there are $4D$ new colors and only at most $2(D-1)$ edges incident with the endpoints of any edge, there must be a choice of C_{new} that will result in dropping at most half the edges of C_{old}. Hence eventually the resulting graph is legally colored, and its average degree is at least $D/4$. Moreover, every distinguished cycle in the new graph (with respect to the new colors) is a distinguished cycle of the original graph (with respect to the original colors). □

We now reach the theorem that implies the proof of item (2) in Theorem 1.6.

Theorem 3.7 *Every legally colored graph on N vertices and average degree $D \geq 32 \log\log 4N$ has a distinguished cycle of length $O(\log N \log\log N)$. Moreover, such a cycle can be found in polynomial time.*

Proof: By Lemma 3.6, instead of the input graph we may consider a new graph G with at most $2N$ vertices, average degree at least $8 \log\log 4N$, and with a legal coloring that uses at most $4D = O(\log\log N)$ colors. Observe that the average degree is at least twice as large as that used in the proof of Theorem 3.4. This allows us to discard half the edges of the graph, and there still would be a distinguished cycle. We shall use this slackness so as to modify the proof of Theorem 3.4 so that after adding each color, no component will have diameter larger than $O(\log N \log\log N)$. This is done by applying Lemma 3.5 after each color class is added. The parameter k in the lemma is chosen to be $8D$, so that even after the lemma is applied $4D$ times (once for each color), at most half of the edges are discarded in total. The proof of Theorem 3.4 still works even though we discard at most half the edges, because there was a factor 2 slackness in the average degree that we started with. Hence eventually a distinguished cycle will be found (when an edge of a new color is placed inside an existing component, or when two edges of a new color join two existing components). The diameter of every component at the time that the distinguished cycle is found is at most $O(\log N \log\log N)$, and the length of the distinguished cycle need never be more than two plus sum of diameters of two components that are connected by two edges of the same color.

All steps of the proof are constructive and give a polynomial time algorithm for finding a distinguished cycle of the appropriate length. □

3.3 A negative example

A question that remains is whether for some constant average degree D and any legal edge coloring there must be a distinguished cycle. The only nontrivial negative result that the author is aware of is the following.

Proposition 3.8 *There are 3-regular graphs whose edges can be colored in such a way that no distinguished cycle exists.*

Proof: Consider an infinite tiling of the plain by hexagons (later we will modify the construction to be finite). This defines a 3-regular graph in a natural way. Legally color its edges by three colors so that every hexagon contains only two colors. This can be done by first coloring any two adjacent edges, and then this determines uniquely the color of every other edge (by alternating two colors along the edges of a hexagon, and using the third color for the other edges incident with the vertices of the hexagon).

We now show that there is no distinguished cycle. W.l.o.g., let red be the color that appears twice on a hypothetical distinguished cycle. (The case that some color appears only once is even simpler, and omitted.) Then the rest of the cycle is composed of paths that have only colors blue and green, and each such path must lie on a single hexagon. This requires two hexagons that are colored by the colors blue and green to be connected by two different edges of color red, but this never happens in the given 3-coloring.

Inspection shows that the coloring is periodic. Hence the infinite graph can be replaced by a finite graph as follows. Picking some orientation of edges as vertical, the hexagons are arranged in rows. Two even rows sufficiently far from each other can be identified to be one row. Each column makes a zigzag pattern. Two such columns sufficiently far from each other (at a distance that is a multiple of three) can be identified to be one column. This results in a finite graph. □

Proposition 3.8 refers to distinguished cycles, but does not imply anything for even covers. The graph there contains cycles of length 6, but it is possible to show that any hypergraph that satisfies the small overlap condition for which the corresponding graph has cycles of length 6 must have an even cover of size at most 12. This leads to the following question.

Question 3.9 *Is there is a 3-uniform hypergraph H satisfying the small overlap condition that on the one hand does not contain any even cover, and on the other hand, the graph G associated with H has a cycle (in which case this cycle must be trivial)?*

If the answer to Question 3.9 is negative, then item 2 of Theorem 1.4 can easily be improved: it would suffice to have $\beta = \Theta(\sqrt{n})$ in order for even covers of size $O(\log n)$ to exist.

3.4 Extensions

The bound on the degree stated in Theorem 3.7 is $32 \log\log 4N$. The leading constant of 32 is a consequence of our attempt to keep the proof simple, rather than optimize the leading constant. It can be drastically reduced, with only a modest loss in the cycle length (which will still asymptotically remain $O(\log N \log\log N)$). Let us mention a few places where there is slackness in our analysis.

The leading constant in the bound on the degree in Theorem 3.4 can be improved with more work. For example, the original reason for having $\log\log 2n$ in the theorem rather than $\log\log n$ is to handle cases that $n \leq 2$ in a unified way. However, this later costs a factor of 2 in the leading constant. Additional savings can be obtained by changing the charging mechanism. Rather than charging both end points of an edge to the smaller of the two components, one can allocate part of the charge to the larger component (not to mention the possibility of propagating the charge to other components).

