Arithmetic Progression Hypergraphs: Examining the Second Moment Method

Michael Mitzenmacher*
michaelm@eecs.harvard.edu

Abstract

In many data structure settings, it has been shown that using "double hashing" in place of standard hashing, by which we mean choosing multiple hash values according to an arithmetic progression instead of choosing each hash value independently, has asymptotically negligible difference in performance. We attempt to extend these ideas beyond data structure settings by considering how threshold arguments based on second moment methods can be generalized to "arithmetic progression" versions of problems. With this motivation, we define a novel "quasi-random" hypergraph model, random arithmetic progression (AP) hypergraphs, which is based on edges that form arithmetic progressions and unifies many previous problems. Our main result is to show that second moment arguments for 3-NAE-SAT and 2-coloring of 3-regular hypergraphs extend to the double hashing setting. We can generalize these results to larger sized hyperedges, when randomly chosen hyperedges satisfy an appropriate notion limited independence. We leave several open problems related to these quasi-random hypergraphs and the thresholds of associated problem variations.

1 Introduction

1.1 Arithmetic Progression Problems

For several multiple-choice hashing data structures, it has been shown that using "double hashing" instead of uniform hashing yields essentially no difference in performance, in both the asymptotic and practical senses of performance. For example, consider the balanced allocation setting [4] where n balls are placed into n bins, but instead of placing them randomly according to a single hash function (assumed to be perfectly random), each ball is placed sequentially in the least loaded of d choices, each obtained independently by a hash function, for some constant integer $d > 2$. The asymptotic fraction of bins of each constant load can be found by fluid limit differential equations [20], and the maximum load is known to be $\log \log n / \log d + O(1)$ with high probability [4]. Now suppose the d choices are chosen in a fashion reminiscent of double hashing, in the following way. Let us assume that n is prime, and each ball is hashed twice, to obtain a chosen uniformly from $[0, n-1]$ and b chosen uniformly from $[1, n-1]$ independently from a. The d choices for a ball are then given by bins $a, a+b, a+2b, \ldots, a+(d-1)b$ modulo n. That is, the d choices are constrained to form an arithmetic progression; as such, only two random numbers modulo n are chosen to determine the choices, instead of d random numbers. It has been shown that, even with this more limited randomness, not only does the maximum load remain $\log \log n / \log d + O(1)$ with high probability, but that the asymptotic fraction of bins of each constant load is the same as when the d choices are all perfectly random [21, 22]. Similar results have been found for other data structures that use multiple hash functions, including Bloom filters [17] and cuckoo hashing [18, 23].

In this paper, we expand on this theme to consider whether similar phenomena occur in "multiple-choice" settings outside of hash-based data structures. Specifically, we attempt to generalize threshold arguments for variations of random constraint problems found using the second moment method, such as variations of k-NAE-SAT and 2-coloring random k-uniform hypergraphs. Formalizing these variations lead us to propose a novel "quasi-random" hypergraph model that we believe is worthy of future study in its own right; we propose several open questions related to this graph model.

To review, the standard k-NAE-SAT problem is the following: given a collection of variables $\{x_1, \ldots, x_n\}$, and a collection of clauses where each clause consists of k literals on these variables (that is, k variables or negations of variables), can we assign true/false values to the variables so that each clause has at least one true and at least one false literal. For the random k-NAE-SAT problem, m clauses are chosen uniformly at random from the collection of all clauses. (There are variations depending on whether literals for each clause are chosen with or without replacement; the choice of

*Harvard University, School of Engineering and Applied Sciences. Supported in part by NSF grants CCF-1320231, CNS-1228598, IIS-0964473, and CCF-0915922. Part of this work was done while visiting Microsoft Research, New England.

such details can simplify analysis, but doesn't affect the asymptotic results we consider here.)

Our proposed variation, k-AP-NAE-SAT, adds the following additional requirements: the number of variables n is prime, and the k variables in each clause must form an arithmetic progression $a, a+b, a+2b, \dots, a+(k-1)b$ modulo n. In the random variation, each clause is chosen independently and uniformly from only $2^k n^2$ possible clauses; that is, the k signs for the literals are chosen independently and uniformly at random, and we choose values for a and b modulo n independently and uniformly at random to choose the variables within a clause. Our primary result is that a second moment method argument of Achlioptas and Moore for random k-NAE-SAT can be applied to random k-AP-NAE-SAT to obtain a lower bound threshold result for the case where $k = 3$ [2].

