Improved Lower Bounds for Testing Triangle-freeness in Boolean Functions via Fast Matrix Multiplication

Hu Fu
Microsoft Research
New England Lab
hufu@microsoft.com

Robert Kleinberg
Cornell University
Dept. of Computer Science
rdk@cs.cornell.edu

August 7, 2013

Abstract

Understanding the query complexity for testing linear-invariant properties has been a central open problem in the study of algebraic property testing. Triangle-freeness in Boolean functions is a simple property whose testing complexity is unknown. Three Boolean functions f_1, f_2 and $f_3 : \mathbb{F}_2^k \to \{0,1\}$ are said to be triangle free if there is no $x, y \in \mathbb{F}_2^k$ such that $f_1(x) = f_2(y) = f_3(x+y) = 1$. This property is known to be strongly testable (Green, 2005), but the number of queries needed is upper-bounded only by a tower of twos whose height is polynomial in $1/\epsilon$, where ϵ is the distance between the tested function triple and triangle-freeness, i.e., the minimum fraction of function values that need to be modified to make the triple triangle free. A lower bound of $(\frac{1}{\epsilon})^{2.423}$ for any one-sided tester was given by Bhattacharyya and Xie (2010). In this work we improve this bound to $(\frac{1}{\epsilon})^{6.619}$. Interestingly, we prove this by way of a combinatorial construction called *uniquely solvable puzzles* that was at the heart of Coppersmith and Winograd (1990)'s renowned matrix multiplication algorithm.

1 Introduction

Property testing studies algorithms using a small number of queries to a large input that decides, with high probability, whether the input satisfies a certain property or is far from it. Typically, the input f is a function mapping from a finite domain D to a range R. A property $\mathcal{P}$ is a subset of all such functions $\{f : D \to R\}$. If we measure the distance between two functions by the Hamming metric, $\mathrm{dist}(f,g) := \mathbf{Pr}_x[f(x) \neq g(x)]$, then the distance from f to the property $\mathcal{P}$ is $\mathrm{dist}(f,\mathcal{P}) := \min_{g\in\mathcal{P}} \mathrm{dist}(f,g)$. Fixing a distance ϵ, an algorithm, called a *tester*, makes randomized queries to f, and outputs YES with probability at least $2/3$ for $f \in \mathcal{P}$, and NO with probability at least $2/3$ if $\mathrm{dist}(f,\mathcal{P}) \geq \epsilon$. A tester is said to be *one-sided* if it outputs YES with probability one for $f \in \mathcal{P}$. The central question studied by property testing, as initiated by Rubinfeld and Sudan (1996) and Goldreich et al. (1998), is to understand the query complexity, i.e., the minimum number of queries needed by a tester, to test various properties.

For example, a property is called *strongly testable* if its query complexity does not depend on the size of the doamin $|D|$ and is only a function of ϵ. For graph and hypergraph properties, strongly testable properties have been exactly characterized (Alon et al., 2006). Among strongly testable properties, it is important to understand which ones admit testers with query complexity polynomial in $1/\epsilon$ and which do not. For example, for undirected graphs and one-sided testers, H-freeness for a fixed subgraph H has polynomial query complexity if and only if H is bipartite (Alon, 2002). Similar characterizations are known for directed graphs and hypergraphs (Alon and Shapira, 2004, 2005; Rödl and Schacht, 2009; Austin and Tao, 2010).

Kaufman and Sudan (2008) suggested that symmetries, or invariance under transformations of a property, play an important role in facilitating efficient testers. As an easy example, a graph property, seen as functions on graph edges, are invariant under graph isomorphisms, i.e. permutations of the nodes. Kaufman and Sudan launched the systematic study of *algebraic* property testing, and in particular singled out *linear-invariant* properties as a natural class of properties to consider. Restricted to the context of Boolean functions, a property $\mathcal{P} \subset \{f : \mathbb{F}_2^k \to \{0,1\}\}$ is said to be linear-invariant if for all $f \in \mathcal{P}$ and linear transformation $L : \mathbb{F}_2^k \to \{0,1\}$, the composition $f \circ L$ is still in $\mathcal{P}$. One may further define a property $\mathcal{P}$ to be *linear* if it is closed under linear operations; for a property $\mathcal{P}$ on Boolean functions, this simply means $f, g \in \mathcal{P}$ entails $f + g \in \mathcal{P}$. Kaufman and Sudan (2008) showed that all properties that are linear-invariant and linear can be tested with query complexity polynomial in $1/\epsilon$. When the linearity condition is relaxed, however, the picture of what is currently understood is less clear. *Triangle-freeness* is one such property.