In Lemma 3.6 we loose a factor of 4 in the average degree. However, there is no need to loose more than a factor of $(1+\epsilon)$ (for some small ϵ), at the cost of having the new legal coloring use more colors (which will eventually translate to longer distinguished cycles). This can be done by allowing G' to have maximum degree roughly D/ϵ, and thereafter using $O(D/\epsilon^2)$ new colors. (Possibly, with tighter analysis, the number of colors would have a better dependency on ϵ than $1/\epsilon^2$.)

In the proof of Theorem 3.7 we may take k to be much larger (say $1/\epsilon$ times the number of colors), which again will reduce the degree requirement at a cost of increasing the diameter of connected components (and hence the length of the distinguished cycle).

Summarizing the above discussion, it should not be difficult to reduce the degree requirement in Theorem 3.7 to $c_1 \log\log 4N$ with c_1 being a constant much smaller than 32 (possibly, nearly 1), at the cost of finding distinguished cycles of length $c_2 \log N \log\log N$, with c_2 being a constant that depends on c_1.

In special families of graphs (that are probably not relevant to the application of refuting semirandom 3SAT instances), we can improve the bounds of Theorem 3.4. We first briefly recall a few well known facts. A graph H is a *minor* of graph G if it can be obtained from G by the operations of contracting edges, deleting edges, and removing isolated vertices. As shown by Robertson and Seymour, every minor closed family of graphs can be characterized by a finite list of forbidden minors. For example, the family of planar graphs is closed under minors, and the two forbidden minors are K_5 and $K_{3,3}$ (this was proved by Wagner, and is related to Kuratowski's theorem). For a minor closed family of graphs, if some graph F on f vertices is a forbidden minor, then so is K_f. Every graph of average degree d must have K_f as a minor, for some $f = \Omega(d/\log d)$ [10, 12]. Hence the average degree of any graph from a minor closed family is bounded by $O(f \log f)$, where f is the size of the smallest forbidden minor. Theorem 3.4 is not interesting (in an asymptotic sense) for minor closed families of graphs, because the degree bound $\log\log 2n$ in the theorem cannot be attained when n is large. For such graphs, the following corollary replaces the dependency on n by a similar dependency on f.

Corollary 3.10 *Let G be any graph with no K_f as a minor and of average degree d, with $d > c \log\log f + O(1)$, where $c \geq 1$ is some universal constant. Then for any legal coloring of the edges of G there is distinguished cycle.*

Proof: The proof follows that of Theorem 3.4, with the following change. When adding color c, let C be a new component formed by connecting some previous components $C_1, C_2, \ldots$, and let $|C_i|$ denote the number of vertices in component C_i. Redistribute the connecting edges so that every one of the original components C_i is incident with at most $\min[|C_i|, s]$ edges, where $s = f \log f$. This is possible, because otherwise G has K_f as a minor. Now apply the charging mechanism of Theorem 3.4.

We now compute how much a vertex v is charged overall. The total charge until its component has size s is $O(\log\log s)$, as in the proof of Theorem 3.4. Thereafter, for the component to grow from size S to fS, vertex v is charged at most s/S. Hence the total charge after size s is reached forms a decreasing geometric series that sums up to $O(1)$. □

4 Nontrivial cycles of length $O(\log n)$

Here we prove item 2 of Theorem 1.4.

We recall some known facts about cycle bases of graphs. Given a connected graph with n vertices and m edges, order the edges in some arbitrary order, and with each set of edges associate an indicator vector in $\{0,1\}^m$ in a natural way. For the purpose of the discussion here, a cycle in a graph will be any collection of edges such that the degree induced on each vertex is even. (In particular, the union of two edge disjoint cycles is a cycle.) The vectors associated with all cycles in a graph form a vector space of dimension $m-n+1$ (with vector addition modulo 2). A basis for this vector space can be obtained as follows. Consider an arbitrary spanning tree T of the graph. For each edge $e \not\in T$, there is a unique cycle (called a *fundamental cycle*) that is a simple cycle containing (some of the) edges of the tree and the edge e. The $m-n+1$ fundamental cycles form a basis for the cycle space.

The radius $R(G)$ of a graph G is defined to be the maximum distance between a center vertex u and any other vertex in the graph, where a center vertex u is any vertex that minimizes this maximum.

Proposition 4.1 *Every graph of radius R has a cycle basis in which every cycle has length at most $2R+1$.*

Proof: Let u be a center vertex for the graph, and consider the spanning tree corresponding to the breadth first search tree rooted at u. Then the fundamental cycles with respect to this tree each has length at most $2R+1$. □

Recall that Lemma 3.5 shows that for every graph with m edges, we may discard half of its edges such that each connected component that remains has radius $O(\log m)$.

Corollary 4.2 *In every graph with m edges one may discard half the edges such that the remaining graph has a cycle basis in which each cycle is of length $O(\log m)$.*

Proof: Use Lemma 3.5 to choose which edges to discard so that each remaining component has radius $O(\log m)$. Thereafter, for each connected component separately, find a cycle basis as in Proposition 4.1. The union of these cycle bases is the desired cycle basis. □

The following lemma motivates our degression to cycle bases of graphs.