The k-NAE-SAT problem is similar to the hypergraph 2-coloring problem of k-uniform hypergraphs: given a collection of m hyperedges each of k vertices, can the vertices be colored with two colors so that each edge has a vertex of each color. Here we consider the variation of 2-coloring hypergraphs, where each hyperedge consists of vertices that form an arithmetic progression. We again show that a second moment argument due to Achlioptas and Moore generalizes to this setting when $k = 3$ [1].

We describe the current impediments to extending these results beyond $k = 3$, and make corresponding conjectures. More generally, we describe a "quasi-random" hypergraph model where edges are constrained to form an arithmetic progression that captures these and many other problems. We also show that we can generalize these results somewhat if we expand this quasi-random hypergraph model to allow for edges to be chosen so that the vertices in an edge are $(k-1)$-wise independent for k-uniform hypergraphs for odd $k > 3$. We suggest several open questions related to these quasi-random hypergraph models.

1.2 Model and Motivation

As described above, we consider variations of random k-NAE-SAT where the variables within a clause form an arithmetic progression, which may appear to be a somewhat unusual object. We offer multiple motivations for why this is an interesting problem to study.

Satisfiability problems are a particularly natural problem to consider in this context, as random k-SAT and variations have been the subject of wide study, with a great deal of work focusing on the thresholds for satisfiability (see, e.g., [2, 8, 10, 15] and references therein). In particular, there has been some belief that random problems at or near the threshold boundary may be where "hard instances" lie for NP-complete variations such as k-SAT [7, 24]. It therefore seems interesting that there may be random k-SAT variations utilizing much less randomness but that seem to have similar threshold properties; such variations may prove useful for understanding the complexity of random k-SAT problems or similar problems, or for testing algorithms for such problems. Indeed, the added structure of arithmetic progressions may potentially make such problems easier to study in some circumstances.

In considering problems with thresholds, the second moment method is a natural line of attack, given that it led the way to key threshold results for many problems relatively recently, and is reasonably well understood. We therefore turn to the seminal results of Achlioptas and Moore [1, 2], which provided a spark to the area.

While not specifically a hashing problem, random k-SAT problems are similar (and closely connected to) hashing problems; see e.g. [9]. As previously noted, there are various settings where using double hashing has essentially the same behavior as hashing with fully random multiple choices. Most of these hashing problems have natural interpretations as problems on random hypergraphs, by viewing elements as hyperedges and buckets as vertices in a hypergraph. For example, the question of whether there is an assignment of elements to buckets in cuckoo hashing is equivalent to the question of whether there is an orientation of hyperedges in a random hypergraph that does not "overload" any vertex by orienting too many edges toward it. It therefore seems natural to ask whether other hypergraph problems behave similarly when using edges based on arithmetic progressions instead of fully random edges.

The double hashing variation of these hashing schemes correspond to the following variations of random hypergraph models, which appear implicit in previous works (see, e.g., [18, 23]) but do not seem to have been formally described. As mentioned previously, in what follows for convenience we allow for n^2 arithmetic progressions on the numbers modulo n for prime $n > 2$; we include the trivial progressions $(a, a, \dots, a)$ where all elements are the same, and we count $(a, a+b, a+2b, \dots, a+(k-1)b)$ and $(a+(k-1)b, \dots, a+b, a)$ as distinct progressions for $a \in [0, n-1]$ and $b \in [1, n-1]$.

Definition 1.1. *For a prime number n and $k < n$, let $H_k^{AP}(n, p)$ be a random k-uniform hypergraph on n vertices where where each of the n^2 possible edges given by $(a, a+b, a+2b, \dots, a+(k-1)b)$ with $a \in [0, n-1]$ and $b \in [0, n-1]$ is included with probability p. Similarly, let $H_k^{AP}(n, m)$ be a random k-uniform hypergraph on n vertices with m edges where the edges are determined by selecting independently and uniformly at random, with* *replacement, each edge from all n^2 possible arithmetic progressions.*

We may also use the phrasing "a hypergraph chosen from $H_k^{AP}(n,m)$" (or $H_k^{AP}(n,p)$) where the meaning is clear. We suggest referring to these types of hypergraphs as *random arithmetic progression hypergraphs*, or more succinctly *random AP hypergraphs*.

One can vary this definition to disallow hyperedges with repeated vertices if needed, or to disallow multi-hyperedges, or to have non-regular edges. For example, if we only allow $b \in [1, n-1]$, then there are only $n(n-1)$ possible hyperedges, and the n hyperedges of the form $(a, a, \ldots, a)$ are not considered. Also, we note that we focus on n prime for convenience, to avoid asymmetries when b is not relatively prime to n. For example, if n is even then there are edges of the form $(a, a+n/2, a, a+n/2, \ldots)$ that can prove inconvenient. (In many settings, this issue may ignored without affecting the asymptotics, or similarly b may instead be restricted to be a randomly chosen number relatively prime to n.)