A function $f : \mathbb{F}_2^k \to \{0,1\}$ is said to be triangle-free if there are no $x, y \in \mathbb{F}_2^k$ such that $f(x) = f(y) = f(x+y) = 1$. More generally, f is said to be (M,σ)-free for a fixed matrix $M \in \mathbb{F}_2^{r\times s}$ and vector $\sigma \in \{0,1\}^s$, if there exists no $x = (x_1, \ldots, x_s) \in (\mathbb{F}_2^k)^s$ such that $Mx = \mathbf{0}$ and $f(x_i) = \sigma_i$ for all $i \in [s]$. Green (2005) showed that $(M,\mathbf{1})$-freeness with rank-one matrix M is strongly testable (which includes triangle-freeness), and started the line of investigations resolving that any $(M,\mathbf{1})$-freeness is strongly testable (Král et al., 2013; Shapira, 2009), and that the intersection of (possibly infinite) (M,σ)-freeness, with rank-one M, is testable (Bhattacharyya et al., 2011, 2010). However, the upper bounds for the number of queries given in these works, though independent of k, are all towers of twos whose heights are polynomial in $1/\epsilon$. The only exception is a result of Bhattacharyya et al. (2012) showing that odd-cycle-freeness can be tested with $\tilde{O}(1/\epsilon^2)$ queries. It was noted by Bhattacharyya et al. that this property is the intersection of *infinite* $(M,\mathbf{1})$-freeness. In fact, it has been conjectured that testing any odd cycle alone takes supernomial number of queries. Prior to this work, the only nontrivial bound for the simplest such property, triangle-freeness, was

given by Bhattacharyya and Xie (2010), who showed that any one-sided tester needs $\Omega(1/\epsilon^{2.423})$ queries. This is in sharp contrast with our complete understanding of the query complexity of testing H-freeness in graphs, the counterpart among graph properties to $(M, \mathbf{1})$-freeness.

Our Results. In this work we improve Bhattacharyya and Xie (2010)'s lower bound and show that any one-sided tester needs $\Omega(1/\epsilon^{6.619})$ queries to test triangle-freeness in Boolean functions. Bhattacharyya and Xie (2010)'s lower bound was built on families of vectors having a combinatorial property called *perfect-matching-free* (PMF, Definition 1). Roughly speaking, a PMF family can be expanded to construct Boolean functions such that for every x with $f(x) = 1$, there exist a small number of y's such that $f(y) = f(x + y) = 1$. Such a function has a number of triangles that is about linear with the number of 1's needed to be flipped to remove all triangles. In other words, the number of triangles is relatively small whereas the distance to triangle-freeness is relatively large, a difficult scenario for a tester. However, Bhattacharyya and Xie were able to find only very small (and hence weak) PMF families by way of numerical calculations. When the dimension of the family exceeds 5 the calculation becomes forbiddingly expensive.

In this work, we are able to construct large PMF families by using a combinatorial structure called *uniquely solvable puzzles* (USP, Definition 2). USP's were defined by Cohn et al. (2005) in their group theoretic approach to fast matrix multiplication. Under their perspective, the most important step in Coppersmith and Winograd (1990)'s famous $O(n^{2.376})$-time algorithm for multiplication of $n \times n$ matrices was a construction of large USP's. Coppersmith and Winograd's algorithm was for a long time the best known algorithm for this fundamental problem, and was improved only recently (Stothers, 2010; Williams, 2012). As we recall in Appendix A, Coppersmith and Winograd's construction crucially relies on large sets of densely populated integers with no three terms in arithmetic progressions (Salem and Spencer, 1942; Behrend, 1946; Elkin, 2010). Seen through the connection we identify here, it may not be a coincidence that the superpolynomial lower bounds for testing nonbipartite H-freeness in graphs also crucially used such sets with no arithmetic progressions (Alon, 2002). However, we were unable to give superpolynomial lower bounds for testing triangle-freeness in Boolean functions.