Lemma 4.3 *Let G be a graph constructed from a hypergraph H as explained in Section 2. Let G' be an edge induced subgraph of G that has a cycle basis in which every cycle is of length at most ℓ. Then if G' has a distinguished cycle (of arbitrary length), then G' must also have a nontrivial cycle of length ℓ.*

Proof: Label every edge of G' by the two hyperedges of H that generate it. Then as we have shown in the proof of Lemma 2.6, there must be some hyperedge e of H that labels only one of the edges of the distinguished cycle. The distinguished cycle can be expressed as a sum (mod 2) of basis cycles. Then it must be the case that at least on one of these basis cycles (which has length at most ℓ), e labels an odd number of its edges. Hence this basis cycle must be nontrivial. □

Note that the nontrivial cycle found in the proof of Lemma 4.3 need not be a distinguished cycle (because addition of basis cycles is done modulo 2 which may lead to cancellations of edges).

Theorem 4.4 *Let G be a graph constructed from a hypergraph H as explained in Section 2. If G has degree $8 \log \log 2n$, then it contains a nontrivial cycle of length $O(\log n)$.*

Proof: Use Corollary 4.2 to remove half the edges and remain with a graph G' of average degree at least $4 \log \log 2n$ and a cycle basis in which each cycle has length $O(\log n)$. Theorem 3.4 implies that G' has a distinguished cycle. Lemma 4.3 implies that G' has a nontrivial cycle of length $O(\log n)$. As G' is a subgraph of G, then also G has a nontrivial cycle of length $O(\log n)$. □

The leading constant of 8 in Theorem 4.4 was chosen for concreteness and simplicity. It can be reduced using arguments similar to those presented in Section 3.4.

Item 2 of Theorem 1.4 follows from Theorem 4.4 in a way similar to the proof of Corollary 2.12.

There is a straightforward correspondence between even covers in hypergraphs and linear dependency modulo 2 in vectors. We note that our proofs, going through the notion of nontrivial cycles, in fact correspond to linear dependencies of $\{0, 1\}$ vectors over any field (and this was also the case in [11]). The reason is as follows. Orient the edges of the nontrivial cycle so that it creates a directed cycle. An edge directed from v_1 to v_2 corresponds to two hyperedges. In the linear dependency we shall add one of them and subtract the other according to the following convention. The hyperedge whose prefix labels the prefix of v_1 is added, and the hyperedge whose prefix labels the prefix of v_2 is subtracted. It is not hard to see that going around the nontrivial cycle, all vertices cancel out.

Corollary 4.5 *For a sufficiently large constant c, in any set of $cn\sqrt{n \log \log n}$ vectors in $\{0, 1\}^n$ of hamming weight 3, there are two disjoint multisets of $O(\log n)$ vectors (the same vector may appear more than once in a multiset and then it is counted more than once) such that the two respective sums of all vectors in the multisets are identical.*

Acknowledgement

The author holds the Lawrence G. Horowitz Professorial Chair at the Weizmann Institute. Part of this work was done at Microsoft Research, Redmond, Washington. I thank Assaf Naor for bringing [11] to my attention, and suggesting that it is related to the work in [6].

References

[1] N. Alon and U. Feige. On the power of two, three and four probes. *Manuscript*, 2008.

[2] N. Alon, S. Hoory, N. Linial, The Moore bound for irregular graphs, *Graphs and Combinatorics* 18 (2002), 53–57.

[3] A. Coja-Oghlan, A. Goerdt and A. Lanka. Strong Refutation Heuristics for Random k-SAT. In *Combinatorics, Probability and Computing*, 16: 5–28, 2007.

[4] U. Feige. Relations between average case complexity and approximation complexity. In *Proc. of the 34th Annual ACM Symposium on Theory of Computing*, pages 534–543, 2002.

[5] U. Feige. Refuting smoothed 3CNF formulas. In *Proceedings of 48th Annual Symposium on Foundations of Computer Science, 407–417, 2007.*

[6] U. Feige, J. H. Kim, E. Ofek. Witnesses for non-satisfiability of dense random 3CNF formulas. In *Proceedings of 47th Annual Symposium on Foundations of Computer Science, 497–508, 2006.*

[7] U. Feige and E. Ofek. Easily refutable subformulas of large random 3CNF formulas. *Theory of Computing*, Volume 3 (2007) Article 2, pp. 25–43. http://theoryofcomputing.org.

[8] J. Friedman, A. Goerdt, and M. Krivelevich. Recognizing more unsatisfiable random k-SAT instances efficiently. *SIAM Journal on Computing*, 35(2): 408–430, 2005.

[9] J. Hastad. Some optimal inapproximability results. *J. ACM 48(4): 798–859 (2001).*

[10] A.V. Kostochka. Lower bound of the Hadwiger number of graphs by their average degree. *Combinatorica 4 (4), pp. 307–316, 1984.*

[11] A. Naor and J. Verstraete. Parity check matrices and product representaions of squares. *Submitted for publication.* A preliminary version appeared in IEEE ISIT 2005.

[12] A. Thomason. An extremal function for contraction of graphs. *Math. Proc. Comb. Phil. Soc. 95, pp. 261–265, 1984.*