The $H_k^{AP}(n,p)$ model is a natural "quasi-random" graph model that has not, to our knowledge, been studied previously. Previous results for hashing schemes suggest it is very like the standard random hypergraph model (where each hyperedge of k distinct vertices is included in the hypergraph with some probability p) in useful ways (at least from the algorithmic perspective). We believe this graph model is worthy of further investigation, as we discuss below. We correspondingly define variations of the k-SAT problem with the addition of an "AP" to denote that the variables must be chosen to form an arithmetic progression. Hence instead of k-SAT and k-NAE-SAT, we may have k-AP-SAT and k-AP-NAE-SAT.

One alternative way of thinking about edges in a hypergraph chosen from $H_k^{AP}(n,m)$ is that the hyperedges themselves are "pairwise independent"; that is, every pair of entries takes on every possible pair of values with equal probability. One could define a model, that here we simply denote by $H_k^j(n,m)$ (or $H_k^j(n,p)$), where each edge holds k vertices that correspond to j-wise independent random variables. While we think this model is less natural than the arithmetic progression model, it does allow us to generalize our results, as we describe briefly where appropriate.

1.3 Result Summary and Some Conjectures

Reviewing previous second moment arguments for threshold behaviors, several of them rely on determining the distribution of the number of monochromatic collections of k variables for a random coloring of the variables with two colors (such as "true" of "false" in the k-SAT settings). In the variations we consider where the variables in our constraints are governed by an arithmetic progression, the relevant question to consider is the distribution of the number of monochromatic collections of k variables, when the variables are in an arithmetic progression. For general k, this is a very challenging problem [5, 19, 25], but for the case of $k = 3$, the distributions are the same as in the case of complete randomness; we discuss and prove this below. This simple fact provides our main results. More generally, for odd k, when the collections of k variables are $(k-1)$-wise independent, the number of monochromatic collections of k variables are the same as for the case of complete randomness.

These results, however, provide some fodder for additional conjectures.

1. Can the second moment method arguments of [1, 2] for k-NAE-SAT, or other second moment arguments (such as in [3, 8]), be made to apply to k-AP-NAE-SAT to obtain corresponding threshold results for all $k \geq 3$? (The same questions can be asked for hypergraph 2-colorability.)

2. There has been further progress on the actual thresholds for k-NAE-SAT and k-SAT using "belief propagation" based analysis (e.g, [10, 11]). Can these techniques be used and these results be shown to also hold in the settings of k-AP-NAE-SAT and k-AP-SAT?

We conjecture all of these questions have positive answers. We discuss some of the challenges to proving the first conjecture below.

2 Second Moment Argument for k-AP-NAE-SAT

2.1 Problem Statement

We consider k-AP-NAE-SAT, following closely the argument of Achlioptas and Moore [2]. We consider a formula F with variables $x_1, x_2, \ldots, x_n$ and $m = rn$ clauses. We use the term literal to refer to a variable or a negated variable. Fitting with our previous discussion, assume n is a prime and F is determined by choosing each clause independently and uniformly at random from all clauses with k literals with replacement, where the k literals are required to be one of the n^2 possible arithmetic progressions modulo n. (The analysis is easily shown to be essentially the same if the clauses cannot include the trivial clauses and/or are chosen without replacement; see the discussion for instance in Section 4.1 of [2] showing such variations affect the threshold by $o(1)$ terms.) We aim to determine as tightly as possible bounds for the threshold for satisfiability; ideally, we would find a constant r^* so that for any constant $\epsilon > 0$, for all $r > (r^* + \epsilon)$ with high probability a randomly selected formula F would not be satisfiable, while for all $r < (r^* - \epsilon)$, with high probability a randomly selected formula F would be satisfiable. Our results will hold only for $k = 3$, but we write the argument for general k and later explain why $k = 3$ is required.

2.2 Adapting the Second Moment Argument

Let $X = X(F)$ be the number of NAE-satisfying truth assignments for a formula F; that is, the truth assignment has the property that for every clause in F, there is at least one satisfied and at least one unsatisfied literal. We apply the second moment method as follows. The first moment is simple to compute, and indeed the fact that the clauses are constrained to be arithmetic progressions does not affect the calculation.

$$\mathbb{E}[X] = 2^n \left(1 - 2^{1-k}\right)^m,$$

since the probability that an assignment NAE-satisfies a randomly chosen clause is $1 - 2^{1-k}$. (In particular, we note the following well-known result for purely random clauses hold here in the setting of random AP clauses as well: if $m = rn$ and $k = 3$, then for $r > \log_{4/3} 2 \approx 2.41$,

$$\mathbb{E}[X] = 2^n (3/4)^{rn} = o(1),$$

so that first order methods give an upper bound on the satisfiability threshold.)