This leads to some fascinating open problems. For example, Cohn et al. (2005) showed that, if large families of a strengthened version of USP's, called strongly uniquely solvable puzzles (SUSP), exist, then the exponent of matrix multiplication is 2, as has long been conjectured. Would a large SUSP imply superpolynomial query complexity for testing any $(M, \mathbf{1})$-freeness in Boolean functions? On the other hand, would such a lower bound imply the success of Cohn et al. (2005)'s campaign on matrix multiplication? We leave these questions for future investigation.

2 Preliminaries

For an integer n, we let $[n]$ denote the set $\{1, 2, \ldots, n\}$. We use $\mathrm{Sym}(S)$ to denote the symmetric group on a set S. We will often identify a Boolean function $f : \mathbb{F}_2^k \to \{0, 1\}$ with the family of subsets in $[k]$ whose indicator function is f.

We will focus on testing triangle-freeness for Boolean function triples. [1] A function triple $f_1, f_2, f_3 : \mathbb{F}_2^k \to \{0, 1\}$ is said to be triangle-free if there is no $x, y \in \mathbb{F}_2^k$ such that $f_1(x) = f_2(y) =$

[1] This is called by Bhattacharyya and Xie (2010) the multiple-function case. Green (2005)'s technique easily generalizes to this case, giving the same bound of tower of twos.

$f_3(x+y) = 1$. Denote by T-FREE the set of function triples that are triangle-free, and the distance of a function triple to T-FREE is defined as

$$\mathrm{dist}((f_1, f_2, f_3), \text{T-FREE}) := \min_{(g_1,g_2,g_3)\in \text{T-FREE}} \mathrm{dist}(f_1, g_1) + \mathrm{dist}(f_2, g_2) + \mathrm{dist}(f_3, g_3).$$

As the following reduction and Theorem 2 shows, the multiple-function and single-function case are essentially equivalent.[2]

Lemma 1 (Xie, 2010)**.** *Given any function triple $f_1, f_2, f_3 : \mathbb{F}_2^k \to \{0,1\}$ which is ϵ-far from* T-FREE *and contains N triangles, there is a single function $f : \mathbb{F}_2^{k+2} \to \{0,1\}$ which is $\frac{\epsilon}{4}$-far from triangle-freeness and contains N triangles.*

Proof. Construct f as follows. For each $x \in \mathbb{F}_2^k$, denote by (a, b, x) the $(k+2)$-dimension vector whose last k coordinates are given by x. For each $x \in \mathbb{F}_2^k$, let $f(0, 0, x)$ be 0, $f(1, 0, x)$ be $f_1(x)$, $f(0, 1, x)$ be $f_2(x)$, and $f(1, 1, x)$ be $f_3(x)$. It is easy to see that any triangle in f has to have its three "vertices" given by entries from f_1, f_2 and f_3, respectively. The lemma follows immediately. □

The *canonical tester* is the naive-looking algorithm that samples $x, y \in \mathbb{F}_2^k$ uniformly at random and returns YES if $f_1(x) = f_2(y) = f_3(x+y) = 1$ and NO otherwise. A tester is said to be *one-sided* if, whenever the input satisfies the property in question, it outputs YES with probability 1. By the following theorem, it is without loss of generality to consider obfuscating the canonical tester.

Theorem 2 (Bhattacharyya and Xie, 2010)**.** *Suppose there is a one-sided tester for* T-FREE *has query complexity $q(\epsilon)$, then the canonical tester has query complexity at most $O(q^2(\epsilon))$. This holds for both the single-function case (when $f_1 = f_2 = f_3$) and the multiple-function case.*

Definition 1 (Perfect-Matching-Free (PMF) Families of Vectors)**.** Let k and m be integers such that $0 < k < m < 2^k$. A (k, m) perfect-matching-free (PMF) family of vectors is a set of vectors $(a_i, b_i, c_i)_{i=1}^m$, where $a_i, b_i, c_i \in \mathbb{F}_2^k$ and $c_i = a_i + b_i$ for all $i \in [m]$, such that for any permutation triple $\pi_1, \pi_2, \pi_3 \in \mathrm{Sym}([m])$, either $\pi_1 = \pi_2 = \pi_3$, or there exists an $i \in [m]$ such that $a_{\pi_1(i)} + b_{\pi_2(i)} \neq c_{\pi_3(i)}$.

One can permute and then concatenate all a_i's in a (k, m) PMF family and obtains $m!$ vectors in $\mathbb{F}_2^{km}$; the same can be done for b_i's and c_i's. By the property of PMF, each new vector obtained from a_i's forms one and only one triangle with two other vectors obtained from b_i's and c_i's, respecitvely, and they are obtained through exactly the same permutation on $[m]$. This means that to remove all $m!$ triangles in the system, one has to remove at least the same number of vectors. This large ratio between the distance to triangle-freeness and the number of triangles is exactly what is needed to obfuscate a tester. One may go further and take multiple copies of a PMF family and repeats this experiment. An asymptotic calculation would give the following theorem.

Theorem 3 (Bhattacharyya and Xie, 2010)**.** *If (k, m) PMF family of vectors exists, then for small enough ϵ and large enough k, there exists a function triple $f_1, f_2, f_3 : \mathbb{F}_2^k \to \{0,1\}$ that is ϵ-far from triangle-freeness, but the canonical tester needs $\Omega((\frac{1}{\epsilon})^\alpha)$ queries to detect a triangle, where $\alpha = (2 - \frac{\log m}{k})/(1 - \frac{\log m}{k})$.* [3]

[2] We acknowledge Xie (2010) for informing us of the possibility of this reduction.

[3] All logarithms in this paper are base 2.

Note that the existence of $(k, 2^{k(1-o_k(1))})$ PMF family would imply a super-polynomial lower bound for any one-sided triangle-freeness tester.

The workhorse of our improved lower bound for testing triangle-freeness is the following combinatorial construction. It was implicitly developed by Coppersmith and Winograd (1990) for their famed $O(n^{2.376})$-time matrix multiplication algrorithm, and Cohn et al. (2005) isolated it and gave it the reinterpretation we use here.

Definition 2 (Uniquely Solvable Puzzles (USP))**.** A *uniquely solvable puzzle* (USP) is a set $U \subset \{1,2,3\}^k$ such that, for all permutation triple $\pi_1, \pi_2, \pi_3 \in \mathrm{Sym}(U)$, either $\pi_1 = \pi_2 = \pi_3$, or there exist a $u \in U$ and an index $i \in [k]$ such that at least two of $(\pi_1(u))_i = 1$, $(\pi_2(u))_i = 2$ and $(\pi_3(u))_i = 3$ hold.

A useful way to look at a USP is to think of it as a set of puzzles having three colors, where each color has m pieces. A solution to the puzzle is an arrangement of the pieces into m rows each of size k, and there cannot be a conflict. The property in Definition 1 requires that there exists a unique solution to this puzzle, up to permutations on rows.

Theorem 4 (Coppersmith and Winograd, 1990; Cohn et al., 2005)**.** *Fixing integer k, the largest USP is of size $\Theta((3/2^{2/3} - o(1))^k)$.*

The upper bound, given by an elegant construction of large USP's in Coppersmith and Winograd's original paper was unfortunately buried in a system of algebraic notations not easy to decipher without a proficiency with that language. For the sake of completeness and to promulgate this beautiful construction, we give its proof, hopefully more accessible, in Appendix A.

3 A Construction of PMF Families via USPs

We now state the main theorem of the paper.

Theorem 5. *For any $\epsilon > 0$ and large enough k, there exists a function triple $f_1, f_2, f_3 : \mathbb{F}_2^k \to \{0,1\}$, such that the triple is ϵ-far from being triangle free, and the canonical tester needs $\Omega((\frac{1}{\epsilon})^{13.239})$ queries to detect a triangle in the triple. In addition, any one-sided tester needs $\Omega((\frac{1}{\epsilon})^{6.619})$ queries.*

By Theorem 3 and Theorem 2, Theorem 5 would be an immediate consequence of the following lemma.