To compute $\mathbb{E}[X^2]$, we consider all possible 4^n ordered pairs of assignments (σ, τ). We recall from [2] that the probability that both σ and τ NAE-satisfy a randomly chosen clause c (from the collection of *all* possible clauses with k literals) depends only on the overlap between the assignments. That is, if σ and τ agree on $z = \alpha n$ variables, then by inclusion-exclusion we find

$$\Pr[\sigma \text{ and } \tau \text{ NAE-satisfy } c] = \\ 1 - 2^{2-k} + 2^{1-k}\left(\alpha^k + (1-\alpha)^k\right).$$

This is because for a clause c to *not* be NAE-satisfied by both σ and τ, the literals of c must evaluate to all true or all false in σ and τ. One way of expressing this is that if we consider truth assignments as bit strings in the natural way (with true being 1 and false being 0), and then consider the exclusive-or $\sigma \oplus \tau$ as a bit string, both σ and τ fail to NAE-satisfy a clause c if and only if the variables in the clause c are *monochromatic* in $\sigma \oplus \tau$; that is, they must all be 0 or all 1. The probability of a randomly chosen clause c being monochromatic in this manner is $\alpha^k + (1-\alpha)^k$. (Technically, this is the probability if the variables are not constrained to be distinct; as mentioned, the difference between allowing or not allowing repeated variables does not affect the overall threshold argument.) This expression for the probability σ and τ NAE-satisfy c forms the basis for the calculations in [2] regarding bounding the ratio $\mathbb{E}[X^2]/\mathbb{E}[X]^2$ for the remainder of their second moment argument.

Let us consider $\mathbb{E}[X^2]$ now where c' is a randomly chosen clause constrained to be an arithmetic progression. As before, we consider all possible 4^n ordered pairs of assignments (σ, τ). Let $\beta(\sigma, \tau)$ be the fraction of arithmetic progressions of length k in the bit string for $\sigma \oplus \tau$ that are monochromatic. The corresponding inclusion-exclusion for this case give us now:

$$\Pr[\sigma \text{ and } \tau \text{ NAE-satisfy } c'] = 1 - 2^{2-k} + 2^{1-k}\beta(\sigma, \tau).$$

For general k, $\beta(\sigma, \tau)$ depends on the actual choices of σ and τ. For the case of $k = 3$, however, we have that if σ and τ agree on $z = \alpha n$ variables, then $\beta(\sigma, \tau)$ is a function of α, and in fact equals $\alpha^3 + (1-\alpha)^3$. In particular, we have the following known result:

Lemma 2.1. *For a prime n, a two-coloring of Z_n where the two color classes consist of αn numbers and $(1-\alpha)n$ numbers has $(1 - 3\alpha + 3\alpha^2)n^2 = (\alpha^3 + (1-\alpha)^3)n^2$ monochromatic arithmetic sequences of length three.*

Although this result is known in the literature [19], for completeness we provide a proof (suggested to us by Yufei Zhao) in the appendix. (This proof nicely generalizes to the setting where edges are chosen to be $(k-1)$-wise independent, as we describe also in the appendix.)

Because the probability a randomly chosen progression of length three is monochromatic is also $(\alpha^3 + (1-\alpha)^3)$, for the second moment calculations of [2], the results when clause variables are constrained to be arithmetic progression of length three remains the same as if the clause variables were randomly chosen. In particular, if we let

$$f(\alpha) = \frac{1}{2} + \frac{1}{4}(\alpha^3 + (1-\alpha)^3),$$

then since

$$\mathbb{E}[X] = 2^n (3/4)^{rn} = o(1),$$

we have

$$\mathbb{E}[X]^2 = 4^n (9/16)^{rn} = \left(4 f(1/2)^r\right)^n,$$

and (letting $H(x)$ be the binary entropy of x)

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{z=0}^{n} \binom{n}{z} f(z/n)^m \\ &\leq (n+1) \max_{\alpha} \left(2^{H(\alpha)} f(\alpha)^r\right)^n. \end{aligned}$$

At $\alpha = 1/2$, we have

$$\left(2 \cdot 2^{H(\alpha)} f(\alpha)^r\right) = \left(4 f(1/2)^r\right).$$

Therefore, when the maximum value for α occurs at $\alpha = 1/2$, this shows that $\mathbb{E}[X^2]$ is within a polynomial factor of $\mathbb{E}[X]^2$. The result of [2] that we require now can be summarized by the following lemma (corresponding to work in Lemmas 2, 3, and 4 of [2]):

Lemma 2.2. *For $r < 3/2$,*

$$\sum_{z=0}^{n} \binom{n}{z} f(z/n)^{rn} < C(2^n)(9/16)^{rn}$$

for some constant C.