Lemma 6. *There exists $(k, \Theta((3/2^{2/3} - o(1))^k))$ PMF family of vectors, for all k.*

Proof of Lemma 6. By Theorem 4, it suffices to construct a $(k, |U|)$ PMF family for any USP $U \subset \{1,2,3\}^k$. Let U be $\{u_1, u_2, \ldots, u_m\}$. We construct $3m$ vectors $a_i, b_i, c_i \in \mathbb{F}_2^k$ for $i = 1, 2, \ldots, m$. For each $i \in [m]$, let $a_{i,j}$ be 1 if $u_{i,j} = 1$, and 0 otherwise; let $b_{i,j}$ be 1 if $u_{i,j} = 2$, and 0 otherwise; let $c_{i,j}$ be 1 if $u_{i,j} \neq 3$, and 0 otherwise. It is clear now that $c_i = a_i + b_i$ for all i.

We now show that $\{a_i, b_i, c_i\}_{i=1}^m$ constitutes a PMF family. Note that a naive translation of the property of USP would not give the desired property for PMF: for $\pi_1, \pi_2, \pi_3 \in \mathrm{Sym}([m])$ that are not all equal and such that $u_{\pi_1(i),j} = 1$, $u_{\pi_2(i),j} = 2$ and $u_{\pi_3(i),j} = 3$ for some $i \in [m], j \in [k]$, we will have that $a_{\pi_1(i),j} = b_{\pi_2(i),j} = 1$ and $c_{\pi_3(i),j} = 0$, which does not prevent the sum of a_i and b_i from being c_i in $\mathbb{F}_2^k$. Instead, we observe that for $\pi_1, \pi_2, \pi_3 \in \mathrm{Sym}([m])$ that are not all equal, there must be an $i \in [m]$ and $j \in [k]$ such that $u_{\pi_1(i),j} \neq 1$, $u_{\pi_2(i),j} \neq 2$ and $u_{\pi_3(i),j} \neq 3$: this is because

of the conservation of the total number of elements in U. The number of 1's and 2's and 3's in U total at mk, and if, by the property according to Definition 1, under permutations of the puzzles there exist conflicts at some position, then there must be some other position that is not covered by a puzzle of any color. For such i and j we would have $a_{\pi_1(i),j} = b_{\pi_2(i),j} = 0$ and $c_{\pi_3(i),j} = 1$, which means that $a_{\pi_1(i)} + b_{\pi_2(i)} \neq c_{\pi_3(i)}$. This shows that we have indeed constructed a $(k, |U|)$ PMF family. □

References

Alon, N. (2002). Testing subgraphs in large graphs. *Random Struct. Algorithms*, 21(3-4):359–370.

Alon, N., Fischer, E., Newman, I., and Shapira, A. (2006). A combinatorial characterization of the testable graph properties: it's all about regularity. In *STOC*, pages 251–260.

Alon, N. and Shapira, A. (2004). Testing subgraphs in directed graphs. *J. Comput. Syst. Sci.*, 69(3):354–382.

Alon, N. and Shapira, A. (2005). A characterization of the (natural) graph properties testable with one-sided error. In *FOCS*, pages 429–438.

Austin, T. and Tao, T. (2010). Testability and repair of hereditary hypergraph properties. *Random Struct. Algorithms*, 36(4):373–463.

Behrend, F. A. (1946). On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci.*, 32:331–332.

Bhattacharyya, A., Chen, V., Sudan, M., and Xie, N. (2011). Testing linear-invariant non-linear properties. *Theory of Computing*, 7(1):75–99.

Bhattacharyya, A., Grigorescu, E., Raghavendra, P., and Shapira, A. (2012). Testing odd-cycle-freeness in boolean functions. *Combinatorics, Probability & Computing*, 21(6):835–855.

Bhattacharyya, A., Grigorescu, E., and Shapira, A. (2010). A unified framework for testing linear-invariant properties. In *FOCS*, pages 478–487.

Bhattacharyya, A. and Xie, N. (2010). Lower bounds for testing triangle-freeness in boolean functions. In *SODA*, pages 87–98.

Cohn, H., Kleinberg, R. D., Szegedy, B., and Umans, C. (2005). Group-theoretic algorithms for matrix multiplication. In *FOCS*, pages 379–388.

Coppersmith, D. and Winograd, S. (1990). Matrix multiplication via arithmetic progressions. *J. Symbolic Computation*, 9(3):250–280.

Elkin, M. (2010). An improved construction of progression-free sets. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 886–905, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.

Goldreich, O., Goldwasser, S., and Ron, D. (1998). Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750.

Green, B. (2005). A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376.

Kaufman, T. and Sudan, M. (2008). Algebraic property testing: the role of invariance. In *STOC*, pages 403–412.

Král, D., Serra, O., and Vena, L. (2013). On the removal lemma for linear systems over abelian groups. *Eur. J. Comb.*, 34(2):248–259.

Rödl, V. and Schacht, M. (2009). Generalizations of the removal lemma. *Combinatorica*, 29(4):467–501.

Rubinfeld, R. and Sudan, M. (1996). Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271.

Salem, R. and Spencer, D. (1942). On sets of integers which contain no three in arithmetic progression. *Proc. Nat. Acad. Sci.*, 28:561–563.

Shapira, A. (2009). Green's conjecture and testing linear-invariant properties. In *STOC*, pages 159–166.

Stothers, A. J. (2010). *On the Complexity of Matrix Multiplication*. PhD thesis, University of Edinburgh.

Williams, V. V. (2012). Multiplying matrices faster than coppersmith-winograd. In *STOC*, pages 887–898.

Xie, N. (2010). Private communication.

A Construction of Large Uniquely Solvable Puzzles

In this appendix we present Coppersmith and Winograd (1990)'s construction of large USP's, isolating it from the matrix multiplication context.

The construction makes use of the following theorem:

Theorem 7 (Salem and Spencer, 1942)**.** *Given $\delta > 0$, for all large enough integer M, there is a set $B \subset [M]$ of size $\Omega(M^{1-\delta})$ such that for all $b_i, b_j, b_k \in B$, $b_i + b_j \equiv 2b_k \mod M$ iff $i = j = k$.*

Such constructions of big sets of integers with no arithmetic progressions constitute an important class of combinatorial objects. Improvements over Salem and Spencer's original construction with slightly larger sizes were given by Behrend (1946) and Elkin (2010), but for our purpose the rougher asymptotic bound of $\Omega(M^{1-\delta})$ suffices.

Now we are ready to describe the construction. We fix a large enough integer N and $M = 2\binom{2N}{N} + 1$. Fix $B \subset [M]$ as given by Theorem 7. Sample $3N$ integers $0 \leq w_j < M$ independently at random for each $j = 0, 1, \cdots, 3N$. We will call these w_j's *weights*. Now consider the set $\mathcal{I}$ of

all subsets $I \subset [3N]$ of size N. Let δ_I be the indicator function of subset I, i.e., for each $j \in [3N]$, $\delta_I(j) = 1$ for $j \in I$, and 0 otherwise. The weights we sampled define three mappings from $\mathcal{I}$ to $\mathbb{Z}_M$:

$$\beta_x(I) \equiv \sum_{j=1}^{3N} \delta_I(j) w_j \mod M; \tag{1}$$

$$\beta_y(I) \equiv w_0 + \sum_{j=1}^{3N} \delta_I(j) w_j \mod M; \tag{2}$$

$$\beta_z(I) \equiv \left(w_0 + \sum_{j=1}^{3N} (1 - \delta_I(j)) w_j \right) /2 \mod M. \tag{3}$$

Note that the operation of division by 2 is well defined for β_z, as M is odd.

With these mappings, we will consider each element $b_i \in B$. First, with each $b_i \in B$ we associate all triples (I, J, K), where $I, J, K \in \mathcal{I}$ are pairwise disjoint, and $\beta_x(I) = \beta_y(J) = \beta_z(K) = b_i$. (A triple (I, J, K) is discarded if the members are not pairwise disjoint, or if they are not mapped to be same b_i.) Second, among all triples associated with the same b_i, we arbitrarily remove all but one triple. To construct our USP $U \subset \{1, 2, 3\}^{3N}$, there will be a puzzle u_i for each b_i associated with a nonempty triple (I_i, J_i, K_i), and for each $j \in [3N]$, $u_i(j) = 1$ for $j \in I_i$, $u_i(j) = 2$ for $j \in J_i$, and $u_i(j) = 3$ for $j \in K_i$.

We first check that we indeed obtain a USP family, before going on to prove its expected size.

Claim 8. *For any $i_1, i_2, i_3 \in [|B|]$, I_{i_1}, J_{i_2} and K_{i_3} are pairwise disjoint iff $i_1 = i_2 = i_3$.*

Note that Claim 8 suffices for the property of USP (Definition 2).