To summarize, for $r < 3/2$, the maximum value for α in the expression $2^{H(\alpha)} f(\alpha)^r$ indeed occurs at $\alpha = 1/2$, and in fact $\mathbb{E}[X^2]$ and $\mathbb{E}[X]^2$ are not just within a factor that is polynomial in n, but we have the stronger result that

$$\mathbb{E}[X^2] \quad \leq \quad C\mathbb{E}[X]^2$$

holds for some constant C. We emphasize that bounding $\mathbb{E}[X^2]$ by $C\mathbb{E}[X]^2$, that is within just a constant factor, requires a significant and lengthy calculation argument, and refer the reader to [2] for the details.

We have the following corollary, which then corresponds to the significant work of Lemma 2 and Theorem 5 of [2]:

Corollary 2.1. *For $k = 3$, when $r < 3/2$ a random k-AP-NAE-SAT formula is satisfiable with probability bounded below by some constant.*

2.3 Remarks on the Result We would like to further say that Corollary 2.1 implies that a random k-AP-NAE-SAT has a solution with probability $1 - o(1)$. In [2], the authors rely on the results of Friedgut [15] to claim that the result with uniformly positive probability can be boosted to with high probability. Specifically, Friedgut's work is in a line of results that show for many problems on random graphs and hypergraphs that if there is a threshold, it must be a "sharp threshold" instead of a "coarse threshold" [6, 13, 14, 15, 16]. In particular, this implies that if the second moment results yield a constant lower bound on the existence of solution, then in fact a solution exists with high probability. While these results intuitively should also apply as well to the settings of $H_k^{AP}(n,p)$ and $H_k^{AP}(n,m)$, it is not immediate that they do so; in particular, these quasi-random hypergraphs do not have the symmetry properties that fully random hypergraphs do, which are used in many of the existing proofs. (Some results here do not require complete symmetry, but have other non-trivial requirements.)

We conjecture that the "sharp threshold" framework of Friedgut (and others) can be generalized to the quasi-random hypergraph models $H_k^{AP}(n,p)$ and $H_k^{AP}(n,m)$; however, we leave this as an open question.

Naturally, we would also like the result to hold for all constant k, not just $k = 3$. Recall that the second moment argument depends on

$$\Pr[\sigma \text{ and } \tau \text{ NAE-satisfy } c'] = 1 - 2^{2-k} + 2^{1-k}\beta(\sigma,\tau),$$

where $\beta(\sigma,\tau)$ is the fraction of arithmetic progressions of length k in the bit string for $\sigma \oplus \tau$ that are monochromatic. For $k = 3$, we found this fraction was

$$f_k(\alpha) = 1 - 2^{2-k} + 2^{1-k}(\alpha^k + (1-\alpha)^k)$$

for every σ and τ. Unfortunately, for $k > 3$, the fraction of arithmetic progressions of length k that are monochromatic varies with σ and τ.

One might hope that, say for randomly chosen σ and τ, the fraction of monochromatic arithmetic progressions in $\sigma \oplus \tau$ would be well concentrated around $f_k(\alpha)$, and this would be sufficient to bound the second moment and apply the second moment method. Chernoff-like concentration bounds for this setting do exist [5, 25], but it is also known that the number of arithmetic progressions can differ significantly from the mean (see, e.g., [19]). The second moment argument of Achlioptas and Moore is, unfortunately for us, rather precise. Recall $\mathbb{E}[X^2]$ is bounded by above by $(n+1)\max_\alpha \left(2^{H(\alpha)} f_k(\alpha)^r\right)^n$, increasing $f_k(\alpha)$ by even an ϵ-size constant increase $\mathbb{E}[X^2]$ exponentially, and the second moment bound is lost. We optimistically conjecture that the second moment method can be generalized for $k > 3$, but it (currently) appears difficult; it may require a different approach to bounding $\mathbb{E}[X^2]$ rather than considering how $\beta(\sigma,\tau)$ depends on the overlap α. Possibly, the second moment method may not be suitable for finding thresholds for general k, and other approaches (as in [10, 11]) may be required.

As mentioned, in the appendix we generalize Lemma 2.1, which in turn yields that Corollary 2.1 generalizes so that when a k-SAT formula is chosen so that the literals for each clause are $(k-1)$-wise independent and the clauses are chosen independently (for odd k), when r is less than the corresponding second moment threshold determined by Achlioptas and Moore [1, 2], the formula is satisfiable with probability bounded below by some constant.

3 Second Moment Argument for Hypergraph 2-Colorability

3.1 Adapting the Second Moment Argument We can similarly consider the second moment method with for hypergraph 2-colorability, another problem considered in [1, 2]. Here we follow the argument of [1].[1]

Let X be the number of 2-colorings of a hypergraph chosen from $H_k^{AP}(n,m)$, with $m = rn$ again. We apply the second moment method as follows, for the case where $k = 3$. The probability that a 2-coloring with $z = \alpha n$ black vertices and $n - z = (1-\alpha)n$ white vertices makes a random hyperedge bichromatic is

$$1 - 3\alpha + 3\alpha^2 = 1 - \alpha^3 - (1-\alpha)^3,$$

based on Lemma 2.1, just as it would be for a standard random hypergraph. Thus

$$\mathbb{E}[X] = \sum_{z=0}^{n} \binom{n}{z} (1 - (z/n)^3 - (1 - z/n)^3)^{rn}.$$

Note that in this settings this expectation result cannot be generalized when $k > 3$, since as we have discussed the number monochromatic hyperedges would not depend only on the fraction of vertices of each color.

To bound $\mathbb{E}[X^2]$, consider a pair of 2-colorings S and T with αn and βn black vertices respectively, and γn vertices that are black in both. We follow the proof as outlined in Section 3 of [1]. For a standard random hypergraph, by a standard inclusion-exclusion argument, the probability a random hyperedge of size k would be bichromatic under both S and T is given by:

$$\begin{aligned} 1 &- \alpha^k - (1-\alpha)^k - \beta^k - (1-\beta)^k \\ &+ \gamma^k + (\alpha-\gamma)^k + (\beta-\gamma)^k + (1-\alpha-\beta+\gamma)^k. \end{aligned}$$

For the case of $k = 3$, the above expression holds also for hypergraphs chosen from $H_k^{AP}(n,m)$, again because of Lemma 2.1. Let the expression above be denoted by $p(\alpha,\beta,\gamma)$. Achlioptas and Moore provide the following expression for $\mathbb{E}[X^2]$ in the setting of standard hypergraphs (see Equation 6 of [1]), which carries over here as well:

$$\mathbb{E}[X^2] = \sum_{z_1,z_2,z_3,z_4} \binom{n}{z_1, z_2, z_3, z_4} p\left(\frac{z_1+z_2}{n}, \frac{z_1+z_3}{n}, \frac{z_1}{n}\right)^{rn},$$

where z_1 is the number of vertices colored black in both assignments, z_2 is the number of vertices colored black in S and white in T, z_3 is the number of vertices colored white in S and black in T, and z_4 is the number of vertices colored white in both assignments.

Given these expressions for $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$, Achlioptas and Moore show that the dominant term when calculating the ratio between $\mathbb{E}[X^2]$ and $\mathbb{E}[X]^2$ for the second moment method is when $\alpha = \beta = 1/2$ and $\gamma = 1/4$ (see Lemma 7 of [1]); they provide additional work that shows that again

$$\mathbb{E}[X^2] \quad \leq \quad C\mathbb{E}[X]^2$$

for a constant C (that depends on k). Their calculations for the standard hypergraph case again carry over to the case of $k = 3$ here, yielding the following final result.

Theorem 3.1. *For a random hypergraph $H_k^{AP}(n,m)$ with $k = 3$, $m = rn$, and $r < 3/2$, the hypergraph is 2-colorable with uniform positive probability.*

3.2 Remarks on the Result Again, we would like the with uniform positive probability result here to yield a with high probability result. The 2-colorability threshold of standard random hypergraphs is considered explicitly by Friedgut [14], who shows that there is a non-uniform sharp threshold for hypergraph 2-colorability, which implies that in the context of standard hypergraphs, the result with uniform positive probability implies a with high probability result. We conjecture that the same holds here; however, the explicit proof by Friedgut depends on a result by Erdős and Simonovits on random t-partite systems [12]. It is an interesting question as to whether that result can generalize to the setting of hyperedges governed by arithmetic progressions.

Also, as with k-AP-NAE-SAT, it remains a challenge to obtain second moment results for $k > 3$. For odd $k > 3$, Theorem 3.1 generalizes when vertices in the edges are $(k-1)$-wise independent, again because Lemma 2.1 generalizes appropriately.