Proof. Suppose I_{i_1}, J_{i_2} and K_{i_3} are pairwise disjoint, we have that

$$b_{i_1} \equiv \beta_x(i_1) \equiv \sum_{j=1}^{3N} \delta_{I_{i_1}}(j) w_j \mod M; \tag{4}$$

$$b_{i_2} \equiv \beta_y(i_2) \equiv w_0 + \sum_{j=1}^{3N} \delta_{J_{i_2}}(j) w_j \mod M; \tag{5}$$

$$b_{i_3} \equiv \beta_z(i_3) \equiv \left(w_0 + \sum_{j=1}^{3N} (1 - \delta_{K_{i_3}}(j)) w_j \right) /2 \equiv \left(w_0 + \sum_{j=1}^{3N} \delta_{I_{i_1} \cup J_{i_2}} w_j \right) /2 \mod M. \tag{6}$$

Straightforwardly, we will have $b_{i_1} + b_{i_2} - 2b_{i_3} \equiv 0 \mod M$. However, since b_{i_1}, b_{i_2} and b_{i_3} are in B, by the property of B, it can only be that $i_1 = i_2 = i_3$. □

We now show that the we indeed have a large USP. This amounts to showing that we have many triples left at the end of the second step of the construction. We first consider the number of triples associated with elements in B in the first step.

Claim 9. *Fixing $b_i \in B$, the expected number of triples (I, J, K) associated with b_i in the first step is $\binom{3N}{N,N,N} M^{-2}$.*

Proof. First, by the same calculation as in Claim 8, we know that if two disjoint $I, J \in \mathcal{I}$ are mapped to the same $b_i \in B$ by β_x and β_y, respectively, then their complement, $K = [3N] - (I \cup J)$, must be mapped to be same b_i by β_z. Now there are $\binom{3N}{N,N,N}$ disjoint triples, the probability that each of the two components is mapped to a fixed b_i is M^{-1}, respectively. Moreover, the two events are independent. The claim follows. □

Claim 10. *Fixing $b_i \in B$, the expected number of triples (I, J, K) associated with b_i that we remove in the second step is at most $\frac{3}{2}\binom{3N}{N,N,N}(\binom{2N}{N} - 1)M^{-3}$.*

Proof. Fixing $b_i \in B$, the expected number of triples (I, J, K) and (I', J', K) $(I \neq I')$ associated with b_i is $\frac{1}{2}\binom{3N}{N,N,N}(\binom{2N}{N} - 1)M^{-3}$. The term $\binom{2N}{N} - 1$ counts the number of I''s disjoint with K and unequal to I. The factor $\frac{1}{2}$ disregards the symmetric case $(I, J, K), (I', J', K)$ and $(I', J', K), (I, J, K)$. The additional M^{-1} here (as compared to the count in Claim 9) indicates the probability of the event $\beta_y(I') = b_i$. Note that this event is independent from the events $\beta_x(I) = b_i$ and $\beta_y(J) = b_i$, even if J' can be equal to I, because of the presence of the weight w_0 in the definition of β_y. Repeat the argument for the cases when two triples coincide on the first or second subset, and the claim follows. (The event that two triples associated with the same b_i disagree on each subset they contain is neglected here, since its probability is significantly smaller than that of the case analyzed here. For large N and M this is easily accommodated.) □

Therefore, by our choice of M, the expected number of triples associated with each b_i remaining after the second step is at least

$$\binom{3N}{N,N,N}M^{-2} - \frac{3}{2}\binom{3N}{N,N,N}\left(\binom{2N}{N} - 1\right)M^{-3} \geq \frac{1}{4}\binom{3N}{N,N,N}M^{-2}.$$

With a standard probabilistic argument, we conclude that there exists a choice of w_j's such that the size of USP we obtain is at least

$$\frac{1}{4}\binom{3N}{N,N,N}M^{-2}|B| = \frac{1}{4}\binom{3N}{N,N,N}M^{-2}M^{1-\delta}.$$

Substituting our choice of M and applying the Stirling's formula, we get the desired asymptotic bound of $(3/2^{2/3} - o(1))^{3N}$ for the size of USP.