4 Conclusion

Motivated by recent results in double hashing, we have introduced quasi-random hypergraphs that we refer to as random AP hypergraphs. Given that random AP hypergraphs have implicitly arisen in the study of

[1] In [2], Achlioptas and Moore provided a more sophisticated argument with a simpler overall calculation, by considering *balanced* colorings, where the number of vertices of each color are the same. In our case, since the number of vertices is odd, we cannot follow their argument completely, although it appears that the natural approach of considering near-balanced colorings $\lfloor n/2 \rfloor$ vertices of the first color and $\lceil n/2 \rceil$ vertices of the second color would allow us to mimic their argument. Because here we are using their mathematical derivations as a black box, it suffices to use the approach of [1], which yields the same result.

multiple-choice hashing data structures, they appear worthy of future study. We have conjectured that random AP hypergraphs share the same thresholds as standard hypergrpahs with regard to 2-colorability, and have correspondingly defined AP versions of random satisfiability problems, which we also conjecture share the same satisfiability thresholds as their standard counterparts. We have shown that second moment arguments for thresholds for these problems generalize when $k = 3$. We have also conjectured that the sharp/coarse threshold results of Friedgut (and others) apply to random AP hypergraphs as well.

At a higher level, it seems worthwhile to understand more clearly how random AP hypergraphs are essentially similar to (and different from) standard hypergraphs in generally usable ways. Perhaps there is some abstractable high-level property shared by random AP and standard hypergraphs that would provide a mechanism for easily showing double hashing versions of hashing data structures have the same performance as standard versions, without an ad hoc analysis for each data structure. At the same time, it would be useful to understand how random AP hypergraphs are different from standard hypergraphs; there should be settings where the use of less randomness and/or specifically the arithmetic progression constraint has a significant algorithmic or complexity-related effect.

Acknowledgments

The author thanks Cris Moore for several helpful discussions, and Yufei Zhao for explaining results regarding monochromatic colorings of arithmetic progressions of length 3.

References

[1] Dimitris Achlioptas and Cristopher Moore. On the 2-colorability of random hypergraphs. In *Proceedings of the 6th International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 953–953. Springer, 2002.

[2] Dimitris Achlioptas and Cristopher Moore. Random k-sat: two moments suffice to cross a sharp threshold. *SIAM Journal on Computing*, 36(3):740–762, 2006.

[3] Dimitris Achlioptas and Yuval Peres. The threshold for random k-sat is $2^k \log 2 - o(k)$. *Journal of the American Mathematical Society*, 17(4):947–973, 2004.

[4] Yossi Azar, Andrei Z Broder, Anna R Karlin, and Eli Upfal. Balanced allocations. *SIAM journal on computing*, 29(1):180–200, 1999.

[5] Bhaswar B Bhattacharya, Shirshendu Ganguly, Xuancheng Shao, and Yufei Zhao. Upper tails for arithmetic progressions in a random set. *arXiv preprint arXiv:1605.02994*, 2016.

[6] Jean Bourgain and Gil Kalai. Influences of variables and threshold intervals under group symmetries. *Geometric and Functional Analysis*, 7(3):438–461, 1997.

[7] Peter C Cheeseman, Bob Kanefsky, and William M Taylor. Where the really hard problems are. In *IJCAI*, volume 1, pages 331–337, 1991.

[8] Amin Coja-Oglan and Konstantinos Panagiotou. Catching the k-naesat threshold. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, pages 899–908. ACM, 2012.

[9] Martin Dietzfelbinger, Andreas Goerdt, Michael Mitzenmacher, Andrea Montanari, Rasmus Pagh, and Michael Rink. Tight thresholds for cuckoo hashing via XORSAT. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, pages 213–225. Springer, 2010.

[10] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, pages 59–68. ACM, 2015.

[11] Jian Ding, Allan Sly, and Nike Sun. Satisfiability threshold for random regular nae-sat. *Communications in Mathematical Physics*, 341(2):435–489, 2016.

[12] Paul Erdős and Miklós Simonovits. Supersaturated graphs and hypergraphs. *Combinatorica*, 3(2):181–192, 1983.

[13] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[14] Ehud Friedgut. Hunting for sharp thresholds. *Random Structures & Algorithms*, 26(1-2):37–51, 2005.

[15] Ehud Friedgut and Jean Bourgain. Sharp thresholds of graph properties, and the k-sat problem. *Journal of the American Mathematical Society*, 12(4):1017–1054, 1999.

[16] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124(10):2993–3002, 1996.

[17] Adam Kirsch and Michael Mitzenmacher. Less hashing, same performance: Building a better bloom filter. *Random Structures & Algorithms*, 33(2):187–218, 2008.

[18] Mathieu Leconte. Double hashing thresholds via local weak convergence. In *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing*, pages 131–137. IEEE, 2013.

[19] Linyuan Lu and Xing Peng. Monochromatic 4-term arithmetic progressions in 2-colorings of z_n. *Journal of Combinatorial Theory, Series A*, 119(5):1048–1065, 2012.

[20] Michael Mitzenmacher. Studying balanced allocations with differential equations. *Combinatorics, Probability and Computing*, 8(5):473–482, 1999.

[21] Michael Mitzenmacher. Balanced allocations and double hashing. In *Proceedings of the 26th ACM Symposium on Parallelism in Algorithms and Architectures*, pages 331–342. ACM, 2014.

[22] Michael Mitzenmacher. More analysis of double hash-

ing for balanced allocations. In *Proceedings of the Thirteenth Workshop on Analytic Algorithmics and Combinatorics*, pages 1–9. SIAM, 2016.

[23] Michael Mitzenmacher and Justin Thaler. Peeling arguments and double hashing. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, pages 1118–1125. IEEE, 2012.

[24] Rémi Monasson, Riccardo Zecchina, Scott Kirkpatrick, Bart Selman, and Lidror Troyansky. Determining computational complexity from characteristic 'phase transitions'. *Nature*, 400(6740):133, 1999.

[25] Lutz Warnke. Upper tails for arithmetic progressions in random subsets. *arXiv preprint arXiv:1612.08559*, 2016.

Appendix: A Key Lemma

We provide here the key lemma regarding two-colorings of Z_n for prime n and length-three arithmetic progressions.

LEMMA 2.1 *For a prime n, a two-coloring of Z_n where the two color classes consist of αn numbers and $(1-\alpha)n$ numbers has $(1-3\alpha+3\alpha^2)n^2$ monochromatic arithmetic sequence of length three.*

Proof. [Proof of Lemma 2.1] Let z_i be a variable for $i \in [0, n-1]$, S be the set of all n^2 arithmetic progressions of length three given as ordered triples, and consider the polynomial

$$\sum_{(i,j,k)\in S} (z_i z_j z_k + (1-z_i)(1-z_j)(1-z_k)).$$

Now consider this polynomial for a given two-coloring of Z_n with z_i being the color of $i \in Z_n$, and where one color class corresponds to $z_i = 0$ and the other corresponds to $z_i = 1$. The summation above then equals the number of monochromatic arithmetic sequences, since each element in the sum is 1 if the triple (i, j, k) corresponds to a monochromatic arithmetic sequence and 0 otherwise. Expanding, we find

$$\begin{aligned}\sum_{(i,j,k)\in S} &(z_i z_j z_k + (1-z_i)(1-z_j)(1-z_k)) \\ &= \sum_{(i,j,k)\in S} (1 - z_i - z_j - z_k + z_i z_j + z_j z_k + z_i z_k) \\ &= n^2 - 3n(\alpha n) + 3(\alpha n)(\alpha n) \\ &= (1 - 3\alpha + 3\alpha^2)n^2.\end{aligned}$$

Note that for the second equality, we use that there are n^2 triples in S; there are $3n$ triples in S including i for the αn values of z_i with $z_i = 1$ (counting the triple (i,i,i) three times); and there are 3 triples in S including i and j for every ordered pair z_i and z_j with $z_i z_j = 1$ (again counting the triple (i,i,i) three times when $i = j$). ■

We note the proof of Lemma 2 generalizes easily to the following setting. Let k be odd, and let S be a set of k-tuples corresponding to a $(k-1)$-wise independent family. Specifically, S consists of k-tuples $(x_1, x_2, \ldots, x_k)$, and for any $1 \le i_1 < i_2 < \ldots < i_{k-1} \le k$, if we choose a k-tuple randomly from S, for any $y_1, y_2, \ldots, y_{k-1} \in [0, n-1]$,

$$\Pr\left(\cap_{j=1}^{k-1} x_{i_j} = y_j\right) = 1/n^{k-1}.$$

In what follows, our sums over S are taken over all $|S|$ many k-tuples $(x_1, x_2, \ldots, x_k) \in S$. We now consider the polynomial

$$\sum_S (z_{i_1} z_{i_2} \cdots z_{i_k} + (1-z_{i_1})(1-z_{i_2})\cdots(1-z_{i_k})).$$

Again this corresponds to the number of monochromatic k-tuples, and since k is odd, we have the term $z_{i_1} z_{i_2} \cdots z_{i_k}$ cancels with the negation of that term from $(1-z_{i_1})(1-z_{i_2})\cdots(1-z_{i_k}))$. It readily follows that if αn of the elements in $[0, n-1)$ are one color and $(1-\alpha)n$ are of the other color, the fraction of monochromatic k-tuples in S is $\alpha^k + (1-\alpha)^k$.